

Imperial College London
Department of Physics

Random Scattering of Surface Plasmons for Sensing and Tracking

Joel Andrew Berk

Submitted in part fulfilment of the requirements
for the degree of Doctor of Philosophy at
Imperial College London

I declare that this thesis and the work reported within it was entirely produced by me. Information derived from the work of others has been referenced in the text and references are given in the bibliography.

Joel Berk (2021)

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

In this thesis, a single particle biosensing setup, capable of sensing and tracking single nanoscale biological particles, is proposed and investigated theoretically. The setup is based on monitoring the speckle pattern intensity distribution arising due to random scattering of surface plasmon polaritons (SPPs) from a metal surface. An analyte particle close to the surface will additionally scatter light, perturbing the speckle pattern. From this speckle pattern perturbation, the analyte particle can be detected and tracked. Theoretical sensitivity analysis predicts a biological particle on the order of 10nm in radius gives a fractional intensity perturbation to the speckle intensity of 10^{-4} , comparable to intensity contrasts used in existing interferometric scattering sensing techniques. A formula for the minimum detectable particle size is derived. In addition, an algorithm is derived capable of extracting the particle trajectory in the single scattering regime from the change to the speckle intensity perturbation over time and shown to be capable of errors $\sim 1\text{nm}$ on simulated data under optimal noise conditions. The effect of multiple scattering on the speckle pattern perturbation is studied, and it is shown that, by tuning the scattering mean free path and individual scatterer properties of a random nanostructure of scatterers on the metal surface, one can increase the magnitude of the speckle field perturbation by up to $\sim 10^2$. A neural network based localisation algorithm is developed to calculate the analyte particle position based on the speckle intensity perturbation and its performance on simulated data is studied. Mean errors on the order of $\sim 20\text{nm}$ were found, depending on the size of the region over which the particle must be tracked. Unlike the single scattering tracking algorithm, the neural network algorithm continues to function in the multiple scattering regime.

Acknowledgements

Producing a PhD thesis in the midst of the global Covid-19 pandemic is not something that I could have achieved without the help and support of many great people. First and foremost, I must thank my supervisor Matthew Foreman. His support and guidance, in physics and beyond, has been invaluable and cannot be overstated. He always tested my ideas and theories with the right questions and a healthy mixture of rigorous scientific scepticism and a willingness to listen to ideas. I am extremely grateful to have worked with him as a PhD student. In a similar vein, the contribution of my second supervisor Carl Paterson, especially to the work in Chapter 4, was crucial. He always managed to steer me away from going down blind alleys and into promising directions, even if I didn't see it at the time.

All the members of the Photonics group in Blackett Laboratory have contributed to creating a productive and engaging work environment which was a pleasure to work in, along with allowing for a healthy dose of after work socialising. In particular, those that I shared an office with over the years, Cheryl, Yu Chen and Niall, all contributed enormously, providing stimulating physics conversations and also whatever the selected lunchtime conversation turned out to be. The self-proclaimed 'core-crew' of PhD students were another essential source of knowledge, especially in the ongoing task of finding out where to get an espresso martini on High Street Kensington (as yet an open question).

Outside of work, I am grateful to the members of Imperial College Hockey Club and Clapham Common Hockey Club for helping provide me with 4 years of 'high quality' hockey, keeping me healthy physically and mentally, along with suitable pre and post match entertainment. I am especially thankful to my teammates over the years in both ICHC and CCHC M3s in this regard.

I also must thank my parents, who had to put up with me through the uncountable lockdowns which meant there was no way to avoid me.

Finally, I have to acknowledge the contribution of KG. She could always cheer me up and make me laugh, even at the hardest times. Her love and affection was essential in helping me finish this thesis and I will be eternally grateful for that.

To Grandpa, Sam and Millie, for living the high life.

Contents

Abstract	3
Acknowledgements	4
Abbreviations and Acronyms	14
1 Introduction	16
1.1 Motivation	16
1.2 Thesis Structure	19
2 Biosensing and Tracking	21
2.1 Evaluating Biosensor Performance	22
2.1.1 Nanopore and Nanowire Sensors	25
2.1.2 Resonance Sensors	26
2.1.3 Label Free Single Molecule Detection	36
2.2 Single Particle Tracking and Localisation	36
2.2.1 Imaging and Microscopy	38
2.2.2 Super-resolution and Localisation Microscopy	38
2.2.3 Interferometric Methods	42
2.3 Conclusion	45
3 Sensitivity Analysis	46
3.1 Surface Plasmon Polaritons	46
3.1.1 Multilayer Structures and SPPs	48
3.2 A Surface Plasmon Scattering Enabled Sensor	51
3.3 Electromagnetic Scattering Theory	53
3.3.1 Lippmann-Schwinger and Dyson Equations	54
3.3.2 Green's Tensor Properties	57
3.3.3 Random Scattering Theory	63
3.4 Scattering of SPPs	67
3.4.1 Scattering from a Rough Surface	68
3.4.2 Scattering from an Analyte Particle	79
3.5 Single Scattering Sensitivity Analysis	85
3.5.1 Field Sensitivity	85
3.5.2 Intensity Sensitivity	89
3.6 Detection Limits	92
3.6.1 Pearson Correlation Coefficient Sensitivity	95
3.7 Conclusion	101

4	Tracking in the Single Scattering Regime	103
4.1	Field Dependence on Particle Position	103
4.2	Tracking Algorithm	106
4.2.1	Minimisation Procedure and Trajectory Consistency Check	110
4.2.2	Sign Ambiguity of Transverse Step	110
4.3	Algorithm Performance	111
4.4	Limitations and Extensions	119
4.5	Conclusion	123
5	Role of Multiple Scattering in Sensitivity	124
5.1	Multiple Scattering Theory and Phenomena	124
5.1.1	Diagrammatic Representations	125
5.1.2	Light Diffusion	126
5.1.3	Multiple Scattering Phenomena	127
5.1.4	Coupled Dipole Model	130
5.2	Enhancement Factors	132
5.2.1	Adding a Scatterer	133
5.2.2	Physical Interpretation	136
5.2.3	Born Expansion of Coupling Matrix	139
5.2.4	Far Field Enhancement Factor	140
5.2.5	Plane Wave Illumination and Reciprocity	141
5.2.6	Intensity Sensitivity Enhancement	143
5.3	Analytic Results	144
5.3.1	Mean Amplitude of Enhancement Factors	150
5.4	Scalar SPP Scattering Results	158
5.4.1	Numerical Results	160
5.4.2	Polarisability Dependence of Peak Enhancement	174
5.4.3	Enhancement Factor Correlations	177
5.5	Conclusion	178
6	Neural Network Based Tracking	180
6.1	Neural Network Background	180
6.1.1	Optimisation Algorithm	183
6.1.2	Neural Networks in Optics and Imaging	185
6.2	Theory of Neural Network Localisation from Speckle	187
6.3	Simulation of Neural Network Tracker	189
6.3.1	Normalisation of the Input	190
6.3.2	Simulation Methods	191
6.3.3	Training	193
6.3.4	Neural Network Performance	193
6.3.5	Comparison to Single Scattering Algorithm	200
6.4	Conclusion	201
7	Conclusion	202
A	Evaluation of Enhancement Factor Integrals	251

List of Tables

2.1	Summary of performance of common label free biosensing methods. The parameter X_{LOD} refers to either the limit of detection in analyte concentration, or analyte particle size for systems with single molecule sensitivity. Sensing volume is expressed as a distance from sensor surface multiplied by sensor area, * indicates analyte must be tightly attached to surface, V_{sens} refers to a single sensing element, multiplexed arrays can increase the effective sensing volume. . .	37
2.2	Summary of noise metrics for dark field and interference methods.	43

List of Figures

2.1	The measured response signal \mathcal{M} as a function of driving frequency before (blue) and after (red) binding of the analyte particle(s). The change in resonance frequency ($\delta\omega_0$), resonance width ($\delta\Delta$) or the change in response signal at a single frequency (δS) can all be used as a sensing signal, or a combination of all of them. While shown as a function of frequency ω , the same principle can be applied to \mathcal{M} as a function of other physical parameters (such as angle in the case of SPR).	28
3.1	Geometries supporting SPP modes (left) a single metal-dielectric interface and (right) a metal film between two dielectrics, the simplest multilayer structure. In both cases, a key feature is the evanescent confinement of the mode to the metal surface, while propagating parallel to the surface.	48
3.2	The sensing setup to be investigated. An SPP propagates on a thin metallic film, and is scattered by surface roughness. An analyte particle, diffusing above the metal film also scatters the SPP when it comes near the surface. This changes the scattered light distribution, and thus measuring the change allows one to deduce the presence of the analyte particle. The red arrows are merely illustrative of scattering and not intended to represent any quantitative information on the distribution of scattered light. While the structure shown here is a glass-metal-water thin film structure, the principle can be applied to any layered structure supporting SPPs.	51
3.3	Representation of the direct, reflected and transmitted contributions to the multilayer Green's tensor $G(\mathbf{r}, \mathbf{r}')$ that propagates a dipolar source at \mathbf{r}' to the field at observation positions above (\mathbf{r}) or below ($\tilde{\mathbf{r}}$) the multilayer structure.	59
3.4	Far field dipole radiation patterns $I_{dp}(\theta, \phi) / \max_{\theta, \phi}(I_{dp})$ normalised to a maximum value of 1 for a dipole oriented in the x , y and z directions 50nm above a dielectric-gold-dielectric thin film structure with $d = 50\text{nm}$ thick gold film. The three variants of the structure plotted here are (top row) glass-gold-water at 650nm, (middle row) water-gold-water at 650nm and (bottom row) glass-gold water at 600nm. The left column shows plots of $\int_0^{2\pi} I_{dp}(\theta, \phi) / \max_{\theta, \phi}(I_{dp}) d\phi$ to give the total intensity radiated in the θ direction, while the right column shows $I_{dp}(\theta, \phi = 0)$ in the (x, z) -plane. Note that the θ -integrated intensity is identical for the in-plane x and y oriented dipoles, and hence only one is plotted.	64
3.5	Diagram showing heuristic argument for scaling of α_{surf} . On the left, a single surface feature is shown occupying a small volume $\sim ha^2$ and thus has a polarisability $\sim (\varepsilon_m - \varepsilon_d)ha^2$. In the right diagram, it can be seen that the entire illuminated region of area $\sim L_y L_{\text{SPP}}$ contains $\sim L_y L_{\text{SPP}}/a^2$ such scattering volumes.	82

3.6	Plots showing the dependence of $ \alpha_a/\alpha_{\text{surf}} $ on the size of the analyte particle (left) and the RMS roughness height (right) for the glass-gold-water structure at $\lambda_0 = 650\text{nm}$. The analyte particle is modelled as a homogeneous sphere of radius R_a and constant permittivity ε_a , with α_a calculated from Eq. (3.71). The three analyte particle compositions plotted are gold ($\varepsilon_a = -13.7 + 1.04i$), polystyrene ($\varepsilon_a = 2.5$) and a ‘biological’ particle ($\varepsilon_a = 1.96$). On the left plot, α_{surf} was held fixed for parameters $h = 6\text{nm}$, $a = 12\text{nm}$, $L_y = 10\lambda_0 = 6.5\mu\text{m}$ while on the right α_a was that of a ‘biological’ particle with the given radius.	83
3.7	(left) The profile of $\langle I_\zeta(\theta = \Theta_{\text{SPP}}, \phi) \rangle$ around the leakage ring for a Gaussian surface correlation function with different values of a , calculated from Eq. (3.75) and normalised to a maximum value of 1 and (right) the value of $\tilde{C}(\Delta\mathbf{k}_\parallel)/\tilde{C}(0)$ around the leakage ring for the same set of correlation lengths.	84
3.8	Plots showing (left) $\max(\mathcal{S}_E)/ \mathbf{A}_{\text{SPP}}^+ $ and $\overline{\mathcal{S}_E}/ \mathbf{A}_{\text{SPP}}^+ $ as a function of z_a with $R_a = 50\text{nm}$ and $r = 100\lambda_0$ fixed and a ‘biological’ analyte and (right) the polar angle θ_{max} at which $\max(\mathcal{S}_E)$ occurs, with the leakage angle Θ_{SPP} shown (red dashed line).	87
3.9	The RMS intensity sensitivity, $\mathcal{S}_I^{\text{RMS}}$, relative to (left) the intensity at the metal surface I_0 (at $r = 100\lambda_0$) and (right) the mean background intensity $\langle I_\zeta \rangle$ (independent of r).	92
4.1	(center) Example scattered intensity in the Fourier plane and the profile around the leakage ring $\tilde{I}_i(\phi)$. (panels) Fringe patterns one would see at different points on the ring from scanning the analyte particle in the x and y directions. In the $\phi = 0$ direction, no fringes are seen as the phase of $\tilde{\mathbf{E}}_a$ does not change, so only decay effects are seen. Due to the random speckle phase, the fringe pattern has an unknown (and different) offset at each ϕ , as depicted by the solid black lines. <i>Figure reprinted with permission from Ref. [273] © 2021 IEEE.</i>	107
4.2	A flow chart of the tracking algorithm including the consistency check and sign flipping steps. The sign switch the transverse component is denoted with a superscript $-$ i.e. if $\delta\mathbf{r}_{ij}^{b,f} = (\delta x_{ij}, \delta y_{ij}, \delta z_{ij})^T$ then $\delta\mathbf{r}_{ij}^{b,f-} = (-\delta x_{ij}, -\delta y_{ij}, \delta z_{ij})^T$. The parameters of the algorithm are L_{thresh} (the distance below which separate estimates of the same step are considered consistent), N_{min} (the initial number of minimisation start points), Δ (the amount to increase the number of start points by each round) and $N_{\text{min}}^{(\text{max})}$, the maximum number of start points to try before terminating.	112
4.3	An example of a simulated trajectory reconstructed using the algorithm. Both noisy and noiseless simulation results are shown. <i>Figure reprinted with permission from Ref. [273] © 2021 IEEE.</i>	114
4.4	Distribution of errors for noiseless coupled dipole simulations with (left) 50 and (right) 100 background dipole scatterers. The 100 background scatterers has more significant multiple scattering effects.	115
4.5	The dependence of the relative frequency of a given fractional error ($ \delta\hat{\mathbf{r}} - \delta\mathbf{r} / \delta\mathbf{r} $) on the ratio of total particle scattered intensity to total background intensity, \bar{I}_a/\bar{I}_b , found without (right) and with (left) the consistency check. <i>Figure reprinted (with alteration, including additional data) with permission from Ref. [273] © 2021 IEEE.</i>	116
4.6	Relative frequency of errors for the algorithm (left) with the consistency check and (right) without the consistency check.	117

4.7	The dependence of the relative frequency of the fractional error distribution in (left) the transverse step, $ \delta\hat{\rho} - \delta\rho / \delta\rho $ and (right) the z step $ \delta\hat{z} - \delta z / \delta z $ on \bar{I}_a/\bar{I}_b	119
5.1	Diagrammatic representation of the Dyson equation. Dotted lines represent unperturbed Green's tensors G_0 while solid lines represent total Green's tensors G including the effect of scatterers. Nodes (\bullet) represent scattering events, introducing a factor of the scattering potential $V(\mathbf{r})$. Intermediate positions are integrated over.	125
5.2	Examples of reciprocal scattering paths contributing to coherent backscattering. In the backscattering direction $\theta = \theta'$, the scattering paths have identical phase and amplitude and add up in phase.	128
5.3	Example multiple scattering paths for each enhancement factor: (γ_1 , left) rescattering of light after scattering from the additional particle en route to the observation point, (γ_2 , center) loop trajectories starting and ending on the additional scatterer and (γ_3 , right) multiple scattering of the illumination field onto the analyte particle. <i>Figure reprinted with permission from Ref. [287] © 2021 American Chemical Society.</i>	137
5.4	Diagrammatic representations of some examples of the interference of scattering paths $\mathcal{L}_i\mathcal{L}_j^*$ contributing to σ_1^2 . Dotted lines with a forward arrow represent a factor of G (freely) propagating the field between two scatterers, while those with a backward arrow represent G^* in the conjugated path. Intermediate scatterer positions are integrated over upon averaging. Dashed lines show shared scatterers between the two paths In the left example, the paths share no scatterers and are therefore statistically independent. For these paths $\langle\mathcal{L}_i\mathcal{L}_j\rangle = \langle\mathcal{L}_i\rangle\langle\mathcal{L}_j\rangle$ and there contribution is included in $ \langle G^{sc}\rangle ^2$. The other two paths share scatterers and therefore are not accounted for in the Drude-Boltzmann approximation.	153
5.5	Diagrammatic representation of how the use of G^{sc} , represented by a solid line with a forward arrow (backward arrow denotes G^{sc*}) to propagate between scatterers accounts for averaging over all scatterer positions not appearing in both \mathcal{L}_i and \mathcal{L}_j . Intermediate scatterer positions are not labelled, but are integrated over. Replacing a dotted unperturbed Green's function line that connects one part of a diagram (represented by X in the figure) with the rest of the diagram (represented by Y) with a solid line representing G^{sc} in a diagram accounts for the whole set of diagrams visiting intermediate scatterers between X and Y, provided those intermediate scatterers do not appear in the rest of the diagram (i.e. in X and Y). As such, connecting scattering events with solid lines and using G^{sc} , one only needs to consider shared scatterers between two paths.	154
5.6	Examples of different kinds of path interferences contributing to σ_1^2 , where every scatterer appears in the both paths.	154
5.7	The first four ladder diagrams which give rise to Eq. (5.71). Solid lines with a forward arrow represent G^{sc} while backwards arrows denote G^{sc*} , intermediate scatterer positions are integrated over. Dashed lines connect shared scatterers.	155
5.8	The three lowest order examples of maximally crossed diagrams, which are responsible for coherent back scattering and weak localisation effects.	156

5.9	Complex means (a) $\langle \gamma_1 \rangle$, (b) $\langle \gamma_2 \rangle$, (c) $\langle \gamma_3 \rangle$ and (d) the mean amplitudes of the individual enhancement factor $\langle \gamma_i \rangle$ and the total enhancement factor $\langle \gamma_1 \gamma_2 \gamma_3 \rangle$ for a gold-water interface at $\lambda_0 = 650\text{nm}$ with 40nm gold sphere scatterers. Vertical dashed lines indicate densities used in Fig. 5.10. <i>Figure reprinted (with alteration) with permission from Ref. [288] © 2021 American Physical Society.</i> .	162
5.10	The relative frequency of γ_1 (top row), γ_2 (middle row) and γ_3 (bottom row) over the complex plane for scatterer densities $n\lambda_0^2 = 0.05$ (left column), 0.67 (middle column) and 8.0 (right column). <i>Figure reprinted with permission from Ref. [288] © 2021 American Physical Society.</i>	163
5.11	Plots showing, as a function of n , (a) the standard deviations σ_{Re} and σ_{Im} of the real and imaginary parts of $\gamma_{1,2,3}$, (b) the ratio $\sigma_{\text{Re}}/\sigma_{\text{Im}}$ and (c) the Pearson correlation coefficient of $ \gamma_{1,2,3} - \langle \gamma_{1,2,3} \rangle $ and $\arg(\gamma_{1,2,3} - \langle \gamma_{1,2,3} \rangle)$	164
5.12	Enhancement factor amplitude statistics for an incident field $E_{0,z} = \exp(ik_{\text{SPP}x} - \kappa_d z)$, (a) the density dependence of $\langle \gamma_i \rangle$ and $\langle \gamma_1 \gamma_2 \gamma_3 \rangle$. The black line indicates the analytic ladder approximation result for $\langle \gamma_1 \rangle$ from Eqs. (5.61) and (5.86), plotted up to the point it diverges. (b) The relative frequency of a given total amplitude enhancement for a different scatterer densities. The vertical dashed coloured lines in (a) indicate the densities plotted in (b). <i>Figure reprinted (with alteration) with permission from Ref. [287] © 2021 American Chemical Society.</i>	165
5.13	The mean amplitudes (a) $\langle \gamma_1 \rangle$ and (b) $\langle \gamma_1 \gamma_2 \gamma_3 \rangle$ for various different observation wavevectors $\mathbf{k}_{\parallel} = k_d(\sin \theta \cos \phi, \sin \theta \sin \phi)$. Note that $\gamma_{2,3}$ do not depend on observation position since they only deal with scattering onto the analyte particle.	167
5.14	The complex mean $\langle \gamma_1 \rangle$ as a function of n for different observation positions (a) $\text{Re}\langle \gamma_1 \rangle$ and (b) $\text{Im}\langle \gamma_1 \rangle$	168
5.15	Equivalent of Fig. 5.9 but with $\alpha = i\alpha_{g1}$	169
5.16	Equivalent of Fig. 5.12 showing but with $\alpha = i\alpha_{g1}$. Histograms are plotted at different densities to Fig. 5.12 since the transitions in distribution occur at different points. <i>Figure reprinted (with alteration) with permission from Ref. [287] © 2021 American Chemical Society.</i>	170
5.17	Equivalent of Fig. 5.10 showing the histograms of the individual enhancement factors on the complex plane but with $\alpha = i\alpha_{g1}$. Densities plotted are (left) $n\lambda_0^2 = 0.05$, (centre left) $n\lambda_0^2 = 0.67$, (centre right) $n\lambda_0^2 = 2.31$ and (right) $n\lambda_0^2 = 8$ which correspond to the same ones as in Fig. 5.10, with the addition of $n = n_{\text{opt,tot}} = 2.31\lambda_0^{-2}$ to show the optimum density distribution is analogous to that of Fig. 5.10.	171
5.18	Equivalent of Fig. 5.9 for the ‘high loss’ case of $\lambda_0 = 600\text{nm}$ and $\alpha = \alpha_{g2}$. <i>Figure reprinted with permission from Ref. [288] © 2021 American Physical Society.</i> .	172
5.19	Equivalent of Fig. 5.10 for the ‘high loss’ $\lambda_0 = 600\text{nm}$ case, with $\alpha = \alpha_{g2}$, showing the probability distribution of the enhancement factors over the complex plane. .	173
5.20	Equivalent to Fig. 5.9 for $\lambda_0 = 600\text{nm}$, $\alpha = -\alpha_{g2}$ and $\mathbf{k}_{\parallel} = -k'_{\text{SPP}}\hat{\mathbf{x}}$. <i>Figure reprinted with permission from Ref. [288] © 2021 American Physical Society.</i> .	175

5.21	The dependence of $\langle \gamma_1 \gamma_2 \gamma_3 \rangle_{\max}$ and $l_{\text{opt,tot}}$ on the phase of α (or equivalently μ) for (a) the high loss $\lambda_0 = 600\text{nm}$ case and (b) the low loss $\lambda_0 = 650\text{nm}$ case. The median value of $ \gamma_1 \gamma_2 \gamma_3 $ at $l_{\text{opt,tot}}$ is also shown. The vertical dashed line indicates the phase of α at which the divergence condition $\text{Im}[\alpha \tilde{G}(\mathbf{k}_{\parallel})] = 0$ from Eq. (5.60) is satisfied. Points for which $\langle \gamma_1 \gamma_2 \gamma_3 \rangle_{\max} \leq 1$ are not plotted but denoted by the shaded blue region, since this indicates no amount of multiple scattering improves sensitivity on average and single scattering (i.e. $n = 0$) provides the best mean sensitivity. <i>Figure reprinted with permission from Ref. [288] © 2021 American Physical Society.</i>	176
5.22	The Pearson correlation coefficients P_{ij} of the enhancement factor amplitudes in the low loss case for (a) $\alpha = \alpha_{g1}$ and (b) $\alpha = i\alpha_{g1}$	178
6.1	A simple example of a fully connected feed forward neural network, with each node in a layer connected to every node in the subsequent layer. The neural network shown consists of a single element input, a hidden layer of 4 nodes, a second hidden layer of 2 nodes and a single element output. Lines indicate the left node output is fed as input to the right node.	181
6.2	Schematic of the structure of the dense neural network used to localise the particle position from the leakage ring speckle pattern intensity profile. The number of nodes shown does not represent the true size of each layer (except for the final output layer).	189
6.3	Mean error as a function of training set size for (left) Gaussian random vector simulated speckle at different ratios of particle scattered intensity to background speckle intensity \bar{I}_a/\bar{I}_b and (right) coupled dipole simulated speckle at different scattering mean free paths l_s (or equivalently scatterer densities).	194
6.4	Mean error as a function of training region size L_x for (left) Gaussian random vector simulated speckle and (right) coupled dipole simulated speckle.	195
6.5	Mean error as a function of SNR_{tot} for (left) Gaussian random vector simulated speckle and (right) coupled dipole simulated speckle.	196
6.6	Mean error as a function of SNR for coupled dipole simulated speckle for different l_s and N_{train}	197
6.7	Mean error as a function of scattering mean free path for coupled dipole simulated speckle. The plots shown are for $N_{\text{train}} = 50,000$ (blue and green lines) and $N_{\text{train}} = 100,000$ (yellow and red lines), with corresponding mean separation of training points $\langle \Delta x \rangle$ shown.	198
6.8	Mean error as a function of analyte particle radius, for different training radii r_{tr} and mean free path l_s	199
A.1	Diagram showing the contour used to evaluate the integral in Eq. (A.1) in the limit $R \rightarrow \infty$ and the location of the poles of the integrand.	252

Abbreviations and Acronyms

CCD	Charge coupled device
CNN	Convolutional neural network
DLS	Dynamic light scattering
DWS	Diffusing wave spectroscopy
FET	Field effect transducer
FWHM	Full width at half maximum
iSCAT	Interferometric scattering microscopy
LDOS	Local density of states
LOD	Limit of detection
LSPR	Localised surface plasmon resonance
MINFLUX	Minimal photon fluxes
MLE	Maximum likelihood estimation
NA	Numerical Aperture
PALM	Photoactivated localisation microscopy
PC	Photonic crystal
PDF	Probability density function
PRAM	Photonic resonance absorbance microscopy
PSF	Point spread function
RIU	Refractive index units
RMS	Root mean square
RTE	Radiative transfer equation
SBR	Signal to background ratio
SGD	Stochastic gradient descent

SLM	Spatial light modulator
SNR	Signal to noise ratio
SPP	Surface plasmon polariton
SPR	Surface plasmon resonance
SPT	Single particle tracking
STED	Stimulated emission depletion
STORM	Stochastic optical reconstruction microscopy
TIR	Total internal reflection
WGM	Whispering gallery mode

Chapter 1

Introduction

1.1 Motivation

The ability to detect and track single nanoscale particles is an important tool within science, and in particular biological and medical sciences, allowing the investigation of processes at a single molecule scale. Live study of biological transport processes, such as the passage of proteins through pores or the motion of biological molecular motors such as myosin, elucidates greater understanding of the mechanisms, informing development of therapeutic responses to disease and diagnostic techniques. Detection of target analyte molecules can be used for diagnosis, while structural information about biological objects at ever smaller scales is similarly valuable.

Ever since early microscopes allowed visualisation of single cells, the use of light has been one of the primary methods to probe biological systems at the smallest scales. A range of microscopy techniques, such as bright field, dark field and phase contrast microscopy, enable images to be formed with a spatial resolution limited to the order of hundreds of nanometres owing to the diffraction limit. The rapidly developing fields of localisation microscopy and super-resolution microscopy extend optical microscopes beyond the diffraction limit, with localisation microscopy capable of extracting an object's position with precision on the order of $\sim 1\text{nm}$ [1] and super-resolution microscopy able to produce images with spatial resolution as low as 5nm [2]. Several factors have driven the advances in localisation microscopy, notably including the increasing availability of computing power enabling the rapid execution of image processing

localisation algorithms, including machine learning based algorithms [3, 4]. Similarly, improved tools to control and shape light such as spatial light modulators (SLMs) have allowed structured light to be used in localisation microscopy [5, 6]. Fluorescent microscopy is a prominent tool in enabling localisation and super-resolution microscopy [7]. The reason for this is that the weak interaction of small biological particles, such as single proteins or viruses, with light means the signal transduced in a microscope by a single particles is small and difficult to detect in the presence of noise. By attaching a fluorophore, which does interact strongly with the light, the signal is enhanced. The problem with this approach (and other labelling methods such as attaching metallic nanoparticles which interact strongly with light) is that the labelling of the analyte particle can affect its function and behaviour, posing a problem for studying the native biological function of the analyte particle [8, 9]. As such, label free approaches are desirable, and use alternative approaches such as interference of scattered light with a strong reference field to enhance the signal from nanoscale biological particles as is done in interferometric scattering microscopy (iSCAT) [10].

In addition to microscopy, light, in particular resonant optical modes, have found application in biosensing [11]. In this sensing modality, the binding of biological particles to receptors on the surface of optical cavities alters the parameters of the resonance, and from measuring this change, binding events can be deduced. Photonic resonance sensing techniques suffers from the same weak interaction with nanoscale biological particles as microscopy, with this limitation overcome by a combination of the confinement of the light mode and the long lifetime of a photon in the mode, both of which increase the sensitivity of the resonance to the presence of the biological particle.

An important class of electromagnetic modes in the context of biosensing are plasmonic modes, which arise when light interacts with metallic structures. Since they were first proposed in 1952 [12], these plasmonic modes, in which light couples to free electrons in the metal, have received much research interest [13, 14]. An important property of such modes is that they are strongly confined to metallic surfaces, whether this be a surface plasmon resonance (SPR) on a planar metal surface or a localised surface plasmon resonance (LSPR) on the surface of metallic nanoparticles. The strong confinement leads to extremely strong electric fields near the metallic

surface and thus strongly enhanced light matter interactions at the surface [15]. As a result, plasmonic modes have seen much use in biosensing [16, 17, 18], with the enhanced interaction overcoming the weak signal limitation mentioned previously. Indeed, SPR sensors have a long established, successful history and are widely commercially available [19]. The enhanced light matter interaction is not without a cost, as the high ohmic losses in metal broaden the resonance, reducing the sensitivity. As such, while LSPR sensors have achieved single particle sensitivity [20], the easily fabricated, commercially available SPR setup has not yet achieved this.

The motivation of this thesis is to theoretically develop and investigate a platform for label free single nanoscale particle sensing and tracking based on a simple experimental setup similar to SPR, to allow for low cost commercial availability. In general, matter can interact with light through two mechanisms: scattering and absorption. In this thesis, scattering is the interaction mechanism used to perform sensing, since the distribution of scattered light contains more information than simply an absorption event. Furthermore, the interference of the scattered light with a larger field can enhance the signal in the manner of iSCAT. Within a coherent detection scheme such as iSCAT, scattering from random disorder can have a significant effect, with disorder arising due to the complex biological media being used, fabrication imperfections in the optical elements or by deliberate design. The random scattering of light is an area of optical physics widely studied, in particular for biological applications. The interference of many randomly scattered light waves gives rise to a distinctive granular pattern of bright and dark spots depending on whether the scattered waves add up constructively or destructively. Statistical properties of speckle patterns and randomly scattered light have been studied and modelled with great success [21]. While appearing random, speckle patterns depend on the properties of the scattering medium and as such, techniques such as dynamic light scattering (DLS) and diffusing wave spectroscopy (DWS) are able to extract information about a sample based on properties of these speckles [22, 23]. The properties of randomly scattered light depend on which scattering processes are dominant, and in particular the degree of multiple scattering, in which light is scattered many times before propagating to a detector. In the single scattering regime, multiple scattering is by definition negligible and light is scattered independently by individual scattering centres. Conversely, multiply scattered light visits many different scatterers

and thus is not simply a coherent sum of the light scattered from each individual scatterer. As a result, multiple scattering gives rise to a rich range of physical phenomena such as localisation and long range correlations that have been studied in depth [24, 25, 26, 27, 28]. The presence of multiple scattering is often seen as a drawback owing to the fact that in the single scattering regime there exists a linear relationship between scattered field and object permittivity that is no longer present when multiple scattering is significant [29]. In certain contexts, however, multiple scattering can in fact provide beneficial features, for example it has been leveraged to overcome the diffraction limit [30] or to give rise to areas of high light intensity for increased interaction with matter [31]. Due to the significantly different properties of scattered light in the single scattering and multiple scattering regimes (and even within multiple scattering, the properties can vary with the degree of multiple scattering, such as between the diffusive and Anderson localised regime), it is important in any scattering system to understand the significance of multiple scattering and to find the optimal working regime for the particular application.

1.2 Thesis Structure

The approach investigated in this thesis, using random scattering of propagating surface plasmon modes, allows for the possibility of plasmonic light matter interaction enhancement, interference effects and multiple scattering, all of which may help to overcome the weak interaction of light with nanoscale biological analyte particles and achieve sensing and tracking, while maintaining an experimental setup similar to the easily fabricated SPR sensor. To provide context and motivation for the work, Chapter 2 reviews previous literature on biosensing, with particular focus on single particle sensitivity, and tracking or localisation of single biological molecules. Common metrics assessing the performance of these methods will be introduced and the relative merits and drawbacks of current techniques discussed. Chapter 3 introduces the proposed sensing setup, motivated by the context provided in Chapter 2, and analyses the sensitivity and detection limits of such a sensor within the single scattering regime. Following this study into the sensing capabilities of the proposed system, Chapter 4 describes how the setup can also

allow for tracking, derives an algorithm, valid in the single scattering regime, to track a single biological molecule based only on light intensity measurements and assesses the algorithm's performance on simulated data. With both the potential sensing and tracking capabilities of the proposed setup investigated in the single scattering regimes, Chapter 5 explores the role of multiple scattering in the setup, and in particular its effect on the sensitivity. Once the effects of multiple scattering on the proposed setup are understood, Chapter 6 extends the tracking capability into the multiple scattering regime via the use of a neural network based localisation algorithm and investigates its performance on simulated data. Finally, Chapter 7 summarises the key findings of this thesis, and briefly discusses possible future work.

Chapter 2

Biosensing and Tracking

The use of physical devices to extract information from biological systems has been ubiquitous in the biological sciences since their inception. Optical microscopy, for example, has played a fundamental role in many biological discoveries from the discovery of the cell to the modern day nanometre scale studies enabled by super resolution microscopy and fluorescent microscopy methods. Developments in photonics and plasmonics have provided a wide range of biosensing technology that utilises light, including systems such as SPR and optical microcavity sensors. Optical technology, both in the form of microscopy and biosensing, has enabled great advances in fundamental biological studies such as improved understanding the movement of proteins on or through membranes, and also in the development of medical technology and techniques. The aim of this chapter is to review the range of methods used in modern biology to sense and track biological objects such as proteins or viruses which have a length scale in the tens of nanometres or smaller and to quantify the sensitivity of these methods. Primarily, the focus will be on methods capable of detecting single nanometre scale biological objects, and also single particle tracking (SPT) or localisation. From this review, the limitations of current methods will become clear, motivating the introduction of new approaches to biosensing and tracking to overcome the limitations, such as the one that will be studied in this thesis. Furthermore, the review of the performance of current methods will provide a context for assessing the proposed biosensing setup to be introduced in Chapter 3.

2.1 Evaluating Biosensor Performance

A biosensor is a device in which biological activity transduces a measurable physical signal [32, 33], for example a change in intensity of light on a camera or electrical current. One must consider several features of any biosensor in quantifying its performance, and it is useful to define some performance metrics to allow comparison of different biosensors. The performance metrics will be defined in a general biosensing context, but also discussed more specifically with regards to how they apply to sensing analyte particles. Firstly, the sensitivity \mathcal{S} quantifies the size of the change in the measured quantity induced by the biological activity one is trying to measure [33]. The sensitivity is defined as

$$\mathcal{S} = \left| \frac{\partial \mathcal{M}}{\partial X} \right|, \quad (2.1)$$

where \mathcal{M} is the measured quantity (e.g. intensity of light) and X is the quantity one is interested in detecting. For example, in particle sensing, X could be the concentration of analyte particles, or for a refractometer, X would be a refractive index. Another important quantity is the limit of detection (LOD), which is the minimum change to \mathcal{M} detectable above the noise in the system, given by [34, 35]

$$\text{LOD} = m\sigma_{\mathcal{M}}, \quad (2.2)$$

where $\sigma_{\mathcal{M}}$ is the standard deviation of the (zero-mean) noise in \mathcal{M} , while m is a constant of order unity that depends on the confidence with which one wants to assign a measurement to a change in X as opposed to noise, and also the probability distribution of the noise. For example, for Gaussian distributed noise at 95% confidence, $m \approx 1.645$ [34], while $m = 3$ is also a common choice [35]. Often, limit of detection is used to refer to the value of X which perturbs \mathcal{M} by an amount equal to the limit of detection. This quantity X_{LOD} is defined by

$$|\mathcal{M}|_{X=X_{\text{LOD}}} - B| = \text{LOD}, \quad (2.3)$$

where $B = \mathcal{M}|_{X=X_0}$ is the background measurement when $X = X_0$, an initial reference value. For example, if X is the concentration of a particular analyte particle, X_0 is the concentration in a blank measurement, often $X_0 = 0$, i.e. in the absence of any analyte particles. It is important to distinguish between the two quantities, as LOD describes a property of the measurement process only, while X_{LOD} additionally depends on the properties of X . The time resolution τ_{res} of a sensor is also an important factor to consider, defined as the time taken to make a measurement of \mathcal{M} . In general, a smaller τ_{res} is better, but often reducing τ_{res} comes at a cost, such as increasing noise $\sigma_{\mathcal{M}}$ due to reduced integration time, while technological constraints also limit the time resolution. It is sufficient, however, to achieve $\tau_{\text{res}} < \tau_X$, where τ_X is the timescale over which X varies. Time resolution beyond this does not provide any additional benefits since the consecutive measurements will be the approximately the same until a time $\sim \tau_X$ has passed. The sensing volume V_{sens} is the volume within which changes to X contribute to perturbing \mathcal{M} . For particle sensing, larger sensing volumes will allow a greater number of particles to affect \mathcal{M} for a given concentration of analyte particles, but additionally there will be more non-specific sensing events, i.e. particles different to the analyte particle which can also change the value of \mathcal{M} . Conversely, a small sensing volume, while reducing the frequency of analyte particles coming into V_{sens} to transduce a change in \mathcal{M} (which can be especially problematic at low concentrations), can more tightly localise the sensing events and allow for compact multiplexed arrays of sensors.

So far, the biosensor performance metrics defined above have been defined in terms of the measured quantity \mathcal{M} . As seen in the definition of X_{LOD} , this measured quantity can contain contributions arising from the background, which contains no information about the quantity of interest X . The information carrying part of the measured quantity can be defined as the signal S , which for small changes in X is given by $S = |(\partial\mathcal{M}/\partial X)X|$. Signal is often used colloquially to refer to the measured quantity \mathcal{M} , however it is important to distinguish S and \mathcal{M} here, in order to fairly compare sensing (or tracking/localisation) methods with differing background levels. When \mathcal{M} is linear in X , the signal is just the background subtracted measurement $S = \mathcal{M} - B$. Two commonly used metrics of the signal, relevant to biosensing, but also to signals measured in SPT/localisation, which should be introduced are the signal to noise ratio

(SNR), $\text{SNR} = S/\sigma_{\mathcal{M}}$, and the signal to background ratio (SBR), defined as $\text{SBR} = S/B$. The SNR is closely related to the LOD, which corresponds to the case when $\text{SNR} = m$ if \mathcal{M} is linear in X , quantifying how easily the signal arising due to X can be distinguished from the system noise. Whereas the LOD is relevant in evaluating the ability to make a binary choice of whether the change in S arises from an actual sensing event or noise, in cases where the measured signal is processed in order to extract further information (e.g. SPT), the SNR is more relevant in quantifying the extent to which noise will degrade the information extracted from the signal. The SBR similarly quantifies the ability to distinguish the perturbation to S from the background, which is valuable in assessing the required dynamic range of any measurement device involved.

The metrics introduced allow for quantitative comparison of the signals transduced within different biosensors. Another important feature to consider in any biosensor is the specificity, meaning the extent to which the signal is transduced solely by the analyte particle of interest, as opposed to other objects present in the sample. A real biological sample will typically contain many different chemical species and molecules, so it is therefore important for a biosensor to have high specificity for the exact analyte species one is aiming to study. Many biosensors achieve both specificity and sufficient signal level through labelling, in which a label particle, for example a fluorophore, quantum dot or metallic nanoparticle, is attached to the analyte particle [36, 37, 38]. The labelled particle transduces the signal, and thus by choice of label, it can be ensured that the SNR and SBR are sufficiently large to avoid signal corruption and give a lower X_{LOD} , regardless of the nature of the native analyte particle. Furthermore, the label provides specificity, since one only labels the analyte species, which can be achieved through using specific chemical properties of the analyte or genetically encoding label particles [8, 39, 40]. Labelling does have significant downsides, in that the label can modify the nature of the analyte particle, possibly altering its biological behaviour [8, 9, 41]. In addition, it may not always be possible or practical to label the analyte particle, for example if the analyte particle cannot be treated before sensing.

The performance and limitations of current biosensing technology will now be considered, with particular attention to the metrics introduced, along with specificity and labelling.

2.1.1 Nanopore and Nanowire Sensors

Nanopore sensors consist of a small (nanometre scale diameter) pore embedded in a surface, either as a protein pore in a membrane [42], or an artificial pore in an insulating surface [43]. A voltage across the membrane drives a current through the pore, which can be measured. The presence of an analyte particle within the pore increases the electrical resistance in the pore, reducing the current through the pore. This current drop is the signal transduced by the analyte. Specificity is provided by the pore design, with engineering of the pore structure (in both the chemical components and physical shape) encouraging the transit of the particular analyte particle through the pore [44]. As the pore design is optimised, both geometrically and chemically, for the native analyte particle, there is no need for labelling. Nanopore technology has found widespread use in DNA sequencing [45, 46]. Different DNA bases have different amplitude current dips and dwell times in the pore (corresponding to the duration of the measured current dip), meaning that as a DNA strand passes through the pore, the pattern of current decreases and dwell times can be converted into a base sequence [47]. In addition, it has found use in detecting the presence of a wide range of biomolecules and toxins [48, 49].

Nanopore sensing provides a label-free, highly specific biosensing technique capable of detecting single nanometre scale particles, but the measured current signal does provide limitations. Specifically, noise in the current arising from nanobubbles in the pore and Johnson noise [50, 51] can reduce the ability to resolve the current dip associated with a translocation of the analyte through the pore. The SNR improves with decreasing pore diameter, with a 3nm diameter pore giving $\text{SNR} \sim 50$ for translocation of a 16.5 μm long DNA strand, whereas a 50nm pore has an SNR below 20, with the exact value dependent on the salt concentration [51]. Furthermore, for DNA bases, the rapid translocation means events can be missed altogether owing to insufficient temporal resolution [52].

Another type of biosensor that follows a similar physical principle to nanopore sensors is the nanowire field effect transistor (FET) sensor [53, 54]. This consists of a wire of nanometre scale diameter, with biological receptors on the surface. The binding of the analyte particle to the receptor changes the conductance of the wire. Measuring the conductance as a function of time, one can deduce binding events from step changes (either increases or decreases) in

the conductance [55]. The receptors bound to the nanowire provide the specificity as they are chosen to have a high affinity for the analyte particle. This also precludes the need for labelling. For DNA strand detection, $\text{SNR} \gtrsim 20$ is typical, comparable to nanopore sensors [56].

A single nanowire or nanopore provides only a very small sensing volume (only detecting analytes on the surface of the nanowire or area of the nanopore, a nanometre scale region). As such, they are unable to provide spatial information, while at low analyte concentrations there is a low frequency of binding/membrane translocation events, increasing the time required to get an event and the probability of missing the presence of the analyte particle altogether. This problem is overcome by using arrays of nanopores/wires [57, 58]. The spatial resolution of a multiplexed array is limited to the size of a single nanopore/wire and the density at which they can be packed.

2.1.2 Resonance Sensors

A wide range of biosensors can be classified as resonance sensors. These share the same underlying physical principle in their operation. The sensor exhibits a resonant response at a particular input parameter (such as frequency of an external driving force) which is sensitive to the presence of the analyte particle. The perturbation caused by the analyte particle changes the physical parameters of the resonance. From measuring the response of the oscillator over time, one can infer the presence of the analyte particle when this response changes.

In general, an oscillator with natural frequency ω_0 loses energy due to damping, so that the energy of the oscillator at time t , assuming no energy is put in, is $U(t) = U_0 \exp(-\omega_0 t/Q)$, where Q is the quality factor (Q-factor), defined as the number of complete oscillations for the energy in the oscillator to decay to $1/e$ of its initial value, U_0 [11]. The energy of the oscillator contained at a frequency ω is $\tilde{U}(\omega)$, the Fourier transform of this exponential decay. The frequency response of an oscillator is therefore Lorentzian, centred on the natural frequency ω_0 , which is expressed

$$\tilde{U}(\omega) = \frac{U_0}{(\omega - \omega_0)^2 + (\omega_0/2Q)^2}. \quad (2.4)$$

As a result, when measuring the frequency response of a resonator, one sees a peak centred on ω_0 ,

but with a finite width, specifically a full width at half maximum (FWHM) of $\Delta\omega = \omega_0/Q$. Note that $\tilde{U}(\omega)$ represents the energy in the resonator at frequency ω , while an actual experiment will often measure energy (or some related quantity) transmitted away from the resonator, which takes the form $U_{bg} - \tilde{U}(\omega)$ for some background energy U_{bg} that would be measured in the absence of the resonator, but is absorbed into the resonator in its presence. The measured quantity would then take the form $U_{bg} - \tilde{U}(\omega)$ which gives an inverted Lorentzian. In order to detect a change in ω_0 , the change must be large enough that it can be resolved from the finite linewidth. Shifts much smaller than $\Delta\omega$ cannot be resolved, with noise levels determining how much smaller than $\Delta\omega$ one can detect [11]. For example, in optical systems, measurements of power are shot noise limited (arising due to the arrival statistics of quantised photons) at best, in which case the minimum detectable shift in resonance frequency, $\delta\omega_{min}$, is given by [59, 60]

$$\frac{\delta\omega_{min}}{\omega_0} = \frac{1}{Q} \sqrt{\frac{\hbar\omega_0}{P_0\nu\beta\tau}} \quad (2.5)$$

where P_0 is the power coupled into the resonator, τ is the integration time of the light detector, β is the cavity transmission efficiency and ν is the quantum efficiency of the detector. While optical resonances often have other limiting noise sources to shot noise such as sample temperature variation, and non-optical resonances will have different noise sources altogether, the principle remains the same, the minimum detectable shift is determined by Q and the properties of the signal noise [61].

The spectral response of oscillators mean there are two approaches in choosing and optimising the resonance used for biosensing. Firstly, one can maximise the shift caused by interaction with the analyte particle, ensuring the large shift is resolvable. The other approach is to reduce the width of the resonance by using high Q resonances, so that even small shifts to the resonance frequency are detectable. In ideal circumstances, one would optimise both factors, but often the two elements are in competition. Indeed, increasing the interaction between the analyte particle and resonator will generally tend to increase the interaction of the oscillator with the environment in general, leading to increased losses. Furthermore, non-specific interactions, i.e. with particles other than the analyte, may also be increased. As such, the design of resonance

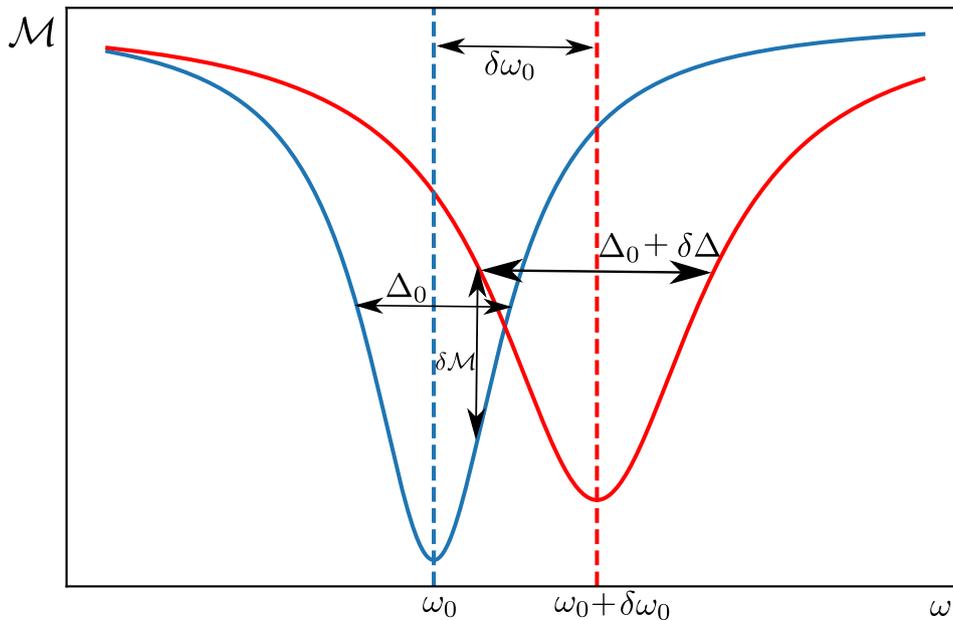


Figure 2.1: The measured response signal \mathcal{M} as a function of driving frequency before (blue) and after (red) binding of the analyte particle(s). The change in resonance frequency ($\delta\omega_0$), resonance width ($\delta\Delta$) or the change in response signal at a single frequency (δS) can all be used as a sensing signal, or a combination of all of them. While shown as a function of frequency ω , the same principle can be applied to \mathcal{M} as a function of other physical parameters (such as angle in the case of SPR).

based biosensors is generally a balance between these two factors. While using the shift in ω_0 is used as an example perturbation through which biosensing can be deduced, other changes to the resonators response, such as broadening of the linewidth, ring down time over which the energy in the oscillator decays or mode splitting in which degenerate modes of the oscillator split into modes at two different resonant frequencies, can also function as an indication of a binding event [60, 62, 63, 64, 65]. Fig. 2.1 shows a schematic example of some of these changing resonance properties, with an inverted Lorentzian lineshape. Regardless of the resonance property being monitored, the same considerations of balancing Q factor and size of the change still apply. The resonator response has been discussed here as a function of frequency, and thus to measure changes in this entire response curve, one would have to measure the response of the oscillator at a range of different frequencies. Some resonance sensors do take this approach, sweeping through frequency of the external force on the oscillator to measure a response curve [66]. Alternatively, the change to a measurement at a single frequency in the response curve could also serve as a sensing signal [20, 67]. Commonly, this measurement frequency will be chosen as $\omega_0 \pm \Delta\omega/2$, as this is where the gradient of a Lorentzian curve is steepest and therefore gives the biggest

signal change for a given shift [68, 69]. While a single measurement is less reliable, being more susceptible to noise, the reduction in number of measurements can provide advantages such as improved time resolution. Finally, it should be noted that a resonant response can also be observed as a function of a parameter other than frequency, for example the angle of incidence of light on an optical resonator [66, 70]. The same principles discussed apply regardless of the parameter, with substitution of the appropriate parameter for ω .

The optimal operating modality for a resonance sensor depends on the exact nature of the resonance being used, and so a few common resonance sensing techniques will be reviewed here.

Mechanical Resonances

Mechanical oscillators have been designed for biosensing applications. In general, they are equivalent to a ‘mass on a spring’ type system with a frequency $\omega_0 = (k/m)^{1/2}$, with k being the systems effective spring constant and m the mass of the oscillator. The presence of the analyte particle changes the mass by Δm , and thus the frequency shifts to $\omega'_0 = (k/m + \Delta m)^{1/2}$ [71]. Carefully designed mechanical oscillators are capable of sensing extremely small mass changes, for example the detection of mercury vapour adsorbed on an oscillator [72]. Achieving such sensitivity in biological samples is significantly more challenging owing to the oscillator being in a fluid with viscous losses, but such biosensing mechanical resonators have been designed [73], capable of measuring the mass of single bacterial cells [71] or viruses [74] and micron scale particles [75]. A major challenge with mechanical resonators is the conversion of the oscillator motion into a physically measurable signal [73, 76]. In comparison, optical signals are relatively simply to measure and thus optical resonances are more common within resonance based biosensing.

Surface Plasmon Resonance

The SPR is an electromagnetic resonance exhibited by metal-dielectric interfaces. The interface supports a propagating mode confined close to the interface called a surface plasmon polariton (SPP) (see Chapter 3 for detailed description of the physics of SPPs). This SPP has a well defined wavevector, k_{SPP} , which depends on the refractive indices of the metal and dielectric

at the interface. In order to couple light into the SPP mode, the transverse component of the wavevector of the incident light must match k_{SPP} , meaning that one sees a large decrease in the reflected light intensity when light is incident at the resonant angle Θ_{SPP} , as the light couples into the SPP mode instead of being reflected. The confinement of the SPP leads to high intensity electric fields near the surface, resulting in strong interaction between the SPP mode and matter near the surface [13, 14]. As a result, k_{SPP} is highly sensitive to the local environment at the surface. It is this principle upon which SPR sensing is based.

The binding of analyte particles to biological receptors on the metal surface shifts k_{SPP} due to the changes in the local refractive index near the surface [19]. The shift in k_{SPP} can be deduced by a range of different measurement schemes. In angular modulation, the reflected intensity is measured for a range of different incident angles. When plotted as a function of angle, the intensity shows a dip at Θ_{SPP} , as in figure 2.1, but as a function of incident angle as opposed to ω . Binding of analyte particles shifts the angle at which the dip occurs. The size of the shift in angle can be converted into a refractive index shift, which allows one to deduce the amount of analyte binding. When measured over time, this information can be translated into kinetic binding rate data [19]. Angular modulation is analogous to measuring $\delta\omega_0$ shown in Fig. 2.1. Intensity modulation is an alternative measurement scheme, analogous to the δS measurement from Fig. 2.1, in which the incident angle is fixed, but the shift in k_{SPP} means the reflected intensity for that incident angle changes, as the resonance reflectivity dip has moved relative to the fixed incident angle [19]. These two methods are examples of the difference between measuring a full response curve (in this case as a function of angle) as opposed to monitoring a single point on the response curve in the case of intensity modulation. There are several other measurement schemes, such as wavelength modulation, polarisation modulation and phase modulation [19, 77], but ultimately all methods monitor a signal that arises due to the change in k_{SPP} .

For an SPR sensor, X in Eq. (2.1) is Δn , in refractive index at the surface caused by binding of analyte particles [78]. The quantity corresponding to \mathcal{M} depends on the modulation used. For example, in angular modulation, \mathcal{M} is Θ_{SPP} , while, in intensity modulation, \mathcal{M} is the intensity of reflected light measured for the fixed incident angle. For a single SPR sensing

channel, angular modulation offers improved LOD owing to multiple measurements being taken, reducing the noise [16, 35], however, the intensity modulation set up allows for SPR microscopy (also called SPR imaging) [79], which provides spatial information on the binding of the analyte particle. By measuring the intensity from an array of many intensity modulation SPR sensors on a surface, one can ascribe different shifts in k_{SPP} , corresponding to different mass of the analyte binding to the surface, to each point on the surface, giving spatial information about the analyte particle binding [80]. The spatial resolution of such sensors is on the micrometre scale.

As SPR sensors ultimately rely on the measurement of light intensity (whether this be for a single incident angle as in intensity modulation, or multiple intensity measurements as in angular or wavelength modulation), there will always be shot noise due to the discrete arrival of photons at the detector, though other noise sources such as thermal variations in refractive index can also affect the SNR. It has been suggested that current SPR sensing technology based on planar metal-dielectric interfaces is approaching the shot noise limit for LOD [78], corresponding to $\sim 10^{-6}$ refractive index units (RIU).

Due to their relatively simple experimental implementation and their capability of label-free sensing, SPR sensors are widely used in studying various biological processes including studies of binding kinetics [81, 82, 83] and detection of toxins [84]. SPR sensors based on planar interfaces, however, are only capable of sensing the analyte particle in bulk, i.e. they can only detect the effect of many binding events. This is because high ohmic losses in the metal, required to form an SPR, result in a relatively low Q factor, and therefore very small changes associated with the binding of a single particle cannot be resolved. As such, they are not appropriate for studies requiring single particle sensitivity, as may be required for example at low analyte concentration.

Localised Surface Plasmon Resonance

An LSPR is a resonance that is closely related to the SPR. Rather than being propagating modes confined to a metallic interface, LSPR modes are localised on the surface of metallic particles [13, 14]. The frequency at which such a localised mode is supported is determined by the optical properties of the metallic particle and its surrounding medium and the particle geometry [14]. For example, the classical expression for the (quasistatic) polarisability of a

sphere of radius R_a and dielectric constant ε_p in a medium of dielectric constant ε_d is [85]

$$\alpha = \varepsilon_d \varepsilon_0 4\pi R_a^3 \left(\frac{\varepsilon_p - \varepsilon_d}{\varepsilon_p + 2\varepsilon_d} \right), \quad (2.6)$$

which gives a resonance condition at a frequency where $|\varepsilon_p + 2\varepsilon_d|$ is minimised, corresponding to the Fröhlich condition $\text{Re}(\varepsilon_p + 2\varepsilon_d) = 0$ provided losses are small [86]. This condition requires the real parts of ε_d and ε_p have opposite signs, and thus can be satisfied for metallic particles embedded in a dielectric medium. Ellipsoidal nanoparticles give alternative resonance conditions [87], while a more accurate resonant condition for the spherical particle can be achieved using Mie theory [86, 88]. In a sensing context, a range of different geometries of metallic nanoparticles, including spheres, nanorods, pyramids and cubes are used [20, 89, 90, 91, 92], but all exhibit a resonant response, and the operating principle is the same for all geometries. Binding of analyte particles to receptors on the surface change the local dielectric constant of the surrounding medium, shifting the resonance frequency. Compared to SPR, the resonant mode in LSPR is confined to a much smaller region, decaying over a length scale of the order $\sim 10\text{nm}$ from the metal surface compared to $\sim 100\text{nm}$ for SPR sensors, with the surface of the nanoparticle also being significantly smaller (e.g. surface of $\sim 50\text{nm}$ sphere) than the area of a SPR mode. As a result, there is a stronger interaction between the analyte particle and an LSPR mode.

Sensing is achieved via modulating the frequency of light incident upon the metallic nanoparticle or an ensemble of nanoparticles. One observes a dip in transmitted power at the resonance frequency, when the incident optical power couples strongly into the localised mode and thus very little is transmitted, with a plot of transmitted power against frequency exhibiting a resonant lineshape. Binding of analyte particles to the receptors can then be deduced by observing a change in the frequency of the dip, with the frequency shift related to a refractive index change or the number of bound analyte particles either through calibration assuming a linear relationship or more detailed electromagnetic modelling of the perturbation to the resonant mode [17, 93]. Intensity modulation based sensors are also used, in which the incident frequency is fixed and changes in intensity are used as a signal [20, 94]. Unlike SPR, LSPR sensors are capable of single particle sensitivity, detecting single proteins [20]. Furthermore, the small size

of a single LSPR nanoparticle sensor allows multiplexing of many sensors in a small region [18]. Metallic nanoparticles can themselves be used as label via attaching them to analyte particles, their strong optical response near the LSPR frequency making them easily detectable [37, 95, 96].

The large number of degrees of freedom controlling LSPRs (optical and geometric properties) mean there is a wealth of research into engineering them for optimised biosensing [18], including use of chiral plasmonic structures [97] for additional specificity. Furthermore, strong plasmonic confinement enables the use of surface enhanced Raman spectroscopy (SERS), an inelastic light scattering process which can allow for particle identification [98, 99, 100]. Particle identification via SERS can alleviate the problem of non-specific interactions in LSPRs.

While the strong mode confinement allows for increased sensitivity, down to the single particle level, the use of precise geometries and nanoscale objects make fabrication more difficult and expensive as compared to SPR. Additionally, the requirement to scan the frequency typically limits time resolution to the millisecond timescale [101, 102], however recent work with an intensity modulation detection scheme has pushed LSPR temporal resolution to below a microsecond ($\sim 10\text{ns}$) [94].

Nanostructured Plasmonic Sensors

Beyond sensors based on simple structures of single plasmonic elements (e.g. a single metal interface or nanoparticle), a range of nanostructured plasmonic sensors, consisting of multiple plasmonic elements arranged together so as to couple their modes, have also been developed [103]. Developments in techniques such as lithography allow fabrication with fine scale control, allowing these nanostructures to be produced with nanometre scale accuracy [104, 105]. Such plasmonic nanostructures can couple different plasmonic modes by clustering metallic nanoparticles in close proximity [106], or metallic nanoparticles or nanoislands on a surface [107, 108, 109]. Another approach to develop nanostructured plasmonic sensors is to insert nanoholes into a metallic surface supporting plasmonic modes [110, 111]. Regular arrays of the individual plasmonic components, in both 2 and 3 dimensions, referred to as plasmonic crystals, have been used for sensing [112, 113, 114, 115, 111], as have random arrays of nanoholes [116, 117]. A major

benefit of such structures is the additional parameters associated with the structure (e.g. lattice spacing) allow for tuning of the resonance properties to optimise sensing, in addition to the tuning of the parameters of the individual components of the array (e.g. nanoparticle size and shape or nanohole depth). Furthermore, such structures can support multiple resonances. By monitoring the full spectral range including multiple resonances, up to 10-fold increase in SNR compared to a single wavelength has been observed [113]. Plasmonic crystals have been shown to achieve sensitivities comparable to conventional SPR, but allow for simpler optical setups and easier coupling to optical fibres [103, 110, 114]. While nanostructured plasmonic sensors can be considered a form of SPR (or LSPR) sensor, the wide range of possible structures, with a much higher dimension parameter space to explore compared to single element plasmonic sensors, means that such structures maintain much research interest, though as yet they have not pushed beyond a single LSPR nanoparticle in terms of single particle sensitivity. Such structures will be relevant to Chapter 5.

High Q Optical Modes

Plasmonic sensors take the approach of improving sensitivity via increasing the strength of the interaction between the resonant system and the analyte particle, achieved via strongly confined electric fields from metals. The use of high Q-factor optical resonators is an alternative approach, in which the focus is placed on designing a high Q resonance in order to be able to measure very small shifts [11]. Whispering gallery modes (WGMs) [60] are one such example, which consist of a dielectric cavity with a closed concave interface. A simple example is a dielectric sphere, though a range of other geometries exist such as rings, discs, toroids and microbubbles [11, 118, 119, 120, 121, 122, 123]. Such cavity resonances can achieve very high Q-factors, up to the order of 10^{12} [124, 125], allowing for sensitive detection. The narrow line widths associated with such high Q-factors mean whispering gallery mode sensors have achieved single particle sensitivity to individual proteins or DNA strands [11, 126, 127, 128]. Sensing is achieved analogously to LSPR sensing, via varying the wavelength of light incident upon the cavity. As with LSPR, this scanning of wavelength limits time resolution, though the narrow linewidth allows faster scanning than LSPR, with $\tau_{\text{res}} \sim 1\mu\text{s}$ typically achievable [65, 129]. The

transmitted power exhibits a resonant lineshape dip at the resonance frequency due to power coupling into the resonant mode. Binding of analyte particles to the cavity can then be deduced by observing a change in the transmitted power vs. frequency plots. In addition to using shifts in the resonance frequency [130], changes to the linewidth [63] can also be used to deduce binding. Restricting detection to just a single feature of a single resonator's frequency response is not required, and in fact using the full spectral response of a system consisting of multiple coupled resonators within so called multimodal sensing allows for greater sensitivity [60]. One such example involves the monitoring the frequency response of two coupled WGMs. The presence of the analyte particle breaks the degeneracy of the modes, causing the frequency response to split into two modes at distinct frequencies. The splitting of the resonance dips into two distinct dips thus indicated the presence of the analyte particle, while the frequency difference between the two split modes can give quantitative information on the number and size of the analyte particles [64].

In addition to the range of WGM geometries, other photonic resonances are used for sensing. For example, photonic crystals (PCs), consisting of periodic structures of dielectric objects, can be designed to have a resonant response which have also been used for biosensing purposes [131, 132]. The degrees of freedom provided by the choice of lattice structure allow for a degree of control over the resonance and therefore optimisation for sensing. Much like in SPR, PC modes can be strongly confined to a surface and give strong light-matter interaction, helping improve sensitivity [133, 134].

In general, photonic modes are far less localised than plasmonic modes, and as such interact less with analyte particles. Hybrid plasmonic-photonic mode sensors aim to combine the narrow linewidth of photonic resonances with the strong light-matter interactions of plasmonic modes to achieve large perturbations to narrow resonances [127, 135]. They operate on the same principle as LSPR or WGM sensors, but combine both plasmonic and photonic resonators, and are capable of extremely sensitive detection, to the level of single DNA strands of molecular mass $\sim 2\text{kDa}$ [127]. Similarly, coupling of localised plasmonic and PC modes is another strategy to achieve high sensitivity [134, 136]. One approach based on coupling these resonances is photonic resonance absorbance microscopy (PRAM), functionalisation of both the plasmonic

nanoparticles and the surface of the PC means that the binding of analyte particle to the plasmonic particle ‘activates’ it to be able to bind to the PC surface. Tuning the PC resonance to coincide with plasmonic resonance of the nanoparticles, the nanoparticles on the PC surface absorb strongly, allowing them to be seen when imaging the PC surface. Thus individual analyte particles can be counted as dark spots in the image of the PC surface. Here, the resonance response provides a large absorbance signal, but the change transduced by the analyte particle is not in the resonance properties, but rather in the binding properties of the functionalised plasmonic nanoparticles [136].

2.1.3 Label Free Single Molecule Detection

Table 2.1 summarises the current sensor performance for the various label free biosensors discussed so far. It can be seen that photonic and localised plasmonic resonances, along with nanowire and nanopore technology, are currently the most suitable techniques for sensing at a single particle level, but as yet only nanopore has joined SPR in becoming widely commercially available. Nanomechanical resonances, while capable of single molecule detection, have even more precise fabrication requirements while not gaining much in terms of either temporal resolution or sensor volume in comparison to the optical resonances or other methods capable of single protein sensitivity. Detection of single, nanoscale molecules remains beyond conventional SPR, and thus, despite the widespread use commercially, the alternative optical resonances have received more research interest in the realm of single molecule sensing.

2.2 Single Particle Tracking and Localisation

The ability to track a single biological particle, reconstructing its trajectory from a physical signal, or localise its position is a valuable tool in studying biological processes, for example studying membrane transport, interaction between proteins and receptors or biological machines and motors [157, 158, 159]. Biosensing approaches using discrete changes in a signal arising from the analyte particle entering a sensing volume or binding to a receptor are not appropriate for tracking since they do not provide a signal that is a function of the particle position. Multiplexed

Table 2.1: Summary of performance of common label free biosensing methods. The parameter X_{LOD} refers to either the limit of detection in analyte concentration, or analyte particle size for systems with single molecule sensitivity. Sensing volume is expressed as a distance from sensor surface multiplied by sensor area, * indicates analyte must be tightly attached to surface, V_{sens} refers to a single sensing element, multiplexed arrays can increase the effective sensing volume.

Method	X_{LOD}	τ_{res}	V_{sens}	Commercially available
SPR	144 virus particles/mL [137, 138]	$\sim 10\text{ns}$ [139, 140]	$\sim 100\text{nm} \times 25\mu\text{m}^2$	Yes [19]
LSPR	1 molecule $\sim 50\text{kDa}$ protein [20] $\sim 10\text{kDa}$ protein [141]	$\sim 10\text{ns}$ [94]	$\sim 10\text{nm} \times 100\text{nm}^2$	No
Photonic Resonance	1 molecule $\sim 2\text{kDa}$ DNA strand [127] $\sim 10\text{kDa}$ DNA strand [136]	$\sim 1\mu\text{s}$ [65, 129]	$\sim 200\text{nm} \times 2000\mu\text{m}^2$ (WGM) [142, 143] $\sim 200\text{nm} \times 10000\text{nm}^2$ (PC) [144, 145]	No
Mechanical Resonance	1 molecule 0.13kDa C_{10}H_8 [146] $\sim 0.5\text{kDa}$ DNA base [147]	$\sim 10\text{ms}$ [146]	$\sim 500\mu\text{m}^2$ * (cantilever) [148] $\sim 500\text{nm}^2$ * (nanotube) [146]	No
Nanowire	1 molecule $\sim 0.5\text{kDa}$ DNA base [149, 150]	$\sim 1\text{ms}$ [151, 152]	$\sim 1000\text{nm}^2$ * [153, 154]	No
Nanopore	1 molecule $\sim 1\text{kDa}$ DNA strand [47]	$\sim 0.01\text{ms}$ [47]	$\sim 4\mu\text{m} \times 100\text{nm}^2$ [155]	Yes [156]

arrays of biosensors, such as used in surface plasmon microscopy, provide a degree of spatial information on binding, but in addition to often lacking the resolution and sensitivity to track a single analyte particle, the binding of the analyte particle to the sensor stops the analyte particle motion. Single particle tracking (SPT) and localisation requires a different, though closely related, approach to biosensing. The main difference is that the physical signal transduced by the analyte particle must vary as a function of particle position in order to allow positional information to be extracted. In assessing the performance of an SPT or localisation technique, many of the same considerations as in biosensing remain. In particular, the SNR and SBR of the signal, specificity (i.e. is the method tracking the correct analyte particle) and labelling requirements are all important criteria. In addition, due to the extra information required compared to biosensing, one must assess a method's spatial and temporal resolution. Spatial resolution refers to how precisely the method can reconstruct a particle's trajectory or localise its position, while temporal resolution, as in the case of sensors, refers to the minimum time between measurements of the analyte position. The exact biological application will determine the requirements on these two resolutions, and there is often a trade-off between them. For example, DNA strands can move through a $\sim 1\text{nm}$ wide membrane pore in $\sim 1\text{ms}$ [160], whereas viral RNA translocation of a cell membrane (on the order of nanometre thickness) takes ~ 3 minutes [161].

2.2.1 Imaging and Microscopy

The most common class of SPT and localisation techniques rely on imaging [162]. A series of images of the analyte particle are taken and then image analysis is used to extract the particle trajectory. The accuracy of these methods is determined by both the imaging method and image analysis algorithm. In the search to optimise both components, a range of approaches and techniques have been used.

Classical optical microscopy and conventional imaging techniques have a long established track record in studying biological systems, but ultimately are limited by diffraction [163]. The spatial resolution is diffraction limited to on the order of $\lambda/2\text{NA}$ where $\text{NA} = n \sin \theta$ is the numerical aperture of the imaging system, n is the refractive index of the imaging medium, θ is the maximum half angle at which light from the object plane is collected and λ is the wavelength of the illuminating light. Since the NA is typically never much larger than 1, imaging using optical wavelengths gives a diffraction limit of the order of hundreds of nanometres. Thus, traditional optical microscopy cannot spatially resolve objects on the scale of tens of nanometres or lower. Decreasing the wavelength to the nanoscale in order to improve spatial resolution is not feasible due to the high energy photons at such small wavelengths, which will damage biological samples, and thus the diffraction limit must be overcome to allow for SPT at a nanoscale.

2.2.2 Super-resolution and Localisation Microscopy

The field of super-resolution microscopy has moved microscopy and imaging beyond the diffraction limit using localisation microscopy. The diffraction limit arises due to the fact that one can only collect plane wave components of light with transverse wavevector components, k , below a cut-off $k < nk_0 \sin \theta$. This Fourier space cut off means the real space image of a point source (an accurate model for nanometre scale single light emitters or scatterers) at (x_0, y_0) will give rise to a point spread function (PSF) centred on (x_0, y_0) in the image [163]. For a simple circular aperture, the PSF is the well known Airy disc [164], however more complicated imaging systems give different PSFs [164, 165]. The PSFs from multiple point sources overlap for separations below the diffraction limit, hence the inability to resolve sources closer than this.

If, however, one knows the light arises from a single source, one can fit a PSF to the image and extract the source position as the centre of the fitted PSF. Thus, while imaging may not be able to resolve objects beyond the diffraction limit, the position of a single isolated emitter can be extracted from an image with precision below the diffraction limit. The uncertainty in the position estimate from this method depends on the noise in the measured intensity distribution. For shot noise limited measurements, the noise standard deviation is proportional to the number of photons contained within the PSF, N_{sig} and the localisation precision scales as $\Delta x \sim N_{sig}^{-1/2}$ [166]. This simple fitting process applied to conventional fluorescence or gold nanoparticle labelled dark field microscopy setups allows for resolutions of $\sim 1\text{nm}$ [167, 168]. While the PSF is distributed over the image of 2D plane, and fitting gives 2D localisation of the source, 3D localisation can be achieved in a variety of ways. This includes simultaneously imaging multiple planes [169], using a z -dependent PSF [165, 170, 171, 172, 173, 174] and holographic approaches in which phase information is extracted to allow z localisation [175]. Incorporation of machine learning techniques into the image analysis process has helped improve localisation precision (see Chapter 6 for a detailed discussion), but ultimately the practical limitations of experimental noise still apply to these methods [176, 177].

In addition to ensuring a sufficient number of photons are present in the PSF, it is also important that the intensity distribution that the PSF is being fitted to arises from a single analyte particle. A collection of nearby sources may be fitted as a single PSF. Distinguishing nearby sources is important in highly dense media where clusters of nearby sources are common. Fluorescent microscopy localisation techniques such as stochastic optical reconstruction microscopy (STORM) or photoactivated localisation microscopy (PALM) use properties of the fluorophores to ensure nearby sources are distinguishable [178, 179]. In PALM, different fluorophores are sequentially activated by incident light into a fluorescing state before bleaching or being switched off. Since only a small fraction of fluorophores are emitting at any one time, the PSFs are sufficiently separated and can be fitted, with the total image built up by combining the images for different fluorophores. Similarly, STORM works via having only a fraction of fluorophores on at any one time, except rather than being switch on/off deliberately by incident light, they randomly switch on and off in a process called photoblinking. Since

nearby fluorophores will be on or off at different times, they can be distinguished and localised separately. These are capable of localising individual fluorophores to nanometre precision, and by attaching fluorophores that emit different wavelengths at different points on a biological structure, one can extract information on orientation. Both PALM and STORM separate the response of nearby fluorophores temporally to allow them to be distinguished. Another example of a fluorescent super-resolution technique is stimulated emission depletion (STED) microscopy, which separates the response of nearby fluorophores spectrally [180]. STED uses the fact that an excited fluorophore, if struck by a photon of the correct energy (or equivalently correct wavelength, the STED wavelength), can be promoted into a higher energy level than the standard fluorescent transition, and thus emits a photon that has a longer (red shifted) wavelength than the standard fluorescent wavelength transitioning from the initial excited state energy level down to a lower state. Thus these photons can be differentiated from those emitted at the pump wavelength [180, 181, 182]. By illuminating at the pump wavelength to excite the fluorophores, and also illuminating with a doughnut shaped beam with low (ideally zero or extremely close to zero) intensity at the centre at the STED wavelength, so that fluorophores surrounding the central spot emit at the red-shifted wavelength, one ensures only the central spot fluoresces at the standard fluorescence wavelength, and thus the position of the light source can be more tightly localised. Scanning this spot across a sample allows an image to be formed. The modified resolution, D , is given by [183]

$$D = \frac{\lambda}{2\text{NA}\sqrt{1 + \frac{I_{max}}{I_{sat}}}} \quad (2.7)$$

where I_{max} is the maximum intensity of the STED light and I_{sat} is the saturation intensity. While Eq. (2.7) indicates increasing the STED intensity allows for theoretically unlimited improvement in resolution, high intensities destroy the fluorophores, known as photobleaching. It can also potentially damage the sample. By measuring the temporal response as the populations of excited fluorophores at the two different excited states and different positions decay, one can improve the resolution [184, 185]. Using plasmonic nanoparticles in STED to give near field intensity enhancements, while also reducing fluorophore excited state lifetime allows for

resolutions $\sim 75\text{nm}$ at significantly lower input intensities than required for the conventional STED setup to achieve such resolutions [186, 187, 188]. Resolutions as low as 2.4nm [189] have been reported, but typical set ups tend to achieve on the order of 30nm [183]. The fact that the intensity minimum spot must be scanned to form an image limits STED, as one must trade off the field of view against temporal resolution owing to the time taken to scan across the desired field of view. The effects of the scanning have been mitigated via illuminating the sample with spatially patterned light consisting of many intensity minimum spots, parallelising the STED image acquisition, allowing for ~ 2000 -fold increased field of view or temporal resolution without loss of spatial resolution [190].

Another fluorescent microscopy technique based on spatially structuring light to improve localisation accuracy is minimal photon fluxes (MINFLUX) [191]. The incident excitation light, much like in STED, has a ring shaped intensity profile with low (ideally zero) intensity at the centre. A fluorophore located at this zero will not emit any fluorescent light. By scanning this point across the sample and finding the location of the intensity minimum where the minimum number of fluoresced photons are detected, fluorophores can be located to a resolution of $\sim 1\text{nm}$ [191, 192]. Several other fluorescence localisation microscopy techniques exist, all working on a similar principle of separating fluorophore response along some variable (e.g. in time) so that nearby PSFs can be distinguished.

Super resolution fluorescence microscopy techniques have pushed spatial resolution down to the nanometre scale, but do come with several limitations. The most obvious is the requirement to use fluorescent labels, the drawbacks of which have been discussed. Another problem is the time resolution of such methods can be limited. The analyte particle must not move significantly within the detector integration time, as this would change the centre of the PSF [1]. Reducing the integration time for improved resolution comes at the cost of reducing the photons received and therefore reduced spatial resolution. As a result, STORM and PALM are restricted to monitoring relatively slow moving processes with a temporal resolution of $\sim 1\text{s}$ [1, 193]. Note that STORM/PALM have their temporal resolution reduced compared to that of the camera taking the measurements owing to the fact several frames corresponding to different fluorophores being ‘switched on’ must be used to form the total image. Due to the fact it can localise with

very low numbers of photons, MINFLUX can achieve much better time resolution on the order $\sim 0.4\text{ms}$ for a 2.4nm spatial resolution, while there is a trade off between improving spatial and temporal resolution [191].

2.2.3 Interferometric Methods

Achieving label free super-resolution imaging is challenging owing to the small number of scattered photons scattered by nanoscale biological particles in their native form. Interferometric scattering (iSCAT) microscopy [10, 194, 195, 196, 197] leverages interference in order to enhance the signal arising from the analyte particle. The electric field scattered by the analyte particle, \mathbf{E}_s interferes with a much larger reference field \mathbf{E}_b . The resulting power measured by a detector is

$$P \propto |\mathbf{E}_b|^2 + |\mathbf{E}_s|^2 + 2 \text{Re}(\mathbf{E}_s \cdot \mathbf{E}_b^*). \quad (2.8)$$

Since $|\mathbf{E}_b| \gg |\mathbf{E}_s|$, the interference term is significantly larger than the intensity scattered directly by the particle, $2 \text{Re}(\mathbf{E}_s \cdot \mathbf{E}_b^*) \gg |\mathbf{E}_s|^2$. As a result, the signal arising due to scattering from the analyte particle is much larger than a technique where only light scattered by the analyte particle is measured. The downside of this interference approach is the signal must be detected against a large background intensity. Furthermore, the noise properties in the two cases are significantly different. In order to elucidate the advantage of an interferometric technique such as iSCAT as compared to a method where only light scattered/emitted from the analyte particle is collected, the SBR and SNR of the two methods should be compared. Considering a detector, e.g. a pixel in charge coupled device (CCD) camera, receiving N_T photons per detector integration time. The major noise sources to consider are shot noise [198] and dark noise (caused by thermal excitation of electrons in the detector) [199, 200]. The measured photon number over any given integration time can be modelled as a Poisson random variable, with mean given by N_T , plus a signal independent Poisson distributed dark count with mean N_d [199]. As a Poisson random variable has a standard deviation equal to its mean, the total noise standard deviation in such a measurement is given by $\sigma = (N_T + N_d)^{1/2}$. There are two limiting cases, firstly the case where detection is shot noise limited, $N_T \gg N_d$ and dark noise is a negligible

Table 2.2: Summary of noise metrics for dark field and interference methods.

	Metric	Exact expression	Shot noise limit $N_T \gg N_d$	Dark noise limit $N_T \ll N_d$
Dark field	S	N_s	N_s	N_s
	σ	$\sqrt{N_s + N_d}$	$\sqrt{N_s}$	$\sqrt{N_d}$
	SBR	N_s/N_d	$N_s/N_d \gg 1$	$N_s/N_d \ll 1$
	SNR	$\frac{N_s}{\sqrt{N_s + N_d}}$	$\sqrt{N_s}$	$\frac{N_s}{\sqrt{N_d}}$
Interference	S	$\sqrt{N_s N_b} \cos \Phi $	$\sqrt{N_s N_b} \cos \Phi $	$\sqrt{N_s N_b} \cos \Phi $
	σ	$\sqrt{N_T + N_d}$	$\sqrt{N_b}$	$\sqrt{N_d}$
	SBR	$\frac{\sqrt{N_s N_b} \cos \Phi }{N_T + N_d}$	$\sqrt{\frac{N_s}{N_b}} \cos \Phi \ll 1$	$\frac{\sqrt{N_s N_b}}{N_d} \cos \Phi \ll 1$
	SNR	$\frac{\sqrt{N_s N_b} \cos \Phi }{\sqrt{N_T + N_d}}$	$\sqrt{N_s} \cos \Phi $	$\frac{\sqrt{N_s N_b}}{\sqrt{N_d}} \cos \Phi $

contribution to the total noise, and conversely dark noise limited detection, in which $N_T \ll N_d$ and dark noise is the dominant noise source. If an analyte particle scattering N_s photons into the detector per integration time, the expression for N_T depends differently on N_s between dark field and interferometric detection. In the non-interferometric case, $N_T = N_s$, whereas the interferometric method has $N_T = N_b + N_s + 2(N_T N_d)^{1/2} \cos \Phi$, where N_b is the number of photons incident on the detector in an integration time due to the reference field and Φ is the phase difference between the analyte particle scattered field and the reference field. It will be assumed that $N_b \gg N_s$, as is the case in iSCAT. The signal, from which information on the analyte particle is deduced, is given (for small N_s) in this case by $S = |\partial N_T / \partial N_s| N_s$. Table 2.2 summarises the signal metrics for a dark field and interference based method.

When both methods are shot noise limited, they achieve comparable SNRs of $\sim N_s^{1/2}$, while the interferometric method suffers from a much lower SBR. The advantage of the interference method is seen in the case $N_s \lesssim N_d \ll N_b$, in which the interference method will be in the shot noise limit, whilst the dark field method will be dark noise limited (or dark noise will be at least comparable to shot noise). The shot noise limited interference method has an increased SNR compared to the dark noise limited dark field SNR, while both have a small SBR. In fact, regardless of whether an interference based method is shot noise or dark noise limited, it achieves improved SNR over dark noise limited dark field measurement when $N_d > N_s^{1/2}$.

The ability to achieve shot noise limited measurements even for small N_s below the dark noise limit is a significant advantage within an interferometric measurement scheme. It should be noted that, while a signal independent Poissonian dark noise was assumed, the same conclusions would be reached with any other statistics of signal independent noise added to the shot noise. Note that the Poisson distribution of the dark noise is not essential to the conclusions, any generic signal independent noise distribution with mean μ_d and standard deviation σ_d would give analogous results in the shot noise $N_T \gg \mu_d, \sigma_d$ and dark noise $N_T \ll \mu_d, \sigma_d$ limits.

The improved SNR that iSCAT achieves through interference enables it to track nanoscale objects, including single viruses and proteins with accuracy down to the order of 2nm [10, 196, 201, 202]. Different illumination and detection schemes have been implemented, including illumination with SPPs and PC modes to benefit from the enhanced light-matter interaction associated with the field confinement [203, 204, 205]. The reference field can share a path with light scattered from the analyte particle to ensure interferometric stability, with no need for a reference arm [196, 197]. For example, for an analyte particle above a dielectric interface scattering light incident from below the interface (e.g. a microscope coverslip), the light reflected from the interface acts as the reference field and interferes with the scattered light from the analyte. Knowledge of the nature of the scattered field (for example highly directional scattering for a particle near an interface) and reference field allows for optimising the illumination and collection of light, for example partially reflective masks have been used to ensure selective blocking of the reflected reference light while still transmitting the majority of the scattered light [202]. It should be noted that such an approach will not improve the SNR in shot noise limited measurements, which is independent of N_b , but rather increases the SBR.

The use of coherent light, required to leverage interference, means other scattering sources also affect the image. Randomly scattered light from impurities such as surface roughness features on glass coverslips, or other biological particles present in the sample, also interfere with the reference and analyte scattered field, causing a so-called speckle pattern consisting of granular bright and dark spots. This speckle pattern can obscure the signal from the analyte particle, and any iSCAT method requires an efficient means of removing the speckle background [10, 194, 202].

2.3 Conclusion

In this chapter, techniques for both biosensing and single particle tracking/localisation have been reviewed. Resonance sensing has been shown to be a powerful label free tool, capable of sensitivity to single proteins and other nanometre scale objects. SPR sensors are simple, widely used and easily fabricated, but lack sensitivity down to the single particle level. Other resonance based sensors can achieve sensitivity at a single particle level, but have more stringent fabrication requirements. Any resonance based sensing scheme ultimately suffers from the fact that one must maximise interaction with the analyte particle while also ensuring interactions with the surrounding environment, causing losses and increasing the resonance width, are minimised. Nanowire and nanopore technology is also capable of single particle sensitivity, but also require precise fabrication [206]. While technological and scientific advances reduce the difficulties of implementing these more precise methods commercially or clinically, any technique that uses a similar or identical sensor design to SPR sensors but with sensitivity at a single particle level would allow for single particle biosensors to be produced at the same scale as SPR devices.

Super-resolution microscopy techniques have greatly advanced SPT/localisation, allowing for nanometre level accuracy in monitoring the motion of single biological objects. Techniques based on fluorescence microscopy, such as STED or PALM, are extremely powerful but require fluorescence labelling. The weak interaction of nanoscale particles with light means super-resolution microscopy of unlabelled analyte particles is challenging, as the low SNR prevents reliable fitting of a PSF. The use of interference to enhance the signal as in iSCAT is a powerful technique, allowing shot noise limited measurements even when the intensity scattered by a particle is smaller than dark noise levels (or other noise sources). As a result, iSCAT microscopy is capable of label free detection and tracking of single viruses and proteins, but the coherent nature of such a technique means it is susceptible to image degrading speckle arising from scattering impurities. Overcoming the issue of speckle could allow interferometric techniques to operate in complex scattering environments which commonly arise in biology.

Chapter 3

Sensitivity Analysis

This chapter introduces the proposed biosensing system that will be investigated in the thesis, and a theoretical sensitivity analysis is performed. In order to do so, several important methods and results from plasmonics and scattering theory (in particular random scattering theory) will be introduced, many of which will be used throughout the thesis. Thus this chapter also serves to lay down the basic modelling framework used to describe the sensing (and, in future chapters, tracking) system. As such, a logical starting place is to introduce plasmonics, and in particular the SPP, the electromagnetic surface wave used in SPR sensing and central to the work in this thesis.

3.1 Surface Plasmon Polaritons

The SPP is an electromagnetic mode that propagates along the interface between a dielectric medium and a metallic medium. For now just considering a single planar interface (see Fig. 3.1), such that the region $z > 0$ is occupied by the dielectric with relative permittivity ε_d and $z < 0$ is occupied by a metal with relative permittivity ε_m (each with corresponding refractive index $n_{d,m} = \varepsilon_{d,m}^{1/2}$), the electric field $\mathbf{E}_{\text{SPP}}(\mathbf{r}, t)$ of SPP mode propagating in the forward x direction

as a function of position $\mathbf{r} = (x, y, z)^T$ and time t is given by [13, 14]

$$\mathbf{E}_{\text{SPP}}(\mathbf{r}, t) = \begin{cases} \mathbf{A}_{\text{SPP}}^+ \exp\{ik_{\text{SPP}}x - \kappa_d z - i\omega t\} & z > 0 \\ \mathbf{A}_{\text{SPP}}^- \exp\{ik_{\text{SPP}}x + \kappa_m z - i\omega t\} & z < 0, \end{cases} \quad (3.1)$$

where ω is the frequency and k_{SPP} is the SPP wavenumber. In order that the field remains finite as $z \rightarrow \pm\infty$, the conditions $\text{Re}(\kappa_{d,m}) > 0$ apply. The amplitude vectors are $\mathbf{A}_{\text{SPP}}^+ = A(\kappa_d, 0, ik_{\text{SPP}})^T$ and $\mathbf{A}_{\text{SPP}}^- = (\varepsilon_d/\varepsilon_m)A(-\kappa_m, 0, ik_{\text{SPP}})^T$, and A is a constant determined by the intensity at (infinitesimally above) the interface, $I_0 = |\mathbf{A}_{\text{SPP}}^+|^2$ (throughout the thesis, the intensity of light shall be taken to be $I = |\mathbf{E}|^2$, ignoring constant prefactors which will normalise out). In general, SPPs must be p-polarised, i.e. the polarisation vector lies in the plane determined by the surface normal and the propagation direction, in this case corresponding to the (x, z) plane [13]. This can be seen by the fact that the vectors $\mathbf{A}_{\text{SPP}}^\pm$ have no y -component. In order to satisfy Maxwell's equations and their associated continuity conditions at material interfaces (namely continuity of parallel components of \mathbf{E} and perpendicular components of $\mathbf{D} = \varepsilon(\mathbf{r})\varepsilon_0\mathbf{E}$ [85]), the following conditions must also be satisfied

$$k_{\text{SPP}}^2 - \kappa_d^2 = \varepsilon_d k_0^2 \quad (3.2)$$

$$k_{\text{SPP}}^2 - \kappa_m^2 = \varepsilon_m k_0^2, \quad (3.3)$$

where $k_0 = \omega/c$ is the free space wavenumber, c is the speed of light *in vacuo*. Furthermore, the interface continuity conditions can only be satisfied if $\text{Re}(\varepsilon_m) < 0$, hence SPPs can not be supported at the interface between two dielectrics [14]. Already, the form of Eq. (3.1) makes an important feature of SPPs clear, specifically the exponential decay of \mathbf{E} away from the interface at $z = 0$. This means, compared to free space optical waves with the same total energy, the energy of the electromagnetic mode is confined to a smaller region and therefore there are higher intensities at the surface. The length scale the intensity decays over away from the surface is given by $L_{d,m} = (2 \text{Re}[\kappa_{d,m}])^{-1}$ in the dielectric (subscript d) or metal (subscript m). Physically, SPPs arise from the coupling of electromagnetic waves to collective oscillations of the free conduction electrons in metals [12]. Taken together, Eqs. (3.2) and (3.3) give the well

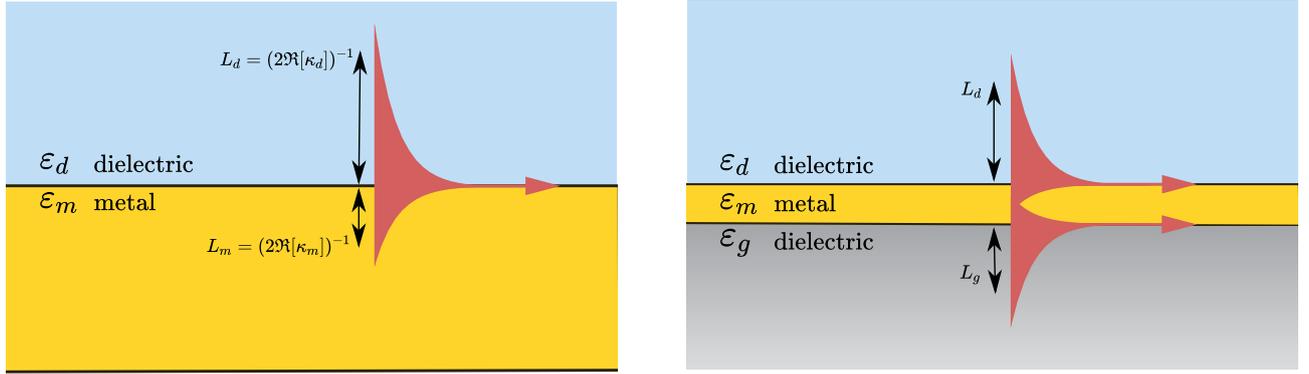


Figure 3.1: Geometries supporting SPP modes (left) a single metal-dielectric interface and (right) a metal film between two dielectrics, the simplest multilayer structure. In both cases, a key feature is the evanescent confinement of the mode to the metal surface, while propagating parallel to the surface.

known SPP dispersion relation for a single interface as [13]

$$k_{\text{SPP}} = \left(\frac{\varepsilon_d \varepsilon_m}{\varepsilon_d + \varepsilon_m} \right)^{\frac{1}{2}} k_0. \quad (3.4)$$

While the use of a metallic substrate enables the interface to support SPPs, metals suffer from ohmic losses, meaning that ε and therefore k_{SPP} are complex quantities. Throughout the thesis, the notation $k'_{\text{SPP}} = \text{Re}(k_{\text{SPP}})$ and $k''_{\text{SPP}} = \text{Im}(k_{\text{SPP}})$ will be used. The real part defines the wavelength of the SPP through $\lambda_{\text{SPP}} = 2\pi/k'_{\text{SPP}}$. Similarly, power is absorbed from the SPP as it propagates and the imaginary part defines the in-plane decay length as $L_{\text{SPP}} = (2k''_{\text{SPP}})^{-1}$. The dispersion relation in Eq. (3.4) means that, at all frequencies, $k'_{\text{SPP}} > n_d k_0$, i.e. the SPP wavenumber is larger than the photon wavenumber in the dielectric [14]. As a result, photons cannot couple directly into SPP modes, and more sophisticated structures are required to enable excitation of SPPs.

3.1.1 Multilayer Structures and SPPs

Multilayer structures can also support SPPs. The simplest example is a thin metallic film between two dielectrics, shown in Fig. 3.1. In general, in a multilayer structure of thickness d , with $z > 0$ occupied by a dielectric medium with permittivity ε_d , $z < -d$ with permittivity ε_N (refractive index $n_N = \varepsilon_N^{1/2}$) and intermediate layers of permittivity ε_i (refractive index $n_i = \varepsilon_i^{1/2}$)

for $i = 1, 2, \dots, N - 1$, an SPP mode takes the form [13, 207]

$$\mathbf{E}_{\text{SPP}}(\mathbf{r}, t) = \begin{cases} \mathbf{A}_{\text{SPP}}^+ \exp\{ik_{\text{SPP}}x - \kappa_d z - i\omega t\} & z > 0 \\ [a_i \mathbf{A}_{\text{SPP}}^{i+} \exp(\kappa_i z) + b_i \mathbf{A}_{\text{SPP}}^{i-} \exp(-\kappa_i z)] \exp\{ik_{\text{SPP}}x - i\omega t\} & \text{layer } i \\ \mathbf{A}_{\text{SPP}}^- \exp\{ik_{\text{SPP}}x + \kappa_N z - i\omega t\} & z < -d, \end{cases} \quad (3.5)$$

where in each layer the polarisation vector is $\mathbf{A}_{\text{SPP}}^{i\pm} \propto (\mp\kappa_i, 0, ik_{\text{SPP}})^T$. The dispersion relations from Eqs. (3.2) and (3.3) now apply in each layer, meaning

$$k_{\text{SPP}}^2 - \kappa_d^2 = \varepsilon_d k_0^2 \quad (3.6)$$

$$k_{\text{SPP}}^2 - \kappa_i^2 = \varepsilon_i k_0^2 \quad (3.7)$$

$$k_{\text{SPP}}^2 - \kappa_N^2 = \varepsilon_N k_0^2. \quad (3.8)$$

Applying the interface continuity conditions to Eq. (3.5) at each interface and simultaneously solving Eqs. (3.6)–(3.8) allows one to find the unknown coefficients a_i and b_i as well as k_{SPP} and the decay constants κ_d , κ_i and κ_N . While the SPP no longer has the dispersion relation given in Eq. (3.4), it still maintains most of the same properties as the single interface SPP, in particular the functional form for $z > 0$ remains identical, but with different wavenumber and out of plane decay constants while still satisfying Eq. (3.2) (since it is equivalent to Eq. (3.6)).

The multilayer structure SPP modes do possess a few additional features. Firstly, such structures can support multiple SPP wavenumbers at a given frequency. For example, in the dielectric-metal-dielectric structure, the SPPs supported at each interface couple together to form a long range and short range mode, where the long range mode has a lower proportion of the intensity distributed in the metal compared to the short range mode, leading to reduced ohmic losses and therefore a longer L_{SPP} [14, 208, 209]. In addition, while Eq. (3.6) still implies that the SPP wavenumber is larger than the photon wavenumber in the upper dielectric, it is possible, by having a permittivity in the lower half-space sufficiently larger than in the upper half-space (i.e. $\varepsilon_N > \varepsilon_d$), that photons in the lower dielectric can couple into the SPP mode when $k_{\text{SPP}} < n_N k_0$. In this case, when light is incident from below the thin film structure

at an angle such that the component of the wavevector parallel to the surface matches the SPP wavenumber, light can couple into the SPP mode. Mathematically, this condition can be expressed as $n_N k_0 \sin \Theta_{\text{SPP}} = k'_{\text{SPP}}$, where Θ_{SPP} is the angle between the incident wavevector and the surface normal at which this coupling occurs. The Kretschmann excitation scheme is based on this principle, consisting of a thin metal film on a dielectric prism [210]. When the incident angle is Θ_{SPP} , light couples into the SPP mode propagating along the surface rather than being specularly reflected. As discussed in Chapter 2, the large decrease in reflected light intensity and the associated strong dependence of the reflected intensity on the angle of incidence near Θ_{SPP} forms the basis for SPR sensing.

While multilayer structures allow for coupling of photons from a dielectric into SPPs, such as via the Kretschmann configuration, they also allow for the reverse process of SPPs coupling into photons in the lower dielectric and propagating away from the interface. This light radiated away from the thin film stack is referred to as leakage radiation [13, 211]. The conservation of the transverse component of the wavevector still applies, and thus leakage radiation is confined to an angle Θ_{SPP} . This leakage radiation can be seen by considering Eqs. (3.5) and (3.8) in the case of lossless media where $k''_{\text{SPP}} = 0$. In this case, Eq. (3.8) gives κ_N as pure imaginary when $n_N k_0 > k_{\text{SPP}}$ and therefore the SPP in the lower half space takes the form $\propto \exp\{i(k_{\text{SPP}}x - i|\kappa_N|z)\}$ of a plane wave propagating away from the surface in a direction Θ_{SPP} to the surface normal. In reality, k_{SPP} is lossy and this allows a small range of angles at which leakage radiation occurs. Leakage radiation will play a significant role in this thesis and will be discussed in greater detail later in this chapter.

Like any other electromagnetic mode, SPPs interact with matter, for example non-planar features on the surface [212] or particles near the surface [213]. Such objects can scatter SPPs into SPP modes propagating in different directions, or into propagating light in the dielectric half spaces, in addition to absorbing some of the light [213]. The scattering interaction with matter near the surface is the basis of the sensing system studied in this thesis.

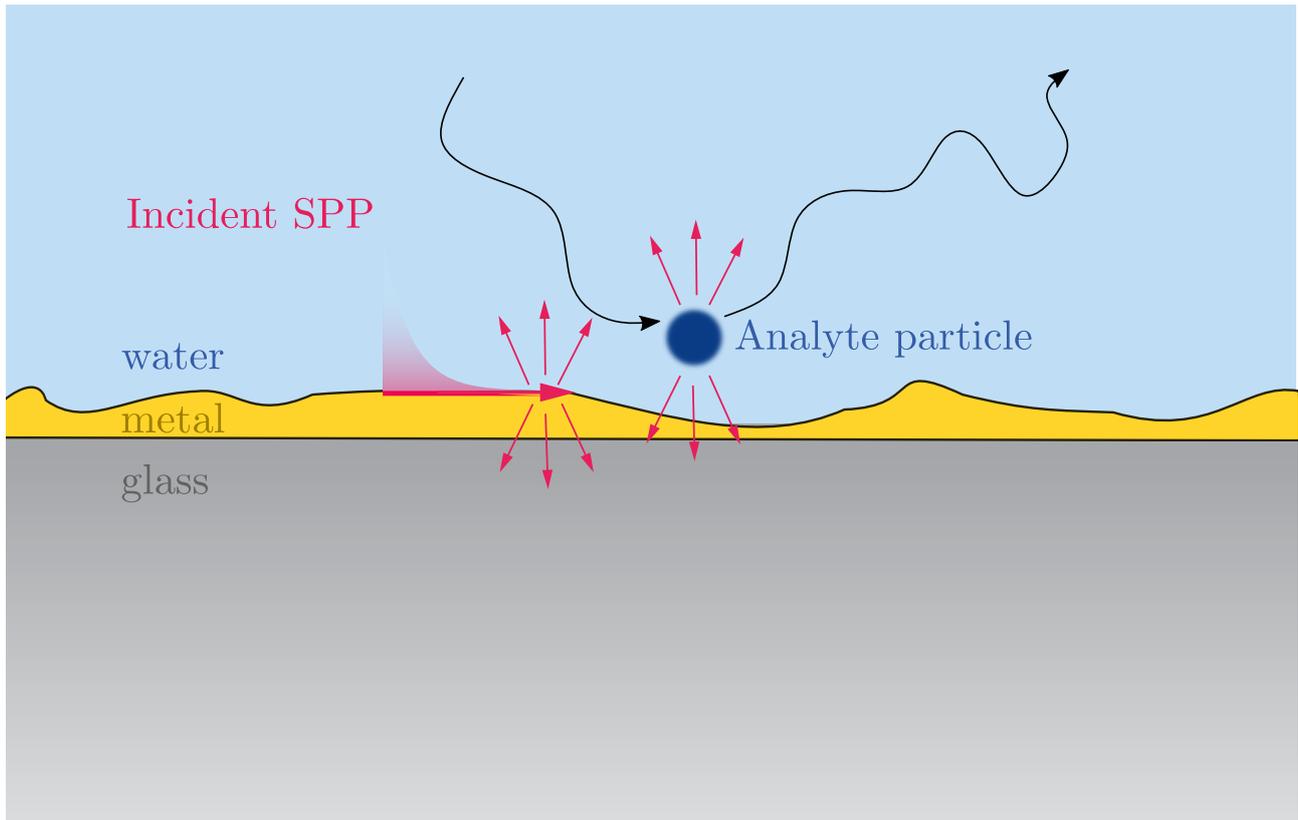


Figure 3.2: The sensing setup to be investigated. An SPP propagates on a thin metallic film, and is scattered by surface roughness. An analyte particle, diffusing above the metal film also scatters the SPP when it comes near the surface. This changes the scattered light distribution, and thus measuring the change allows one to deduce the presence of the analyte particle. The red arrows are merely illustrative of scattering and not intended to represent any quantitative information on the distribution of scattered light. While the structure shown here is a glass-metal-water thin film structure, the principle can be applied to any layered structure supporting SPPs.

3.2 A Surface Plasmon Scattering Enabled Sensor

The principle of the proposed sensing system is shown in Fig. 3.2. The solution containing the analyte particles is above a metallic surface supporting SPPs. The surface has fixed scatterers on the surface, either arising from surface roughness or produced by deliberate fabrication of scattering objects on the surface [214, 215, 216, 217]. An SPP excited on the metal surface will be scattered by these surface scatterers to give a light distribution which can be measured. When an analyte particle comes close to the surface, it too scatters the SPP. The additional scattered light changes the measured light distribution, from which one can deduce the presence of the analyte particle. By using a confined plasmonic field, the method benefits from the high intensity at the surface, resulting in enhanced scattering from scatterers near the surface

(including the analyte particle) compared to other evanescent wave scattering approaches such as total internal reflection (TIR) microscopy [218].

While the principle is easily explained, the use of fixed background scatterers should be explained in more detail. Indeed, one could base a sensor on an identical principle in the absence of this background random scattering, with the analyte particle illuminated by an evanescent field analogous to TIR microscopy, except with a plasmonic as opposed to photonic field. Firstly, however, it should be considered that imperfect fabrication and surface abrasion means the metallic surface will never be perfectly planar and will have random surface roughness features, which will scatter SPPs. Since the analyte particles are small and only scatter light weakly, even a small amount of scattering from the surface will be comparable to the ‘darkfield’ signal from the particle. As a result, even in studying the ‘darkfield’ version of such a sensor, one would have to consider the effects of such surface roughness scattering. In addition to some level of background scattering being unavoidable, the intensity of light measured at a detector is the coherent sum of light scattered from the fixed background scatterers and light scattered from the analyte, and thus there is interference between the two. Therefore, a method using background scattered light can benefit from the advantages of interferometric detection discussed in Section 2.2.3. While iSCAT often uses the reflected field as a reference field to interfere with the field scattered from an analyte particle, the coupling into the confined SPP mode means very little light is reflected. The fixed random surface scattering, however, will scatter light away from the surface which then can interfere with the analyte particle scattered field.

Modelling the sensor system requires a description of scattering from both the analyte particle and the background scatterers. This can be done using electromagnetic scattering theory, and it is important to introduce some fundamental principles from this. Furthermore, since scattering from a randomly rough surface (or a surface with randomly distributed scatterers fabricated on it) will be studied, random scattering theory in particular must be considered.

3.3 Electromagnetic Scattering Theory

Fundamentally, electromagnetic scattering theory involves solving Maxwell's equations in order to find the total electromagnetic field when a given incident field illuminates an arrangement of scattering objects. The physical arrangement of scattering objects and a background dielectric environment is described by a position dependent dielectric function, $\varepsilon(\mathbf{r})$ (one can also consider position dependent permeability for magnetic scattering media, but in the context of this thesis only non-magnetic media will be considered). From Maxwell's equations, in the absence of sources, one can derive a wave equation for the electric field $\mathbf{E}(\mathbf{r}, t)$ [26, 85]

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}, t) + \frac{\varepsilon(\mathbf{r})}{c^2} \partial_t^2 \mathbf{E}(\mathbf{r}, t) = \mathbf{0}. \quad (3.9)$$

Throughout this thesis, a harmonic time dependence of the form $\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r})e^{-i\omega t}$ will be assumed, where ω is the frequency, meaning Eq. (3.9) reduces to

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}) - \frac{\varepsilon(\mathbf{r})\omega^2}{c^2} \mathbf{E}(\mathbf{r}) = \mathbf{0}. \quad (3.10)$$

Once Eq. (3.10) has been solved for \mathbf{E} , the magnetic field follows from using $\mathbf{H} = \nabla \times \mathbf{E}/(i\omega\mu\mu_0)$. The dielectric function $\varepsilon(\mathbf{r})$ describes both the background and the scatterers, and it is useful to separate the two so that $\varepsilon(\mathbf{r}) = \varepsilon_b(\mathbf{r}) + \varepsilon_s(\mathbf{r})$, where $\varepsilon_b(\mathbf{r})$ describes the background dielectric environment and $\varepsilon_s(\mathbf{r})$ describes the scatterers. The distinction between background and scatterers is to some extent arbitrary, in the sense that one can always add an arbitrary function to one of the contributions and subtract the same function from the other without changing the problem. Generally, the background dielectric environment is one for which the solution to Eq. (3.10) with $\varepsilon(\mathbf{r}) = \varepsilon_b(\mathbf{r})$ is known, either analytically or can be calculated numerically relatively easily. This is because the background solution is used to build the full solution, as will be described here. Usually, the distinction follows physical intuition as to what one would call a scatterer, however, different contexts will call for different choices of background, even within the same problem. In many scattering problems, the background dielectric environment will be some homogeneous medium such that $\varepsilon_b(\mathbf{r}) = \varepsilon_b$ is constant, however, in the case of SPP

scattering it is more useful to consider the multilayer structure with planar interfaces as the background. In this case, $\varepsilon_b(\mathbf{r}) = \varepsilon_b(z)$ is a piecewise constant function where $\varepsilon_b(z) = \varepsilon_i$ when z is in the i th layer. The scattering part can take a variety of forms depending on the scatterers in question. A common and simple example is a collection of objects where the j th object is made up of a medium of permittivity ε_j and occupies a volume V_j , in which case the scattering part of the dielectric function is given by

$$\varepsilon_s(\mathbf{r}) = \begin{cases} \varepsilon_j - \varepsilon_b(\mathbf{r}) & \mathbf{r} \in V_j \\ 0 & \mathbf{r} \notin V_j. \end{cases} \quad (3.11)$$

Given the potentially complicated form of $\varepsilon(\mathbf{r})$, very few scattering scenarios permit an exact solution to Eq. (3.10), but there are general approaches that can tackle a wide range of different scattering geometries to allow approximate or numerical solutions.

3.3.1 Lippmann-Schwinger and Dyson Equations

A powerful and widely used tool in scattering theory to find solutions to Eq. (3.10) is the Green's tensor. The Green's tensor $G(\mathbf{r}, \mathbf{r}')$ for a general dielectric environment $\varepsilon(\mathbf{r})$ is defined for Eq. (3.10) in the same way as for a general linear differential operator by [85, 213, 219, 220]

$$\nabla \times \nabla \times G(\mathbf{r}, \mathbf{r}') - \frac{\varepsilon(\mathbf{r})\omega^2}{c^2}G(\mathbf{r}, \mathbf{r}') = \mathbb{I}\delta(\mathbf{r} - \mathbf{r}') \quad (3.12)$$

with the boundary conditions that the field either decays or takes the form of an outgoing wave as \mathbf{r} goes to infinity. As Eq. (3.10) is a vector partial differential equation (PDE), $G(\mathbf{r}, \mathbf{r}')$ is a tensor quantity and \mathbb{I} is the 3×3 rank 2 identity tensor (this notation for the identity tensor or matrix shall be used throughout the thesis). While the tensor nature of G is important, it shall be left implicit, with the distinction from scalar quantities clear from context. Green's functions are widely used throughout physics, with the primary utility lying in the fact that one can use it to find the solution to an inhomogeneous differential equation with arbitrary source term. In the context scattering theory, Eq. (3.10) is a homogeneous equation, with the inhomogeneous

form given with a source function $\mathbf{j}(\mathbf{r})$ on the right hand side of Eq. (3.10) such that

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}) - \varepsilon(\mathbf{r}) \frac{\omega^2}{c^2} \mathbf{E}(\mathbf{r}) = \mathbf{j}(\mathbf{r}). \quad (3.13)$$

The general solution to this inhomogeneous form of the wave equation is given by

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_0(\mathbf{r}) + \int d\mathbf{r}' G(\mathbf{r}, \mathbf{r}') \mathbf{j}(\mathbf{r}'), \quad (3.14)$$

where $\mathbf{E}_0(\mathbf{r})$ satisfies Eq. (3.10) with no source term. While Eq. (3.10) for a scattering problem does not contain a source term, by bringing the scattering part ε_s of the dielectric function to the right hand side, Eq. (3.10) can be written in a form analogous to Eq (3.13) as

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}) - \varepsilon_b(\mathbf{r}) \frac{\omega^2}{c^2} \mathbf{E}(\mathbf{r}) = \varepsilon_s(\mathbf{r}) \frac{\omega^2}{c^2} \mathbf{E}(\mathbf{r}). \quad (3.15)$$

The term on the right containing ε_s can now be considered to act as a source and Eq. (3.14) applied. As a result, the total electric field satisfies the following integral equation, known as the Lippmann-Schwinger equation [221, 222]

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_0(\mathbf{r}) + \int d\mathbf{r}' G_0(\mathbf{r}, \mathbf{r}') \varepsilon_s(\mathbf{r}') \frac{\omega^2}{c^2} \mathbf{E}(\mathbf{r}'), \quad (3.16)$$

where G_0 is the Green's tensor for the background dielectric environment (i.e. satisfies Eq. (3.12) with $\varepsilon(\mathbf{r}) = \varepsilon_b(\mathbf{r})$). Physically, the first term \mathbf{E}_0 corresponds to the illuminating field as in the absence of scatterers $\varepsilon_s = 0$ and $\mathbf{E} = \mathbf{E}_0$ is just the illuminating field, while the second term corresponds to the scattered field. Note that the Lippmann-Schwinger equation, while based on Eq. (3.14), is fundamentally different in that, while $\mathbf{j}(\mathbf{r})$ is a fixed external source of the electromagnetic field from which one calculates the electric field, the effective source in Eq. (3.16) contains the electric field that one wants to solve for. Thus, rather than being a solution, Eq. (3.16) is a reformulation of Eq. (3.10) as an integral equation. One approach to solve this integral equation is to iteratively substitute \mathbf{E} from Eq. (3.16) back in to the integrand on the

right hand side, resulting in the Born series [24, 223]

$$\begin{aligned} \mathbf{E}(\mathbf{r}) = & \mathbf{E}_0(\mathbf{r}) + \int d^3\mathbf{r}' G_0(\mathbf{r}, \mathbf{r}') \frac{\varepsilon_s(\mathbf{r}')\omega^2}{c^2} \mathbf{E}_0(\mathbf{r}') \\ & + \int d^3\mathbf{r}' d^3\mathbf{r}'' G_0(\mathbf{r}, \mathbf{r}') \frac{\varepsilon_s(\mathbf{r}')\omega^2}{c^2} G_0(\mathbf{r}', \mathbf{r}'') \frac{\varepsilon_s(\mathbf{r}'')\omega^2}{c^2} \mathbf{E}_0(\mathbf{r}'') + \dots \end{aligned} \quad (3.17)$$

Truncating a Born series at a finite number of terms gives an approximate expression for the total electric field in terms of \mathbf{E}_0 . Terminating the series at the first order term in ε is known as the Born approximation, in which case the field is calculated via

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_0(\mathbf{r}) + \int d^3\mathbf{r}' G_0(\mathbf{r}, \mathbf{r}') \frac{\varepsilon_s(\mathbf{r}')\omega^2}{c^2} \mathbf{E}_0(\mathbf{r}'). \quad (3.18)$$

Since the integrand now contains the known field \mathbf{E}_0 rather than the total field being solved for, this can be considered an approximate solution, expressing \mathbf{E} in terms of known quantities. Terminating the series at the n th order term in ε_s is known as the n th Born approximation, and while often including more terms increases the accuracy of the method, this is not always true [223]. The various terms in the Born series can be given an intuitive physical interpretation. A factor of $\varepsilon_s(\mathbf{r}')$ corresponds to a scattering event at \mathbf{r}' and a factor of $G_0(\mathbf{r}', \mathbf{r}'')$ describes free propagation in the background from \mathbf{r}'' to \mathbf{r}' . Thus, the integrand in the n th order term describes the incident field being scattered n times before propagating to the observation point, with the integration then summing over all positions from which scattering can occur. The Born series can therefore be interpreted as a sum over all scattering events, with the (first order) Born approximation corresponding to a single scattering approximation. Following analogous steps to deriving Eq. (3.16), an integral equation relating the Green's tensor G for the full system with the scatterers ($\varepsilon = \varepsilon_b + \varepsilon_s$) to the Green's tensor G_0 for the background environment without the scatterers ($\varepsilon = \varepsilon_b$) can be derived from Eq. (3.12) and is given as follows

$$G(\mathbf{r}, \mathbf{r}') = G_0(\mathbf{r}, \mathbf{r}') + \int d^3\mathbf{r}'' G_0(\mathbf{r}, \mathbf{r}'') \varepsilon_s(\mathbf{r}'') \frac{\omega^2}{c^2} G(\mathbf{r}'', \mathbf{r}'). \quad (3.19)$$

While almost identical in form to Eq. (3.16), in the context of Green's functions, Eq. (3.19) is referred to as the Dyson equation [224, 225]. Similarly, the iterative infinite series of Eq. (3.19)

formed by repeated substitution of G in the integrand is referred to as a Dyson series, analogous to the Born series of Eq. (3.17), and can be interpreted as a sum over all scattering paths in going from \mathbf{r}' to \mathbf{r} . Other scattering quantities, such as the t-matrix [221] or propagator [226] also have a Lippmann-Schwinger equation and corresponding Born series. The iterative equation structure of Eqs. (3.16) and (3.19) (and the associated infinite series) comes up repeatedly in scattering theory and the Green's tensor G_0 propagating between scattering events plays a fundamental role in these equations.

3.3.2 Green's Tensor Properties

Due to the importance of G and G_0 in scattering theory, it is useful to discuss some properties of Green's tensors. In an electromagnetic context, the Green's tensor physically corresponds to the field $\mathbf{E}_{\text{dip}}(\mathbf{r})$ radiated by a point electric dipole at position \mathbf{r}' with dipole moment \mathbf{p} , through $\mathbf{E}_{\text{dip}} = (k_0^2/\varepsilon_0)G(\mathbf{r}, \mathbf{r}')\mathbf{p}$. For this reason, the arguments \mathbf{r} and \mathbf{r}' will be referred to as the observation and source positions. The Green's tensor G_{dir} for 3D homogeneous space (constant $\varepsilon(\mathbf{r}) = \varepsilon$) has been constructed [221, 227] and is given by

$$G_{\text{dir}}(\mathbf{r}, \mathbf{r}') = \left(\mathbb{I} + \frac{1}{\varepsilon k_0^2} \nabla \nabla \right) g(\mathbf{r}, \mathbf{r}') \quad (3.20)$$

where $g(\mathbf{r}, \mathbf{r}') = \exp(i\varepsilon^{1/2}k_0|\mathbf{r} - \mathbf{r}'|)/(4\pi|\mathbf{r} - \mathbf{r}'|)$ is the scalar Helmholtz Green's function and the ∇ operator acts on the \mathbf{r} coordinate. Owing to the translational invariance, $G_{\text{dir}}(\mathbf{r}, \mathbf{r}') = G_{\text{dir}}(\mathbf{r} - \mathbf{r}')$ depends only on the difference between the source and observation positions. Similarly, the homogeneous nature means that the Green's tensor only depends on the magnitude $|\mathbf{r} - \mathbf{r}'|$ of the separation.

While the planar multilayer Green's tensor does not have a simple analytic form like G_{dir} , there are still several properties it possesses that should be emphasised. While the variation in $\varepsilon(z)$ along the z axis breaks the full 3D translational invariance, the system still maintains transverse translational invariance in the $\boldsymbol{\rho} = (x, y)$ plane. This implies that G only depends on the separation in this plane, i.e. $G(\boldsymbol{\rho}, z, \boldsymbol{\rho}', z') = G(\boldsymbol{\rho} - \boldsymbol{\rho}', z, z')$ where $\boldsymbol{\rho} = (x, y)$ and $\boldsymbol{\rho}' = (x', y')$ are the transverse components of \mathbf{r} and \mathbf{r}' respectively while z and z' are the

respective z -components. Excluding a small number of simple cases (for example the interface of a dielectric and a perfect conductor), the Green's function for a multilayer structure does not have a simple analytic expression in real space, but the 2D (x, y) Fourier transform can be found analytically using Fresnel coefficients [228]. Throughout the thesis, the (2D) Fourier transform pair of a function $F(\boldsymbol{\rho} - \boldsymbol{\rho}')$ shall be denoted $\tilde{F}(\mathbf{k}_{\parallel})$ with the pair related according to

$$F(\boldsymbol{\rho} - \boldsymbol{\rho}') = \int \tilde{F}(\mathbf{k}_{\parallel}) \exp [i\mathbf{k}_{\parallel} \cdot (\boldsymbol{\rho} - \boldsymbol{\rho}')] \frac{d^2 \mathbf{k}_{\parallel}}{(2\pi)^2} \quad (3.21)$$

$$\tilde{F}(\mathbf{k}_{\parallel}) = \int F(\boldsymbol{\rho} - \boldsymbol{\rho}') \exp [-i\mathbf{k}_{\parallel} \cdot (\boldsymbol{\rho} - \boldsymbol{\rho}')] d^2 \boldsymbol{\rho}. \quad (3.22)$$

Note that \mathbf{k}_{\parallel} is a 2D wavevector with dimensions of inverse length. Considering a source position above the thin film stack, $z' > 0$, the Green's tensor for a multilayer stack can be expressed in terms of direct, reflected and transmitted components as follows

$$G(\mathbf{r}, \mathbf{r}') = \begin{cases} G_{\text{dir}}(\mathbf{r}, \mathbf{r}') + G_{\text{ref}}(\mathbf{r}, \mathbf{r}') & z > 0 \\ G_{\text{tr}}(\mathbf{r}, \mathbf{r}') & z < -d \end{cases} \quad (3.23)$$

where G_{dir} is given by Eq. (3.20) for the upper dielectric medium, G_{ref} describes reflection from the multilayer stack and G_{tr} describes transmission through the multilayer stack, as shown in Fig. 3.3. By taking the plane wave (Weyl) expansion of G_{dir} , decomposing into s - and p -polarised components and reflecting or transmitting each component using the generalised Fresnel coefficients, one finds that, for a source position z' above the multilayer stack, the Fourier space direct, reflected and transmitted part of the Green's tensor are [213, 228]

$$\tilde{G}_{\text{dir}}(\mathbf{k}_{\parallel}; z, z') = \frac{i}{2k_z} \left[e^{ik_z|z-z'|} (\hat{\mathbf{e}}_s(\mathbf{k}_{\parallel}) \hat{\mathbf{e}}_s^{\dagger}(\mathbf{k}_{\parallel}) + \hat{\mathbf{e}}_{p\pm}(\mathbf{k}_{\parallel}) \hat{\mathbf{e}}_{p\pm}^{\dagger}(\mathbf{k}_{\parallel})) \right] \quad (3.24)$$

$$\tilde{G}_{\text{ref}}(\mathbf{k}_{\parallel}; z, z') = \frac{i}{2k_z} \left[e^{ik_z(z+z')} (r_s(\mathbf{k}_{\parallel}) \hat{\mathbf{e}}_s(\mathbf{k}_{\parallel}) \hat{\mathbf{e}}_s^{\dagger}(\mathbf{k}_{\parallel}) + r_p(\mathbf{k}_{\parallel}) \hat{\mathbf{e}}_{p+}(\mathbf{k}_{\parallel}) \hat{\mathbf{e}}_{p-}^{\dagger}(\mathbf{k}_{\parallel})) \right] \quad (3.25)$$

$$\tilde{G}_{\text{tr}}(\mathbf{k}_{\parallel}; z, z') = \frac{i}{2k_z} \left[e^{ik_z z' - ik_{zN}(z+d)} (t_s(\mathbf{k}_{\parallel}) \hat{\mathbf{e}}_s(\mathbf{k}_{\parallel}) \hat{\mathbf{e}}_s^{\dagger}(\mathbf{k}_{\parallel}) + t_p(\mathbf{k}_{\parallel}) \hat{\mathbf{e}}_{p-}^N(\mathbf{k}_{\parallel}) \hat{\mathbf{e}}_{p-}^{\dagger}(\mathbf{k}_{\parallel})) \right]. \quad (3.26)$$

The parameters $k_z = (k_d^2 - k_{\parallel}^2)^{1/2}$ and $k_{zN} = (k_N^2 - k_{\parallel}^2)^{1/2}$ correspond to the z component of the wavevector with transverse component \mathbf{k}_{\parallel} in the upper and lower dielectric respectively,

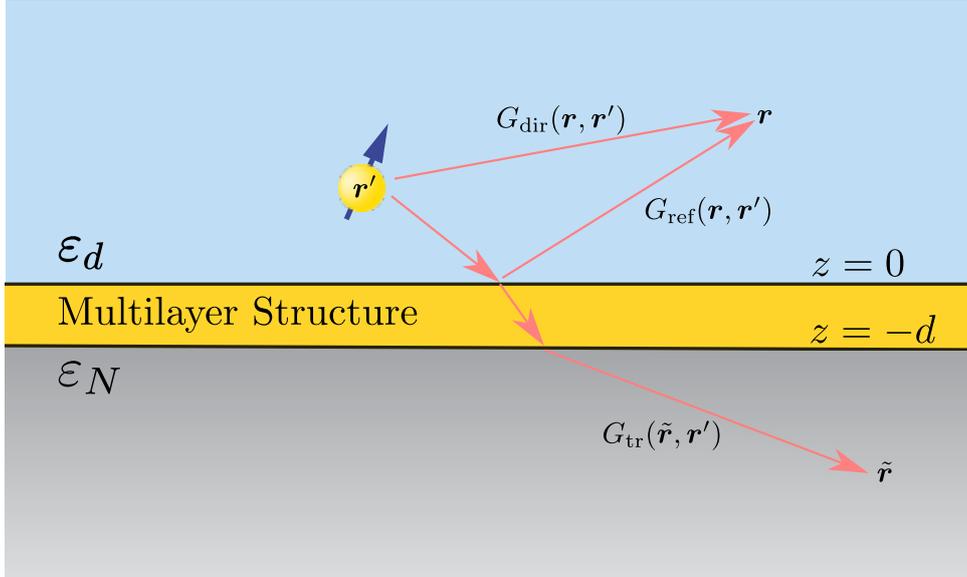


Figure 3.3: Representation of the direct, reflected and transmitted contributions to the multilayer Green's tensor $G(\mathbf{r}, \mathbf{r}')$ that propagates a dipolar source at \mathbf{r}' to the field at observation positions above (\mathbf{r}) or below ($\tilde{\mathbf{r}}$) the multilayer structure.

with $k_d = \varepsilon_d^{1/2} k_0$ and $k_N = \varepsilon_N^{1/2} k_0$ being the wavenumbers in the uppermost and lowermost media. The unit vectors $\hat{\mathbf{e}}_s(\mathbf{k}_{\parallel})$ and $\hat{\mathbf{e}}_{p\pm}(\mathbf{k}_{\parallel})$ correspond to the s - and p -polarised unit vectors for a plane wave in the upper dielectric with wavevector $\mathbf{k}_{\parallel} \pm k_z \hat{\mathbf{z}}$ (the $+$ sign corresponds to waves travelling in the positive z direction and the $-$ sign to those travelling in the negative z direction) and are defined by

$$\hat{\mathbf{e}}_s(\mathbf{k}_{\parallel}) = \hat{\mathbf{k}}_{\parallel} \times \hat{\mathbf{z}} \quad (3.27)$$

$$\hat{\mathbf{e}}_{p\pm}(\mathbf{k}_{\parallel}) = \frac{1}{k_d} (k_{\parallel} \hat{\mathbf{z}} \mp k_z \hat{\mathbf{k}}_{\parallel}). \quad (3.28)$$

Similarly, the vector $\hat{\mathbf{e}}_{p\pm}^N$ is the (upward and downward propagating) p -polarised unit vector in the lowermost medium and is given identically to Eq. (3.28) but with k_d and k_z replaced by k_N and k_{zN} . The s -polarised vector depends only on the transverse component \mathbf{k}_{\parallel} of the wavevector and thus is the same for upward and downward propagating waves as well as in different media. As such, no distinction is required for different media or propagation direction. Note that in Eq. (3.24), the sign chosen for the $\hat{\mathbf{e}}_{p\pm}$ factors depends on the relative position of the source and observation points. Namely, the upper sign is chosen for $z > z'$ and the lower sign for $z < z'$. The quantities $r_{s,p}(\mathbf{k}_{\parallel})$ and $t_{s,p}(\mathbf{k}_{\parallel})$ are the generalised Fresnel reflection and

transmission coefficients for s - and p -polarised plane waves in the uppermost medium. These describe the amplitude of the reflected and transmitted wave from the whole multilayer stack and can be found by applying the interface continuity conditions of Maxwell's equations at each interface [228, 229]. Note that the form of G considered here is only valid for source and observation points outside the multilayer structure (while the case here has assumed $z' > 0$, the $z' < -d$ case is clearly of the same form by simply inverting the z -axis, with the Fresnel coefficients for transmission/reflection from the other side of the stack). It is possible to find \tilde{G} for sources and observation points inside the layers $-d < z, z' < 0$ by the same process of reflecting/transmitting the plane wave components of G_{dir} from the layer in which the source lies to the observation point. For such source or observation points, however, the Green's tensor cannot be expressed in terms of one overall generalised reflection and transmission coefficients for the whole system. For the applications considered in the thesis, the source and observation point will lie outside the layers.

The Fresnel coefficients depend on the permittivity and thickness of all the layers and can provide valuable information on the optical response of a given structure, in addition to their use in Eqs. (3.25) and (3.26) to calculate the Green's tensor. In particular, the poles of the Fresnel coefficients on the complex plane correspond to the wavevectors of the guided modes of the multilayer structure [230, 231, 232]. In the case of structures supporting SPPs, this means r_p and t_p have poles at $k_{\parallel} = k_{\text{SPP}}$ (the poles occur in the p -polarisation coefficients as SPPs are p -polarised). In addition to providing an approach to calculating k_{SPP} for arbitrary planar structures [231], this fact has significant consequences on the spatial distribution of scattered light in systems supporting SPPs, especially in the far field.

Far Field Green's Tensor

Using the inverse Fourier transform, $G(\mathbf{r}, \mathbf{r}')$ can be found from Eqs. (3.24)–(3.26) via

$$\begin{aligned}
 G(\mathbf{r}, \mathbf{r}') &= \int \frac{d^2 \mathbf{k}_{\parallel}}{(2\pi)^2} \tilde{G}(\mathbf{k}_{\parallel}; z, z') e^{i\mathbf{k}_{\parallel} \cdot (\boldsymbol{\rho} - \boldsymbol{\rho}')} \\
 &= \begin{cases} \int \frac{d^2 \mathbf{k}_{\parallel}}{(2\pi)^2} \left[\tilde{G}_{\text{dir}}(\mathbf{k}_{\parallel}; z, z') + \tilde{G}_{\text{ref}}(\mathbf{k}_{\parallel}; z, z') \right] e^{i\mathbf{k}_{\parallel} \cdot (\boldsymbol{\rho} - \boldsymbol{\rho}')} & z > 0 \\ \int \frac{d^2 \mathbf{k}_{\parallel}}{(2\pi)^2} \tilde{G}_{\text{tr}}(\mathbf{k}_{\parallel}; z, z') e^{i\mathbf{k}_{\parallel} \cdot (\boldsymbol{\rho} - \boldsymbol{\rho}')} & z < -d \end{cases} \quad (3.29)
 \end{aligned}$$

In all but the most simple of cases (e.g. a single perfectly reflecting interface), the inverse Fourier transforms given in Eqs. (3.25) and (3.26) required to calculate the real space Green's tensor cannot be calculated analytically. The Fourier space expressions are still valuable in allowing numerical calculation of $G(\mathbf{r}, \mathbf{r}')$ via numerically performing the inverse Fourier transform in Eq. (3.29). Beyond this, the Fourier space representation is important in deriving an analytic form of the real space Green's tensor in the far field. In the context of multilayer structures, far field is defined as the region where observation point is far away from the upper or lower interface ($z = 0$ and $z = -d$ planes respectively), on the scale of the wavelength, $|z| \gg \lambda_0$. By performing a stationary phase approximation [233] to Eq. (3.29), one finds G^∞ is given by [228]

$$G^\infty(\mathbf{r}, \mathbf{r}') = \begin{cases} -ik_d \cos \theta \tilde{G}(\mathbf{k}_{o\parallel}; 0^+; z') \frac{e^{ik_d r}}{r} e^{-i\mathbf{k}_{o\parallel} \cdot \boldsymbol{\rho}'} & z > 0 \\ ik_N \cos \theta \tilde{G}(\mathbf{k}_{o\parallel}, -d^-; z') \frac{e^{ik_N r}}{r} e^{-i\mathbf{k}_{o\parallel} \cdot \boldsymbol{\rho}'} & z < -d, \end{cases} \quad (3.30)$$

where θ is the polar angle of the position \mathbf{r} in spherical coordinates and $\mathbf{k}_o = \varepsilon(z)^{1/2} k_0 \hat{\mathbf{r}} = \varepsilon(z)^{1/2} k_0 (x/r, y/r, z/r)^T$ is a wavevector in the direction of the observation point with magnitude given by the wavenumber in the medium the observation point lies in. The notation h^\pm in the argument of the Green's tensor denotes the limit as the argument approaches h from above (plus sign) or below (minus sign). Since the far field Green's tensor is proportional to the Fourier space representation, the Fresnel reflection and transmission coefficients determine the distribution of light in the far field radiated by a dipole. In particular, a dipole radiates a higher intensity of light to an observation position with $k_{o\parallel}$ near a pole of $r_{s,p}$ and $t_{s,p}$ since $|G^\infty \mathbf{p}|^2 \propto |\tilde{G} \mathbf{p}|^2$ grows large near a pole. Physically this occurs as light radiated from the

dipole with the transverse component of the wavevector matching that of a guided mode couples strongly into the mode (e.g. SPPs), which then couples strongly into outgoing photons with the same transverse component, propagating in the direction of \mathbf{k}_o . Note, however, that just because a multilayer stack supports a guided mode and there are poles of the Fresnel coefficient does not necessarily imply such directional dipole intensity patterns. Firstly, the poles are in the complex plane, while $\mathbf{k}_{o\parallel}$ is a real vector. Therefore, if the poles are not near the real axis on the complex plane (i.e. they have large imaginary parts), then no observation point can be said to be near the pole and therefore there is no direction with an especially large value of the dipolar intensity $\propto |G^\infty \mathbf{p}|^2$. More generally, the squared magnitude of a function $f(z) \propto (z - a - ib)^{-1}$ with a (simple) complex pole $a + ib$ is Lorentzian along the real line centred on a and width b

$$|f(x)|^2 \propto \frac{1}{(x - a)^2 + b^2},$$

meaning that as a mode becomes more lossy and the imaginary part of the mode wavenumber increases, the high intensity peak in the dipole radiation pattern broadens over a wider range of angles. A second reason that not all guided modes give highly directional dipolar radiation is that the guided modes may not couple to propagating modes. To see this mathematically, note that $k_{o\parallel} \leq k_o = \varepsilon(z)^{1/2} k_0$. This means that, for a multilayer system supporting a guided mode, if the pole in $r_{s,p}$ and $t_{s,p}$ at k_{guided} occurs at a real part much larger than $\varepsilon(z)^{1/2} k_0$ (i.e. $k'_{\text{guided}} > \varepsilon(z)^{1/2} k_0$), the intensity peak will not be seen. Physically, guided modes with transverse wavenumbers larger than the photon wavenumber couple to evanescent modes in the dielectric, which exponentially decay and are not transmitted to the far field. In the case of SPPs ($k_{\text{guided}} = K_{\text{SPP}}$), it has already been established that $k'_{\text{SPP}} > k_d$, meaning that the pole does not give rise to a peak in the dipole radiation pattern in the upper dielectric. As discussed in Section 3.1.1, a multilayer configuration with sufficiently large ε_N does allow coupling of SPPs out into the lower dielectric in the form of leakage radiation, and thus the dipole radiates strongly in the direction of $\theta = \Theta_{\text{SPP}}$ corresponding to the pole of t_p . This ring has finite width $\Delta(\cos \theta) \sim k''_{\text{SPP}}$ since the mode is lossy and the pole does not lie exactly on the real axis. The ring is known as the leakage radiation ring. This feature and its dependence on the properties

of the Fresnel coefficient poles is demonstrated in Fig. 3.4, which shows the dipole intensity patterns $I_{dp}(\theta, \phi; \mathbf{r}') = (k_0^4/\varepsilon_0^2)|G_0(\mathbf{r}, \mathbf{r}')\mathbf{p}|^2$ (normalised to a maximum of 1) for some different gold film structures with $\mathbf{r}' = (0, 0, 50\text{nm})^T$. In the top row of Fig. 3.4, the z oriented dipole shows strongly directional radiation into the angle $\theta = \Theta_{\text{SPP}}$ (shown by the dashed blue line) at which $k_{o\parallel} = k'_{\text{SPP}}$ and there is a pole in t_p . The peak is less prominent in the in-plane dipole case, as while the z oriented dipole only radiates p -polarised light, some of the light radiated by the in-plane dipole is s -polarised and doesn't couple to SPPs. The radiation in the upper half space is more diffusely distributed over a wider angle since the SPP mode cannot couple into this half space. The second row shows an equivalent plot for $\lambda_0 = 600\text{nm}$. At this wavelength, gold absorbs light more strongly, giving a more lossy SPP (larger k''_{SPP}) and as a result, the peak around Θ_{SPP} is broader. In the water-gold-water case (bottom row of Fig. 3.4), the SPPs cannot couple into either space and the scattering is diffuse in both half spaces. The confinement of the dipolar radiated light to a narrow angular range, a manifestation of leakage radiation, has significant consequences for the scattering of SPPs, and therefore the proposed sensing setup.

3.3.3 Random Scattering Theory

The scattering theories described so far give a framework to calculate the scattered field from a known scattering configuration. It is very common, however, that the exact scattering configuration is not known. For example, in the context of the proposed sensor, the exact surface roughness profile or arrangement of scatterers on the surface is not generally known and would be impractical to measure for any given sensor surface. A more practical approach is to relate statistical properties of the scattering configuration, for example the mean square height deviation of the surface roughness or the density of scatterers on the surface, to statistical properties of the scattered light. Statistical approaches to describe light scattering are well established [21, 25, 234]. The same approaches, such as Eqs. (3.16) and Eq. (3.19), can be used, but rather than having a fixed scattering configuration from which a field is calculated, the scattering configuration (i.e. $\varepsilon_s(\mathbf{r})$) is treated as a random variable, and therefore the scattered field is also a random variable. In this context, by modelling $\varepsilon_s(\mathbf{r})$ as a random variable, scattering theory equations such as Eqs. (3.16) and (3.19) now provide a route to

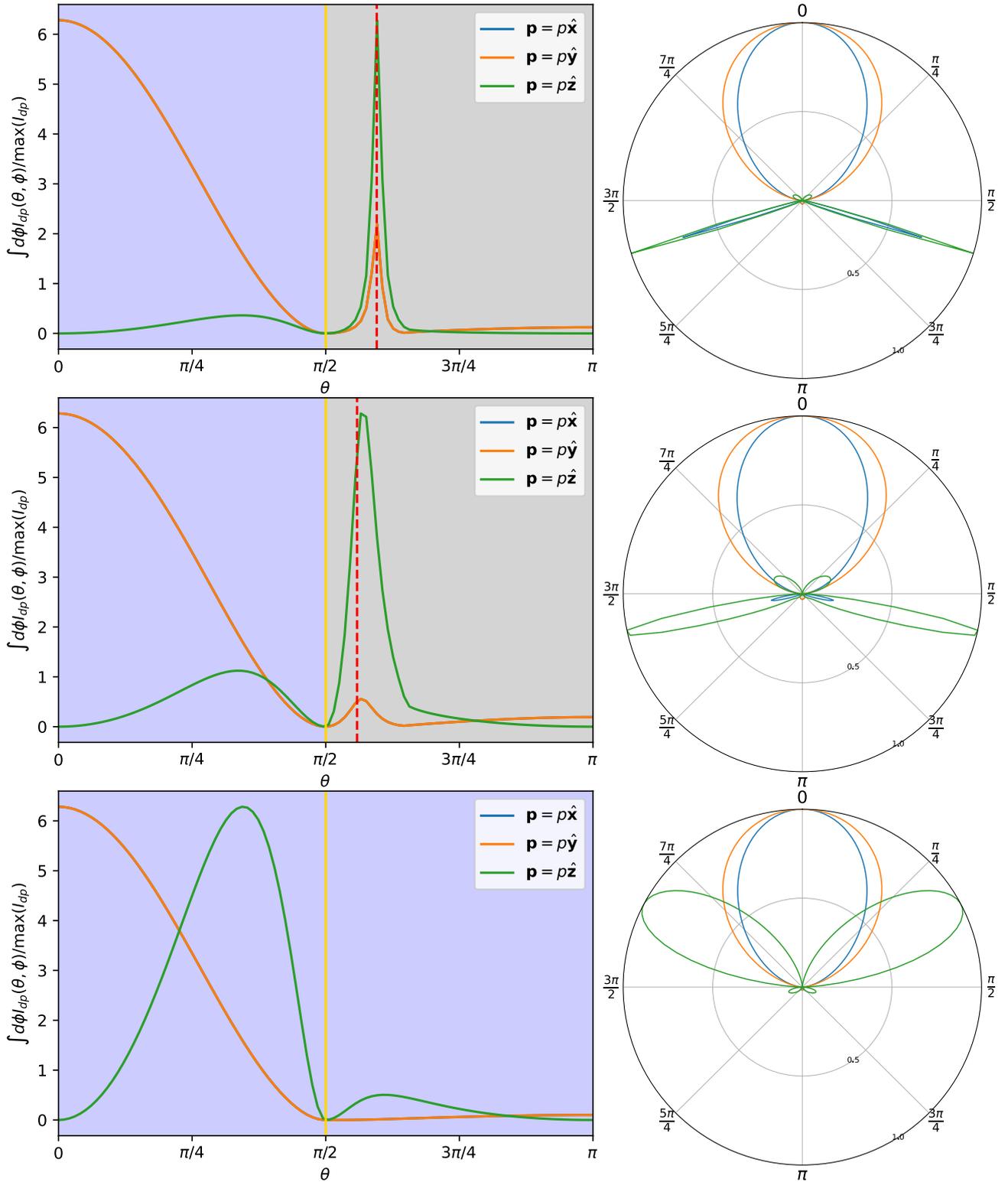


Figure 3.4: Far field dipole radiation patterns $I_{dp}(\theta, \phi) / \max_{\theta, \phi}(I_{dp})$ normalised to a maximum value of 1 for a dipole oriented in the x , y and z directions 50nm above a dielectric-gold-dielectric thin film structure with $d = 50\text{nm}$ thick gold film. The three variants of the structure plotted here are (top row) glass-gold-water at 650nm, (middle row) water-gold-water at 650nm and (bottom row) glass-gold-water at 600nm. The left column shows plots of $\int_0^{2\pi} I_{dp}(\theta, \phi) / \max_{\theta, \phi}(I_{dp}) d\phi$ to give the total intensity radiated in the θ direction, while the right column shows $I_{dp}(\theta, \phi = 0)$ in the (x, z) -plane. Note that the θ -integrated intensity is identical for the in-plane x and y oriented dipoles, and hence only one is plotted.

link the statistics of the field random variables to the statistics of the scattering configuration, defining the statistics of the field or Green's tensor, and hence also the statistics of quantities such as the intensity $I = |\mathbf{E}|^2$ derived from them. In most cases, deriving an exact statistical distribution for such quantities (e.g. \mathbf{E} , G or I) is not possible analytically, and instead methods have been developed to find (approximate) expressions for the moments of the field (or Green's tensor) [221, 24, 26]. Since $\mathbf{E}(\mathbf{r})$ is a random complex vector function, the definition of moments needs to be properly defined, as opposed to a real scalar random variable X where one can simply define the n th moment as $\langle X^n \rangle$ (throughout this thesis, angled brackets shall denote averaging over realisations of scattering disorder). The n th order moment of the field (or indeed any other function) will be taken to refer to any average of products of n copies of components of \mathbf{E} and its complex conjugate \mathbf{E}^\dagger , possibly evaluated at different positions. While this leaves many possible different moments at a given order, only some moments are of interest physically, such as the mean intensity $\langle I(\mathbf{r}) \rangle = \langle |\mathbf{E}(\mathbf{r})|^2 \rangle$, field autocorrelation $\langle \mathbf{E}(\mathbf{r})\mathbf{E}^\dagger(\mathbf{r}') \rangle$ or intensity autocorrelation $\langle I(\mathbf{r})I(\mathbf{r}') \rangle = \langle |\mathbf{E}(\mathbf{r})|^2 |\mathbf{E}(\mathbf{r}')|^2 \rangle$. In order to calculate these moments using Eq. (3.17), the Born expansion is substituted for each occurrence of the field, and thus the moment can be expressed as the sum of integrals of products of Green's tensors and some moment of $\varepsilon_s(\mathbf{r})$. These moments are known from the statistical model of $\varepsilon_s(\mathbf{r})$ (a common choice is a Gaussian random process, which will be discussed in Section 3.4.1). Even though the individual terms contributing to the moment can be expressed in terms of integrals of known functions, there are an infinite number of different terms, which are often conveniently represented in diagrammatic form. In certain limits (e.g. the average distance propagated between scattering events being much larger than a wavelength [26]), one can neglect classes of diagrams and find approximate analytic results for the moments. These approximations and the diagrammatic representations will be discussed in full detail in Chapter 5, which deals with multiple scattering effects. Since this chapter considers only the single scattering regime, it suffices to say that, since the Born series is truncated with a single scattering Born approximation, there are only a finite number of integrals to be summed to find any moments. While quantities averaged over scattering realisations can provide information about a scattering system, it should be noted, however, that a given static scattering configuration deterministically results in the

spatial dependence of the field $\mathbf{E}(\mathbf{r})$, and that this spatial distribution need not be similar to the average over many scattering configurations. The need for statistical modelling arises from the lack of knowledge of the scattering configuration as opposed to something fundamentally random in the system. In ergodic systems, averaged quantities (either spatially or temporally averaged in dynamic random media in which the scattering configuration changes with time) are equivalent to the averages over realisations [234]. While temporal ergodicity applies to dynamic random media [235, 236], spatial ergodicity applies much less commonly [237, 238]. Even in the absence of ergodicity, the averaged quantities can be useful to understand the distribution of values one could observe such as the spread and central tendency.

It can be useful to link the statistics of the scattering configuration to those of the scattered light, especially when the scattering medium is well understood such that an accurate statistical model of it can be developed, however there are many universal features of randomly scattered light that apply almost independent of the detailed statistics of the scattering process. Such randomly scattered light has a distinctive granular pattern of bright and dark spots known as a speckle pattern. This arises due to the fact that the randomly scattered waves can either add up in phase (giving a bright spot) or out of phase (giving a dark spot). Owing to this random interference effect, many statistical properties of speckle patterns can be understood and modelled through random phasor theory without the need to resort to modelling of the scattering process. To see this, consider the i th component E_i of the electric field randomly scattered from some arbitrary random scattering configuration, which can be expressed [25]

$$E_i(\mathbf{r}) = \sum_s A_s(\mathbf{r}) e^{i\Phi_s(\mathbf{r})} \quad (3.31)$$

where A_s and ϕ_s are the phase and amplitude gained by a wave propagating through scattering sequence s (i.e. a sequence of n scatterers with positions $\{\mathbf{r}_1 \dots \mathbf{r}_n\}$, with the sum over all possible n and all positions) and are random variables). Under the approximation that the phase and amplitude of different scattering paths are independent, and also that A_s and Φ_s are uncorrelated and Φ_s is uniformly distributed from 0 to 2π [21], Eq. (3.31) becomes a sum over a large number of independent random variables. One can therefore derive, by applying the Central

Limit Theorem, a Gaussian distribution for the field amplitude component E_i . For polarised light, the Gaussian field distribution results a exponential distribution for the intensity, with probability density function (PDF) $P(I) \propto \exp(-I/\langle I \rangle)$, $\langle I \rangle$ being the mean intensity averaged over speckle realisations [21]. Partially polarised light results in a modified intensity PDF [21]. A feature of this negative exponential intensity distribution, known as Rayleigh statistics, is that the standard deviation of the intensity is equal to the mean $\sigma_I = \langle (I - \langle I \rangle)^2 \rangle^{1/2} = \langle I \rangle$. This result is often stated in terms of the contrast \mathcal{C} defined by $\mathcal{C} = \sigma_I / \langle I \rangle$. Under Rayleigh statistics, fully polarised speckle has $\mathcal{C} = 1$, while partially polarised speckle has reduced contrast of the form $\mathcal{C} = [(1 + \mathcal{P}^2)/2]^{1/2}$ where \mathcal{P} is the degree of polarisation [21]. A wide range of speckle patterns exhibit Rayleigh statistics, even when generated by scattering environments of significantly different nature. It is possible, however, for effects such as strong multiple scattering or absorption to break down the assumptions on the phasor sum and give rise to non-Rayleigh statistics within the speckle pattern [239, 240, 241, 242, 243].

3.4 Scattering of SPPs

Now that the theoretical groundwork has been established, attention can be turned to describing the scattering occurring in the proposed sensing setup (Fig. 3.2). In this context, the background environment is defined to be the multilayer structure with planar interfaces, and the background Green's tensor G_0 is defined with respect to this. Where possible, the multilayer structure will be left general, but specific structures will be introduced and studied where required. The incident field \mathbf{E}_0 is taken as a plane wave SPP propagating in the x direction. Since it is important to consider the effects of absorption within plasmonic systems, two forms of the incident wave will be considered, a lossless and lossy form. These can both be expressed in the form

$$\mathbf{E}_0(\mathbf{r}) = A(x)\mathbf{E}_{\text{SPP}}(\mathbf{r}) \quad (3.32)$$

In the lossless case, $A(x) = 1$, however, for lossy waves, the exponential decay $\exp(-k''_{\text{SPP}}x)$ leads to a divergence in the field as $x \rightarrow -\infty$. To avoid this, it is assumed the wave is launched from $x = 0$, and therefore vanishes for $x < 0$. Thus, $A(x) = \Theta(x)$ for lossy SPPs with $k''_{\text{SPP}} \neq 0$,

where $\Theta(x)$ is the standard Heaviside step function defined to be 1 for $x \geq 0$ and 0 otherwise.

Two scattering configurations must be considered; scattering with no analyte particle near the surface and scattering with an analyte particle near the surface. Since measurement of light in the near field close to the surface of the multilayer stack is much more challenging than in the far field, most of the focus will be on studying the scattered light in the far field.

3.4.1 Scattering from a Rough Surface

With no analyte particle, the scattering from the rough surface is the only effect to be considered. The uppermost rough interface is defined by the surface $z = \zeta(\boldsymbol{\rho})$, rather than the smooth case where the interface is defined by the $z = 0$ plane, where $\zeta(\boldsymbol{\rho})$ is the surface roughness profile. As a result, the dielectric function $\varepsilon(\mathbf{r})$ is now given by

$$\varepsilon(\mathbf{r}) = \varepsilon_b(z) + \varepsilon_r(\mathbf{r}). \quad (3.33)$$

The background $\varepsilon_b(z)$ corresponds to the planar multilayer structure with ε_i being the permittivity in the i th layer such that

$$\varepsilon_b(z) = \varepsilon_i \quad z \in \text{layer } i. \quad (3.34)$$

The scattering perturbation due to surface roughness, $\varepsilon_r(\mathbf{r})$, is [244, 245, 246]

$$\varepsilon_r(\mathbf{r}) = (\varepsilon_m - \varepsilon_d) (\Theta(\zeta(x, y) - z) - \Theta(-z)) \quad (3.35)$$

Here, the $\varepsilon_{d,m}$ are the dielectric constants of the dielectric and metal respectively on either side of the interface. It has been assumed that $\zeta > -d_1$, where d_1 is the thickness of the uppermost metallic layer, such that the surface never dips below the first layer. Additionally, it is assumed there are no overhangs so that ζ is a single valued function. The exact form of ε_r is not very easy to work with mathematically, and it is convenient to expand to first order in ζ to give

$$\varepsilon_r(\mathbf{r}) \approx (\varepsilon_m - \varepsilon_d) \delta(z) \zeta(x, y) + \mathcal{O}(\zeta^2). \quad (3.36)$$

The expansion to first order in ζ is significantly easier to work with mathematically, and is a good approximation provided the typical height variation is much smaller than the free space wavelength $\lambda_0 = 2\pi/k_0$, as is usually the case for rough metal surfaces [247, 248]. Approaches avoiding this expansion have been developed, for example using curvilinear coordinate transforms [249]. Such methods, however tend to mix the effects up to a given order of ζ into different terms, whereas the expansion here ensures all effects up to $\mathcal{O}(\zeta)$ are accounted for, while extending the expansion to higher order terms offers a systematic approach to account for different orders [244]. Using this first order expansion in Eq. (3.36), the field scattered from the surface, \mathbf{E}_ζ , is given by Eq. (3.18) with $\varepsilon_s = \varepsilon_r$ as

$$\mathbf{E}_\zeta(\mathbf{r}) = (\varepsilon_m - \varepsilon_d) \frac{\omega^2}{c^2} \int d^2 \boldsymbol{\rho}' G_0(\mathbf{r}, \boldsymbol{\rho}', z' = 0) \zeta(\boldsymbol{\rho}') \mathbf{E}_0(\boldsymbol{\rho}', z' = 0). \quad (3.37)$$

Due to the $\delta(z)$ in Eq. (3.36), the integral is evaluated over the 2D $z' = 0$ plane, and from now on the $z' = 0$ will be implicitly implied by the $\boldsymbol{\rho}'$ 2D argument. While this restriction to a planar interface simplifies the integration, there is some subtlety with regards to the value of the Green's tensor and incident field at the interface when $z' = 0$. Since these quantities (specifically their normal components) are discontinuous at the interface, the value at $z' = 0$ is not well defined. It has been shown that, correct to first order in the roughness, that one should use the limit from above, $z \rightarrow 0^+$, for one of G_0 or \mathbf{E}_0 and the limit from below, $z \rightarrow 0^-$, for the other [250]. It does not matter whether one uses $G_0(\mathbf{r}, \mathbf{0}^-) \mathbf{A}_{\text{SPP}}^+$ or $G_0(\mathbf{r}, \mathbf{0}^+) \mathbf{A}_{\text{SPP}}^-$, since the boundary conditions on G_0 and \mathbf{E}_0 ensure they are equivalent. Using the far field form of the Green's tensor from Eq. (3.30) and substituting for \mathbf{E}_0 results in

$$\mathbf{E}_\zeta(\mathbf{r}) = (\varepsilon_m - \varepsilon_d) \frac{\omega^2}{c^2} G_0(\mathbf{r}, \mathbf{0}^-) \mathbf{A}_{\text{SPP}}^+ \int d^2 \boldsymbol{\rho}' A(x') \zeta(\boldsymbol{\rho}') \exp(-i\Delta \mathbf{k}_\parallel \boldsymbol{\rho}') \exp(-k''_{\text{SPP}} x') \quad (3.38)$$

where $\Delta \mathbf{k} = \varepsilon_b(z)^{1/2} k_0 \hat{\mathbf{r}} - k'_{\text{SPP}} \hat{\mathbf{x}}$ is the wavevector shift between the incident SPP wavevector and \mathbf{k}_o (the outgoing wavevector propagating towards \mathbf{r} in the far field) and $\Delta \mathbf{k}_\parallel$ is its transverse component. The scattered field is dipolar like, arising from the $G(\mathbf{r}, \mathbf{0}^-) \mathbf{A}_{\text{SPP}}^+$ factor, but modified by the random nature of the surface due to the integral. This integral is a continuous analog of the random phasor sum in Eq. (3.31). In the lossless case, $k''_{\text{SPP}} = 0$ and $A(x') = 1$,

this integral reduces to a Fourier transform of the surface profile, $\tilde{\zeta}(\Delta\mathbf{k}_{\parallel})$. With a relatively simple link between the surface profile and the scattered field established, a statistical model of the surface profile can be introduced.

Surface Roughness Statistics

The surface is assumed to be a zero-mean Gaussian random process, a common model random for surface roughness [220, 244, 251]. This means the surface profile statistics are defined by

$$\langle \zeta(\boldsymbol{\rho}) \rangle = 0 \quad (3.39)$$

$$\langle \zeta(\boldsymbol{\rho})\zeta(\boldsymbol{\rho}') \rangle = h^2 C(\boldsymbol{\rho} - \boldsymbol{\rho}'), \quad (3.40)$$

where angled brackets denote averaging over an ensemble of random surface realisations and $h^2 = \langle \zeta(\boldsymbol{\rho})^2 \rangle$ is the root mean square (RMS) height deviation, parameterising the typical height of a peak or depth of a trough in the surface profile. The function $C(\boldsymbol{x})$ is the surface correlation function, normalised so that $C(\mathbf{0}) = 1$. As $C(\boldsymbol{x})$ depends only on the separation $\boldsymbol{\rho} - \boldsymbol{\rho}'$, statistical homogeneity of the surface has been assumed, meaning that all points have the same probability distribution for the height and translational invariance is restored after averaging over realisations. One form of C commonly used is $C(\boldsymbol{x}) = \exp\left(-\frac{|\boldsymbol{x}|^2}{2a^2}\right)$ [14, 251, 252], which has built in the additional assumption of statistical isotropy, so that $C(|\boldsymbol{x}|)$ only depends on the magnitude of the separation, though anisotropic forms may be relevant [253]. The width of the function is given by the surface correlation length, a , which parameterises the typical width of a peak/trough on the surface. For typical rough metal surfaces, the correlation length is much shorter than the optical wavelength, $a \ll \lambda_0$ [247, 248]. In this case, a δ -function limit can be taken so that $C(\boldsymbol{\rho} - \boldsymbol{\rho}') = 2\pi a^2 \delta(\boldsymbol{\rho} - \boldsymbol{\rho}')$. The convenience of this form in deriving analytic results means it will be commonly used in this chapter. For a Gaussian process, Eqs. (3.39) and (3.40) entirely define the statistics of the surface, as any higher order moment with p factors of ζ either vanishes (p odd) or can be split into a sum over all possible products of pairwise correlations by Wick's theorem (also referred to as Isserlis' theorem) (p even) [24, 254, 255]. For

example, $\langle \zeta(\boldsymbol{\rho}_1)\zeta(\boldsymbol{\rho}_2)\zeta(\boldsymbol{\rho}_3) \rangle = 0$ while the fourth order moment can be split according to

$$\langle \zeta(\boldsymbol{\rho}_1)\zeta(\boldsymbol{\rho}_2)\zeta(\boldsymbol{\rho}_3)\zeta(\boldsymbol{\rho}_4) \rangle = \langle \zeta(\boldsymbol{\rho}_1)\zeta(\boldsymbol{\rho}_2) \rangle \langle \zeta(\boldsymbol{\rho}_3)\zeta(\boldsymbol{\rho}_4) \rangle + \langle \zeta(\boldsymbol{\rho}_1)\zeta(\boldsymbol{\rho}_3) \rangle \langle \zeta(\boldsymbol{\rho}_2)\zeta(\boldsymbol{\rho}_4) \rangle \quad (3.41)$$

$$+ \langle \zeta(\boldsymbol{\rho}_1)\zeta(\boldsymbol{\rho}_4) \rangle \langle \zeta(\boldsymbol{\rho}_2)\zeta(\boldsymbol{\rho}_3) \rangle. \quad (3.42)$$

Since Eq. (3.38) gives a relationship between ζ and \mathbf{E}_ζ , the statistics of \mathbf{E}_ζ are also defined by Eqs. (3.39) and (3.40). More precisely, the fact that Eq. (3.38) is linear in ζ means any moment of \mathbf{E}_ζ is a linear combination of moments of ζ , which can be calculated through use of Wick's theorem and Eq. (3.40).

Scattered Field Moments

The linear relationship between \mathbf{E}_ζ and ζ means \mathbf{E}_ζ is also a Gaussian random process defined by its first and second moments. Therefore it is valuable to calculate these moments. From Eqs. (3.37) and (3.39), it immediately follows that $\langle \mathbf{E}_\zeta(\mathbf{r}) \rangle = \mathbf{0}$. This, of course, does not mean there is no scattered field, but rather, since \mathbf{E}_ζ is a complex quantity, the field at \mathbf{r} for different realisations of the surface cancel out on average. This is true also for a fully developed speckle field [25]. Since $\mathbf{E}_\zeta(\mathbf{r})$ is a complex random vector, there exist two second moment matrices given by $\langle \mathbf{E}(\mathbf{r})\mathbf{E}^\dagger(\mathbf{r}') \rangle$ and $\langle \mathbf{E}(\mathbf{r})\mathbf{E}^T(\mathbf{r}') \rangle$. The first of these moments is known as the mutual coherency matrix and finds use in the study of partially polarised light [256, 257]. The (i, j) elements of these matrices, corresponding to the mean of the product of the i th component and j th component of the field, are

$$\langle E_{\zeta,i}(\mathbf{r})E_{\zeta,j}^*(\mathbf{r}') \rangle = |B| \int d^2\boldsymbol{\rho}_1 d^2\boldsymbol{\rho}_2 \sum_k G_{0,ik}(\mathbf{r}, \boldsymbol{\rho}_1) E_{0,k}(\boldsymbol{\rho}_1) C(\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2) \sum_l G_{0,jl}^*(\mathbf{r}', \boldsymbol{\rho}_2) E_{0,l}^*(\boldsymbol{\rho}_2) \quad (3.43)$$

$$\langle E_{\zeta,i}(\mathbf{r})E_{\zeta,j}(\mathbf{r}') \rangle = B \int d^2\boldsymbol{\rho}_1 d^2\boldsymbol{\rho}_2 \sum_k G_{0,ik}(\mathbf{r}, \boldsymbol{\rho}_1) E_{0,k}(\boldsymbol{\rho}_1) C(\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2) \sum_l G_{0,jl}(\mathbf{r}', \boldsymbol{\rho}_2) E_{0,l}(\boldsymbol{\rho}_2). \quad (3.44)$$

The constant prefactor $B = (\varepsilon_m - \varepsilon_d)^2 h^2 k_0^4$ has been introduced. Index notation is used, with $E_{\zeta,i}$ denoting the i th Cartesian component of the vector \mathbf{E}_ζ (and similar for $E_{0,i}$) while $G_{0,ij}$

denotes the (i, j) element of the tensor G_0 . Using the far field result of Eq. (3.38), Eqs. (3.43) and (3.44) when \mathbf{r} and \mathbf{r}' are in the far field can be expressed as

$$\langle E_{\zeta,i}(\mathbf{r})E_{\zeta,j}^*(\mathbf{r}') \rangle = |B| \sum_{k,l} G_{0,ik}(\mathbf{r}, \mathbf{0}^-) A_{\text{SPP},k}^+ G_{0,jl}^*(\mathbf{r}', \mathbf{0}^-) A_{\text{SPP},l}^{+*} \mathcal{I}(\Delta \mathbf{k}_{\parallel}, \Delta \mathbf{k}'_{\parallel}) \quad (3.45)$$

$$\langle E_{\zeta,i}(\mathbf{r})E_{\zeta,j}(\mathbf{r}') \rangle = B \sum_{k,l} G_{0,ik}(\mathbf{r}, \mathbf{0}^-) A_{\text{SPP},k}^+ G_{0,jl}(\mathbf{r}', \mathbf{0}^-) A_{\text{SPP},l}^+ \mathcal{I}(\Delta \mathbf{k}_{\parallel}, -\Delta \mathbf{k}'_{\parallel}), \quad (3.46)$$

where the integral function $\mathcal{I}(\mathbf{q}, \mathbf{q}')$ has been defined as

$$\mathcal{I}(\mathbf{q}, \mathbf{q}') = \int d^2 \boldsymbol{\rho}_1 d^2 \boldsymbol{\rho}_2 C(\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2) A(x_1) A(x_2) \exp[-i(\mathbf{q} \cdot \boldsymbol{\rho}_1 - \mathbf{q}' \cdot \boldsymbol{\rho}_2)] \exp[-k''_{\text{SPP}}(x_1 + x_2)] \quad (3.47)$$

The transverse wavevector shift $\Delta \mathbf{k}_{\parallel}$, as defined in Eq. (3.38), is a function of observation position, $\Delta \mathbf{k}_{\parallel}(\mathbf{r})$, though this dependence is left implicit. Similarly, $\Delta \mathbf{k}'_{\parallel}$ is defined in the same manner but for the second observation position \mathbf{r}' , i.e. $\Delta \mathbf{k}'_{\parallel} = \Delta \mathbf{k}_{\parallel}(\mathbf{r}')$. The results of Eqs. (3.45) and (3.46) will form the building blocks to study the statistical properties of the speckle pattern formed by scattering from random roughness. Even on their own, these results reveal some interesting properties about the speckle pattern which bear commenting on.

Non-Circular Statistics

A complex random variable Z is said to be circular if $e^{i\theta}Z$ and Z are identically distributed for any phase shift θ [258]. A speckle field that obeys Rayleigh statistics is a relevant example of a circular complex random variable. Circularity implies several further properties of Z , for example the independence of the phase and amplitude distribution, and a uniform phase distribution over $[0, 2\pi]$. Due to the fact $\langle Z^2 \rangle = \exp(2i\theta) \langle Z^2 \rangle$ for any θ , the pseudo covariance vanishes, $\langle Z^2 \rangle = 0$, for any circular complex random variable. From Eq. (3.46), it can be concluded that $\langle E_{\zeta,i} E_{\zeta,j} \rangle \neq 0$ unless $\mathcal{I}(\Delta \mathbf{k}_{\parallel}, -\Delta \mathbf{k}'_{\parallel}) = 0$, ignoring the trivial case where the Green's tensor factors vanish which simply means no light is scattered to \mathbf{r} or \mathbf{r}' . Circularity of the speckle field is a necessary (though not sufficient) condition for Rayleigh statistics. It has been shown that absorption [241, 242] or multiple scattering effects [239, 240] can lead to

non-circular or non-Rayleigh statistics. Clearly, as Eq. (3.46) was derived under the single scattering Born approximation, the non-circularity here cannot arise due to multiple scattering effects. To understand the physical origin of the non-circularity here, it is useful to consider the simplification of Eq. (3.47) under two approximations. Firstly, the short surface correlation length δ -function approximation, with a lossy wave ($A(x) = \Theta(x)$), which simplifies to integral to

$$\mathcal{I}(\Delta\mathbf{k}_{\parallel}, \Delta\mathbf{k}'_{\parallel}) = 2\pi a^2 \int_0^{\infty} dx_1 \int_{-\frac{L_y}{2}}^{\frac{L_y}{2}} dy_1 \exp \{ [-i(\Delta k_x - \Delta k'_x) - 2k''_{\text{SPP}}] x_1 - i(\Delta k_y - \Delta k'_y) y_1 \}. \quad (3.48)$$

Note that the y_1 integral has been restricted over a finite width L_y . In general, the plane wave SPP is considered to be infinite in width, and $L_y \rightarrow \infty$ would render the y_1 integral as a δ -function. Any real physical system, however, would have a finite width, either due to the finite illumination area or finite roughness patch (i.e. $\zeta(\mathbf{r}) = 0$ outside a given region). As such, a finite limit is kept on the integral in order to capture this effect, while also avoiding any divergence problems associated with the δ -function. The limit $L_y \rightarrow \infty$ can always be taken where appropriate. Performing the integration results in

$$\mathcal{I}(\Delta\mathbf{k}_{\parallel}, \pm\Delta\mathbf{k}'_{\parallel}) = 2\pi a^2 \frac{L_y \operatorname{sinc} [(\Delta k_y \mp \Delta k'_y)L_y/2]}{-i(\Delta k_x \mp \Delta k'_x) - 2k''_{\text{SPP}}}, \quad (3.49)$$

where the sinc function is defined by $\operatorname{sinc}(x) = \sin(x)/x$. In this short surface correlation length approximation, it can be explicitly seen that the integral in Eq. (3.46) does not vanish and the non-circularity of the field is made explicit. To quantify the degree of non-circularity, a quantity $f_C(\mathbf{r})$ can be defined as

$$f_C(\mathbf{r}) = \frac{|\langle E_{\zeta,i}(\mathbf{r}) E_{\zeta,j}(\mathbf{r}) \rangle|}{\langle E_{\zeta,i}(\mathbf{r}) E_{\zeta,j}^*(\mathbf{r}) \rangle}. \quad (3.50)$$

This acts as a measure of the size of the pseudo-covariance of two components of the electric field at a position \mathbf{r} relative to the covariance of the same two components. If the field is circular, $f_C = 0$, while $f_C \ll 1$ indicates that it is in some sense close to circular and could be

approximated as such. Using Eqs. (3.49) and (3.49), f_C takes the form

$$f_C(\mathbf{r}) = \frac{|\text{sinc}(\Delta k_y L_y)|}{\sqrt{1 + \left(\frac{\Delta k_x}{k''_{\text{SPP}}}\right)^2}}. \quad (3.51)$$

Notably, for $\Delta k_x \neq 0$, $f_C(\mathbf{r}) \rightarrow 0$ in the low loss limit $k_{\text{SPP}} \rightarrow 0$. In general, increased loss (i.e. increased k''_{SPP}) increases f_C and thus it can be concluded that absorption plays a role in the non-circular statistics. For $\Delta k_x = 0$, even the lossless limit does not give a circular variable. The second limiting case to further the understanding is to consider a general correlation function, but in the lossless case, with the incident SPP illuminating a finite area of roughness A (assumed to be a rectangular area of sides L_x and L_y) that is sufficiently large that all integrals over $\boldsymbol{\rho}_{1,2}$ are well approximated by being over an infinite area. In this case, Eq. (3.47) becomes a Fourier transform and simplifies to

$$\begin{aligned} \mathcal{I}(\Delta \mathbf{k}_{\parallel}, \pm \Delta \mathbf{k}'_{\parallel}) &= \int d^2 \boldsymbol{\rho}_1 d^2 \boldsymbol{\rho}_2 C(\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2) e^{-i \Delta \mathbf{k}_{\parallel} \cdot \boldsymbol{\rho}_1} e^{i \Delta \mathbf{k}'_{\parallel} \cdot \boldsymbol{\rho}_2} \\ &= \tilde{C}(\Delta \mathbf{k}_{\parallel}) (2\pi)^2 \delta_L(\Delta \mathbf{k}_{\parallel} \mp \Delta \mathbf{k}'_{\parallel}), \end{aligned} \quad (3.52)$$

where δ_L is a nascent delta function defined by $(2\pi)^2 \delta_L(\mathbf{q}) = L_x \text{sinc}(q_x L_x/2) L_y \text{sinc}(q_y L_y/2)$, which tends to a δ function as $L_{x,y} \rightarrow \infty$. From this, $f_C(\mathbf{r})$ in the lossless case is given by

$$f_C(\mathbf{r}) = |\text{sinc}(\Delta k_x L_x) \text{sinc}(\Delta k_y L_y)|. \quad (3.53)$$

It can be seen that f_C is maximised around $\Delta \mathbf{k}_{\parallel} = \mathbf{0}$, and decays over a scale $\sim 1/L_{x,y}$ determined by the size of the illuminated rough area. Outside of this region, the speckle field is approximately circular. Absorption acts to limit the illuminated area to a length scale L_{SPP} , and this is why absorption also gives non-circular statistics, with Eq. (3.51) decaying with $\Delta k_x L_{\text{SPP}}$, though now as a Lorentzian as opposed to sinc type decay.

Intensity Statistics

The intensity statistics can be derived from the moments above. The mean intensity $\langle I_\zeta(\mathbf{r}) \rangle = \langle |\mathbf{E}_\zeta(\mathbf{r})|^2 \rangle = \sum_i \langle |E_{\zeta,i}(\mathbf{r})|^2 \rangle$ is simply

$$\langle I_\zeta(\mathbf{r}) \rangle = |B| |G_0(\mathbf{r}, \mathbf{0}^-) \mathbf{A}_{\text{SPP}}^+|^2 \mathcal{I}(\Delta \mathbf{k}_\parallel, \Delta \mathbf{k}_\parallel). \quad (3.54)$$

The average intensity radiated to the far field can be seen to match that of a dipole at the surface oriented along the SPP polarisation, modified by the surface correlation integral. This already points to a significant feature of random SPP scattering. As discussed in Section 3.3.2 and shown in Fig. 3.4, the dipole intensity pattern is strongly confined to a narrow range of angles near Θ_{SPP} due to the leakage radiation condition. As a result, the randomly scattered light, while still a speckle, is confined to this leakage radiation ring. The ring shaped speckle is in fact evident in Eq. (3.38), with the Green's tensor factor giving the ring like profile, and the ζ dependent integral being a random function of \mathbf{r} giving the random speckle behaviour.

To further characterise the intensity statistics at a point \mathbf{r} , the variance defined by $\sigma_\zeta^2(\mathbf{r}) = \langle I_\zeta^2(\mathbf{r}) \rangle - \langle I_\zeta(\mathbf{r}) \rangle^2$ can be calculated via calculating the second moment of the intensity as

$$\begin{aligned} \langle I_\zeta^2(\mathbf{r}) \rangle &= \sum_{i,j=1}^3 \langle E_{\zeta,i}(\mathbf{r}) E_{\zeta,i}^*(\mathbf{r}) E_{\zeta,j}(\mathbf{r}) E_{\zeta,j}^*(\mathbf{r}) \rangle \\ &= \langle I_\zeta(\mathbf{r}) \rangle^2 + \sum_{i,j=1}^3 |\langle E_{\zeta,i}(\mathbf{r}) E_{\zeta,j}(\mathbf{r}) \rangle|^2 + |\langle E_{\zeta,i}(\mathbf{r}) E_{\zeta,j}^*(\mathbf{r}) \rangle|^2, \end{aligned} \quad (3.55)$$

where Wick's theorem has been used to split up the mean of the product of four fields into a sum of pairwise partitions. Using Eqs. (3.43) and (3.44) evaluated at $\mathbf{r}' = \mathbf{r}$, the variance is found to be

$$\begin{aligned} \sigma_\zeta^2(\mathbf{r}) &= \sum_{i,j=1}^3 |\langle E_{\zeta,i}(\mathbf{r}) E_{\zeta,j}^*(\mathbf{r}) \rangle|^2 + |\langle E_{\zeta,i}(\mathbf{r}) E_{\zeta,j}(\mathbf{r}) \rangle|^2 \\ &= \langle I_\zeta(\mathbf{r}) \rangle^2 (1 + |f_C(\mathbf{r})|^2). \end{aligned} \quad (3.56)$$

The speckle contrast, here given by $\sigma_\zeta / \langle I_\zeta \rangle$, deviates from the classic Rayleigh statistics value

of unity owing to the non-circularity. When $|f_C(\mathbf{r})| \ll 1$, unity contrast is observed as in fully developed speckle. Another quantity that helps characterise the speckle pattern properties is the intensity autocorrelation defined as $F_\zeta(\mathbf{r}, \mathbf{r}') = \langle I_\zeta(\mathbf{r})I_\zeta(\mathbf{r}') \rangle - \langle I_\zeta(\mathbf{r}) \rangle \langle I_\zeta(\mathbf{r}') \rangle$. Since this is a fourth order (central) moment of the field, it can be calculated with the same approach as Eq. (3.55) using

$$\begin{aligned} \langle I_\zeta(\mathbf{r})I_\zeta(\mathbf{r}') \rangle &= \sum_{i,j=1}^3 \langle E_{\zeta,i}(\mathbf{r})E_{\zeta,i}^*(\mathbf{r})E_{\zeta,j}(\mathbf{r}')E_{\zeta,j}^*(\mathbf{r}') \rangle \\ &= \langle I_\zeta(\mathbf{r}) \rangle \langle I_\zeta(\mathbf{r}') \rangle + \sum_{i,j=1}^3 \left[|\langle E_{\zeta,i}(\mathbf{r})E_{\zeta,j}(\mathbf{r}') \rangle|^2 + |\langle E_{\zeta,i}(\mathbf{r})E_{\zeta,j}^*(\mathbf{r}') \rangle|^2 \right], \end{aligned} \quad (3.57)$$

The autocorrelation is therefore given by

$$F_\zeta(\mathbf{r}, \mathbf{r}') = |B|^2 |G_0(\mathbf{r}, \mathbf{0}^-) \mathbf{A}_{\text{SPP}}^+|^2 |G_0(\mathbf{r}', \mathbf{0}^-) \mathbf{A}_{\text{SPP}}^+|^2 \left[|\mathcal{I}(\Delta \mathbf{k}_\parallel, \Delta \mathbf{k}'_\parallel)|^2 + |\mathcal{I}(\Delta \mathbf{k}_\parallel, -\Delta \mathbf{k}'_\parallel)|^2 \right]. \quad (3.58)$$

The Green's tensor factors arise from the different dipolar intensities radiated to points \mathbf{r} and \mathbf{r}' , rather than any statistical properties. A normalised autocorrelation $\mathcal{F}_\zeta(\mathbf{r}, \mathbf{r}') = F_\zeta(\mathbf{r}, \mathbf{r}') / (\langle I_\zeta(\mathbf{r}) \rangle \langle I_\zeta(\mathbf{r}') \rangle)$ eliminates this dependence on the mean intensity at each point and can be expressed

$$\mathcal{F}_\zeta(\mathbf{r}, \mathbf{r}') = \frac{|\mathcal{I}(\Delta \mathbf{k}_\parallel, \Delta \mathbf{k}'_\parallel)|^2 + |\mathcal{I}(\Delta \mathbf{k}_\parallel, -\Delta \mathbf{k}'_\parallel)|^2}{\mathcal{I}(\Delta \mathbf{k}_\parallel, \Delta \mathbf{k}_\parallel)^2}. \quad (3.59)$$

In the lossless case, using Eqs. (3.52) yields

$$\begin{aligned} \mathcal{F}_\zeta(\mathbf{r}, \mathbf{r}') &= \text{sinc}^2 \left(\frac{(\Delta k_x - \Delta k'_x)L_x}{2} \right) \text{sinc}^2 \left(\frac{(\Delta k_y - \Delta k'_y)L_y}{2} \right) \\ &\quad + \text{sinc}^2 \left(\frac{(\Delta k_x + \Delta k'_x)L_x}{2} \right) \text{sinc}^2 \left(\frac{(\Delta k_y + \Delta k'_y)L_y}{2} \right). \end{aligned} \quad (3.60)$$

Similarly, the lossy short surface correlation length case (Eqs. (3.49) and (3.49)) gives

$$\mathcal{F}_\zeta(\mathbf{r}, \mathbf{r}') = \frac{\text{sinc}^2 [(\Delta k_y - \Delta k'_y)L_y/2]}{1 + (\Delta k_x - \Delta k'_x)^2 L_{\text{SPP}}^2} + \frac{\text{sinc}^2 [(\Delta k_y + \Delta k'_y)L_y/2]}{1 + (\Delta k_x + \Delta k'_x)^2 L_{\text{SPP}}^2} \quad (3.61)$$

As noted in studying the non-circularity function f_C , the effect of absorption is similar to that of having a system of finite length L_{SPP} , in that in both cases the autocorrelation decays with the outgoing wavevector difference on the scale $1/L_{x,y}$ (with the $L_x = L_{\text{SPP}}$ in presence of absorption) though again absorption gives a Lorentzian as opposed to sinc decay. The peak in Eqs. (3.60) and (3.61) at $\Delta\mathbf{k}_{\parallel} = \Delta\mathbf{k}'_{\parallel}$ is unsurprising as this corresponds to the correlation of light scattered in the same direction, which are by definition statistically correlated (in the far field). The secondary peak at $\Delta\mathbf{k}_{\parallel} = -\Delta\mathbf{k}'_{\parallel}$ has a slightly more detailed explanation. From Eq. (3.38), one can see that, since ζ is real, \mathbf{E}_{ζ} at the observation point corresponding to $-\Delta\mathbf{k}_{\parallel}$ is related to \mathbf{E}_{ζ} at $\Delta\mathbf{k}_{\parallel}$ via the Hermitian symmetry of the Fourier transform integral. Specifically, they are related via

$$\mathbf{E}_{\zeta}(-\Delta\mathbf{k}_{\parallel}) = M_G \mathbf{E}_{\zeta}^*(\Delta\mathbf{k}_{\parallel}), \quad (3.62)$$

where the argument of \mathbf{E}_{ζ} has been set to \mathbf{k}_{\parallel} for clarity here, noting that \mathbf{r} fully determines \mathbf{k}_{\parallel} , and the matrix M_G is required to relate the different phase, amplitude and polarisation of the Green's tensor factors, defined such that

$$(\varepsilon_m - \varepsilon_d) \frac{\omega^2}{c^2} G_0(\mathbf{r}', \mathbf{0}^-) \mathbf{A}_{\text{SPP}}^+ = M_G (\varepsilon_m^* - \varepsilon_d) \frac{\omega^2}{c^2} G_0^*(\mathbf{r}, \mathbf{0}^-) \mathbf{A}_{\text{SPP}}^{+*}. \quad (3.63)$$

Crucially, M_G is entirely determined by deterministic functions that do not depend on the random surface roughness ζ . As such, $\mathbf{E}_{\zeta}(-\Delta\mathbf{k}_{\parallel})$ and $\mathbf{E}_{\zeta}(\Delta\mathbf{k}_{\parallel})$ are mutually dependent and thus the correlation between these two points is maximised.

The length scale over which F_{ζ} decays is a measure of a speckle size: points closer than this length are correlated as they lie within the same speckle, while points separated by more than this are uncorrelated (or have a much weaker correlation) since they are not in the same speckle [259]. From Eqs. (3.60) and (3.61), the speckle size is inversely proportional to the size of the illuminated area, a property widely observed in speckle patterns [21]. Taking \mathbf{r} and \mathbf{r}' in the leakage ring, $\Delta k_y = k'_{\text{SPP}} \sin \phi$, $\Delta k_x = k'_{\text{SPP}} (\cos \phi - 1)$ and similar for primed coordinate, the

leakage radiation ring autocorrelation is found to be, in the lossless and lossy cases respectively

$$\begin{aligned} \mathcal{F}_\zeta(\mathbf{r}, \mathbf{r}') &= \text{sinc}^2 \left(\frac{(\cos \phi - \cos \phi') k'_{\text{SPP}} L_x}{2} \right) \text{sinc}^2 \left(\frac{(\sin \phi - \sin \phi') k'_{\text{SPP}} L_y}{2} \right) \\ &+ \text{sinc}^2 \left(\frac{(\cos \phi + \cos \phi' + 2) k'_{\text{SPP}} L_x}{2} \right) \text{sinc}^2 \left(\frac{(\sin \phi + \sin \phi') k'_{\text{SPP}} L_y}{2} \right) \end{aligned} \quad (3.64)$$

$$\mathcal{F}_\zeta^{LR}(\phi, \phi') = \frac{\text{sinc}^2 [(\sin \phi - \sin \phi') k'_{\text{SPP}} L_y / 2]}{1 + (\cos \phi - \cos \phi')^2 k_{\text{SPP}}'^2 L_{\text{SPP}}^2} + \frac{\text{sinc}^2 [(\sin \phi + \sin \phi') k'_{\text{SPP}} L_y / 2]}{1 + (\cos \phi + \cos \phi' - 2)^2 k_{\text{SPP}}'^2 L_{\text{SPP}}^2}. \quad (3.65)$$

Within the leakage ring, there are no two points with $\Delta \mathbf{k}'_{\parallel} = -\Delta \mathbf{k}_{\parallel}$, as it would require $\cos \phi = \cos \phi'$ and $\sin \phi = \sin \phi'$ simultaneously. Assuming that $L_{x,y} \gg \lambda_{\text{SPP}}$ and $L_{\text{SPP}} \gg \lambda_{\text{SPP}}$, this means the second term becomes negligible and one has

$$\mathcal{F}_\zeta(\mathbf{r}, \mathbf{r}') \approx \text{sinc}^2 \left(\frac{(\cos \phi - \cos \phi') k'_{\text{SPP}} L_x}{2} \right) \text{sinc}^2 \left(\frac{(\sin \phi - \sin \phi') k'_{\text{SPP}} L_y}{2} \right) \quad (3.66)$$

$$\mathcal{F}_\zeta^{LR}(\phi, \phi') \approx \frac{\text{sinc}^2 [(\sin \phi - \sin \phi') k'_{\text{SPP}} L_y / 2]}{1 + (\cos \phi - \cos \phi')^2 k_{\text{SPP}}'^2 L_{\text{SPP}}^2}, \quad (3.67)$$

indicating an angular speckle size $\sim (k'_{\text{SPP}} \max(L_x, L_y))^{-1}$.

Unsurprisingly, the light scattered from the random rough surface has many properties similar to that of fully developed speckle, but there are a few notable differences. The non-circularity of the field and the fact the contrast is not unity are examples of this, but the fact that the mean intensity varies in space according to a dipolar intensity profile is perhaps the most significant. As discussed in Section 3.3.2, these can exhibit strong directional dependence in systems supporting SPPs. This position dependence of the mean intensity shows that the speckle pattern is not statistically homogeneous, and therefore spatial averaging is not equivalent to averaging over realisations of surface roughness. As a result, in this context, care must be taken when taking averages of speckle patterns that it is clear what is being averaged over, and that an experimentally measured spatially averaged quantity obtained from the speckle pattern scattered from a given fixed rough surface differs from the same quantity averaged over realisations of the rough surface, which would require measurements over multiple different surfaces to calculate experimentally. In order to make the distinction clear, quantities spatially averaged over some region \mathcal{A} shall be denoted with an overbar, e.g. $\overline{F} = \int_{\mathcal{A}} F d\mathbf{r} / \mathcal{A}$, while the

angled brackets will continue to denote averaging over realisations of the scattering configuration. The statistics of the background speckle arising from scattering from a rough surface have been characterised, and so the light scattered from the analyte particle must now be considered.

3.4.2 Scattering from an Analyte Particle

The first thing to note is that, working in the single scattering regime, the field scattered from the analyte particle, $\mathbf{E}_a(\mathbf{r})$, is independent of $\mathbf{E}_\zeta(\mathbf{r})$. The total scattered field can therefore be found by simply adding the two scattered fields. A general analyte particle can be modelled by a potentially position dependent dielectric function $\varepsilon_a(\mathbf{r})$ occupying a volume \mathcal{V}_a in the dielectric halfspace $z > 0$. The scattered field for this scattering configuration can be expressed in the single scattering Born approximation as

$$\mathbf{E}_a(\mathbf{r}) = \frac{\omega^2}{c^2} \int_{\mathbf{r}' \in \mathcal{V}_a} d^3\mathbf{r}' G_0(\mathbf{r}, \mathbf{r}') (\varepsilon_a(\mathbf{r}') - \varepsilon_d) \mathbf{E}_0(\mathbf{r}'). \quad (3.68)$$

While this form allows for a completely general size, shape and composition of the analyte particle, it does not allow for much general analysis. To simplify the expression, it is assumed that the analyte particle volume is small such that $G_0(\mathbf{r}, \mathbf{r}')$ and $\mathbf{E}_0(\mathbf{r}')$ do not vary significantly over \mathcal{V}_a . As a result, these can be approximated as constant within the integral and replaced with their values at a point \mathbf{r}_a in \mathcal{V}_a , yielding

$$\mathbf{E}_a(\mathbf{r}) \approx k_0^2 G_0(\mathbf{r}, \mathbf{r}_a) \left[\int_{\mathbf{r}' \in \mathcal{V}_a} d^3\mathbf{r}' (\varepsilon_a(\mathbf{r}') - \varepsilon_d) \right] \mathbf{E}_0(\mathbf{r}_a). \quad (3.69)$$

It is important to consider exactly what constitutes a ‘small’ analyte particle volume. The condition for \mathbf{E}_0 to not vary significantly over \mathcal{V}_a can be seen from Eq. (3.5). The SPP field varies on the length scales λ_{SPP} and L_d in the x and z directions respectively. In order for the field to not vary significantly over \mathcal{V}_a regardless of orientation of the analyte particle, the longest length of the analyte particle must be smaller than $\min(\lambda_{\text{SPP}}, L_d)$. This restricts the approximation to particles smaller than $\sim 100\text{nm}$ at optical wavelengths. For example, the glass-gold-water thin film structure at 650nm gives $L_d = 102\text{nm}$ and $\lambda_{\text{SPP}} = 454\text{nm}$. The

condition to treat G_0 as constant is less straightforward. Firstly, the condition depends on the observation position \mathbf{r} , since the Green's tensor varies with both \mathbf{r}' and \mathbf{r} . For example, in the very near field where $|\mathbf{r} - \mathbf{r}'| < \lambda_0$, the Green's tensor varies much quicker as it approaches the singularity at $\mathbf{r} = \mathbf{r}'$. When \mathbf{r} is in the far field, however, the condition is quite simple, since, from Eq. (3.30), $G_0(\mathbf{r}, \mathbf{r}')$ varies with \mathbf{r}' on the length scale $\lambda_d = \lambda_0/\varepsilon_d^{1/2}$. Therefore, the analyte particle's longest length scale must be smaller than this. This condition is generally less restrictive than that arising from the field variation as $\lambda_d > L_d$ and $\lambda_d > \lambda_{\text{SPP}}$, so that provided the analyte particle is smaller than L_d , the approximations made in deriving Eq. (3.69) are valid. Recalling the relation between the Green's tensor and the field radiated by a dipole moment \mathbf{p} , it is informative to express Eq. (3.69) as

$$\mathbf{E}_a(\mathbf{r}) = \frac{k_0^2}{\varepsilon_0} G_0(\mathbf{r}, \mathbf{r}_a) \alpha_a \mathbf{E}_0(\mathbf{r}_a). \quad (3.70)$$

where α_a is a constant determined by $\varepsilon_a(\mathbf{r})$ and the shape of the analyte particle. Thus, the analyte particle, provided it is sufficiently small, radiates a dipole field with dipole moment $\mathbf{p} = \alpha_a \mathbf{E}_0(\mathbf{r}_a)$. Physically, α_a can be seen to correspond to the polarisability of the analyte particle. Based on the Born approximation, one finds $\alpha_a = \varepsilon_0 \int_{V_a} d^3\mathbf{r}' (\varepsilon_a(\mathbf{r}') - \varepsilon_d)$, however, more accurate models, accounting for depolarisation effects and the fact the internal field in the analyte particle differs from \mathbf{E}_0 , such as the renormalised Born approximation, give different expressions for α_a [142, 260, 261]. Rather than attempting to calculate α_a from integrating $\varepsilon_a(\mathbf{r})$ over the volume, α_a will be treated as a parameter of the analyte particle. Standard results exist for calculating the polarisability of an object such as the quasi-static polarisability (Eq. (2.6)) or from Mie theory [262]. The Mie theory polarisability is given by [88, 263]

$$\alpha_{\text{Mie}} = i \frac{6\pi}{k_d^3} a_1 \varepsilon_0 \varepsilon_d, \quad (3.71)$$

where $k_d = \varepsilon_d^{1/2} k_0$ is the wavenumber of light in the dielectric and a_1 is the electric dipole Mie coefficient, which can be calculated from the analyte particle radius R_a and permittivity ε_a from standard Mie theory [221, 264]. Strictly speaking, this result (and Eq. (2.6)) only applies for a sphere in a homogeneous space and not one near interfaces. Provided the sphere is sufficiently

small, however, the effect of the interaction between the analyte particle and the interface on the polarisability (known as surface dressing) can be ignored. This effect will be discussed in greater detail in Chapter 5.

The same result as Eq. (3.70) could be arrived upon within a point scatterer model [265] in which the dielectric function associated with the analyte particle is $\varepsilon_a(\mathbf{r}) = (\alpha_a/\varepsilon_0)\delta(\mathbf{r} - \mathbf{r}_a)$. This point scatterer model will be useful in future chapters. In general, α_a is a complex tensor quantity reflecting the fact the analyte particle may have an anisotropic response and that the dipole moment need not be in phase with the incident field. The intensity scattered directly from the analyte particle is $I_a(\mathbf{r}) = |\mathbf{E}_a(\mathbf{r})|^2$. In the case of an isotropic scatterer for which α_a is a scalar, this can be expressed

$$I_a(\mathbf{r}) = \frac{k_0^4}{\varepsilon_0^2} |\alpha_a|^2 |G_0(\mathbf{r}, \mathbf{r}_a) \mathbf{A}_{\text{SPP}}^+|^2 e^{-x_a/L_{\text{SPP}}} e^{-z_a/L_d}. \quad (3.72)$$

Comparing Eqs. (3.54) and (3.72) the similarities allow an effective surface polarisability to be defined. Firstly, consider the short surface correlation length lossy case in which Eq. (3.49) applies. In this case, Eq. (3.54) reduces to

$$\langle I_\zeta(\mathbf{r}) \rangle = 2\pi |\varepsilon_m - \varepsilon_d|^2 a^2 h^2 L_y L_{\text{SPP}} k_0^4 |G_0(\mathbf{r}, \mathbf{0}^-) \mathbf{A}_{\text{SPP}}^+|^2. \quad (3.73)$$

From this, the effective surface polarisability is defined as $\alpha_{\text{surf}}^2 = 2\pi \varepsilon_0^2 |\varepsilon_m - \varepsilon_d|^2 a^2 h^2 L_y L_{\text{SPP}}$. The scaling of this term can be explained heuristically by considering surface scattering as arising from scattering from many individual surface roughness features (i.e. peaks and troughs) on the surface. Each scattering feature has a volume $\sim ha^2$ (see Fig. 3.5). Since each feature is small, it can be treated as a dipole scatterer with polarisability $\sim \Delta\varepsilon ha^2$ proportional to the volume of the feature times the difference in dielectric from the background. These many dipoles add up incoherently due to the random phase differences so that the total intensity is proportional to the number of such roughness features N_{feat} scattering light multiplied by the average intensity scattered from a feature $\sim |G_0 \Delta\varepsilon ha^2 \mathbf{E}_0|^2$. The number of scattering features is $N_{\text{feat}} \sim \text{total area illuminated}/\text{average area of feature} = L_y L_{\text{SPP}}/a^2$. Thus the scattered light is proportional to $I_\zeta \sim N_{\text{feat}} |G_0 \Delta\varepsilon h^2 a \mathbf{E}_0|^2 = L_y L_{\text{SPP}} a^2 h^2 \Delta\varepsilon |G_0 \mathbf{E}_0|^2$, matching the scaling of

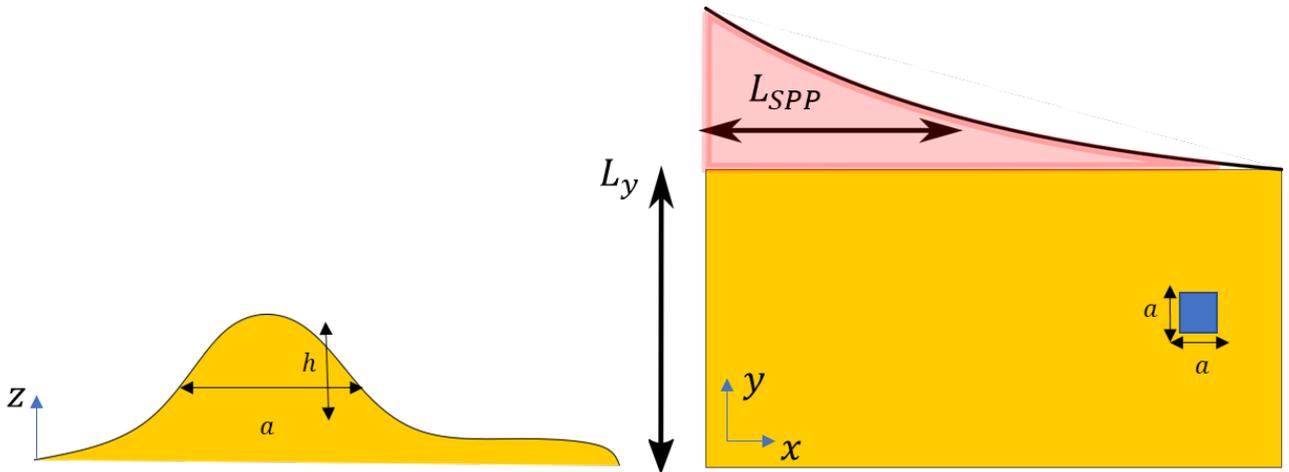


Figure 3.5: Diagram showing heuristic argument for scaling of α_{surf} . On the left, a single surface feature is shown occupying a small volume $\sim ha^2$ and thus has a polarisability $\sim (\varepsilon_m - \varepsilon_d)ha^2$. In the right diagram, it can be seen that the entire illuminated region of area $\sim L_y L_{SPP}$ contains $\sim L_y L_{SPP}/a^2$ such scattering volumes.

Eq. (3.73). The value of this effective polarisability is that the mean intensity scattered to a point \mathbf{r} in the far field by the rough surface is equivalent to that of a single point scatterer of polarisability α_{surf} located at $\mathbf{0}^-$ oriented in the direction of \mathbf{A}_{SPP}^+ . While the different positions of the dipole source mean the dipolar intensity patterns of Eqs. (3.72) and (3.73) differ, the effective polarisability can be compared to α_a to give an idea of the comparative strength with which the surface and analyte particle scatter light. Fig. 3.6 shows the relative size of the surface and analyte particle polarisabilities where the analyte particle has been modelled as a homogeneous sphere and the Mie polarisability expression (Eq. (3.71)) has been used. Unless otherwise stated, the surface roughness parameters used in this chapter from now on are $L_y = 10\lambda_0$, $h = 6\text{nm}$ and $a = 12\text{nm}$. It can be seen that the effective surface polarisability is $\sim 10^3$ times larger than that of a 50nm sphere of refractive index 1.4 (chosen as typical of biological objects such as viruses or proteins [266, 267, 268]). The typical surface scattered intensities are therefore expected to be $\sim 10^6$ larger than the light scattered from a 50nm analyte particle even for surface roughness with parameters typical of a metal film, with the surface still scattering an order of magnitude more strongly even as the analyte particle radius approaches λ_0 , though for radii as large as this, the point dipole approximation breaks down and higher order multipoles contribute to scattering. Here it can also be seen why a dark field version of the proposed sensing set up, in which there is no background scattered light, is significantly

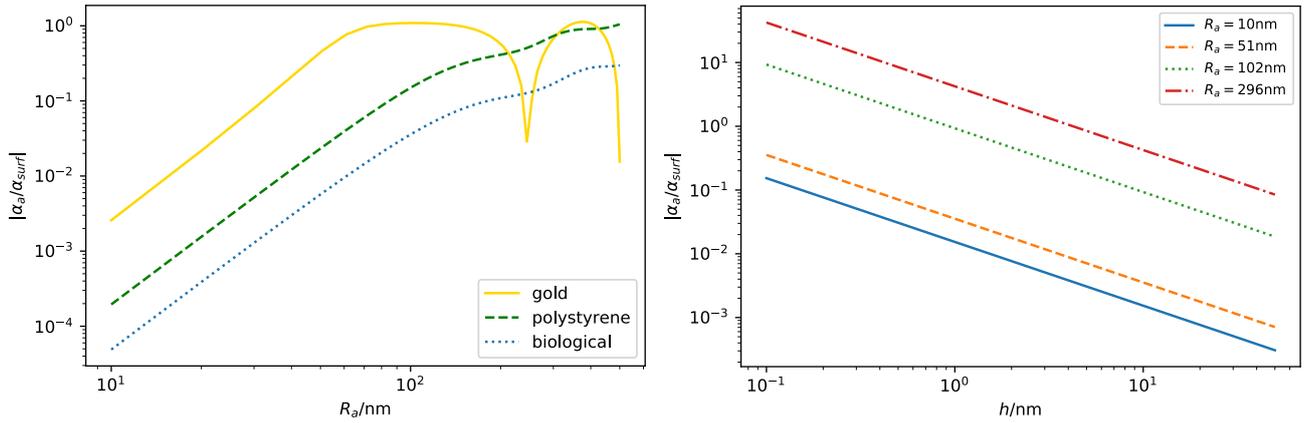


Figure 3.6: Plots showing the dependence of $|\alpha_a/\alpha_{\text{surf}}|$ on the size of the analyte particle (left) and the RMS roughness height (right) for the glass-gold-water structure at $\lambda_0 = 650\text{nm}$. The analyte particle is modelled as a homogeneous sphere of radius R_a and constant permittivity ε_a , with α_a calculated from Eq. (3.71). The three analyte particle compositions plotted are gold ($\varepsilon_a = -13.7 + 1.04i$), polystyrene ($\varepsilon_a = 2.5$) and a ‘biological’ particle ($\varepsilon_a = 1.96$). On the left plot, α_{surf} was held fixed for parameters $h = 6\text{nm}$, $a = 12\text{nm}$, $L_y = 10\lambda_0 = 6.5\mu\text{m}$ while on the right α_a was that of a ‘biological’ particle with the given radius.

more challenging, requiring RMS surface roughness deviations of the order $\sim 0.1\text{nm}$ or less simply to get the surface scattered light down to a comparable level to the light scattered from an analyte particle of radius 51nm . In the gold case, at large radius $R_a \gtrsim 100\text{nm}$, dips in α_a as a function of R_a are seen, which arise due to a known oscillatory dependence of the Mie coefficient on R_a when $|(n_a/n_d)k_d R_a| \gg 1$ [269]. For a gold sphere analyte particle, the analyte particle refractive index n_a has a significantly larger modulus than n_d , and thus this oscillatory behaviour is seen in Fig. 3.6. For the other two analyte particle compositions plotted, the refractive index contrast with the background is sufficiently small to ensure $|(n_a/n_d)k_d R_a| \lesssim 1$ and thus the dips are not observed.

The effective polarisability was derived under the assumption of a short surface correlation length, but when this does not hold, the mean scattered intensity is no longer purely dipolar as it is modified by $\mathcal{I}(\Delta\mathbf{k}_{\parallel}, \Delta\mathbf{k}_{\parallel})$. Using the lossless result from Eq. (3.52) results in Eq. (3.54) becoming

$$\langle I_{\zeta}(\mathbf{r}) \rangle = |\varepsilon_m - \varepsilon_d|^2 a^2 h^2 L_y L_x k_0^4 |G_0(\mathbf{r}, \mathbf{0}^-) \mathbf{A}_{\text{SPP}}^+|^2 \tilde{C}(\Delta\mathbf{k}_{\parallel}). \quad (3.74)$$

The short surface correlation length δ -function approximation is equivalent to taking $\tilde{C}(\Delta\mathbf{k}_{\parallel}) = \tilde{C}(\mathbf{0})$ to be constant. Using a Gaussian correlation function allows assessment of when that

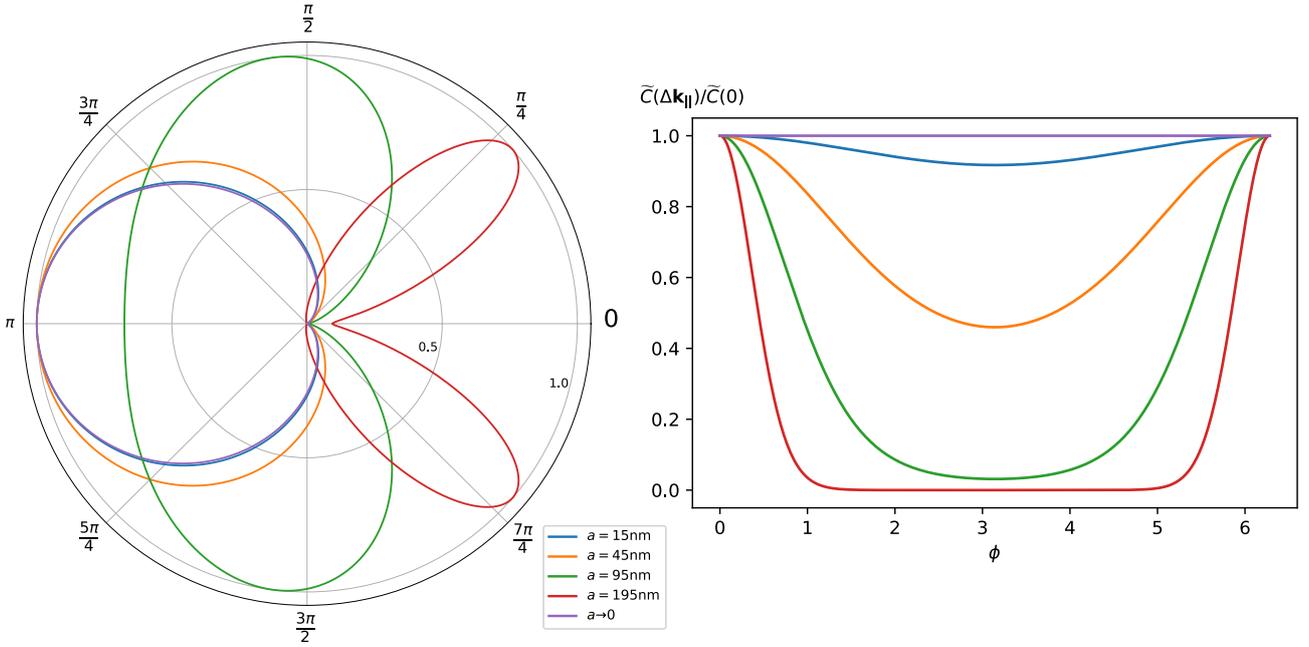


Figure 3.7: (left) The profile of $\langle I_C(\theta = \Theta_{\text{SPP}}, \phi) \rangle$ around the leakage ring for a Gaussian surface correlation function with different values of a , calculated from Eq. (3.75) and normalised to a maximum value of 1 and (right) the value of $\tilde{C}(\Delta \mathbf{k}_{\parallel})/\tilde{C}(0)$ around the leakage ring for the same set of correlation lengths.

approximation breaks down. Recalling that the normalisation is such that $C(\mathbf{0}) = 1$, the Fourier space correlation function becomes $C(\mathbf{q}) = 2\pi a^2 \exp(-q^2 a^2/2)$. It follows that the mean surface scattered intensity is

$$\langle I_C(\mathbf{r}) \rangle = \frac{k_0^4}{\varepsilon_0^2} \alpha_{\text{surf}}^2 |G_0(\mathbf{r}, \mathbf{0}^-) \mathbf{A}_{\text{SPP}}^+|^2 \exp(-\Delta k_{\parallel}^2 a^2/2). \quad (3.75)$$

Note that the effective surface polarisability still appears, though now with L_{SPP} replaced by L_x , and therefore still provides a heuristic tool to compare the scattering strength of the surface to that of a small particle. The result is modified from Eq. (3.73) by a function $C(\Delta \mathbf{k}_{\parallel}) \leq 1$. For observation points in the leakage ring, $\Delta \mathbf{k}_{\parallel} = k'_{\text{SPP}}(\cos \phi - 1, \sin \phi)^T$. Thus the short surface correlation length result is modified by a factor $\exp\{-k_{\text{SPP}}'^2 a^2(1 - \cos \phi)\}$. The smallest value this can take is in the backscattered direction when $\phi = \pi$, which is $\exp(-2k_{\text{SPP}}'^2 a^2)$. In order for this correction to be significant, it is required that $k_{\text{SPP}}'^2 a^2 \gtrsim 1$. Taking $a = 10\text{nm}$ and $\lambda_{\text{SPP}} = 454\text{nm}$ for the glass-50nm gold-water thin film structure at $\lambda_0 = 650\text{nm}$ one finds that the correction value is 0.96 at smallest. Fig. 3.7 shows the modified mean intensity around the ring for a few different surface correlation lengths. For $a \lesssim 20\text{nm}$, the short range correlation

approximation $a \rightarrow 0$ is fairly accurate. Long correlation lengths cause the scattering to be strongly in the forward direction. This is because long surface correlation lengths correspond to smoother surfaces, and therefore the roughness deflects the scattered light through much smaller angles compared to steeper gradient surfaces.

3.5 Single Scattering Sensitivity Analysis

With a model established describing all single scattering processes involved in the sensing setup, a single scattering sensitivity analysis can now be performed. Recalling from Chapter 2 that sensitivity is defined as the derivative of the measured signal with respect to the quantity of interest, an appropriate definition here, in the context of sensing a single particle, is the change in the measured signal on addition of the analyte particle. Different sensing signals define different sensitivities, and a few possible signals will be studied here.

3.5.1 Field Sensitivity

Taking the electric field as a measured signal, the field sensitivity can be defined as

$$\mathcal{S}_E = |\Delta \mathbf{E}| = \mathbf{E} - \mathbf{E}_\zeta. \quad (3.76)$$

Direct measurement of the electric field at optical frequencies is challenging and rarely done in practical experimental setups [270], thus it is unlikely that the electric field would actually act as a sensing signal. It is still useful to define and study \mathcal{S}_E , primarily because any actual measured signal (for example intensity) ultimately derives from a change in the electric field, and therefore sensitivities of other metrics can be related to \mathcal{S}_E . In the single scattering regime, the change to the field is simply the additional field scattered from the analyte particle so that

$$\begin{aligned} \mathcal{S}_E(\mathbf{r}) &= |\mathbf{E}_a(\mathbf{r})| \\ &= \left| \frac{k_0^2}{\varepsilon_0} G_0(\mathbf{r}, \mathbf{r}_a) \alpha_a \mathbf{E}_0(\mathbf{r}_a) \right| \\ &= \frac{k_0^2}{\varepsilon_0} \left| G_0(\mathbf{r}, \mathbf{r}_a) \alpha_a \mathbf{A}_{\text{SPP}}^+ \right| e^{-\frac{x_a}{2L_{\text{SPP}}}} e^{-\frac{z_a}{2L_a}}. \end{aligned} \quad (3.77)$$

Since $\mathcal{S}_{\mathbf{E}} \propto |\mathbf{E}_0(\mathbf{r}_a)|$, the sensitivity decays exponentially with the analyte particle's distance away from the surface z_a and with distance along the propagation direction x_a , while also being proportional to the field strength at the surface $|\mathbf{A}_{\text{SPP}}^+|$. This is determined by the illumination power, and clearly one can increase this power to improve sensitivity. Ultimately, however, practical considerations such as the damage to the sample limit how much power one can put into the system. Due to this proportionality, the sensitivity will be measured relative to the incident field directly above the surface $z = 0^+$. Dependence on \mathbf{r}_a also appears in the Green's tensor. In the far field, Eq. (3.30) reveals that the transverse position $\boldsymbol{\rho}_a$ only changes the phase of G_0^∞ and this does not affect $\mathcal{S}_{\mathbf{E}}$. Thus, the only transverse dependence arises from the incident field decay. On the other hand, from Eqs. (3.24)–(3.26), $G_0^\infty(\mathbf{r}, \mathbf{r}_a)$ has a different variation with z_a depending on the observation position. For \mathbf{r} in the lower half space (and in the far field), only the transmission component \tilde{G}_{tr} contributes to $G_0^\infty(\mathbf{r}, \mathbf{r}_a)$, which exhibits an $\exp(ik_z z_a)$ dependence on z_a . For observation points with $k_{o\parallel} < k_d$, z_a only changes the Green's tensor phase. On the other hand, observation points with $k_{o\parallel} > k_d$, possible in the $z < -d$ halfspace provided $\varepsilon_N > \varepsilon_d$, result in an exponential decay of G_0^∞ as $k_z = i\kappa_z$ where $\kappa_z = (k_{o\parallel}^2 - k_d^2)^{1/2}$ is real. This introduces an additional decay $\exp\{-\kappa_z z'\}$ to $\mathcal{S}_{\mathbf{E}}$. Such a decay arises due to the fact the light at these angles arises from coupling of evanescent components of the radiated field through the multilayer structure into propagating components in the lower half space, and as such the coupling decays evanescently as the dipole is moved further from the surface. Importantly, this range of angles includes the leakage radiation ring in structures with SPP modes. Overall, the dependence on $\mathbf{r}_a = (x_a, y_a, z_a)^T$ can be summarised for observation points below the multilayer stack via

$$\mathcal{S}_{\mathbf{E}}(\mathbf{r}) = \frac{k_0^2}{\varepsilon_0} |G_0^\infty(\mathbf{r}, \mathbf{0}^+) \alpha_a \mathbf{A}_{\text{SPP}}^+| e^{-\frac{x_a}{2L_{\text{SPP}}}} e^{-\frac{z_a}{2L_a}} \times \begin{cases} 1 & \theta > \theta_c \\ \exp(-\kappa_z z_a) & \frac{\pi}{2} < \theta \leq \theta_c. \end{cases} \quad (3.78)$$

where $\theta_c = \pi - \sin^{-1}(n_d/n_N)$ is the critical angle at which the modes coupling from the dipole through the multilayer stack to the far field in direction θ switch from evanescent to propagating modes. For observation positions in the upper half plane, the z dependence is more complicated

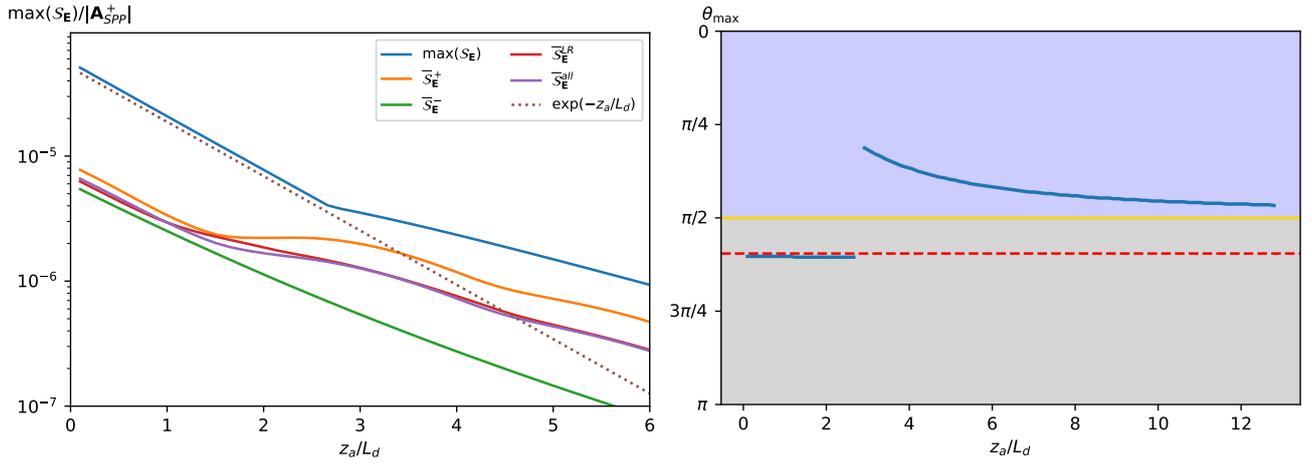


Figure 3.8: Plots showing (left) $\max(\mathcal{S}_{\mathbf{E}})/|\mathbf{A}_{\text{SPP}}^+|$ and $\overline{\mathcal{S}_{\mathbf{E}}}/|\mathbf{A}_{\text{SPP}}^+|$ as a function of z_a with $R_a = 50\text{nm}$ and $r = 100\lambda_0$ fixed and a ‘biological’ analyte and (right) the polar angle θ_{\max} at which $\max(\mathcal{S}_{\mathbf{E}})$ occurs, with the leakage angle Θ_{SPP} shown (red dashed line).

due to the fact both G_{dir} and G_{ref} contribute and have different z_a dependence. As such, the exact z_a dependence of $\mathcal{S}_{\mathbf{E}}$ depends on the interference between the direct and reflected Green’s tensor, which is decided by the Fresnel coefficient. This means the z_a dependence of $\mathcal{S}_{\mathbf{E}}$ in the upper half space is strongly dependent on the parameters of the multilayer stack, but will in general oscillate as the analyte particle moves away from the surface. The length scale of the variation of $|G_{\text{dir}} + G_{\text{ref}}|$ with z_a will be $\sim 1/k_z = 1/(k_d \cos \theta)$. When $k_z < \kappa_d$, the $\exp(-\kappa_d z_a)$ decay of the incident field will dominate over the z_a dependence arising from the Green’s tensor. Finally, due to the r^{-1} dependence of G_0^∞ (see Eq. (3.30)) $\mathcal{S}_{\mathbf{E}} \propto r^{-1}$, where r is the distance of the observation point from the origin. The maximum field sensitivity for isotropic scatterers is generally achieved in the leakage radiation ring due to the directionality of the dipole radiation, but this comes at the cost of a more rapid decay of sensitivity with the distance of the particle from the surface. In fact, the decay in the leakage ring is approximately doubled compared to outside the critical region, since Eq. (3.2) gives $\kappa_d = (k_{\text{SPP}}^2 - k_d^2)^{1/2}$ which for $k_{\text{SPP}}'' \ll k_{\text{SPP}}'$ is approximately the same decay as $\kappa_z = (k_{\text{SPP}}^2 - k_d^2)^{1/2}$. Fig. 3.8 shows the dependence of $\max(\mathcal{S}_{\mathbf{E}})$ on z_a for the case study structure at $\lambda_0 = 650\text{nm}$ for a 50nm radius spherical ‘biological’ analyte particle at a radial coordinate $r = 100\lambda_0$ ($\mathcal{S}_{\mathbf{E}}$ at any other distance r in the far field can be calculated from the results in Fig. 3.8 via the $\propto r^{-1}$ scaling), as well as the observation polar angle at which the maximum occurred. The leakage ring provides maximum sensitivity until the analyte moves sufficiently far from the surface. Once above a certain height ($z_p = 2.8L_d$ in

this case), the evanescent coupling to SPPs and into the leakage ring is sufficiently weak that the maximum sensitivity arises from light diffusely scattered into the water, and the sensitivity decay with height is proportional to the field decay $\propto \exp(-z_a/2L_d)$. Since the incident field intensity has decayed to a fraction $e^{-2.8} \approx 0.06$ of its value at the metal surface, it is reasonable to consider this outside the sensing volume. As a result, it can be concluded that the electric field is the most sensitive at a point in the leakage ring whenever sensing can occur. The sensitivity here is a pointwise metric, in that each point \mathbf{r} a measurement is taken has a defined sensitivity. In general, no sensing setup would involve the measurement at a single point as this would be susceptible to noise, and so it is also useful to consider average sensitivities $\overline{\mathcal{S}}_{\mathbf{E}}^{\Omega}$ over a particular d -dimensional region Ω .

$$\overline{\mathcal{S}}_{\mathbf{E}}^{\Omega} = \int_{\Omega} \frac{d^d \mathbf{r}}{|\Omega|} \mathcal{S}_{\mathbf{E}}(\mathbf{r}), \quad (3.79)$$

where $|\Omega|$ is the d -dimensional volume of Ω . Additionally, actual light detectors will be of finite size and therefore any measurement will be averaged over the detector size. Generally, light is measured on a 2D surface, so Ω will most commonly be taken as an area. Fig. 3.8 shows these spatially averaged quantities for a few different choices of Ω , these being the surface of a sphere in the far field averaging over propagating light in all directions ($\overline{\mathcal{S}}^{\text{all}}$), a hemisphere in the far field collecting all light propagating in the far field in either the upper ($\overline{\mathcal{S}}^+$) or lower ($\overline{\mathcal{S}}^-$) halfspace and finally averaging around the leakage ring ($\overline{\mathcal{S}}^{LR}$). Explicitly, these are defined by

$$\overline{\mathcal{S}}_{\mathbf{E}}^{\text{all}} = \frac{1}{4\pi r^2} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \mathcal{S}_{\mathbf{E}}(\mathbf{r}) r^2 \sin \theta d\theta d\phi \quad (3.80)$$

$$\overline{\mathcal{S}}_{\mathbf{E}}^+ = \frac{1}{2\pi r^2} \int_{\theta=0}^{\frac{\pi}{2}} \int_{\phi=0}^{2\pi} \mathcal{S}_{\mathbf{E}}(\mathbf{r}) r^2 \sin \theta d\theta d\phi \quad (3.81)$$

$$\overline{\mathcal{S}}_{\mathbf{E}}^- = \frac{1}{2\pi r^2} \int_{\theta=\frac{\pi}{2}}^{\pi} \int_{\phi=0}^{2\pi} \mathcal{S}_{\mathbf{E}}(\mathbf{r}) r^2 \sin \theta d\theta d\phi \quad (3.82)$$

$$\overline{\mathcal{S}}_{\mathbf{E}}^{LR} = \frac{1}{2\pi r^2 [\cos(\Theta_{\text{SPP}} - \frac{\delta}{2}) - \cos(\Theta_{\text{SPP}} + \frac{\delta}{2})]} \int_{\Theta_{\text{SPP}} - \frac{\delta}{2}}^{\Theta_{\text{SPP}} + \frac{\delta}{2}} \int_{\phi=0}^{2\pi} \mathcal{S}_{\mathbf{E}}(\mathbf{r}) r^2 \sin \theta d\theta d\phi \quad (3.83)$$

where δ is the FWHM of the leakage ring. These averaged sensitivities maintain the same r^{-1} as the pointwise $\mathcal{S}_{\mathbf{E}}(\mathbf{r})$ as they are all over regions with fixed r . Unsurprisingly, the metal film

reflects most of the light into the upper half space so that $\mathcal{S}_{\mathbf{E}}^+ > \mathcal{S}_{\mathbf{E}}^-$. For an analyte particle very near the surface, with strong evanescent coupling, $\mathcal{S}_{\mathbf{E}}^+$ and $\mathcal{S}_{\mathbf{E}}^-$ are much closer together, with the leakage ring being much more sensitive than the rest of the lower half space. While the pointwise sensitivity is largest in the leakage ring, when averaged around the whole ring it is actually comparable to and slightly less than the average sensitivity over the whole upper half space. This is because the dipolar profile around the ring means the field sensitivity in the backward directions is much smaller than the forward directions and hence the average sensitivity around the ring is much smaller than $\max(\mathcal{S}_{\mathbf{E}})$. The balance between high sensitivity and the size of the area with large sensitivity (i.e. the number of highly sensitive points) is important to consider, especially in the context of the practical experimentally measurable signals.

As it is not being considered as a likely sensing signal, there is no need to study the other common sensing metrics such as LOD introduced in Chapter 2. Instead, these should be considered in the context of signals that can be realistically measured in a biosensing setup.

3.5.2 Intensity Sensitivity

A more practical measurement signal is the intensity of the light, which can be measured with simple photodetectors or cameras. In this case, the pointwise sensitivity metric is defined by

$$\mathcal{S}_I(\mathbf{r}) = |\Delta I(\mathbf{r})| = I(\mathbf{r}) - I_b(\mathbf{r}), \quad (3.84)$$

where $I(\mathbf{r}) = |\mathbf{E}_0(\mathbf{r}) + \mathbf{E}_\zeta(\mathbf{r}) + \mathbf{E}_a(\mathbf{r})|^2$ is the intensity when the analyte particle is in the sensing volume (i.e. close to the surface) and $I_b(\mathbf{r}) = |\mathbf{E}_0(\mathbf{r}) + \mathbf{E}_\zeta(\mathbf{r})|^2$ is the intensity in the absence of the analyte particle. The intensity perturbation is given by

$$\begin{aligned} \Delta I(\mathbf{r}) &= I_a(\mathbf{r}) + 2 \operatorname{Re}[\mathbf{E}_b^*(\mathbf{r}) \cdot \mathbf{E}_a(\mathbf{r})] \\ &= \mathcal{S}_{\mathbf{E}}^2(\mathbf{r}) + 2\mathcal{S}_{\mathbf{E}}(\mathbf{r})\sqrt{I_b(\mathbf{r})} \left| \hat{\boldsymbol{\xi}}_b^*(\mathbf{r}) \cdot \hat{\boldsymbol{\xi}}_a(\mathbf{r}) \right| \cos[\Phi(\mathbf{r})]. \end{aligned} \quad (3.85)$$

The initial speckle field $\mathbf{E}_b = \mathbf{E}_0 + \mathbf{E}_\zeta$ has been defined, while $\hat{\boldsymbol{\xi}}_{a,b}(\mathbf{r}) = \mathbf{E}_{a,b}(\mathbf{r})/|\mathbf{E}_{a,b}(\mathbf{r})|$ are the unit polarisation vectors for the analyte scattered and initial speckle field and $\Phi(\mathbf{r})$ is the phase difference between \mathbf{E}_a and \mathbf{E}_b . Noting that $\langle \text{Re}(\mathbf{E}_\zeta^* \cdot \mathbf{E}_a) \rangle = 0$, the mean intensity change is just the intensity scattered from the analyte plus the interference with the incident field

$$\langle \Delta I(\mathbf{r}) \rangle = I_a(\mathbf{r}) + 2 \text{Re}[\mathbf{E}_0^*(\mathbf{r}) \cdot \mathbf{E}_a(\mathbf{r})]. \quad (3.86)$$

The reason the interference term $2 \text{Re}[\mathbf{E}_\zeta^* \cdot \mathbf{E}_a]$ vanishes is because the random phase of the speckle field means it is equally likely \mathbf{E}_a will increase the intensity (adding in phase with \mathbf{E}_ζ) or decrease the intensity (adding out of phase), and these effects cancel out when averaging over realisations. From a sensing perspective, however, a decrease in intensity is providing the same information about the presence of the analyte as an intensity increase, and it is the magnitude of the change that is of interest. One way to measure this is to calculate the root mean square (RMS) sensitivity, defined as $\mathcal{S}_I^{RMS} = \langle \Delta I^2 \rangle^{1/2}$. The advantage this has over $\langle \mathcal{S}_I \rangle$ is that it can be calculated analytically from the statistical model of the surface. Wick's theorem means the odd order terms in \mathbf{E}_ζ vanish and $\langle \Delta I^2 \rangle = \langle \Delta I \rangle^2 + \langle [2 \text{Re}(\mathbf{E}_\zeta \cdot \mathbf{E}_a^*)]^2 \rangle$, with the average of the squared interference term given by Wick's theorem to be

$$\begin{aligned} \langle [2 \text{Re}(\mathbf{E}_\zeta(\mathbf{r}) \cdot \mathbf{E}_a^*(\mathbf{r}))]^2 \rangle &= 2 \sum_{i,j=1}^3 \left\{ \text{Re} \left[\langle E_{\zeta,i}(\mathbf{r}) E_{\zeta,j}(\mathbf{r}) \rangle E_{a,i}^*(\mathbf{r}) E_{a,j}^*(\mathbf{r}) \right] \right. \\ &\quad \left. + \langle E_{\zeta,i}(\mathbf{r}) E_{\zeta,j}^*(\mathbf{r}) \rangle E_{a,i}^*(\mathbf{r}) E_{a,j}(\mathbf{r}) \right\}. \end{aligned} \quad (3.87)$$

The second term on the right hand side of Eq. (3.87), using Eqs. (3.45) and (3.70), can be expressed

$$\begin{aligned} &\sum_{i,j=1}^3 \langle E_{\zeta,i}(\mathbf{r}) E_{\zeta,j}^*(\mathbf{r}) \rangle E_{a,i}^*(\mathbf{r}) E_{a,j}(\mathbf{r}) \\ &= \frac{k_0^4}{\varepsilon_0^2} |B| \left| [G_0(\mathbf{r}, 0^-) \mathbf{A}_{\text{SPP}}^+] [G_0(\mathbf{r}, \mathbf{r}_a) \alpha_a \mathbf{A}_{\text{SPP}}^+] \right|^2 e^{-x_a/L_{\text{SPP}}} e^{-z_a/L_a} \mathcal{I}(\Delta \mathbf{k}_\parallel, \Delta \mathbf{k}_\parallel) \\ &= \left| \mathbf{E}_{dp}^{\text{surf,eff}}(\mathbf{r}) \cdot \mathbf{E}_a(\mathbf{r}) \right|^2. \end{aligned} \quad (3.88)$$

This corresponds to the magnitude squared of the dot product of the effective dipole field from the surface scattering, $\mathbf{E}_{dp}^{\text{surf,eff}}(\mathbf{r}) = (k_0^2/\varepsilon_0)G_0(\mathbf{r}, \mathbf{0}^-)\alpha_{\text{surf}}\mathbf{A}_{\text{SPP}}^+$, with the dipole field scattered from the particle. Similarly, the first term of Eq. (3.87), using Eq. (3.46), becomes

$$\begin{aligned} & \sum_{i,j=1}^3 \text{Re} [\langle E_{\zeta,i}(\mathbf{r})E_{\zeta,j}(\mathbf{r}) \rangle E_{a,i}^*(\mathbf{r})E_{a,j}^*(\mathbf{r})] \\ &= \frac{k_0^4}{\varepsilon_0^2} \text{Re} \left\{ B [(G_0(\mathbf{r}, \mathbf{0}^-)\mathbf{A}_{\text{SPP}}^+) (G_0(\mathbf{r}, \mathbf{r}_a)\alpha_a\mathbf{A}_{\text{SPP}}^+)^* e^{ik_{\text{SPP}}x_p} e^{-\kappa_d z_p}]^2 \mathcal{I}(\Delta\mathbf{k}_{\parallel}, -\Delta\mathbf{k}_{\parallel}) \right\}. \end{aligned} \quad (3.89)$$

Both terms are proportional to $|\alpha_{\text{surf}}\alpha_a|^2$, whereas $I_a \propto |\alpha_a|^4$ is much smaller. As discussed earlier (e.g. see Fig. 3.6), $\alpha_a \ll \alpha_{\text{surf}}$ for typical biological particle and surface roughness parameters and therefore the mean interference term dominates over I_a . Furthermore, since the incident SPP field \mathbf{E}_0 decays with distance from either surface (see Eq. (3.5)), it vanishes in the far field and can be ignored in the calculation of $\mathcal{S}_I^{\text{RMS}}$. As a result, $\langle \Delta I \rangle = I_a \ll \langle [2 \text{Re}(\mathbf{E}_{\zeta} \cdot \mathbf{E}_a^*)]^2 \rangle$. In addition, in the leakage ring $|f_C| \ll 1$ so that the first term in Eq. (3.87) is much smaller than the second and can be ignored. Within these approximations, $\mathcal{S}_I^{\text{RMS}}$ becomes

$$\begin{aligned} \mathcal{S}_I^{\text{RMS}} &\approx \sqrt{2} \frac{k_0^4}{\varepsilon_0^2} |\alpha_{\text{surf}} [G_0(\mathbf{r}, 0^-)\mathbf{A}_{\text{SPP}}^+] [G_0(\mathbf{r}, \mathbf{r}_a)\alpha_a\mathbf{A}_{\text{SPP}}^+]| e^{-\frac{x_a}{2L_{\text{SPP}}}} e^{-\frac{z_a}{2L_a}} \\ &= \sqrt{2} \left| \mathbf{E}_{dp}^{\text{surf,eff}}(\mathbf{r}) \cdot \mathbf{E}_a(\mathbf{r}) \right|. \end{aligned} \quad (3.90)$$

Unlike $\mathcal{S}_{\mathbf{E}}$, which only depends on the analyte particle, the statistics of the surface scattered light are important in $\mathcal{S}_I^{\text{RMS}}$. As one would expect from an interference effect, $\mathcal{S}_I^{\text{RMS}} \propto \langle I_{\zeta} \rangle^{1/2}$, meaning that increasing the scattering from the surface increases the sensitivity. Much like the field sensitivity, $\mathcal{S}_I^{\text{RMS}} \propto I_0$ where $I_0 = |\mathbf{A}_{\text{SPP}}^+|^2$ is the intensity at the metal surface. As before, a sensitivity normalised to this intensity $\mathcal{S}_I^{\text{RMS}}/I_0$ can be considered. It is also useful to consider the sensitivity in terms of the fractional change to the intensity, i.e. the mean size of the intensity change relative to the mean intensity before the analyte particle binds $\mathcal{S}_I^{\text{frac}} = \mathcal{S}_I^{\text{RMS}}/\langle I_{\zeta}(\mathbf{r}) \rangle$. Fig. 3.9 shows the $\max(\mathcal{S}_I^{\text{RMS}})$ and spatially averaged RMS sensitivity over the leakage ring $\overline{\mathcal{S}}_I^{\text{RMS,LR}}$ defined analogously to Eq. (3.83). The maximum occurs in the leakage ring in all cases. The fractional intensity change in the leakage ring is on average of the order of 10^{-4} in the leakage ring for a 10nm radius ‘biological’ analyte particle. This level of fractional intensity

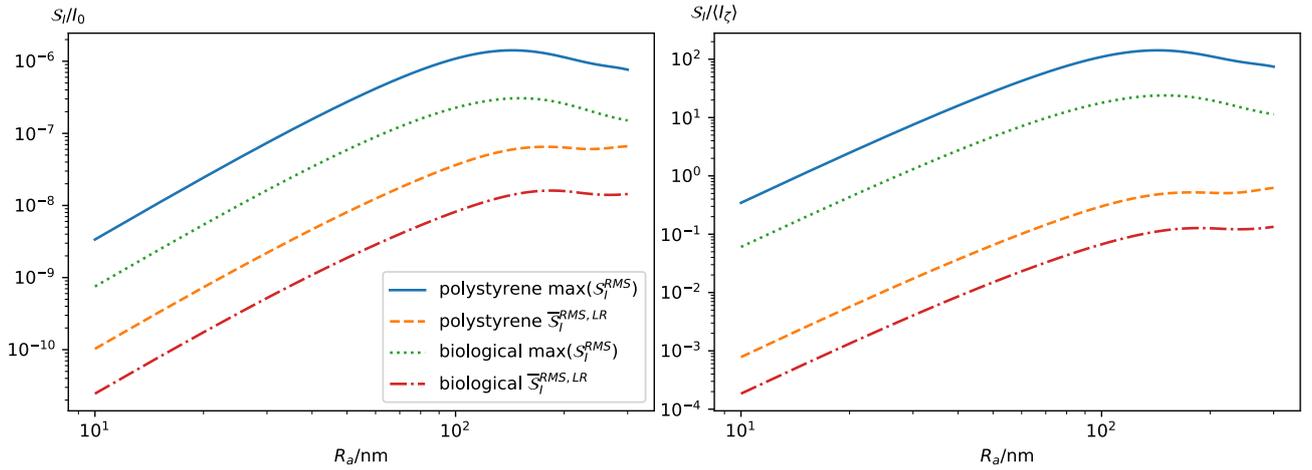


Figure 3.9: The RMS intensity sensitivity, \mathcal{S}_I^{RMS} , relative to (left) the intensity at the metal surface I_0 (at $r = 100\lambda_0$) and (right) the mean background intensity $\langle I_\zeta \rangle$ (independent of r).

change is comparable to those involved in detection of single proteins using iSCAT [271]. Note that the both I_a and I_ζ scale as r^{-2} in the far field, so while $\mathcal{S}_I^{RMS} \propto r^{-2}$ (as indeed is \mathcal{S}_I), this cancels out in $\mathcal{S}_I^{RMS}/\langle I_\zeta \rangle$ so that the fractional sensitivity is independent of distance in the far field. In fact, more generally, since \mathbf{E}_a and \mathbf{E}_b propagate together to a detector, the any form of intensity attenuation, in addition to the r^{-2} decay from the Green's tensor owing to the energy being spread over a wider surface, for example due to optical elements in the detection setup, apply equally to both the surface and analyte particle scattered contributions and therefore the fractional sensitivity results of Fig. 3.9 still hold. On the other hand, $\mathcal{S}_I/I_0 \propto r^{-2}$ does depend on r and is shown for $r = 100\lambda_0$ as in Fig. 3.8, with the r^{-2} scaling giving the value for any other r from the results in Fig. 3.9.

3.6 Detection Limits

The limit of detection is defined in Eq. (2.2). For intensity measurements, the limit of detection is when the intensity change ΔI is equal to the noise standard deviation (multiplied by some constant m of order unity that depends on the confidence interval one requires to classify a measurement as a binding event as opposed to noise). Using the shot noise model as in Chapter

2, the standard deviation in the background intensity is

$$\sigma_I = \sqrt{I_\zeta I_{1\gamma}} \quad (3.91)$$

where $I_{1\gamma}$ is a parameter of the photon detector (corresponding to the intensity measured when a single photon is incident on the detector in integration time τ). It is given by

$$I_{1\gamma} = \frac{\gamma \hbar \omega}{A_{\text{det}} \tau} \quad (3.92)$$

where \hbar is the reduced Planck's constant, A_{det} is the detector area and γ is the detector quantum efficiency giving the fraction of photons incident on the detector that are converted into a count. The intensity on a detector is related to the number of photons incident in one integration time, N_{phot} via $I = I_{1\gamma} N_{\text{phot}}$. Using Eq. (3.91), the limit of detection for intensity is

$$\text{LOD}_I = m \sqrt{I_\zeta I_{1\gamma}}. \quad (3.93)$$

To detect a particle, it is required $\mathcal{S}_I \geq \text{LOD}_I$, which yields equality when

$$\left| I_a^{\text{LOD}} + 2 \hat{\boldsymbol{\xi}}_b^* \cdot \hat{\boldsymbol{\xi}}_a \sqrt{I_a^{\text{LOD}} I_\zeta} \cos \Phi \right| = \sqrt{I_\zeta I_{1\gamma}}, \quad (3.94)$$

where I_a^{LOD} is the minimum detectable analyte particle intensity. When $I_\zeta \gg I_a^{\text{LOD}}$, this reduces to

$$I_a^{\text{LOD}} = \frac{m^2 I_{1\gamma}}{2 |\hat{\boldsymbol{\xi}}_b^* \cdot \hat{\boldsymbol{\xi}}_a| |\cos \Phi|}. \quad (3.95)$$

It can be seen that the closer the polarisation of the analyte scattered field is to the background field (i.e. larger $|\hat{\boldsymbol{\xi}}_b^* \cdot \hat{\boldsymbol{\xi}}_a|$), the lower the LOD, as the interference term is larger. This reveals another advantage of the leakage ring, in which the scattered light is predominantly p -polarised and therefore the analyte and surface scattered fields are (approximately) co-polarised. Furthermore, the limit of detection on particle scattered intensity (and therefore also on particle polarisability/size) is independent of the background intensity (provided it is significantly larger

than the analyte scattered intensity). This intensity LOD should be contrasted with a dark field type measurement, where one attempts to detect the direct scattered term. Given it has been established that, for metal surfaces, even surface roughness of RMS height $\lesssim 1\text{nm}$ scatters comparably to biological particles, such a setup is not realistic for the proposed system without very high quality smooth metal surfaces. Considering a dark noise level (i.e. in the absence of any analyte particle near the surface) of σ_d , assuming this is sufficiently larger than the shot noise level in I_a that it can be neglected (i.e. $\sigma_d \gg (I_a I_{1\gamma})^{1/2}$). This can arise from numerous sources, including detector dark noise, temperature variations or surface scattering. The dark noise analyte intensity limit, $I_a^{\text{LOD,dark}}$, is given by

$$I_a^{\text{LOD,dark}} = m\sigma_d. \quad (3.96)$$

From this, it can be seen $I_a^{\text{LOD}} < I_a^{\text{LOD,dark}}$ (i.e. the proposed setup has a lower detection limit than a dark field TIR microscopy measurement) when $\sigma_d > mI_{1\gamma}/(2|\hat{\boldsymbol{\xi}}_b^* \cdot \hat{\boldsymbol{\xi}}_a| |\cos \Phi|)$. Assuming the analyte and speckle field are co-polarised (in the leakage radiation ring, both will be predominantly p -polarised) so that $|\hat{\boldsymbol{\xi}}_b^* \cdot \hat{\boldsymbol{\xi}}_a| = 1$ and replacing $|\cos \Phi|$ by its mean value over uniformly distributed phase from 0 to 2π , $\langle |\cos \Phi| \rangle = 2/\pi$ gives an approximate condition $\sigma_d > (m\pi/4)I_{1\gamma}$. Given m is of order unity, the proposed method has reduced detection limits compared to a dark field approach if σ_d is larger than the intensity signal measured from ~ 1 photon per integration time (assuming it is also larger than the intensity scattered from the analyte particle). This finding is analogous to the results presented in Table 2.2, except in the context of LOD as opposed to SNR.

The minimum detectable size of analyte as determined by shot noise depends only on $I_{1\gamma}$, i.e. the detector properties, and the power incident on the particle. Specifically, using $I_a^{\text{LOD}} = |\mathcal{S}_{\mathbf{E}}|^2$, the limit of detection on polarisability α_a^{LOD} is found by putting $I_a = |\mathcal{S}_{\mathbf{E}}|^2 = I_a^{\text{LOD}}$. Substituting from Eqs. (3.77) and (3.95) yields

$$|\alpha_a^{\text{LOD}}| = \frac{m\sqrt{I_{1\gamma}}}{\frac{k_0^2}{\varepsilon_0} |G_0^\infty \hat{\mathbf{A}}_{\text{SPP}}^+| \sqrt{2I_0 |\hat{\boldsymbol{\xi}}_b^* \cdot \hat{\boldsymbol{\xi}}_a \cos \Phi|}}. \quad (3.97)$$

In order to estimate the minimum detectable size of analyte particle, Eq. (2.6) is used with $\alpha = \alpha_a^{\text{LOD}}$ and $R_a = R_a^{\text{LOD}}$, which results in

$$R_a^{\text{LOD}} = \left(\frac{m}{k_0^2 |Q(\varepsilon_a, \varepsilon_d)| |G_0^\infty \hat{\mathbf{A}}_{\text{SPP}}^+|} \right)^{\frac{1}{3}} \left(\frac{I_{1\gamma}}{2I_0 |\hat{\boldsymbol{\xi}}_b^* \cdot \hat{\boldsymbol{\xi}}_a \cos \Phi|} \right)^{\frac{1}{6}} \quad (3.98)$$

where Q is a function of the material parameters

$$Q(\varepsilon_a, \varepsilon_d) = 4\pi\varepsilon_d \left(\frac{\varepsilon_a - \varepsilon_d}{\varepsilon_a + 2\varepsilon_d} \right) \quad (3.99)$$

Importantly, $R_a^{\text{LOD}} \propto I_0^{1/6}$ so that increasing the intensity at the surface (i.e. power coupled into the SPP mode) monotonically decreases the minimum detectable particle size, which is therefore only limited by the amount of power one can couple into the SPP mode and the possibility of photodamage at high intensity. Similarly, the $|G_0^\infty \hat{\mathbf{A}}_{\text{SPP}}^+|^{-1/3}$ scaling shows the benefit of the highly directional dipolar radiation pattern in reducing R_a^{LOD} .

3.6.1 Pearson Correlation Coefficient Sensitivity

Another sensing signal that can be used is the speckle intensity Pearson correlation coefficient $C_p(\tau)$ defined as [272]

$$C_p(\tau) = \frac{\text{Cov}[I(0), I(\tau)]}{\sigma(0)\sigma(\tau)} \quad (3.100)$$

where $I(t)$ is the speckle pattern at time t , consisting of the intensities $I_i(t)$ measured at M points \mathbf{r}_i . In the context of the SPP scattering system considered in this thesis, this could be the intensity measured at points around the leakage radiation ring $I_i(t) = I(\phi_i; t)$, but more generally can apply to any speckle pattern measured over a set of points. The (spatial) standard deviation is $\sigma^2(t) = \sum_{i=1}^M (I_i(t) - \overline{I(t)})^2 / (M - 1) = \overline{I(t)^2} - \overline{I(t)}^2$ and $\overline{f(I)} = \sum_{i=1}^M f(I_i) / M$. The covariance is defined as $\text{Cov}[I(0), I(\tau)] = \sum_{i=1}^M (I_i(0) - \overline{I(0)})(I_i(\tau) - \overline{I(\tau)}) / (M - 1)$. While the sample mean and covariance use $M - 1$ as a divisor to give an unbiased estimate of the population mean and variance, it is assumed throughout that M is sufficiently large that the

difference between using M and $M - 1$ as a divisor is negligible. The inequality $-1 \leq C_p \leq 1$ applies. If the scattering configuration remains unchanged up to $t = \tau$, the speckle remains unchanged and, in the absence of noise, $C_p = 1$. Conversely, if the analyte particle enters the sensing volume in the time interval, the scattering from the analyte particle changes the speckle pattern and hence $C_p < 1$. Thus, a step in C_p indicates the presence of the analyte particle. The Pearson correlation coefficient benefits from using the entire speckle pattern, allowing for averaging out of noise over intensity measurements at different positions. Furthermore, C_p is unchanged by scaling and addition of constant intensity to the speckle pattern, so that global intensity fluctuations, e.g. due to fluctuating incident power, leave C_p unchanged. The sensitivity for C_p is defined as

$$\mathcal{S}_C = |C_p^{\text{bind}} - C_p(0)| = 1 - C_p^{\text{bind}}, \quad (3.101)$$

where C_p^{bind} is $C_p(\tau)$ in the case where an analyte particle has bound in the time interval τ and remained bound. The LOD is defined as

$$\text{LOD}_C = m\Delta C_p^{\text{noise}} \quad (3.102)$$

where $\Delta C_p^{\text{noise}}$ is the mean amplitude of the change to C_p due to noise. In order to calculate these two quantities, the effect of perturbations to the intensity I_i on C_p must be considered

Ergodic Rayleigh Statistics Speckle

It is important to note that all the averaged quantities (i.e. $\text{Cov}[I(0), I(t)]$ and σ) are spatially averaged for a single speckle realisation. It has already been demonstrated that the speckle arising from scattering from a rough surface is generally not statistically homogeneous owing to the dipolar like directional scattering and therefore the speckle patterns are not ergodic. As such, these spatial averages cannot be directly related to the averages over realisations. It is, however, worth studying the case of a speckle pattern obeying Rayleigh statistics while also being spatially ergodic (i.e. averaging over space is equivalent to averaging over realisations), partially due to the simplification it provides. Furthermore, Eq. (3.56) shows that when

$|f_C| \ll 1$, as is the case in the leakage radiation ring, the contrast with respect to averaging over realisations is unity, i.e. $\sigma_\zeta(\mathbf{r}) = \langle I_\zeta(\mathbf{r}) \rangle$. Provided the speckle size is much smaller than the scale on which the dipole intensity envelope of $\langle I_\zeta(\mathbf{r}) \rangle$ varies, the spatial average over a region containing many speckles, but without significant variation in $\langle I_\zeta(\mathbf{r}) \rangle$, can be considered to be ergodic since different speckles are statistically independent (at least within the single scattering mode). As a result, each speckle is an independent sampling of the same intensity probability distribution (since $\langle I_\zeta(\mathbf{r}) \rangle$ is approximately the same) over disorder realisation and, provided there are sufficient number of speckles, the spatial averages will converge to the average over realisations. Thus, there are practical conditions where spatial ergodicity approximately applies. Under the assumption of the unity contrast with respect to the spatial mean and standard deviation, one has $\sigma(0) = \overline{I(0)}$. The background speckle will be denoted $I_b = I(0)$ and its standard deviation $\sigma_b = \sigma(0)$. The presence of the analyte particle perturbs the intensity at \mathbf{r}_i by ΔI_i , so that the covariance can be expressed, using the properties of the covariance, as $\text{Cov}(I_b, I_b + \Delta I) = \text{Cov}(I_b, I_b) + \text{Cov}(I_b, \Delta I) = \sigma_b^2 + \text{Cov}(I_b, \Delta I)$. As a result, the sensitivity is

$$\mathcal{S}_C = 1 - \frac{\sigma_b}{\sigma(t)} - \frac{\text{Cov}(I_b, \Delta I)}{\sigma_b \sigma(\tau)}. \quad (3.103)$$

Additionally, $\sigma^2(\tau) = \sigma_b^2 + \sigma_\Delta^2 + 2 \text{Cov}(I_b, \Delta I)$, where $\sigma_\Delta = \sum_{i=1}^M (\Delta I_i - \overline{\Delta I})^2 / M$. Replacing σ_b with $\overline{I_b}$ results in

$$\mathcal{S}_C = 1 - \left(1 + \frac{\sigma_\Delta^2}{\overline{I_b}^2} + 2 \frac{\text{Cov}(I_b, \Delta I)}{\overline{I_b}^2} \right)^{-\frac{1}{2}} - \frac{\text{Cov}(I_b, \Delta I)}{\overline{I_b} \sqrt{\sigma_b^2 + \sigma_\Delta^2 + 2 \text{Cov}(I_b, \Delta I)}}. \quad (3.104)$$

The intensity change arising from an analyte particle scattering intensity $I_{a,i}$ into pixel i is $\Delta I_i = I_{a,i} + (I_{a,i} I_{b,i})^{1/2} \cos \Phi_i$ and includes the interference term. The phase difference between the background speckle and the analyte scattered field Φ_i is uniformly distributed between 0 and 2π and independent of the speckle amplitude $I_{b,i}$ (and also $I_{a,i}$) for fully developed speckle. As a result, using ergodicity, $\overline{\cos \Phi} = \langle \cos \Phi \rangle = 0$ and, from the independence of the phase and amplitude, $\overline{f(I_a, I_b)g(\Phi)} = \overline{f(I_a, I_b)} \overline{g(\Phi)}$ for arbitrary functions f and g . This allows some

terms in Eq. (3.104) to be calculated. For example, the covariance becomes

$$\begin{aligned}
 \text{Cov}(I_b, \Delta I) &= \overline{I_b(I_a + \sqrt{I_a I_b} \cos \Phi_i)} - \bar{I}_b(\bar{I}_a + \sqrt{I_a I_b} \cos \Phi) \\
 &= \bar{I}_b \bar{I}_a - \bar{I}_b \bar{I}_a \\
 &= 0
 \end{aligned} \tag{3.105}$$

where all terms proportional to $\cos \Phi$ vanish upon averaging and the independence of I_b and I_a means $\overline{I_a I_b} = \bar{I}_a \bar{I}_b$. In addition, $\overline{\Delta I} = 0$ and therefore the σ_Δ^2 is given by

$$\sigma_\Delta^2 = \frac{4}{M} \sum_{i=1}^M I_{a,i} I_{b,i} \cos^2 \Phi_i \tag{3.106}$$

$$= 2\bar{I}_a \bar{I}_b \tag{3.107}$$

where $\overline{\cos^2 \Phi} = \langle \cos^2 \Phi \rangle = 1/2$ arises from the uniform distribution of Φ . Substituting these results into Eq. (3.104) gives

$$\begin{aligned}
 \mathcal{S}_C &= 1 - \left(1 + \frac{2\bar{I}_a}{\bar{I}_b}\right)^{-\frac{1}{2}} \\
 &\approx \frac{\bar{I}_a}{\bar{I}_b} + \mathcal{O}\left(\frac{\bar{I}_a^2}{\bar{I}_b^2}\right).
 \end{aligned} \tag{3.108}$$

In order to find LOD_C , the noise in intensity measurements must be translated into the noise perturbation to C_p . The noise perturbs C_p in exactly the same form as Eq. (3.104), except ΔI is replaced by the perturbation due to noise, as opposed to the perturbation from the analyte particle, so that $\Delta C_p^{\text{noise}}$ is given analogously to Eq. (3.103) as

$$\Delta C_p^{\text{noise}} = 1 - \frac{\sigma_b^2 + \text{Cov}(I_b, \Delta_{\text{noise}})}{\sigma_b \sqrt{\sigma_b^2 + \sigma_{\text{noise}}^2 + \text{Cov}(I_b, \Delta_{\text{noise}})}}, \tag{3.109}$$

where $\Delta_{\text{noise},i}$ is the noise in the i th pixel. Under the shot noise model, $\Delta_{\text{noise},i}$ is Poisson distributed. For large enough photon numbers ($\gtrsim 1000$ per integration time), the noise is well approximated as a zero-mean Gaussian random variable with standard deviation matching that of the underlying Poisson distribution (i.e. Eq. (3.91)), so that $\Delta_{\text{noise},i} \sim \mathcal{N}(0, I_{b,i} I_{1\gamma})$.

Note that in addition to distributions over space and disorder (i.e. speckle) realisations, there is now distribution over noise realisations. Assuming the noise in each pixel is independent, $\overline{\Delta_{\text{noise}}}$, as a linear combination of zero mean Gaussian random variables, is distributed (over noise realisations) according to a zero mean Gaussian distribution with standard deviation $\sum_{i=1}^M I_{b,i} I_{1\gamma} / M^2 = \overline{I_b} I_{1\gamma} / M$. Similarly, the factor $\overline{I_b \Delta_{\text{noise}}}$ is also a linear combination of zero mean Gaussian random variables, and thus is also Gaussian distributed as follows

$$\begin{aligned} \overline{I_b \Delta_{\text{noise}}} &= \frac{1}{M} \sum_{i=1}^M I_{b,i} \Delta_{\text{noise},i} \\ \implies \overline{I_b \Delta_{\text{noise}}} &\sim \mathcal{N} \left(0, \sum_{i=1}^M \left(\frac{I_{b,i}}{M} \right)^2 I_{b,i} I_{1\gamma} \right), \end{aligned} \quad (3.110)$$

where again the distribution is over noise realisations. As a result, the covariance of the background speckle and the noise perturbation $\text{Cov}(I_b, \Delta_{\text{noise}}) = \overline{I_b \Delta_{\text{noise}}} - \overline{I_b} \overline{\Delta_{\text{noise}}}$ vanishes averaging over noise realisations as follows

$$\begin{aligned} \mathbb{E}(\text{Cov}(I_b, \Delta_{\text{noise}})) &= \underbrace{\mathbb{E}(\overline{I_b \Delta_{\text{noise}}})}_{=0} - \overline{I_b} \underbrace{\mathbb{E}(\overline{\Delta_{\text{noise}}})}_{=0} \\ &= 0, \end{aligned} \quad (3.111)$$

where $\mathbb{E}(\cdot)$ denotes averaging over noise realisations and $\overline{I_b}$ does not depend on noise realisations. The (spatial) variance of the noise perturbation, σ_{noise}^2 is given by

$$\sigma_{\text{noise}}^2 = \frac{1}{M} \sum_{i=1}^M (\Delta_{\text{noise},i} - \overline{\Delta_{\text{noise}}})^2 \quad (3.112)$$

Considering the distribution over noise realisations, due to the fact that both $\Delta_{\text{noise},i}$ and $\overline{\Delta_{\text{noise}}}$ have normal distributions, it can be deduced that $\Delta_{\text{noise},i} - \overline{\Delta_{\text{noise}}} \sim \mathcal{N}(0, (I_{b,i} - \overline{I_b}/M) I_{1\gamma})$. As

a result, the noise average spatial standard deviation $\mathbb{E}(\sigma_{\text{noise}}^2)$ is

$$\begin{aligned}
 \mathbb{E}(\sigma_{\text{noise}}^2) &= \frac{1}{M} \sum_i \mathbb{E} [(\Delta_{\text{noise},i} - \bar{\Delta}_{\text{noise}})^2] \\
 &= \frac{1}{M} \sum_i [(I_{b,i} - \bar{I}_b/M)I_{1\gamma}] \\
 &= \bar{I}_b I_{1\gamma} \left(1 - \frac{1}{M}\right) \approx \bar{I}_b I_{1\gamma}
 \end{aligned} \tag{3.113}$$

where it has been assumed $M \gg 1$ for the final approximate result, and the result $\mathbb{E}(X^2) = \sigma^2$ for a zero mean Gaussian distribution of standard deviation σ has been used. Replacing σ_{noise}^2 with the noise averaged value $\mathbb{E}(\sigma_{\text{noise}}^2)$ and $\text{Cov}(I_b, \Delta_{\text{noise}})$ with $\mathbb{E}[\text{Cov}(I_b, \Delta_{\text{noise}})]$ in Eq. (3.109) and also using $\sigma_b = \bar{I}_b$ results in

$$\begin{aligned}
 \Delta C_p^{\text{noise}} &= 1 - (1 + I_{1\gamma}/\bar{I}_b)^{-1/2} \\
 &\approx \frac{I_{1\gamma}}{2\bar{I}_b} + \mathcal{O}\left(\frac{I_{1\gamma}^2}{\bar{I}_b^2}\right).
 \end{aligned} \tag{3.114}$$

The binomial expansion is valid under the assumption $\bar{I}_b \gg I_{1\gamma}$, i.e. the average number of photons incident on a pixel in one integration time averaged over all pixels is much larger than one. Setting $\mathcal{S}_C = \text{LOD}_C = m\Delta C_p^{\text{noise}}$ gives \bar{I}_a^{LOD} for a correlation coefficient based sensing setup as

$$\bar{I}_a^{\text{LOD}} = \frac{mI_{1\gamma}}{2}. \tag{3.115}$$

Compared to Eq. (3.95), both sensing metrics give comparable forms for I_a^{LOD} , in the sense that both are proportional to the detector property $I_{1\gamma}$, though for the correlation coefficient, this is the average around the whole leakage ring (or entire speckle pattern), as opposed to just a given point for a pointwise intensity measurement limit. Note that Eq. (3.115) is intended as an approximate scaling rule for the detection limits in terms of the intensities involved, and the prefactors may not give an accurate result in any experiment owing to approximations made in replacing quantities in $\Delta C_p^{\text{noise}}$ with averaged quantities.

3.7 Conclusion

In this section, a random SPP scattering system has been introduced as a route to single particle sensing. In order to facilitate study of this biosensing platform, a review of some fundamental principles and techniques of plasmonics and (random) scattering theory was undertaken. The statistical properties of the speckle arising from SPP scattering from a rough metal surface were studied in the single scattering regime and linked to the surface roughness property. Importantly, the statistics of the speckle pattern were shown to be strongly dependent on position, and the far field intensity is predominantly confined to the leakage radiation ring. Similarly, light scattered from an analyte particle is also strongly confined to the leakage radiation ring. This confinement is shown to have benefits in terms of improved sensitivity at points in the leakage ring. The single scattering sensitivity analysis using the scattering model developed revealed that binding of a ‘biological’ particle with radius of 10nm gives a fractional intensity change on the order 10^{-4} , consistent with the intensity contrasts in detection of single proteins in iSCAT. In the shot noise case, the LOD for intensity measurements was found to be independent of the surface scattered intensity and proportional to the intensity scattered from the analyte particle. As such, it is maximised in the leakage ring. Furthermore, $R_a^{\text{LOD}} \propto I_0^{-1/6}$ means one can decrease the minimum detectable particle size via large intensity at the surface, so that tightly confined SPP modes provide greater sensitivity for the same total mode energy.

Overall, the benefits of the proposed system can be summarised as

- simple SPR-like experimental setup
- strong interaction with analyte particle due to plasmonic confinement
- interferometric signal enhancement as discussed in Section 2.2.3
- no requirement for interferometric stability or external reference field
- stronger signal in the leakage radiation ring due to highly directional scattering
- no need for high quality surfaces due to random roughness

This list will be revisited later in Chapter 7 after more in-depth study.

The single scattering analysis in this chapter considered simply the change caused by the addition of an analyte particle and analysed this in the context of sensing. As the single scattering regime was assumed, the change was found to be a simple interference of the analyte particle scattered field with a (fixed) speckle field. Due to the dependence on phase, interference effects are very sensitive to small changes in environment, including scatterer position. Thus, it might be wondered whether further information on the particle may be extracted from the signal, in particular on the analyte particle position. The next chapter investigates this and explores how the proposed system enables SPT in the single scattering regime.

Chapter 4

Tracking in the Single Scattering Regime

Some of the work presented in this chapter has been published in Ref. [273].

In order to perform single particle tracking using the proposed random SPP scattering system, there has to exist a method to extract the analyte particle trajectory from a time series of speckle patterns. This chapter explores how the speckle pattern depends on the analyte particle position in the single scattering regime and hence develops and studies an algorithm to extract the analyte particle trajectory from the far field speckle intensity pattern.

4.1 Field Dependence on Particle Position

Starting from the Born approximation of Eq. (3.68), the field $\mathbf{E}_a(\mathbf{r}; \mathbf{r}_a)$ scattered from an analyte particle centred at position \mathbf{r}_a occupying a volume \mathcal{V}_a (of arbitrary shape) above the thin film structure is given by

$$\mathbf{E}_a(\mathbf{r}; \mathbf{r}_a) = \frac{\omega^2}{c^2} \int_{\mathbf{r}' \in \mathcal{V}_a} [\varepsilon_a(\mathbf{r}') - \varepsilon_d] G_0(\mathbf{r}, \mathbf{r}') \mathbf{E}_0(\mathbf{r}') d^3 \mathbf{r}' \quad (4.1)$$

where $G_0(\mathbf{r}, \mathbf{r}')$ is the Green's tensor for the thin film geometry given by Eq. (3.29). Upon shifting the analyte by $\delta \mathbf{r} = (\delta x, \delta y, \delta z)$, the scattered field can be calculated in the same way but with the integral over a different volume \mathcal{V}'_a , corresponding to the original volume \mathcal{V}_a

occupied by the analyte translated by $\delta\mathbf{r}$, with a translated dielectric function $\varepsilon'_a(\mathbf{r})$. Assuming the particle is spherically symmetric or the particle does not rotate upon translation (see later for a discussion of analyte particle rotation), the transformed dielectric function is related to the original one through $\varepsilon'_a(\mathbf{r} + \delta\mathbf{r}) = \varepsilon_a(\mathbf{r})$. As a result, the scattered field from the shifted particle, $\mathbf{E}_a(\mathbf{r}; \mathbf{r}_a + \delta\mathbf{r})$, is given by

$$\mathbf{E}_a(\mathbf{r}; \mathbf{r}_a + \delta\mathbf{r}) = \frac{\omega^2}{c^2} \int_{\mathbf{r}'' \in V'} [\varepsilon'_a(\mathbf{r}'') - \varepsilon_d] G_0(\mathbf{r}, \mathbf{r}'') \mathbf{E}_0(\mathbf{r}'') d^3\mathbf{r}'' . \quad (4.2)$$

A change of variables $\mathbf{r}'' = \mathbf{r}' + \delta\mathbf{r}$ translates the integration volume from V' back to V , and the integral becomes

$$\mathbf{E}_a(\mathbf{r}; \mathbf{r}_a + \delta\mathbf{r}) = \frac{\omega^2}{c^2} \int_{\mathbf{r}' \in V} [\varepsilon_a(\mathbf{r}') - \varepsilon_d] G_0(\mathbf{r}, \mathbf{r}' + \delta\mathbf{r}) \mathbf{E}_0(\mathbf{r}' + \delta\mathbf{r}) d^3\mathbf{r}' . \quad (4.3)$$

Compared to Eq. (4.1), the integrals are closely related, the only change being the shift of \mathbf{r}' to $\mathbf{r}' + \delta\mathbf{r}$ in the arguments of G_0 and \mathbf{E}_0 . For a plane wave SPP illumination, it follows from Eq. (3.5) that, for an analyte particle above the thin film stack, the fields evaluated at different positions are related by $\mathbf{E}_0(\mathbf{r}' + \delta\mathbf{r}) = e^{ik_{\text{SPP}}\delta x} e^{-\kappa_d\delta z} \mathbf{E}_0(\mathbf{r}')$, assuming the analyte particle remains in the same medium. Recalling from Chapter 3 that the far field is proportional to the 2D (x, y) Fourier transform of the field at $z = 0$ or $z = -d$ depending on whether the observation point is above or below the multilayer stack, the Fourier transform of Eq. (4.3) is considered. The scattered field below the multilayer stack in the Fourier plane $\tilde{\mathbf{E}}_a(\mathbf{k}_{\parallel}; \mathbf{r}_a)$ (the tilde will denote Fourier plane quantities throughout the chapter) is given by the 2D Fourier transform with respect to the transverse position vector $\boldsymbol{\rho} = (x, y)$ of Eq. (4.3) evaluated at the glass metal interface. This can be observed using Fourier plane imaging [274], while the scattered field in the far field is proportional to it. From the properties of the Green's tensor established in Chapter 3, the Fourier space Green's tensor \tilde{G}_0 evaluated at different source positions for observation points below the multilayer stack can be related using Eqs. (3.26) and (3.30) through

$$\tilde{G}_0(\mathbf{k}_{\parallel}; \boldsymbol{\rho}' + \delta\boldsymbol{\rho}, z, z' + \delta z) = \tilde{G}_0(\mathbf{k}_{\parallel}; \boldsymbol{\rho}', z, z') e^{-i\mathbf{k}_{\parallel} \cdot \delta\boldsymbol{\rho}} e^{ik_z\delta z} , \quad (4.4)$$

where k_z is defined as in Eqs. (3.24)-(3.26) and $\delta\boldsymbol{\rho} = (\delta x, \delta y)$ is the transverse component of $\delta\mathbf{r}$. Substituting Eq. (4.4) into the Fourier transform of Eq. (4.3), the field scattered by the translated analyte particle in the far field in the direction defined by \mathbf{k}_{\parallel} is hence found to be

$$\tilde{\mathbf{E}}_a(\mathbf{k}_{\parallel}; \mathbf{r}_a + \delta\mathbf{r}) = \underbrace{e^{ik_{\text{SPP}}\delta x} e^{-\kappa_d\delta z}}_{\text{change to } \mathbf{E}_0} \underbrace{e^{-i\mathbf{k}_{\parallel}\cdot\delta\boldsymbol{\rho}} e^{ik_z\delta z}}_{\text{change to } \tilde{\mathbf{G}}_0} \tilde{\mathbf{E}}_a(\mathbf{k}_{\parallel}; \mathbf{r}_a). \quad (4.5)$$

Considering an observation point in the leakage radiation ring $\theta = \Theta_{\text{SPP}}$, such that $k_{\parallel} = k'_{\text{SPP}}$, then it follows from Eq. (4.5) that the scattered field after the translation of the analyte particle at azimuthal coordinate ϕ (measured anticlockwise from $+x$ axis) on the leakage radiation ring, $\tilde{\mathbf{E}}(\phi; \mathbf{r}_a + \delta\mathbf{r})$, is related to the field at the same point before the translation $\tilde{\mathbf{E}}(\phi; \mathbf{r}_a)$ through a phase and amplitude shift as follows

$$\tilde{\mathbf{E}}_a(\phi; \mathbf{r}_a + \delta\mathbf{r}) = e^{i\Psi(\phi; \delta\boldsymbol{\rho}, \delta z)} e^{-\Lambda(\delta x, \delta z)} \tilde{\mathbf{E}}_a(\phi; \mathbf{r}_a). \quad (4.6)$$

The phase shift and decay functions $\Psi(\phi; \delta\boldsymbol{\rho}, \delta z)$ and $\Lambda(\delta x, \delta z)$ are defined as

$$\Psi(\phi; \delta\boldsymbol{\rho}, \delta z) = k'_{\text{SPP}}\delta x(1 - \cos\phi) - k'_{\text{SPP}}\delta y \sin\phi - \kappa_d''\delta z \quad (4.7)$$

$$\Lambda(\delta x, \delta z) = k''_{\text{SPP}}\delta x + \left[\kappa_d' + (k_{\text{SPP}}'^2 - \varepsilon_d k_0^2)^{1/2} \right] \delta z. \quad (4.8)$$

Note that since the SPP wavenumber is larger than the wavenumber in the upper dielectric (aqueous solution), $k_z = ik_z = i(k_{\text{SPP}}'^2 - \varepsilon_d k_0^2)^{1/2}$ is imaginary and thus the $\exp(ik_z\delta z) = \exp(-\kappa_z z\delta z)$ factor in Eq. (4.5) contributes to the decay factor rather than the phase shift. From Eq. (4.6) it can thus be seen that as the analyte particle moves, the direct scattered field in the leakage ring acquires an additional phase shift Ψ with respect to the background speckle reference field, in addition to a change in amplitude. Accordingly, in the single scattering regime, the total field $\tilde{\mathbf{E}} = \tilde{\mathbf{E}}_b + \tilde{\mathbf{E}}_a$ including the light scattered from the random surface $\tilde{\mathbf{E}}_b$ changes in a predictable manner with the position of the analyte particle. As a result, the changes in the speckle pattern as the analyte moves around contain information about the trajectory of the analyte particle.

4.2 Tracking Algorithm

In a tracking scenario, multiple frames of the speckle intensity pattern are collected over time so as to capture the dynamics of the analyte particle. When the i th frame ($i = 1, 2, \dots$) is measured at time t_i , the particle is located at \mathbf{r}_i which gives rise to a speckle pattern $\tilde{I}_i(\phi)$ around the leakage ring. It is assumed that the time over which the measurements are made is small compared to the timescale over which the analyte particle moves, so that the particle can be considered to be in the same position throughout the measurement. Within the single scattering regime, $\tilde{I}_i(\phi)$ is given by

$$\tilde{I}_i(\phi) = |\tilde{\mathbf{E}}(\phi; \mathbf{r}_i)|^2 \quad (4.9)$$

$$= |\tilde{\mathbf{E}}_b(\phi) + \tilde{\mathbf{E}}_a(\phi; \mathbf{r}_i)|^2 \quad (4.10)$$

$$= \tilde{I}_b(\phi) + \tilde{I}_a(\phi; \mathbf{r}_i) + 2 \operatorname{Re} \left[\tilde{\mathbf{E}}_b^*(\phi) \cdot \tilde{\mathbf{E}}_a(\phi; \mathbf{r}_i) \right], \quad (4.11)$$

where $\tilde{I}_b = |\tilde{\mathbf{E}}_b(\phi)|^2$ and $\tilde{I}_a = |\tilde{\mathbf{E}}_a(\phi; \mathbf{r}_i)|^2$ are the background speckle intensity and analyte particle scattered intensity respectively. As established in Chapter 3, small biological particles scatter weakly compared to the light scattered from a rough metal surface and so it may be assumed $\tilde{I}_a \ll \tilde{I}_b$, allowing \tilde{I}_i to be expressed

$$\tilde{I}_i(\phi) \approx \tilde{I}_b(\phi) + 2 \operatorname{Re} \left[\tilde{\mathbf{E}}_b^*(\phi) \cdot \tilde{\mathbf{E}}_a(\phi; \mathbf{r}_i) \right]. \quad (4.12)$$

The interference term depends on particle position, tracing out different fringe patterns at different points, ϕ , on the ring (see Fig. 4.1 for a visualisation of these fringe patterns). As the phase of the background speckle $\tilde{\mathbf{E}}_b(\phi)$ is random and unknown, the initial fringe displacement is unknown. As a result, there is insufficient information in two frames of intensity measurements to extract the change in position $\delta \mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ between the frames. For example, consider taking two frames, for which the background speckle subtracted intensity in the i th frame,

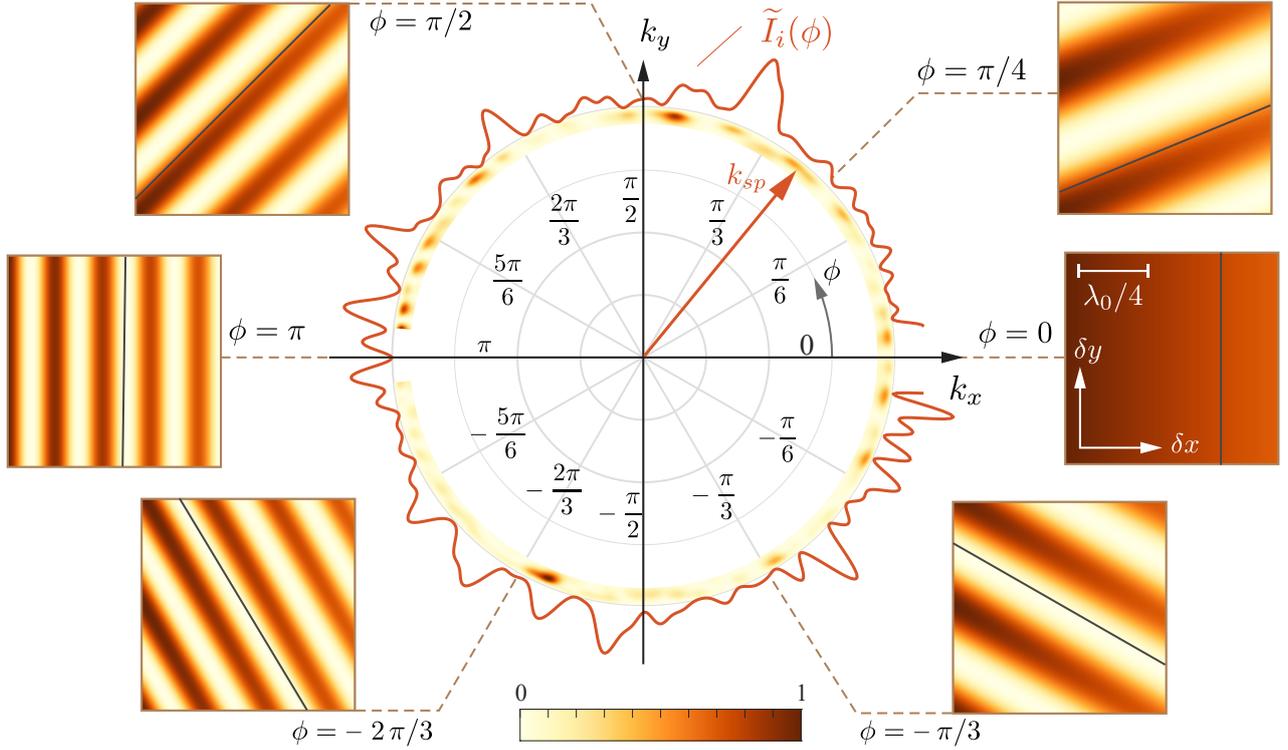


Figure 4.1: (center) Example scattered intensity in the Fourier plane and the profile around the leakage ring $\tilde{I}_i(\phi)$. (panels) Fringe patterns one would see at different points on the ring from scanning the analyte particle in the x and y directions. In the $\phi = 0$ direction, no fringes are seen as the phase of $\tilde{\mathbf{E}}_a$ does not change, so only decay effects are seen. Due to the random speckle phase, the fringe pattern has an unknown (and different) offset at each ϕ , as depicted by the solid black lines. *Figure reprinted with permission from Ref. [273] © 2021 IEEE.*

$\Delta\tilde{I}_i = \tilde{I}_i - \tilde{I}_b$, is an interference term given in each frame in the j th pixel by

$$\text{Frame 1: } \Delta\tilde{I}_1 = A_j \cos \Phi_j \quad (4.13)$$

$$\text{Frame 2: } \Delta\tilde{I}_2 = A_j e^{-\Lambda(\delta\mathbf{r}_{12})} \cos [\Phi_j + \Psi(\phi_j; \delta\mathbf{r}_{12})]. \quad (4.14)$$

where $\delta\mathbf{r}_{12}$ is the shift in analyte position between the two frames. With N pixels, there are $2N$ measurements, while there are N unknown amplitudes A_i , N unknown phases Φ_i and three unknown components of the shift $\delta\mathbf{r}_{12}$, giving $2N + 3 > 2N$ unknowns. As there are more unknowns than measurements, finding $\delta\mathbf{r}_{12}$ from two frames of data is an underdetermined problem. To overcome this issue, the tracking algorithm can use three frames in an analogous manner to phase shifting interferometry [275, 276], in which three or more known reference phases are used to calculate an unknown wavefront phase. The interference term in the three

frames is given by

$$\text{Frame 1: } \Delta \tilde{I}_1 = A_j \cos \Phi_j \quad (4.15)$$

$$\text{Frame 2: } \Delta \tilde{I}_2 = A_j e^{-\Lambda(\delta \mathbf{r}_{12})} \cos [\Phi_j + \Psi(\phi_j; \delta \mathbf{r}_{12})] \quad (4.16)$$

$$\text{Frame 3: } \Delta \tilde{I}_3 = A_j e^{-\Lambda(\delta \mathbf{r}_{13})} \cos [\Phi_j + \Psi(\phi_j; \delta \mathbf{r}_{13})]. \quad (4.17)$$

Now there are $3N$ measurements, but since A_i and Φ_i are fixed for all three frames, the number of unknowns increases to $2N + 6$ due to the three components of $\delta \mathbf{r}_{13}$. Provided $N \geq 6$, the problem is now solvable for the two shifts in position. Assuming that a reference measurement of $\tilde{I}_b(\phi)$, which is fixed for a given surface roughness profile, is taken before the analyte particle moves into the sensing volume, the background subtracted intensities $\Delta \tilde{I}_i(\phi) = \tilde{I}_i(\phi) - \tilde{I}_b(\phi)$ for three data frames (labelling them as $i = 1, 2, 3$) are given by

$$\Delta \tilde{I}_1(\phi) = 2 \operatorname{Re} \left[\tilde{\mathbf{E}}_b^*(\phi) \cdot \tilde{\mathbf{E}}_a(\phi; \mathbf{r}_2) e^{i\Psi_{21}} e^{-\Lambda_{21}} \right] \quad (4.18)$$

$$\Delta \tilde{I}_2(\phi) = 2 \operatorname{Re} \left[\tilde{\mathbf{E}}_b^*(\phi) \cdot \tilde{\mathbf{E}}_a(\phi; \mathbf{r}_2) \right] \quad (4.19)$$

$$\Delta \tilde{I}_3(\phi) = 2 \operatorname{Re} \left[\tilde{\mathbf{E}}_b^*(\phi) \cdot \tilde{\mathbf{E}}_a(\phi; \mathbf{r}_2) e^{i\Psi_{23}} e^{-\Lambda_{23}} \right], \quad (4.20)$$

where Eq. (4.6) has been used to express $\tilde{\mathbf{E}}_a(\phi; \mathbf{r}_1)$ and $\tilde{\mathbf{E}}_a(\phi; \mathbf{r}_3)$ in terms of $\tilde{\mathbf{E}}_a(\phi; \mathbf{r}_2)$, Ψ_{ij} and Λ_{ij} are $\Psi(\phi; \delta \mathbf{r}_{ij})$ and $\Lambda(\delta \mathbf{r}_{ij})$ respectively. Expanding the complex exponential in Eq. (4.18) into real and imaginary parts gives

$$\Delta \tilde{I}_1(\phi) = e^{-\Lambda_{21}} \cos \Psi_{21} \Delta \tilde{I}_2(\phi) - e^{-\Lambda_{21}} \sin \Psi_{21} \tilde{K}(\phi), \quad (4.21)$$

where Eq. (4.19) has been used and $\tilde{K}(\phi) = 2 \operatorname{Im}[\tilde{\mathbf{E}}_b^*(\phi) \cdot \tilde{\mathbf{E}}_a(\phi; \mathbf{r}_2)]$ defined. Following the same process with Eq. (4.20) results in

$$\Delta \tilde{I}_3(\phi) = e^{-\Lambda_{23}} \cos \Psi_{23} \Delta \tilde{I}_2(\phi) - e^{-\Lambda_{23}} \sin \Psi_{23} \tilde{K}(\phi). \quad (4.22)$$

Multiplying Eq. (4.21) by $e^{-\Lambda_{23}} \sin \Psi_{23}$ and Eq. (4.22) by $e^{-\Lambda_{21}} \sin \Psi_{21}$ and subtracting to eliminate $\tilde{K}(\phi)$ gives a single equation for each point ϕ on the ring, which takes the form

$$\mathbf{u} \cdot \mathbf{\Delta} = 0 \quad (4.23)$$

where

$$\mathbf{u} = \begin{pmatrix} e^{-\Lambda_{23}} \sin \Psi_{23} \\ e^{-\Lambda_{23}-\Lambda_{21}} \sin(\Psi_{21} - \Psi_{23}) \\ -e^{-\Lambda_{21}} \sin \Psi_{21} \end{pmatrix}$$

and $\mathbf{\Delta} = (\Delta\tilde{I}_1, \Delta\tilde{I}_2, \Delta\tilde{I}_3)^T$. The trigonometric identity $\cos \Psi_{23} \sin \Psi_{21} - \cos \Psi_{21} \sin \Psi_{23} = \sin(\Psi_{21} - \Psi_{23})$ has been used to simplify the form of \mathbf{u} . Note that Eq. (4.23) involves only the measured intensity differences and the two shift vectors $\delta\mathbf{r}_{21}$ and $\delta\mathbf{r}_{23}$ and holds for any value of ϕ on the ring. Assuming then that the intensity is sampled at N discrete angles ϕ_k , $k = 1, 2, \dots, N_\phi$, Eq. (4.23) can form a set of N_ϕ equations with 6 unknowns (three components from each of the shift vectors). Finding a solution to this set of non-linear equations would give the steps $\delta\mathbf{r}_{12}$ and $\delta\mathbf{r}_{23}$. By solving for each set of 3 consecutive frames, one can build up the analyte particle trajectory via finding the shifts between each frame. A least squares approach to solving Eq. (4.23) corresponds to extracting estimates of the $\delta\mathbf{r}_{21}$ and $\delta\mathbf{r}_{23}$ given by the solution to the 6D (global) minimisation

$$(\delta\hat{\mathbf{r}}_{21}, \delta\hat{\mathbf{r}}_{23}) = \arg \min_{(\delta\mathbf{r}_{21}, \delta\mathbf{r}_{23})} \sum_{k=1}^N \left(\frac{\mathbf{u}(\phi_k) \cdot \mathbf{\Delta}(\phi_k)}{|\mathbf{u}(\phi_k)|} \right)^2, \quad (4.24)$$

where $\delta\hat{\mathbf{r}}_{ij}$ denotes an estimate of the true step $\delta\mathbf{r}_{ij}$. The vector \mathbf{u} has been normalised by its magnitude in order to exclude the $\mathbf{u}(\phi_k) = \mathbf{0} \forall \phi_k$ solution, which is always present. This least squares approach to finding the particle shifts provides robustness against small perturbations to the solution from the correct values of $\delta\mathbf{r}_{21}$ and $\delta\mathbf{r}_{23}$ due to experimental noise and the fact that the small \tilde{I}_a term neglected in Eq. (4.12) is actually present. Successive application of the algorithm to sets of 3 frames allows the trajectory of a particle from frame to frame over an

arbitrary number of frames may be reconstructed.

4.2.1 Minimisation Procedure and Trajectory Consistency Check

An important component of the tracking algorithm is the choice of minimisation procedure. The solution corresponds to the global minimum, as opposed to a local one. As such, it is important to avoid converging to a local minimum. The global minimisation procedure used in this thesis is a multi-start minimisation algorithm [277], in which a local minimisation algorithm is repeated for multiple random start points and the lowest minimum is taken. While it is possible that none of the starting points converge to the global minimum, the fact the tracking algorithm uses three frames to simultaneously extract two steps allows for mitigation of this problem. Overlapping sets of three frames allow separate estimates for the same step generated from minimisation of different functions arising from different measurements. For example, $\delta\mathbf{r}_{23}$ can be estimated from frames 1,2,3 and from frames 2,3,4. By requiring the two estimates to agree within some small margin for error, a consistency check indicates when the minimisation is failing. In particular, inconsistencies can occur if either the minimum found is not the global minimum, or the effects of noise or a non-negligible direct scattering term mean that the true step no longer corresponds to the global minimum. In the first case, the minimisation procedure can be rerun for the two relevant sets of frames with new additional random starting points until consistent global minima for each are found. In the second case, while the algorithm cannot give the correct step (regardless of the minimisation procedure), the consistency check nevertheless indicates when the algorithm is failing.

4.2.2 Sign Ambiguity of Transverse Step

From Eq. (4.7) it can be seen that $\Psi(\phi; -\delta\boldsymbol{\rho}, -\delta z) = -\Psi(\phi; \delta\boldsymbol{\rho}, \delta z)$. This means that, in the absence of the decay factors, $\mathbf{u}(-\delta\mathbf{r}_{21}, -\delta\mathbf{r}_{23}) = -\mathbf{u}(\delta\mathbf{r}_{21}, \delta\mathbf{r}_{23})$ and the negative of the true steps would satisfy Eq. (4.23). In reality, the decay of the field in the x and z directions ensures $-\delta\mathbf{r}$ is not a solution. If the decay is weak (e.g. small δz) the exact backward step may turn out to be either a very low local minimum in the minimisation landscape, or, if noise is strong enough to sufficiently perturb the landscape, a global minimum. Since the decay

in the z direction is over a much shorter length scale than in the x direction ($\kappa'_d \gg k''_{\text{SPP}}$), and the phase shift depends only weakly on δz , in practice, this incorrect sign error shows up occasionally but is usually only seen in the estimate of the transverse step. In order to reduce the occurrence of these sign ambiguities, an extra step is added to the algorithm. Firstly, the minimisation algorithm is run as usual to find the global minimum, and then a second minimisation is performed using the retrieved step with the transverse step reversed as the initial start point albeit . If the second minimisation converges to a lower minimum, the new sign flipped step estimate is taken, otherwise the initial estimate is retained. Sign errors therefore only remain after this step if the flipped transverse step truly corresponds to a lower minimum than the true step as a consequence of noise. When this occurs, the consistency check can be used. Applying the consistency check to both the estimate from the global minimum, one can pick the sign most consistent with the estimates from the overlapping sets of frames. With this, the sign error can only persist if every set of three frames give the sign flipped result as the global minimum, significantly less likely than for just a single set of three frames, in which case the trajectory would be the exact opposite (at least in the transverse direction) of the true trajectory. It should be emphasised, however, that in the majority of cases, the decay factor ensures that there is no sign error and the true step corresponds to the global minimum. Figure 4.2 shows a schematic outline of the full algorithm including the consistency and sign ambiguity checks.

4.3 Algorithm Performance

The tracking algorithm performance was studied using simulated data. In particular, the value of the background field at each pixel was constructed by first generating a complex zero mean Gaussian random number for each azimuthal pixel ϕ_i . This vector of N_ϕ Gaussian random numbers was then convolved with a Gaussian smoothing function to yield a ring of speckle with finite speckle size with a desired average angular width, where the convolution was cyclic to preserve the periodic ring (i.e. that $\tilde{I}(\phi) = \tilde{I}(\phi + 2\pi)$). The chosen speckle size was found to have no significant effect on the algorithm, but was chosen to be 0.05 radians, corresponding

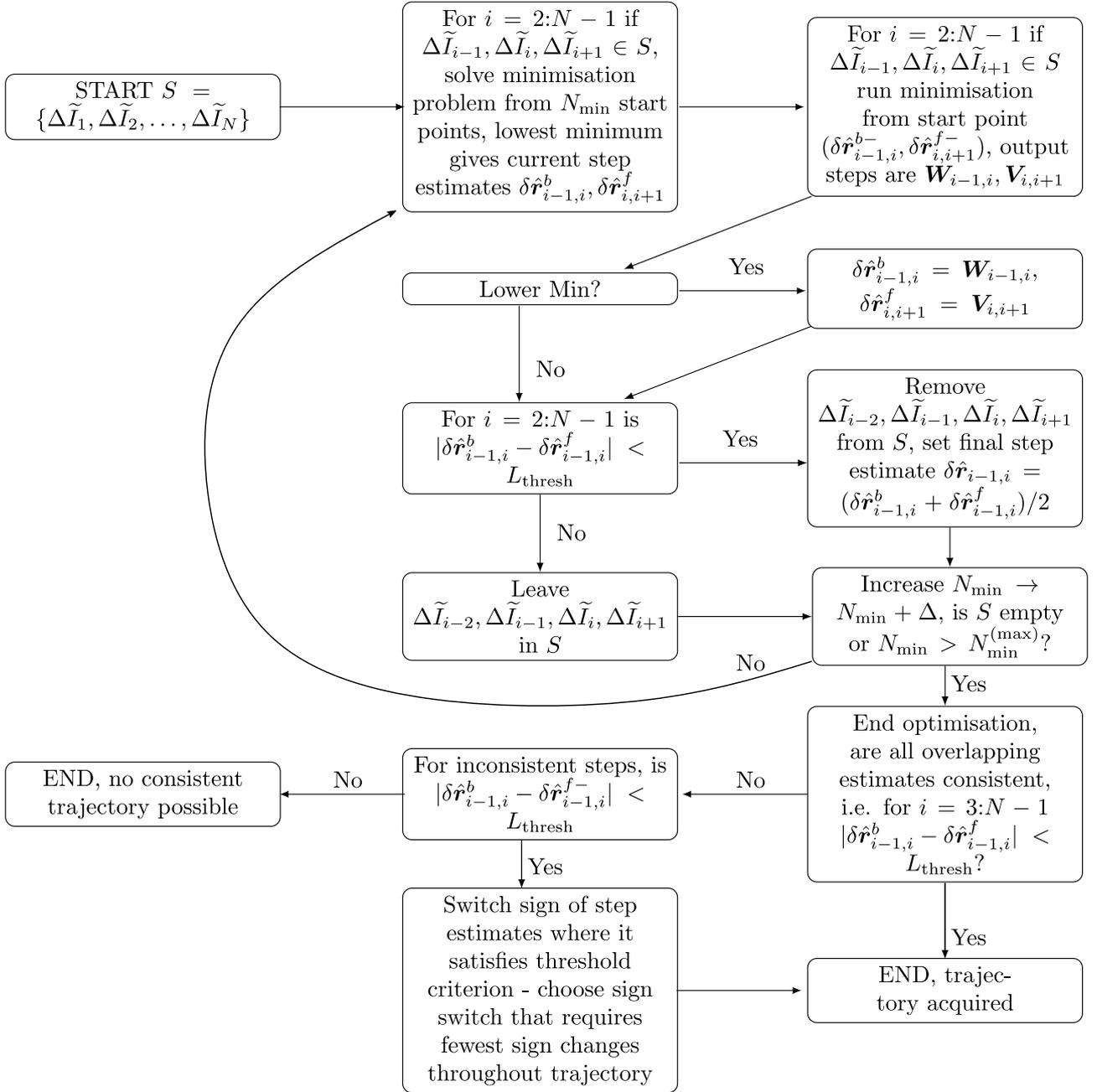


Figure 4.2: A flow chart of the tracking algorithm including the consistency check and sign flipping steps. The sign switch the transverse component is denoted with a superscript $-$ i.e. if $\delta\mathbf{r}_{ij}^{b,f} = (\delta x_{ij}, \delta y_{ij}, \delta z_{ij})^T$ then $\delta\mathbf{r}_{ij}^{b,f-} = (-\delta x_{ij}, -\delta y_{ij}, \delta z_{ij})^T$. The parameters of the algorithm are L_{thresh} (the distance below which separate estimates of the same step are considered consistent), N_{min} (the initial number of minimisation start points), Δ (the amount to increase the number of start points by each round) and $N_{\text{min}}^{(\text{max})}$, the maximum number of start points to try before terminating.

to observed leakage ring speckle patterns [278]. The analyte particle was modelled as a dipole scatterer with dipole moment aligned with the incident SPP field, i.e. the particle scattered field is calculated using Eq. (3.70). The derivation of the tracking algorithm does not, however, require this approximation to be valid, it is merely used to reduce the simulation time. The particle was assumed to undergo a 3D random walk, except it was not allowed to cross the $z = 0$ interface. If a randomly generated step did take the particle below the interface to a z coordinate $-z_a$, it was instead reflected off the interface to $z_a > 0$. The medium in which the particle was diffusing was taken as water (refractive index $n_d = 1.33$ [279]), while the particle trajectory started 60 nm above a 40 nm thick gold film (refractive index $n_m = 0.28 + 2.93i$ [280]) on top of a glass substrate (refractive index $n_g = 1.5$). The tracking algorithm does not require the particle to be undergoing a random walk. Indeed, in most cases, the analyte particle will be diffusing under the influence of a potential due to optical forces arising from the SPP field [281], convection and thermophoresis caused by heating effects [282], in addition to any interaction or surface potential that the analyte particle may experience from bound receptors or surface charges [283]. The choice of a (reflected) random walk was to ensure an unbiased sampling of propagation directions so that the algorithm performance could be more uniformly tested. The simulation free space wavelength was $\lambda_0 = 600$ nm, while the random walk was generated with a fixed step size of $0.15\lambda_0$. With these physical parameters, the short range SPP wavenumber and decay constant are $k_{\text{SPP}} = (1.50 + 0.04i)k_0$ and $\kappa_d = (0.68 + 0.08i)k_0$ respectively. The minimisation procedure used a multi-start trust-region-reflective method (using inbuilt Matlab functions, specifically the ‘lsqnonlin’ function), with 100 random starting points and the search space was bound to step sizes below $\lambda_0/2$. Larger search spaces require significantly more start points to find the global minimum at the same success rate as a smaller search space since the probability of starting at a point that converges to the true global minimum reduces with search space hypervolume. Minimisation was performed on all sets of consecutive frames $i - 1, i, i + 1$, with the result of the minimisation for each set of three frames providing estimates $\delta\hat{\mathbf{r}}_{ij}$ of $\delta\mathbf{r}_{ij}$ for $j = i - 1$ and $j = i + 1$. The consistency check threshold was set so that a step for which the two step estimates differ by more than 10% was deemed inconsistent. When steps were found to be inconsistent, the minimisation procedure was repeated on the two relevant inconsistent

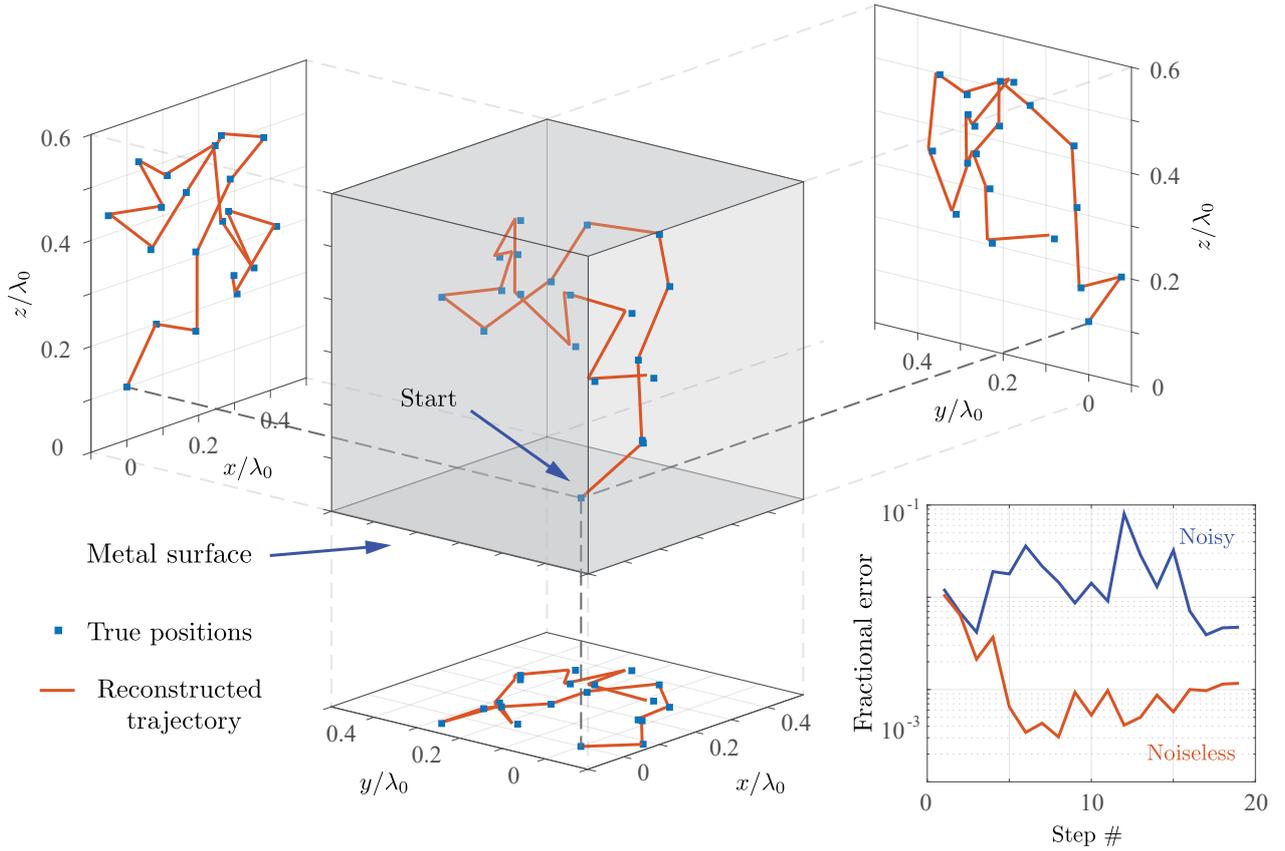


Figure 4.3: An example of a simulated trajectory reconstructed using the algorithm. Both noisy and noiseless simulation results are shown. *Figure reprinted with permission from Ref. [273] © 2021 IEEE.*

sets of three frames, for 200 new random starting points, with new step estimates replacing the previous estimate if a lower minimum was found. The consistency check was repeated with the new estimates, and global minimisation repeated if the consistency check failed, adding 100 random starting points after each failure, up to a maximum of five times before the algorithm was terminated.

For noiseless simulations, the algorithm was able to accurately reconstruct the particle trajectory, for sub-wavelength step sizes up to $\lambda_0/2$, with fractional errors (defined as $|\delta\hat{\mathbf{r}}_{ij} - \delta\mathbf{r}_{ij}|/|\delta\mathbf{r}_{ij}|$) below 1%, corresponding to sub nanometre precision (see Fig. 4.3 for an example trajectory reconstruction). In addition to a randomly generated background speckle, the method was also tested on simulated speckle patterns generated by coupled dipole simulations (see Chapter 5 for a detailed description of the coupled dipole simulation method) in which the background surface scatterers consisted of 50 and 100 gold spheres of radius 60nm randomly distributed in a $4L_{\text{SPP}} \times 4L_{\text{SPP}}$ area on the metal surface, while the analyte particle was

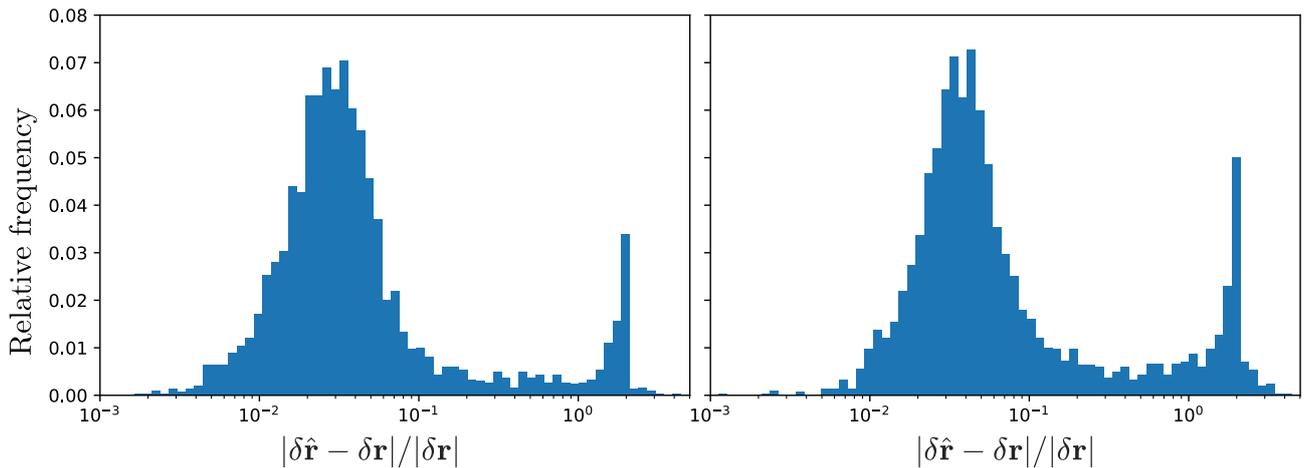


Figure 4.4: Distribution of errors for noiseless coupled dipole simulations with (left) 50 and (right) 100 background dipole scatterers. The 100 background scatterers has more significant multiple scattering effects.

modelled as a 25nm radius sphere of refractive index 1.4. The relative frequency distribution, in logarithmically scaled bin widths, of fractional error in the step estimate for three step trajectories in coupled dipole simulations shown in Fig. 4.4. These simulations include multiple scattering effects, and use numerical integration of Eq. (3.29) to find the Green’s tensor in the near field (required within coupled dipole simulations). While the majority of fractional errors are between 10^{-2} and 10^{-1} (corresponding to errors on the order of 1nm to 10nm), there is a second prominent peak for fractional errors close to 2. This arises due to the sign ambiguity, since the exact sign flipped step results in a fractional error of exactly 2. In a three step trajectory, the sign flipping part of the algorithm still has an ambiguity in that, if switching the sign of one of the sets of estimates (i.e. either from frames 1,2,3 or from frames 2,3,4) improves the consistency of the two estimates of $\delta\mathbf{r}_{23}$, there is no way of determining which set of estimates should be flipped. For longer trajectories, however, the choice can be made based on the consistency with subsequent sets of estimates. For example, if one has a four step trajectory where the consistency of the overlapping estimates of $\delta\mathbf{r}_{23}$ and $\delta\mathbf{r}_{34}$ is improved by switching the sign of the estimates from frames 2,3,4, this is the set of estimates one should sign flip, rather than switching the sign on the estimates from both frames 1,2,3 and frames 3,4,5, which would also equally improve the consistency. This is based on the fact it is less likely for two of the minimisation procedures to give the incorrect sign compared to one minimisation problem giving the incorrect sign. Indeed, from Fig. 4.4, the frequency of sign errors is $\sim 5\%$,

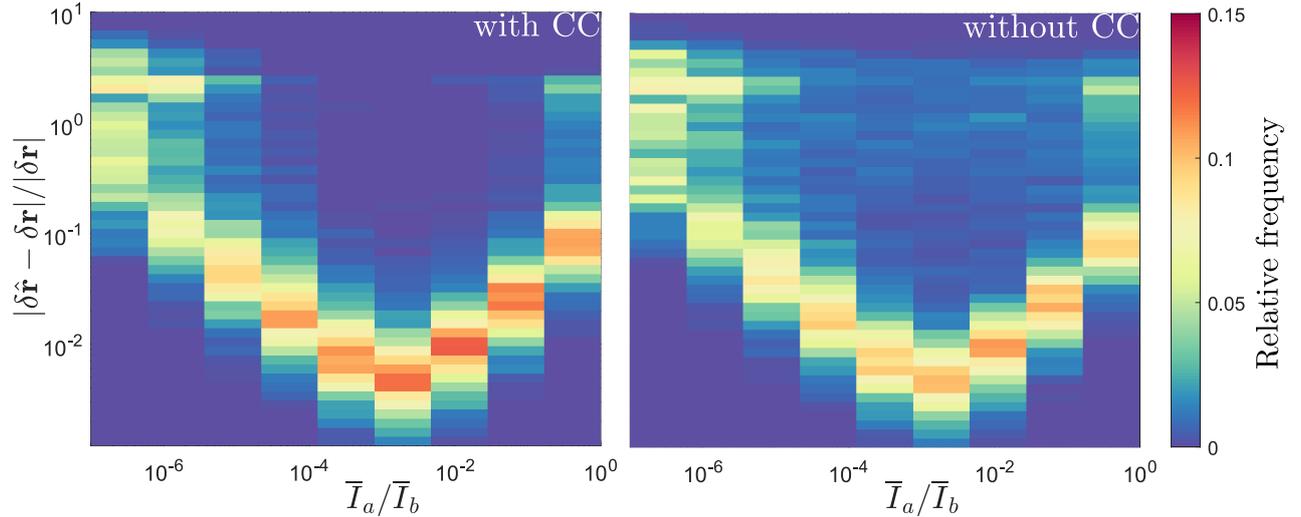


Figure 4.5: The dependence of the relative frequency of a given fractional error ($|\delta\hat{\mathbf{r}} - \delta\mathbf{r}|/|\delta\mathbf{r}|$) on the ratio of total particle scattered intensity to total background intensity, \bar{I}_a/\bar{I}_b , found without (right) and with (left) the consistency check. *Figure reprinted (with alteration, including additional data) with permission from Ref. [273] © 2021 IEEE.*

so that two separate minimisation problems from different sets of three frames giving sign errors would have a probability of ~ 0.0025 . With 100 background surface scatterers, the stronger multiple scattering reduces the algorithm performance, both in terms of the typical size of error and in the frequency of the sign error, as seen by the larger relative frequency of fractional errors close to 2 in Fig. 4.4(b) compared to (a).

It is important to consider the effect of noise on the algorithm performance. Simulations (based on Gaussian random background speckle as opposed to coupled dipole simulations) were performed assuming shot noise limited intensity measurements. Accordingly, the simulated intensity measured in each azimuthal pixel was corrupted with noise. The noisy intensity was derived from a Poisson distributed random variable with mean and variance corresponding to the number of photons $\tilde{N}_k = \tilde{I}_i(\phi_k)/I_{1\gamma}$ incident on that pixel. Fig. 4.5 shows the relative frequency of a given fractional error in the estimated step as a function of the ratio of the average total intensity scattered by the particle in the first frame and the average total background intensity, i.e. \bar{I}_a/\bar{I}_b , where $\bar{I}_a = \sum_{k=1}^{N_\phi} \langle \tilde{I}_a(\phi_k) \rangle / N_\phi$ and $\bar{I}_b = \sum_{k=1}^{N_\phi} \langle \tilde{I}_b(\phi_k) \rangle / N_\phi$ are the spatial averages around the ring of the analyte and background scattered intensities. The tildes denoting Fourier plane quantities have been dropped from the spatial averages $\bar{I}_{a,b}$ for notational convenience. Realisation averages, denoted with angled brackets, here are taken over an

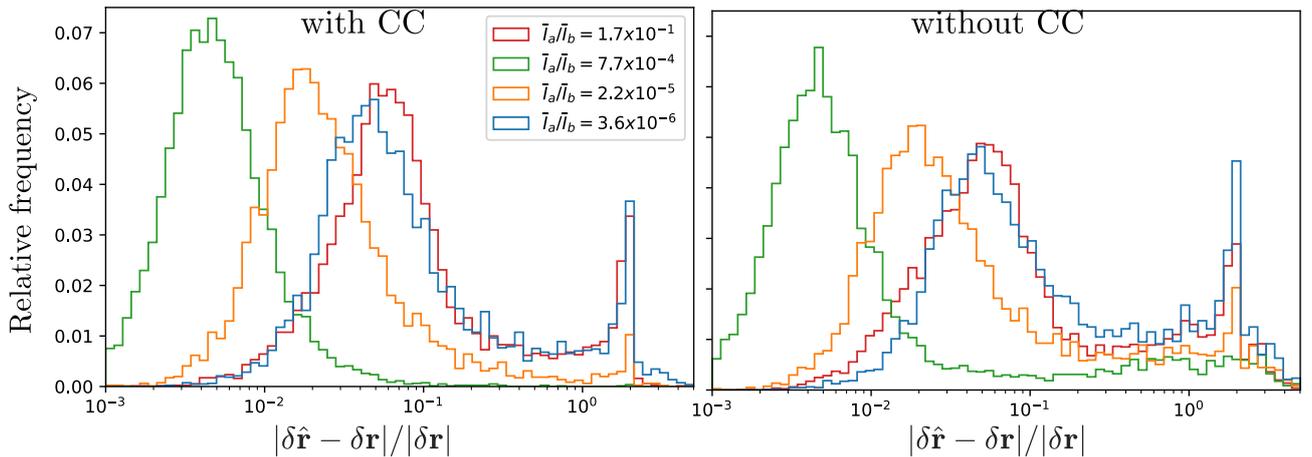


Figure 4.6: Relative frequency of errors for the algorithm (left) with the consistency check and (right) without the consistency check.

ensemble of 2700 different noise realisations, speckle realisations and particle trajectories. For the simulations the total number of photons in the background speckle integrated around the ring ($\tilde{N}_b = \sum_{k=1}^{N_\phi} \langle \tilde{I}_b(\phi_k) \rangle / I_{1\gamma}$) was assumed to be fixed at 10^{10} , corresponding to a power of $3.3 \times 10^{-7} \text{W}$ for a 10ms integration time. There is an optimal regime (here corresponding to an intensity ratio of 7.7×10^{-4}) where the errors are comparable to the noiseless results. As the particle scattered intensity decreases below this optimal regime, the change in intensity as the particle moves is obscured by the shot noise, increasing the error in the step estimate. On the other hand, as the particle scattered intensity becomes comparable to the background intensity the errors increase due to the direct scattering term ignored in Eq. (4.12) becoming significant and thus the approximation used in deriving the algorithm no longer holds. The consistency check part of the algorithm can be seen to improve performance. When the check is not used the algorithm can occasionally converge to a local minima corresponding to an incorrect step. The reconstructed trajectory for shot noise limited measurements (at an intensity ratio $\sum_{k=1}^N \bar{I}_a / \bar{I}_b = 10^{-3}$) is also shown in Fig. 4.3. Whilst reconstruction errors larger than the noiseless case can be seen, the errors remain $\lesssim 1 \text{ nm}$. The relative frequency histograms at selected scattered intensity ratios (i.e. vertical cross sections of Fig. 4.5) are shown in Fig. 4.6. Note Fig. 4.6 has different size bins compared to Fig. 4.5, and thus the values of the relative frequency are different. Comparing Figs. 4.4 and 4.6, it can be seen that the effects of multiple scattering, noise (see the smallest intensity ratio cases, $\bar{I}_a / \bar{I}_b = 2.2 \times 10^{-5}$ and 3.6×10^{-6} , of Fig. 4.6) and the direct scattering \tilde{I}_a contribution (see the largest intensity ratio

case, $\bar{I}_a/\bar{I}_b = 1.7 \times 10^{-1}$, of Fig. 4.6) are broadly the same. As any of these factors cease to be negligible, the larger peak in the error distribution (corresponding to the best estimates), shifts to the right, due to the fact the global minimum is perturbed away from the point in the minimisation space corresponding to the true step $\delta\mathbf{r}_{ij}$. Furthermore, the sign errors start to appear with increasing frequency as any of the aforementioned contributions to the measured intensity increase, evidenced by the growth of the peak around a fractional error of 2. As described in relation to Fig. 4.4, this sign error is increasingly unlikely with longer trajectories, though the larger the frequency in the three step trajectory case shown, the more likely this error is to persist in longer trajectories. The $\bar{I}_a/\bar{I}_b = 7.7 \times 10^{-4}$ case demonstrates that when intensity noise, the multiply scattered intensity and the direct scattered intensity are sufficiently small in comparison to the single scattering interference term, the sign error occurs with negligible probability and errors of $\sim 1\%$ or below are overwhelmingly likely. From both Figs. 4.5 and 4.6, it is clear that the consistency check part of the algorithm improves average performance, increasing the frequency of low error results. In the absence of the consistency check, even the optimal $\bar{I}_a/\bar{I}_b = 7.7 \times 10^{-4}$ case still has a non negligible probability of large errors, arising due to the minimum found being a local, as opposed to global, minimum. These errors arise purely from the minimisation procedure chosen. For the multi-start algorithm chosen, one would expect similar improvements from simply running the minimisation from more start points initially without any consistency check. The advantage with the consistency check approach is that it reduces the computation time by preventing running unnecessary minimisation runs after the true minimum has been found. Similar procedures could be implemented for alternative global minimisation algorithms to inform when to terminate the minimisation process to save computation time. The consistency check cannot, however, reduce the error of the algorithm if the global minimum has been found, save for identifying if sign reversed estimate improves consistency. In addition to helping reduce the occurrence of local minima related errors, the consistency check flags up when no consistent estimates of a step are possible, thus informing the user that one or both of the estimates of this step are likely to be unreliable. The dependence of Ψ and Λ on the transverse component $\delta\boldsymbol{\rho}_{ij} = (\delta x_{ij}, \delta y_{ij})^T$ of the shift is very different from their δz_{ij} dependence. As such, the algorithm may have significant differences in the errors for

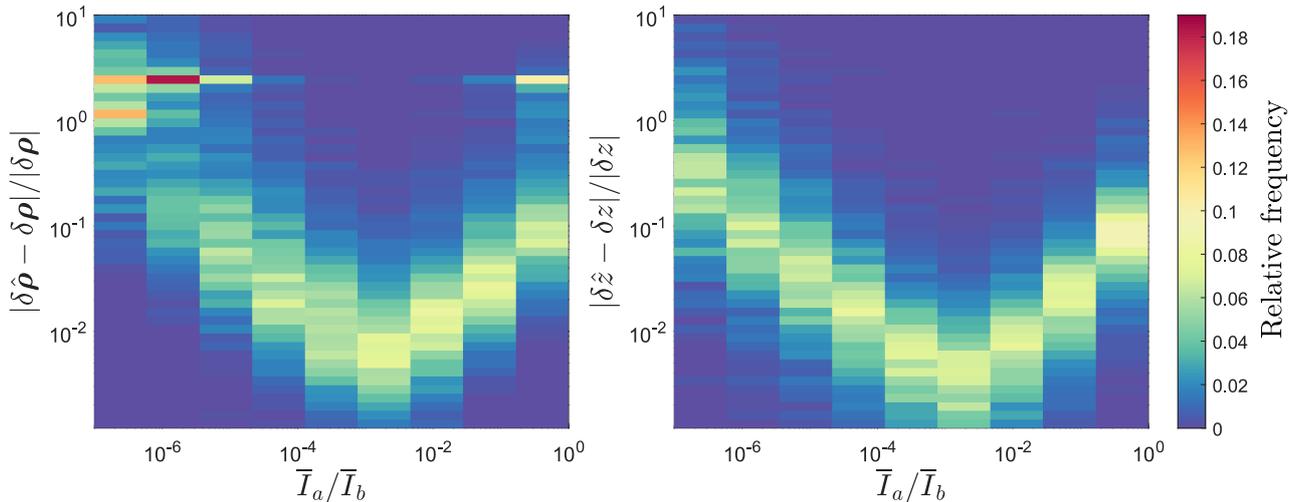


Figure 4.7: The dependence of the relative frequency of the fractional error distribution in (left) the transverse step, $|\delta\hat{\rho} - \delta\rho|/|\delta\rho|$ and (right) the z step $|\delta\hat{z} - \delta z|/|\delta z|$ on \bar{I}_a/\bar{I}_b .

each component, and Fig. 4.7 shows the distributions of the fractional errors in the estimates of both the transverse and z step. The dependence on \bar{I}_a/\bar{I}_b matches that seen in Fig. 4.5, as would be expected. The significant difference between the transverse and z errors is the absence of the peak in frequency at a fractional error of 2 associated with the sign error. As was discussed in Section 4.2.2, this can be attributed to the fact that the sign error arises from the fact that Ψ is even, so is insensitive to the sign, while $\Lambda(-\mathbf{r}) \neq \Lambda(\mathbf{r})$ in general. Since Ψ only depends weakly on z while Λ depends on z strongly (since $\kappa_d''/\kappa_d' \approx 0.1$), the algorithm is much more sensitive to the sign of z .

Since the measured light corresponds to the interference between the background and analyte scattered signal, parameters such as the SNR and SBR follow the analysis in Chapter 2 (in particular the results presented in table 2.2).

4.4 Limitations and Extensions

In addition to the effects of noise, a few other factors must be considered in assessing the tracking algorithm. Firstly, as the Born approximation is used, the method only applies in the single scattering regime. The coupled dipole simulations include multiple scattering effects, but these are sufficiently weak in the results of Fig. 4.4 that the method still works fairly well. To understand the effects of multiple scattering on the algorithm, it is useful to categorise different

multiple scattering paths based on the role the analyte particle plays in them. There are paths which do not involve any scattering from the analyte particle, and the light is multiply scattered from the surface only. Another set of paths only involve a single scattering event from the analyte particle, with scattering from the surface only before and after this scattering event, and the final set of scattering paths are those that are scattered multiply from the analyte particle, revisiting (via surface scattering) the analyte particle after being scattered from it, possibly multiple times. The general effects of different scattering paths in the context of sensing/tracking are discussed in greater detail in Chapter 5. Here, their effects in the context of this single scattering algorithm are briefly considered. Multiple scattering from the surface only (i.e. scattering paths not involving the analyte particle) has no effect on the method, since this only affects the background field $\tilde{\mathbf{E}}_b$, and the algorithm does not rely on any particular property of the background speckle. Additionally, for small particles, the third class of paths involving multiple scattering events from the particle has negligible effect on the scattered field. The second class, involving multiple scattering from the surface before and after a single scattering event from the analyte particle, does have an effect on the validity of the algorithm. In particular, the field experienced by the particle includes the light scattered from the surface, so that Eq. (4.1), rather than just containing the incident field, takes the form

$$\mathbf{E}_a(\mathbf{r}; \mathbf{r}_a) = \frac{\omega^2}{c^2} \int_{\mathbf{r}' \in \mathcal{V}_a} [\varepsilon_a(\mathbf{r}') - \varepsilon_d] G(\mathbf{r}, \mathbf{r}') (\mathbf{E}_0(\mathbf{r}') + \mathbf{E}_b(\mathbf{r}')) d^3 \mathbf{r}', \quad (4.25)$$

where $\mathbf{E}_b(\mathbf{r})$ is the speckle field scattered from the rough surface, but here evaluated in the near field in the position of the analyte particle, rather than in the leakage radiation ring in the far field. The phase and amplitude variation of this near field speckle contribution is fixed but unknown for a given realisation of the surface scattering potential. As such, if $|\mathbf{E}_b|$ is comparable to or larger than $|\mathbf{E}_0|$ at the particle's position, the algorithm breaks down. In addition, scattering from the surface after the scattering from the analyte particle means $G \neq G_0$, i.e. the Green's tensor is not that of a flat interface multilayer structure, but instead includes the rough surface.

As well as multiple scattering, the effect of the rotation of the particle, which was ignored in

the algorithm derivation, must be considered. In particular, it was assumed that \mathcal{V}'_a was related to \mathcal{V}_a by a translation by $\delta\mathbf{r}$ only. For a non-spherically symmetric particle, if the particle rotates between two frames, the volumes are also oriented differently. Provided the optical response of the analyte particle is close to isotropic (i.e. $\alpha_a \propto \mathbb{I}$), the effects of rotation on the scattered field are negligible. If the particle has a significantly different response along a particular axis, as for example may be the case for a long thin rod shaped analyte particle, then the particle orientation can strongly affect the scattered field. In theory, provided one has a model for the particle shape, the effect can be accounted for within the algorithm. By calculating, whether numerically or analytically, the integral $\int G_0(\mathbf{r}, \mathbf{r}')\mathbf{E}_0(\mathbf{r}')$ over all the orientations of the particle, one can calculate the phase and amplitude changes for a given rotation. As a result, the phase and amplitude shift functions are now also dependent on particle orientation $\boldsymbol{\xi}$ (parameterised by for example the orientation of the principal axes of the analyte particle), $\Lambda(\delta\mathbf{r}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$ and $\Psi(\delta\mathbf{r}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$, where $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are 3D vectors consisting of the orientation of the three analyte particle principal axes before and after the shift respectively. From here, the algorithm can proceed in the same way as outlined above in solving Eq. (4.24), except with the new phase and amplitude functions. Indeed, the tracking algorithm can be applied to any form of the phase and amplitude function. Such an approach, however, adds several complications. Firstly, the calculation of Λ and Ψ is more involved and may not be possible analytically, depending on the particle shape, which may slow down the algorithm. Perhaps more significantly, the extra parameters that have to be minimised over increase the dimensionality of the minimisation problem. In this case, the unknowns are the two 3D shift vectors between the frames, and the three 3D orientation vectors at each frame, giving a 15 dimensional minimisation problem (as opposed to 6 dimensional in the no rotation case). The significant increase in dimensionality makes the minimisation search challenging, and possibly unfeasible. Furthermore, such a method requires an explicit model for the analyte particle shape, and is only valid as far as this model can be applied. The effect of a non plane wave SPP illumination can similarly be incorporated into the model. Again, provided one has a model for $\mathbf{E}_0(\mathbf{r})$ which allows for the calculation of the integral $\int G_0(\mathbf{r}, \mathbf{r}')\mathbf{E}_0(\mathbf{r}')$, the resulting phase and amplitude functions $\Lambda(\mathbf{r}_1, \mathbf{r}_2, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$ and $\Psi(\mathbf{r}_1, \mathbf{r}_2, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$ can be fed into the algorithm. Due to the non-uniform illumination, the

phase and amplitude shifts are no longer a function of the difference in positions and therefore there are three additional unknown parameters (i.e. \mathbf{r}_1 and \mathbf{r}_2 as opposed to $\delta\mathbf{r}_{12}$) one must minimise over. In order to implement the higher dimensional versions of the algorithm, the minimisation procedure would most likely have to be altered from the brute force multiple start point approach, but provided the spherically symmetric particle approximation holds and the amplitude envelope for a non-uniform illumination varies over a much longer length scale than the distance the analyte particle moves between frames, the simplest version of the algorithm presented in this chapter will remain accurate. Note that the tracking algorithm does not require a specific minimisation algorithm in solving Eq. (4.24), and could be implemented using alternative algorithms to the multiple start point one used here, such as simulated annealing [284].

It is worth also mentioning that the phase and amplitude functions can be relatively easily modified to include different observation positions and different plane wave illuminations to the SPP leakage ring based approach investigated in this thesis. These alternative forms follow from Eq. (4.5) by choosing the \mathbf{k}_{\parallel} and k_z for the appropriate observation positions and replacing the $\exp(ik_{\text{SPP}}\delta x - \kappa_d\delta z)$ factor with $\exp(i\mathbf{k}_{\text{inc}} \cdot \delta\mathbf{r})$ for a given incident wavevector \mathbf{k}_{inc} .

Another limitation of the algorithm is that it relies on the assumption that only a single analyte particle is moving in the sensing volume. When there are multiple moving scatterers, the method breaks down even in the single scattering regime, since the intensity changes are the sum of many interference terms from each moving scatterer. As such, the tracking algorithm is only applicable at low concentrations where there is unlikely to be more than one analyte particle in the sensing volume at any one time. Therefore, the algorithm may be appropriate for tracking single molecular machines such as myosin on a substrate [285, 286].

Finally, the algorithm requires the analyte particle to be sufficiently near the surface that the SPP field is significant. If it diffuses more than a few decay lengths L_d away from the surface, the perturbation to the speckle pattern becomes too weak to detect. As such, the method is best suited to monitoring processes occurring at the surface, such as the interaction of an analyte protein with receptors on the surface or a molecular machine bound to the surface. The in-plane decay on the length scale L_{SPP} introduces a similar limitation on the region of the surface over

which the particle can be tracked, though the fact that generally $L_{\text{SPP}} \gg L_d$ means this is a much looser restriction. While the restricted volume within which the particle can be tracked is a limitation of the method, it does also increase the probability that only a single analyte particle is diffusing within the sensing volume at any one time.

4.5 Conclusion

In this section, an algorithm has been developed capable of extracting a particle's trajectory as it moves near the surface from the speckle pattern formed by the random scattering of SPPs. Simulations show it to be capable of nanometre level precision, depending on the noise parameters. In addition to noise, other factors which can affect the performance of the algorithm are considered such as rotation or non uniform illumination, and possible extensions to mitigate the effect of such factors on the tracking algorithm are considered. The generalisation of the algorithm to a wider range of experimental setups is also discussed.

The tracking algorithm relies on the predictable changes to the interference of the background speckle and the analyte scattered field that occurs in the single scattering regime, and thus only works when multiple scattering effects are negligible. It is important to consider the role multiple scattering has on the perturbation to the speckle pattern caused by the analyte particle, both in the context of the proposed tracking algorithm and more generally in the entire sensing system.

Chapter 5

Role of Multiple Scattering in Sensitivity

Some of the work presented in this chapter has been published in Refs. [287] and [288].

Until now, the system introduced in Chapter 3 has been studied in the single scattering regime, in which the effects of multiple scattering are considered negligible. In this chapter, the role of multiple scattering in the sensing setup is considered, and in particular the effect it has on the sensitivity of the system

5.1 Multiple Scattering Theory and Phenomena

Multiple scattering increases the analytic complexity required in modelling light scattering, but it is often essential as a diverse range of phenomena are observed within multiple scattering environments. While all orders of scattering (i.e. single, double etc.) occur in any given scenario, the relative importance of different terms varies with the parameters of the system. A parameter to measure the strength of multiple scattering is the scattering mean free path l_s , the average distance travelled between scattering events. Depending on how this compares to other length scales of the system, one can see different properties of scattered light. For example, when $l_s > L$, where L is the length scale determined by the size of the scattering region, single scattering is dominant. While multiple scattering effects are important whenever $l_s < L$, there are significant

differences between $l_s > \lambda$ and $l_s < \lambda$, where λ is the wavelength of the scattered waves. Thus, it is important to consider different regimes within multiple scattering, and this section aims to review key features and phenomena that occur within these different multiple regimes.

5.1.1 Diagrammatic Representations

As discussed in Chapter 3, a multiple scattering problem can be expressed as a Lippmann-Schwinger or Dyson integral equation (Eqs. (3.16) and (3.19) for the electric field and Green's tensor respectively), with similarly structured equations existing for other quantities also [24, 26, 27, 221, 289]. The iterative nature of these equations allows them to be expressed as infinite series. It is often convenient to represent these equations and series diagrammatically, analogous to Feynman diagrams commonly used in quantum field theory. A typical diagrammatic representation of the Dyson series is shown in Fig. 5.1. While different conventions exist within different contexts, generally lines represent Green's functions, either unperturbed (G_0) or perturbed (G), and nodes represent a factor of the scattering potential $V(\mathbf{r})$, which for Maxwell's wave equation is $V(\mathbf{r}) = \varepsilon(\mathbf{r})\omega^2/c^2$. Each node has a position associated with it, as do the start and end points, which provide the arguments for the Green's tensor connecting the nodes. Positions associated with intermediate nodes are integrated over (but not the start and end points).

$$\begin{array}{c} \mathbf{r}' \quad \mathbf{r} \\ \cdots \rightarrow \cdots \end{array} + \begin{array}{c} \mathbf{r}' \quad \mathbf{r} \\ \cdots \rightarrow \bullet \rightarrow \cdots \end{array} + \begin{array}{c} \mathbf{r}' \quad \mathbf{r} \\ \cdots \rightarrow \bullet \rightarrow \bullet \rightarrow \cdots \end{array} + \dots = \begin{array}{c} \mathbf{r}' \quad \mathbf{r} \\ \longrightarrow \longrightarrow \end{array}$$

Figure 5.1: Diagrammatic representation of the Dyson equation. Dotted lines represent unperturbed Green's tensors G_0 while solid lines represent total Green's tensors G including the effect of scatterers. Nodes (\bullet) represent scattering events, introducing a factor of the scattering potential $V(\mathbf{r})$. Intermediate positions are integrated over.

In addition to providing a shorthand notation for the integrals and allowing an intuitive physical interpretation to the equations in terms of scattering trajectories, certain diagrams can be grouped together based on similar structure and approximate solutions based on combining certain diagrams can be found [24, 26, 27]. The exact rules for averaging diagrams (i.e. the terms in a Born or Dyson series) depend on the statistics of the scattering potential $V(\mathbf{r}) = \varepsilon_s(\mathbf{r})\frac{\omega^2}{c^2}$. A commonly used model is to treat $V(\mathbf{r})$ as a zero mean Gaussian process [24, 254]. This allows

a solution to the Dyson equation for the perturbed Green's function, averaged over random scattering realisations, in terms of a quantity known as the self energy Σ , which is the sum from a subset of diagrams with a connected structure. For a scalar wave satisfying the scalar Helmholtz equation in a homogeneous background, the solution can be conveniently expressed in Fourier space as [24]

$$\langle \tilde{G}(\mathbf{q}) \rangle = \frac{-1}{k_0^2 - q^2 + \Sigma(\mathbf{q})}. \quad (5.1)$$

The form of Eq. (5.1) indicates one common feature of multiple scattering, namely that it shifts the dispersion relation. From comparison to the unperturbed Green's function $\tilde{G}_0 = -1/(k_0^2 - q^2)$, for which the dispersion relation for a mode of wavevector \mathbf{q} is $q = k_0 = \omega/c$, the dispersion relation with multiple scattering present becomes $q^2 = k_0^2 + \Sigma(\mathbf{q})$. This dispersion relation shift due to multiple scattering has been studied in the context of SPP scattering specifically [245, 290, 291, 292, 293]. It is worth noting that, in this scalar wave model, the elastic scattering mean free path can be related to the self-energy via $l_s^{-1} = -\text{Im}[\Sigma(\mathbf{q})]/k_0$ [24]. A form of l_s applicable to a collection of scatterers with scattering cross-section σ_s and number density n , valid at small n , is $l_s = (n\sigma_s)^{-1}$ [24]. This form matches intuitively with what one would expect the mean distance travelled between collisions to be for a particle moving through a collection of objects of physical cross section σ_s . A related quantity is the transport mean free path l_t , which defines the length over which the direction of propagation of a photon is randomised [294]. For isotropic scattering, this is identical to l_s , however they differ for anisotropic scattering due to the fact a single scattering event does not completely randomise the direction in which a photon is scattered and it retains some memory of the initial propagation direction.

5.1.2 Light Diffusion

Analytical solutions to the Dyson or Lippmann-Schwinger type equations of multiple scattering require approximations to which sets of diagrams can be neglected and which provide significant contributions, and as such are limited to restricted range of parameters, while numerical solutions can be computationally demanding, especially when requiring Monte Carlo simulations to study

statistical properties over many realisations. An alternative approach often taken is to use transport theory [294, 295], in which one considers a differential equation for the specific intensity (intensity per unit solid angle in a given direction) known as the radiative transport equation (RTE) [27], equivalent to the Boltzmann transport equation applied to light transport [295]. When the system is much larger than the transport mean free path l_t (e.g. the thickness of a slab of random media $z \gg l_t$, when considering the transmission of intensity through a wide slab) such that light undergoes many scattering events before leaving the scattering media, absorption is weak in comparison to scattering ($l_t > L_{\text{abs}}$ where L_{abs} is the absorption length in the scattering media) and scattering is close to isotropic, the RTE simplifies to the much simpler diffusion equation, and thus the well developed theory of the diffusion equation can be used to model intensity transport [296]. The transport theory/diffusion approach is valid when the many scattering events randomise the propagation of the scattered light, and the random walk of many photons through multiple scattering events gives rise to diffusion in the same way Brownian motion of particles results in diffusion. While these equations can be understood heuristically in terms of random walks, it is possible to derive the RTE and a diffusion equation from the Lippmann-Schwinger equation under the approximation that $l_s \gg \lambda_0$ so that the effects due to interference between different diagrams average out [27]. Within such an approach, it is possible to define a diffusion coefficient D for the transport of light [297].

5.1.3 Multiple Scattering Phenomena

One of the key differences introduced by multiple scattering compared to single scattering is that the field incident on a given scatterer is the sum of the fields scattered from all other scatterers, plus the illumination field. In a random scattering configuration, this means that the field incident on a particle can be a (near field) speckle pattern arising from the interference of many randomly scattered waves. As such, even when the incident illumination is homogeneous, the field incident on a scatterer can be highly dependent on its position. Certain regions, where scattered light adds in phase give rise to high intensity regions known as hotspots [298]. Scatterers in such hotspots scatter more light due to the higher intensity incident upon them. In the context of sensors, this hotspot mechanism enables greater sensitivity when the analyte particle is in a hotspot [31].

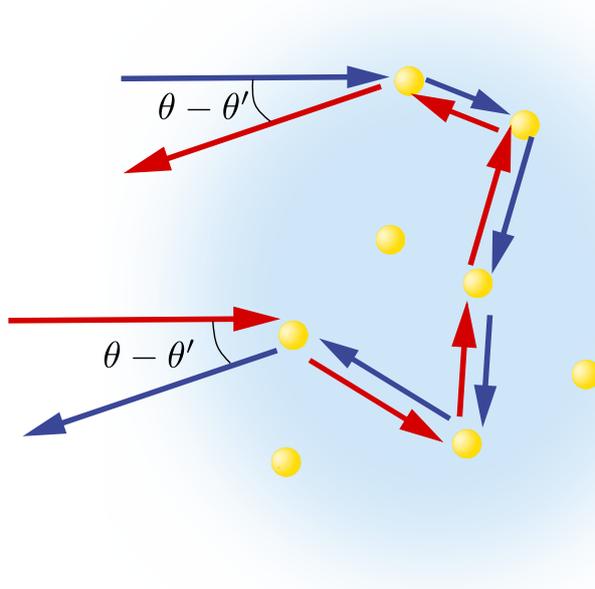


Figure 5.2: Examples of reciprocal scattering paths contributing to coherent backscattering. In the backscattering direction $\theta = \theta'$, the scattering paths have identical phase and amplitude and add up in phase.

In addition to elastic scattering, the hotspot effect is used to enhance fluorescence and Raman scattering [298, 299, 300]. While random scattering environments such as rough surfaces give rise to hotspots [298]), specifically engineered structures such as metallic nanoparticles near metal interfaces are also used to generate hotspots [301, 302]. Another phenomenon arising from the inhomogeneous nature of the scattered light, is fluctuations in the local density of states (LDOS). The number of available modes a given scatterer can scatter light into, described by the LDOS, depends on the arrangement of nearby scatterers, and as such can vary significantly. The LDOS fluctuations introduced by multiple scattering have significant consequences for fluorescence, with emitters at a position with high LDOS having reduced lifetime [303, 304, 305]. Furthermore, LDOS fluctuations introduce intensity correlations to speckle patterns known as C_0 correlations [306, 307, 308]. As well as C_0 correlations, multiple scattering introduces further intensity correlations into speckle patterns. In Chapter 3, the intensity correlation function decays over the length scale given by the speckle size, and this correlation within a speckle (C_1 correlation) is present within the single scattering regime. Significant multiple scattering introduces additional, long range correlations (C_2 and C_3 correlations) between individual speckle spots [24, 28, 309, 310, 311]. Coherent backscattering, weak localisation and Anderson localisation (also called strong localisation) are related phenomena that arise due

to strong multiple scattering and have received much research interest [312]. They all arise due to the interference of reciprocal scattering paths, where reciprocal scattering paths are those that visit the same scatterers in opposite order. Reciprocity means the light propagating along the paths acquires the same phase and amplitude shifts [219, 313, 314]. Theoretical descriptions of localisation can become very involved, but the principle is relatively intuitive to understand, especially in the context of coherent backscattering. If one considers scattering of light incident at an angle θ into an outgoing angle $\pi + \theta'$ (relative to the same fixed axis) by two reciprocal scattering paths, as shown in Fig. 5.2, one sees the only difference in phase and amplitude of the two contributions arises from propagation to the first scatterer and from the final scatterer. In the backscattering direction $\theta' = \theta$, the two paths contribute a scattered field with identical phase and amplitude and therefore the contributions add up in phase. As such, the backscattered intensity is doubled relative to what one would expect from assuming all scattering paths add with random phase and one sees an intensity peak in this direction. Outside a narrow angular range $\sim \lambda/l_s$ around the backscattering range, the phase difference between the incoming and outgoing legs of the respective paths is sufficient to ensure the contributions add up with random phase. Note that the single scattering contribution does not have a coherent backscattering peak as there is no reciprocal scattering path to a path with a single scatterer. As such, coherent backscattering is not seen in the single scattering regime. It has, however, been experimentally observed for a range of scattering environments, including in the scattering of SPPs [315, 316, 317, 318, 319, 320].

More generally, weak and Anderson localisation arise as loop paths add in phase due to reciprocity and increase the intensity of light scattered back to a given point. Initially, in the weak localisation regime, this serves to reduce the rate of intensity diffusion (i.e. a smaller effective diffusion coefficient D). As disorder decreases, the Anderson localisation regime is achieved, $D \rightarrow 0$ and scattered light modes are confined to a region on the scale of the localisation length ξ . Anderson localisation is a wave interference phenomena, and thus is not restricted to light but can also occur in the random scattering of any waves, including electromagnetic waves such as SPPs [212, 220, 252, 312, 321, 322]. Experimentally, Anderson localisation has been observed for a range of wave phenomena including electrons, cold atoms and sound waves

[323, 324, 325, 326, 327, 328]. Anderson localisation of light has also been experimentally realised using disordered photonic structures [329, 330, 331] and disordered optical fibres [332, 333], as has Anderson localisation of SPPs [334, 335, 336]. There is evidence to suggest that, while Anderson localisation of scalar waves by a random 3D arrangement of point scatterers is possible, the near field interaction in the scattering of vector electromagnetic waves can prevent it in three dimensional light scattering by point scatterers [337, 338]. Importantly, it arises due to the specific form of the electromagnetic Green's tensor at small distances, rather than the vector nature, and similar models of elastic vector waves, which take a different near field form, do exhibit Anderson localisation [337, 339].

5.1.4 Coupled Dipole Model

The coupled dipole model is a scattering model that accounts for all orders of multiple scattering by a collection of point dipole scatterers. Thus it can describe systems where the individual scatterers are sufficiently small to be described as point scatterers, and additionally, in what is referred to as the discrete dipole approximation, as an approximation to larger scattering objects which have been discretised into small elements which can be modelled as point scatterers. Each point dipole has a dipole moment \mathbf{p}_j , allowing the total field $\mathbf{E}(\mathbf{r})$ outside the volume of the scatterers, when illuminated by an incident field \mathbf{E}_0 , to be expressed [340]

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_0(\mathbf{r}) + \frac{k_0^2}{\varepsilon_0} \sum_{j=1}^N G(\mathbf{r}, \mathbf{r}_j) \mathbf{p}_j, \quad (5.2)$$

where G is the Green's tensor for the background medium (note for the rest of this chapter, for notational convenience, G as opposed to G_0 denotes the background Green's tensor in the absence of any scatterers), \mathbf{r}_j is the position of the j th scatterer and N is the number of scatterers. The dipole moment is then given by $\mathbf{p}_j = \tilde{\alpha}_j \mathbf{E}_{\text{exc}}(\mathbf{r}_j)$, where $\tilde{\alpha}_j$ is the 'bare' polarisability of the j th scatterer (the meaning of this will become clear) and $\mathbf{E}_{\text{exc}}(\mathbf{r}_j)$ is the exciting electric field incident on the j th scatterer, distinct from the total electric field [340, 341]. In particular, \mathbf{E}_{exc} includes the field radiated by all other dipoles and the field reflected from

background inhomogeneities, and can be expressed

$$\mathbf{E}_{\text{exc}}(\mathbf{r}_j) = \mathbf{E}_0(\mathbf{r}_j) + \frac{k_0^2}{\varepsilon_0} G_{\text{ref}}(\mathbf{r}_j, \mathbf{r}_j) \mathbf{p}_j + \frac{k_0^2}{\varepsilon_0} \sum_{i \neq j} G(\mathbf{r}_j, \mathbf{r}_i) \mathbf{p}_i \quad (5.3)$$

where $G_{\text{ref}}(\mathbf{r}_j, \mathbf{r}_j)$ is the part of the Green's tensor describing reflection of interfaces or inhomogeneities. The first term is the incident field, the second term describes reflection of the light scattered by the j th scatterer back onto itself while the final term is the light scattered from all other dipoles onto the j th scatterer. Substituting Eq. (5.3) into $\mathbf{p}_j = \tilde{\alpha}_j \mathbf{E}_{\text{exc}}(\mathbf{r}_j)$ allows the formation of a self-consistent equation for the dipole moments \mathbf{p}_i as

$$\mathbf{p}_j = \tilde{\alpha}_j \mathbf{E}_0(\mathbf{r}_j) + \frac{k_0^2}{\varepsilon_0} \tilde{\alpha}_j G_{\text{ref}}(\mathbf{r}_j, \mathbf{r}_j) \mathbf{p}_j + \frac{k_0^2}{\varepsilon_0} \sum_{i \neq j} \tilde{\alpha}_j G(\mathbf{r}_j, \mathbf{r}_i) \mathbf{p}_i. \quad (5.4)$$

Defining a dressed polarisability $\alpha_j = (I - (k_0^2/\varepsilon_0) \tilde{\alpha}_j G_{\text{ref}}(\mathbf{r}_j, \mathbf{r}_j))^{-1} \tilde{\alpha}_j$, Eq. (5.4) becomes

$$\mathbf{p}_j = \alpha_j \mathbf{E}_0(\mathbf{r}_j) + \frac{k_0^2}{\varepsilon_0} \sum_{i \neq j} \alpha_j G(\mathbf{r}_j, \mathbf{r}_i) \mathbf{p}_i. \quad (5.5)$$

The difference between the bare and dressed polarisability is that the bare polarisability describes the polarisability of the scatterer in a homogeneous background, while the dressed polarisability accounts for interaction with the background dielectric environment. From now on, polarisability shall refer to the dressed polarisability unless otherwise stated. The resulting set of linear coupled dipole equations can be expressed as

$$\sum_{j=1}^N M_{ij} \mathbf{p}_j = \mathbf{p}_{0,i}, \quad (5.6)$$

for $i = 1, 2, \dots, N$, where $\mathbf{p}_{0,i} = \alpha_i \mathbf{E}_0(\mathbf{r}_i)$ is the dipole moment induced by the incident field in the i th scatterer, the matrix elements M_{ij} are defined by

$$M_{ij} = \begin{cases} \mathbb{I} & i = j \\ -\frac{k_0^2}{\varepsilon_0} \alpha_i G_{ij} & i \neq j, \end{cases} \quad (5.7)$$

for $i = 1, 2, \dots, N$ and the notation $G_{ij} = G(\mathbf{r}_i, \mathbf{r}_j)$ has been introduced. Solving Eq. (5.6) for the N dipole moments then allows $\mathbf{E}(\mathbf{r})$ to be found at any position outside the volume of the scatterers using Eq. (5.2).

5.2 Enhancement Factors

Studies have established that light scattered by a multiple scattering environment is sensitive to the addition of a single scatterer [342, 343, 344], and also the position of a single scatterer, enabling localisation [345, 346]. From a sensing or tracking perspective, it is useful to understand how these perturbations in the multiple scattering regime differ from the single scattering case, and in particular which regime is more sensitive to the addition of a single scatterer. This would allow for optimising of sensor design by ensuring the scattering environment is in the correct regime. While the hotspot mechanism is an example of a multiple scattering effect that gives rise to increased sensitivity, it is also possible for a scatterer to be in a ‘dark spot’ in which destructive interference of scattered fields gives rise to low intensity regions, resulting in reduced scattering. Thus, the comparison of the two regimes requires more in depth study to quantify the effect of these competing phenomena, along with other multiple scattering effects. In addition to determining which regime is more sensitive, it is also useful to study what magnitude of difference is achievable, and the dependence on the strength of multiple scattering (i.e. the scatterer density or scattering mean free path) or other scattering parameters (such as the individual scatterer properties).

In order to quantify the effects of multiple scattering within the sensing system, consider the coupled dipole equation for an arbitrary arrangement of N point scatterers with dressed polarisabilities α_i in positions \mathbf{r}_i for $i = 1, 2, \dots, N$ in a background with Green’s tensor G , which, when illuminated by an incident field \mathbf{E}_0 , gives rise to a total electric field $\mathbf{E}(\mathbf{r})$ given by Eq. (5.2). At this point, beyond the validity of the point scatter approximation and the absence of magnetic media, no assumptions have been made about the scatterers or their positions. Before studying the effect of adding an analyte particle to this system, it is worth pointing out a few features of the matrix M defined in Eq. (5.7). Firstly, within the single scattering

regime, the off-diagonal terms are negligible, meaning that $M_{ij} \approx I\delta_{ij}$ and $\mathbf{p}_i \approx \mathbf{p}_{0,i}$, i.e. there is negligible coupling between dipoles. Furthermore, if all the scatterers are identical, such that $\alpha_i = \alpha \forall i$, then M is a symmetric matrix, $M^T = M$. This follows from the Onsager reciprocity of the Green's tensor $G(\mathbf{r}, \mathbf{r}') = G^T(\mathbf{r}', \mathbf{r})$ [219]. Importantly, however, M is complex and not Hermitian, i.e. $M^\dagger \neq M$ in general. This distinction is emphasised as some properties of complex symmetric matrices are very different to real symmetric matrices and also Hermitian matrices (for example Hermitian and real symmetric matrices have real eigenvalues and are normal matrices while complex symmetric matrices do not have these properties in general). Finally, if, in addition to identical scatterers, the positions \mathbf{r}_i are random, M is a special class of random matrix known as a Euclidean random matrix. An N dimensional Euclidean random matrix A is defined by a deterministic function $f(\mathbf{r}, \mathbf{r}')$ and a set of N random points \mathbf{r}_i in Euclidean space, with the (i, j) element given by $A_{ij} = f(\mathbf{r}_i, \mathbf{r}_j)$. In this case, $f(\mathbf{r}, \mathbf{r}') = -(k_0^2/\varepsilon_0)\alpha G(\mathbf{r}, \mathbf{r}')$. The properties of Euclidean random matrices have been studied in a wide range of physical contexts including optical scattering [347, 348, 349, 350]. When the scatterers are not identical, M is not strictly a Euclidean random matrix as defined above, however many of the theoretical approaches applied to Euclidean random matrices can still be used.

5.2.1 Adding a Scatterer

The addition of an analyte particle can now be considered by perturbing the scattering configuration with the introduction of an additional point scatterer with polarisability α_{N+1} at position \mathbf{r}_{N+1} . The perturbed system is described by a set of coupled dipole equations $\sum_{j=1}^{N+1} M'_{ij}\mathbf{p}'_j = \mathbf{p}_{0,i}$ ($i = 1, 2, \dots, N + 1$) analogous to Eq. (5.6) but for the new arrangement of scatterers with the additional scatterer. The solution to this set of equations gives the modified dipole moments \mathbf{p}'_j after the addition of the particle. The matrix elements M'_{ij} for the perturbed system are again given by Eq. (5.7) except now the indices i and j run from 1 to $N + 1$ (hence $M'_{ij} = M_{ij}$, for $i, j \leq N$). The new set of $N + 1$ dipole moments gives rise to the perturbed field \mathbf{E}' found using Eq. (5.2) to be

$$\mathbf{E}'(\mathbf{r}) = \mathbf{E}_0(\mathbf{r}) + \frac{k_0^2}{\varepsilon_0} \sum_{j=1}^{N+1} G(\mathbf{r}, \mathbf{r}_j)\mathbf{p}'_j, \quad (5.8)$$

where the sum now has $N + 1$ elements due to the additional scatterer and the perturbed dipole moments \mathbf{p}'_j are used instead of the original ones. Combining Eqs. (5.2) and (5.8), the field perturbation $\Delta \mathbf{E} = \mathbf{E}' - \mathbf{E}$ due to the addition of the scatterer is hence

$$\Delta \mathbf{E}(\mathbf{r}) = \frac{k_0^2}{\varepsilon_0} G(\mathbf{r}, \mathbf{r}_{N+1}) \mathbf{p}_{N+1} + \frac{k_0^2}{\varepsilon_0} \sum_{j=1}^N G(\mathbf{r}, \mathbf{r}_j) \delta \mathbf{p}_j, \quad (5.9)$$

where $\delta \mathbf{p}_j = \mathbf{p}'_j - \mathbf{p}_j$ is the perturbation to the j th dipole moment and, since there is no $(N + 1)$ th scatterer in the unperturbed system, the prime has been dropped from \mathbf{p}_{N+1} . The first term of Eq. (5.9) is the field radiated by the added dipole \mathbf{p}_{N+1} , while the second term is the change to the field arising from the perturbation to the other N initial dipole moments due to addition of the analyte particle. In the single scattering regime, the coupling between dipoles is negligible and $\mathbf{p}_i = \mathbf{p}'_i = \mathbf{p}_{0,i}$ meaning that $\mathbf{p}_{N+1} = \mathbf{p}_{0,N+1}$, while $\delta \mathbf{p}_i = \mathbf{0}$ and the second term vanishes. As a result, the single scattering perturbation $\Delta \mathbf{E}_{ss}(\mathbf{r})$ reduces to

$$\Delta \mathbf{E}_{ss}(\mathbf{r}) = \frac{k_0^2}{\varepsilon_0} G(\mathbf{r}, \mathbf{r}_{N+1}) \mathbf{p}_{0,N+1}. \quad (5.10)$$

Note this matches Eq. (3.70), derived in the single scattering Born approximation. The coupled dipole equations for the perturbed $N + 1$ scatterer system can be split into the N equations for $i = 1, \dots, N$ and the $i = N + 1$ equation as follows

$$\sum_{j=1}^N M_{ij}(\mathbf{p}_j + \delta \mathbf{p}_j) - \frac{k_0^2}{\varepsilon_0} \alpha_i G_{i,N+1} \mathbf{p}_{N+1} = \mathbf{p}_{0,i}, \quad (5.11)$$

$$\mathbf{p}_{N+1} - \frac{k_0^2}{\varepsilon_0} \sum_{j=1}^N \alpha_{N+1} G_{N+1,j}(\mathbf{p}_j + \delta \mathbf{p}_j) = \mathbf{p}_{0,N+1}, \quad (5.12)$$

where the $j = N + 1$ in the sums has been separated from the rest of the sum. Using Eq. (5.6), Eq. (5.11) can be rearranged to yield

$$\delta \mathbf{p}_i = \sum_{j=1}^N M_{ij}^{-1} \frac{k_0^2}{\varepsilon_0} \alpha_j G_{j,N+1} \mathbf{p}_{N+1}, \quad (5.13)$$

where M_{ij}^{-1} is here used to denote the (i, j) th 3×3 block (corresponding to rows $3i - 2$ to $3i$ and columns $3j - 2$ to $3j$) of the inverse of the entire $3N \times 3N$ matrix M , as opposed to $(M_{ij})^{-1}$, the inverse of the 3×3 sub-matrix M_{ij} . Substituting Eq. (5.13) into Eq. (5.9) then gives

$$\Delta \mathbf{E}(\mathbf{r}) = \frac{k_0^2}{\varepsilon_0} G(\mathbf{r}, \mathbf{r}_{N+1}) \gamma_1(\mathbf{r}) \mathbf{p}_{N+1}. \quad (5.14)$$

where the enhancement factor γ_1 has been introduced, defined by

$$\gamma_1(\mathbf{r}) = \mathbb{I} + \frac{k_0^2}{\varepsilon_0} G(\mathbf{r}, \mathbf{r}_{N+1})^{-1} \sum_{i,j=1}^N G(\mathbf{r}, \mathbf{r}_i) M_{ij}^{-1} \alpha_j G_{j,N+1}. \quad (5.15)$$

Expressing $\Delta \mathbf{E}$ in this form allows for comparison with the single scattering result in Eq. (5.10). Specifically, it is evident that the perturbation to the dipole moments of the N initial scatterers from introduction of an additional scatterer is described by the factor γ_1 . The coupling of dipoles arising from multiple scattering acts to modify the effective dipole moment of the additional scatterer such that $\mathbf{p}_{N+1} \rightarrow \gamma_1(\mathbf{r}) \mathbf{p}_{N+1}$. The tensor nature of γ_1 reflects the fact that the polarisation of the field perturbation can be modified by multiple scattering. Similarly, γ_1 is a complex quantity, implying multiple scattering can affect both the phase and amplitude of $\Delta \mathbf{E}$.

Not all multiple scattering effects are captured in γ_1 however. In addition to the additional scatterer perturbing the other dipole moments, the local field experienced by the additional scatterer contains a contribution from scattering of the illumination field by the N initial scatterers, in addition to the contribution from the illumination field on the particle. Substituting Eqs. (5.13) and (5.6) into Eq. (5.12) results in

$$\begin{aligned} \mathbf{p}_{N+1} &= \mathbf{p}_{0,N+1} + \frac{k_0^2}{\varepsilon_0} \sum_{i,j=1}^N \alpha_{N+1} G_{N+1,i} M_{ij}^{-1} \mathbf{p}_{0,j} \\ &+ \left(\frac{k_0^2}{\varepsilon_0} \right)^2 \sum_{i,j=1}^N \alpha_{N+1} G_{N+1,i} M_{ij}^{-1} \alpha_j G_{j,N+1} \mathbf{p}_{N+1}. \end{aligned} \quad (5.16)$$

Rearranging for \mathbf{p}_{N+1} yields

$$\begin{aligned} \mathbf{p}_{N+1} &= \left[\mathbb{I} - \left(\frac{k_0^2}{\varepsilon_0} \right)^2 \sum_{i,j=1}^N \alpha_{N+1} G_{N+1,i} M_{ij}^{-1} \alpha_j G_{j,N+1} \right]^{-1} \\ &\quad \times \left[\mathbf{p}_{0,N+1} + \frac{k_0^2}{\varepsilon_0} \sum_{i,j=1}^N \alpha_{N+1} G_{N+1,i} M_{ij}^{-1} \mathbf{p}_{0,j} \right]. \end{aligned} \quad (5.17)$$

By defining two further enhancement factors, $\Delta \mathbf{E}$ can be expressed as

$$\Delta \mathbf{E}(\mathbf{r}) = \frac{k_0^2}{\varepsilon_0} G(\mathbf{r}, \mathbf{r}_{N+1}) \gamma_1 \gamma_2 \gamma_3 \mathbf{p}_{0,N+1}, \quad (5.18)$$

where

$$\gamma_2 = \left[\mathbb{I} - \left(\frac{k_0^2}{\varepsilon_0} \right)^2 \sum_{i,j=1}^N \alpha_{N+1} G_{N+1,i} M_{ij}^{-1} \alpha_j G_{j,N+1} \right]^{-1} \quad (5.19)$$

$$\gamma_3 = \mathbb{I} + \frac{k_0^2}{\varepsilon_0} \sum_{i,j=1}^N \alpha_{N+1} G_{N+1,i} M_{ij}^{-1} \frac{\mathbf{p}_{0,j} \mathbf{p}_{0,N+1}^\dagger}{|\mathbf{p}_{0,N+1}|^2}. \quad (5.20)$$

Expressed in this way, the effect of multiple scattering is equivalent to changing the dipole moment from $\mathbf{p}_{0,N+1}$ to $\gamma_1 \gamma_2 \gamma_3 \mathbf{p}_{0,N+1}$. In general, as with γ_1 , the enhancement factors γ_2 and γ_3 are complex tensors, meaning multiple scattering can change the phase, amplitude and polarisation of $\Delta \mathbf{E}$.

5.2.2 Physical Interpretation

Each enhancement factor can be associated with a class of multiple scattering paths involving the additional scatterer as shown in Fig. 5.3. The various factors appearing in Eqs. (5.15), (5.19) and (5.20) have simple physical interpretations. A factor of G_{ij} describes free propagation from \mathbf{r}_j to \mathbf{r}_i (i.e. with no scattering in between), while α_i describes scattering from the i th scatterer. Finally, M_{ij}^{-1} describes the sum of all scattering paths propagating from the j th scatterer to the i th scatterer that do not include the additional scatterer. The reasoning behind this interpretation of M_{ij}^{-1} will become clear in Section 5.2.3. With these physical interpretations of the various factors appearing in the enhancement factors, one can infer the

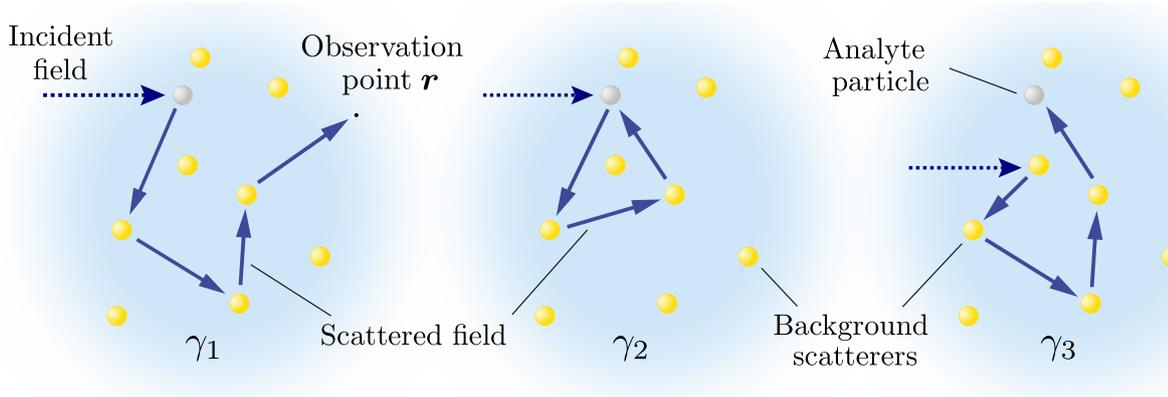


Figure 5.3: Example multiple scattering paths for each enhancement factor: (γ_1 , left) rescattering of light after scattering from the additional particle en route to the observation point, (γ_2 , center) loop trajectories starting and ending on the additional scatterer and (γ_3 , right) multiple scattering of the illumination field onto the analyte particle. *Figure reprinted with permission from Ref. [287] © 2021 American Chemical Society.*

scattering paths described by each. Firstly, γ_1 describes the effect of rescattering of the field as it propagates to the observation point \mathbf{r} after being scattered by the additional scatterer. The factor of $\alpha_j G_{j,N+1}$ freely propagates the scattered field from \mathbf{r}_{N+1} to a scattering event at the j th scatterer, while M_{ij}^{-1} propagates the field from the j th scatterer to the i th scatterer via all possible scattering paths involving the initial N scatterers. Free propagation from the i th scatterer to the observation point \mathbf{r} is then described by $G(\mathbf{r}, \mathbf{r}_i)$. Secondly, γ_2 describes the effect of loop scattering paths in which the scattered light returns (possibly multiple times) to the additional dipole via multiple scattering from the N initial dipoles. As in γ_1 , the factor of $M_{ij}^{-1} \alpha_j G_{j,N+1}$ propagates the scattered field from the additional scatterer to the i th scatterer via all possible scattering paths not including the additional scatterer. Following this, the field then propagates the field back to the additional scatterer from which it is scattered again, represented by a factor $\alpha_{N+1} G_{N+1,i}$, completing the loop. Thus, the sum in Eq. (5.19) is a single loop factor representing the contribution of all single loop paths. Summing over the number of loops yields a geometric series in terms of the single loop factor, and hence γ_2 can be expressed as a matrix inverse of the form $(\mathbb{I} - \text{single loop factor})^{-1}$. The loop contribution is a self-interaction effect analogous to the surface dressing of polarisability discussed previously. Finally, γ_3 accounts for the effect of scattering of the incident field onto the additional scatterer. The incident field at the j th scatterer (represented by the factor \mathbf{p}_i) is multiply scattered (by the initial N scatterers) to the i th scatterer, described by M_{ij}^{-1} , and then propagated to the additional scatterer at \mathbf{r}_{N+1} ,

as is described by the final factor of $\alpha_{N+1}G_{N+1,i}$. The previously discussed hotspot mechanism is described by this class of scattering paths [31, 298]. An immediate consequence of these interpretations is seen in the α_{N+1} dependence of the factors. Since γ_1 only includes scattering after scattering from the additional particle, it has no dependence on α_{N+1} . Similarly, for scalar α_{N+1} , the α_{N+1} factors cancel in Eq. (5.20), so that γ_3 shows no dependence on α_{N+1} . This is because γ_3 deals with scattering before the analyte particle, and thus does not depend on the strength of scattering from the analyte particle. It does, however, depend on a tensor α_{N+1} , since an anisotropic additional scatterer results in different responses to different polarisations of the scattered field. Finally, γ_2 exhibits a dependence on α_{N+1} even if it is scalar, with $\mathbb{I} - \gamma_2^{-1} \propto \alpha_{N+1}$. This is due to the fact it deals with scattering loops visiting the additional scatterer, so intrinsically depends on its properties, in particular the strength with which it scatters light.

The enhancement factors describe any and all multiple scattering effects within the coupled dipole model, and apply generally to any sort of background media by appropriate choice of Green's tensor. In addition to this, the enhancement factors can also describe multiple scattering of other waves, such as acoustic waves, with point scatterer models. The derivations of $\gamma_{1,2,3}$ assumed a (3 dimensional) vector field, however, the results hold equally for waves which can be described by a scalar model, with replacement of quantities with their scalar equivalent (e.g. G is a scalar Green's function, \mathbf{E} becomes a scalar field, M is an $N \times N$ matrix with scalar elements M_{ij} and M_{ij}^{-1} is the (i, j) scalar element of M^{-1} , so that $\gamma_{1,2,3}$ are complex scalars, rather than tensors).

The interpretation of $\gamma_{1,2,3}$ as enhancement factors is perhaps seen most intuitively in the scalar version. Here, the magnitude of the field perturbation, relative to the single scattering perturbation is $|\delta E/\delta E_{ss}| = |\gamma_1\gamma_2\gamma_3|$, and thus the magnitudes of each factor gives the enhancement to the field sensitivity \mathcal{S}_E arising from corresponding set of scattering paths. The phases give the phase shift due to multiple scattering. In the full vector case, $|\Delta \mathbf{E}|/|\Delta \mathbf{E}_{ss}|$ can be at least bounded through $\|G(\mathbf{r}, \mathbf{r}_{N+1})\gamma_1\gamma_2\gamma_3G(\mathbf{r}, \mathbf{r}_{N+1})^{-1}\| \leq \kappa_G\|\gamma_1\gamma_2\gamma_3\| \leq \kappa_G\|\gamma_1\|\|\gamma_2\|\|\gamma_3\|$ where κ_G is the condition number of $G(\mathbf{r}, \mathbf{r}_{N+1})$ given by the ratio of the maximal and minimal singular values [351].

For a given background environment, the enhancement factors are determined by the scattering configuration before the addition of the analyte particle (α_i and \mathbf{r}_i for $i = 1, \dots, N$) and the polarisability and position of the added analyte particle (α_{N+1} and \mathbf{r}_{N+1}). While the exact scattering configuration is unknown for a random sensing setup such as the one in Fig. 3.2, the statistics over an ensemble of random configurations can be studied. In order to get an analytic description of the statistics, the inverse matrix M^{-1} must be expressed in a more tractable way in terms of the known quantities

5.2.3 Born Expansion of Coupling Matrix

By expressing $M = \mathbb{I} - P$, the Neumann series $(\mathbb{I} - P)^{-1} = \sum_{l=0}^{\infty} P^l$, valid when $\|P\| < 1$ for any matrix norm, can be used to expand M_{ij}^{-1} that appears in the enhancement factors as

$$M_{ij}^{-1} = \mathbb{I}\delta_{ij} + \sum_{k=1}^{\infty} P_{ij}^k. \quad (5.21)$$

As with M_{ij}^{-1} , P_{ij}^k denotes the (i, j) 3×3 block (vector field) or scalar element (scalar field) of the matrix P^k , as opposed to $(P_{ij})^k$, the (i, j) element raised to the power k . The elements P_{ij} of P are given by

$$P_{ij} = \begin{cases} 0 & i = j \\ \frac{k_0^2}{\varepsilon_0} \alpha_i G_{ij} & i \neq j \end{cases}, \quad (5.22)$$

from which the k th order term in the Neumann series can be found as

$$P_{ij}^k = \left(\frac{k_0^2}{\varepsilon_0} \right)^k \sum_{\substack{l_1, l_2, \dots, l_{k-1}=1 \\ l_{r+1} \neq l_r \\ l_1 \neq i \\ l_{k-1} \neq j}} \alpha_i G_{il_1} \alpha_{l_1} G_{l_1 l_2} \alpha_{l_2} G_{l_2 l_3} \dots \alpha_{l_{k-1}} G_{l_{k-1} j}. \quad (5.23)$$

The expansion of M^{-1} , along with Eq. (5.23) makes clear the physical interpretation of M_{ij}^{-1} as the sum over all scattering paths starting at the j th scatterer and ending at the i th scatterer. The term P_{ij}^k corresponds to the contribution from all paths visiting exactly k scatterers, where

each factor of $\alpha G_{l_r, l_{r+1}}$ propagates the field to the next scattering event. The $l_{r+1} \neq l_r$ exclusion arises because a scattering path does not visit the same scatterer consecutively (as the self interaction is accounted for in the polarisabilities α_r). Summing over all k then accounts for all scattering paths. The expansion can be thought of as a discretised version of the Born or Dyson series such as presented in Eq. (3.17), where instead of integration over a continuous scattering potential, there is a sum over discrete point scatterers. The expansion of M^{-1} allows for the enhancement factors to be expressed as products of Green's functions, allowing the application of diagrammatic methods.

5.2.4 Far Field Enhancement Factor

While γ_2 and γ_3 are independent of observation position \mathbf{r} , $\gamma_1(\mathbf{r})$ is a function of \mathbf{r} and thus can be simplified when \mathbf{r} is in the far field. Taking the Green's tensors $G(\mathbf{r}, \mathbf{r}_{N+1})$ and $G(\mathbf{r}, \mathbf{r}_i)$ from Eq. (5.15) to be the far field form for a general thin film background, as given in Eq. (3.30), and that all scatterers are above the thin film structure, γ_1 simplifies to

$$\gamma_1(\mathbf{k}_{\parallel}) = \mathbb{I} + \frac{k_0^2}{\varepsilon_0} \sum_{i,j=1}^N [R^{\pm}(z_i, z_{N+1}) e^{-i\mathbf{k}_{\parallel} \cdot (\boldsymbol{\rho}_i - \boldsymbol{\rho}_{N+1})} e^{-ik_z(z_i - z_{N+1})} M_{ij}^{-1} \alpha_j G_{j,N+1}], \quad (5.24)$$

where the change of argument from \mathbf{r} to \mathbf{k}_{\parallel} emphasises that this form of γ_1 is considered in the far field. The function $R^{\pm}(z_i, z_{N+1})$ is defined through the equation

$$G_{\infty}(\mathbf{r}, \mathbf{r}_{N+1})^{-1} G_{\infty}(\mathbf{r}, \mathbf{r}_i) = R^{\pm}(z_i, z_{N+1}) e^{-i\mathbf{k}_{\parallel} \cdot (\boldsymbol{\rho}_i - \boldsymbol{\rho}_{N+1})} e^{-ik_z(z_i - z_{N+1})}, \quad (5.25)$$

where the \pm denotes whether \mathbf{r} is above (+) or below (−) the thin film structure. Using Eq. (3.30), R^{\pm} can be simplified under certain assumptions. Firstly, when $z_{N+1} = z_i$, $R^{\pm}(z_i, z_{N+1}) = \mathbb{I}$. Similarly, for observation points below the multilayer structure, $R^{-}(z_i, z_{N+1}) = \mathbb{I}$. The remaining case, for observation points above the thin multilayer stack (i.e. on the same side as the scatterers) at different heights, the relative phase shifts to the direct and reflected components mean R^{+} depends on the heights non-trivially. In this case, using Eqs. (3.24),

(3.25) and (3.30), one finds

$$R^+(z_i, z_{N+1}) = \tilde{H}(\mathbf{k}_{\parallel}, z_{N+1})^{-1} \tilde{H}(\mathbf{k}_{\parallel}, z_i). \quad (5.26)$$

where

$$\tilde{H}(\mathbf{k}_{\parallel}, z_i) = \left[(1 + r_s(k_{\parallel})e^{2ik_z z_i}) \hat{\mathbf{e}}_s(\mathbf{k}_{\parallel}) \hat{\mathbf{e}}_s^{\dagger}(\mathbf{k}_{\parallel}) + \hat{\mathbf{e}}_{p+}(\mathbf{k}_{\parallel}) \hat{\mathbf{e}}_{p+}^{\dagger}(\mathbf{k}_{\parallel}) D_p(k_{\parallel}, z_i) \right], \quad (5.27)$$

$D_p(k_{\parallel}, z_i)$ is a diagonal matrix given by

$$D(k_{\parallel}, z_i) = \mathbb{I} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} r_p(k_{\parallel}) e^{2ik_z z_i}. \quad (5.28)$$

Here, it can be seen how the different propagation phases of the direct and reflected light are included. This form of R^{\pm} is significantly more complicated, and most of the results in this chapter will apply to cases where $R^{\pm} = \mathbb{I}$. Such an assumption is not too restrictive in the context of the proposed sensing setup. Firstly, it applies to all observation points below the multilayer stack, including the leakage ring. In addition, scatterers are located on or near the upper surface, and thus can be well approximated as all lying at the same height. Also, in a homogeneous background where the Green's tensor is given by Eq. (3.20), $R^{\pm} = \mathbb{I}$ for any scatterer arrangement and far field observation position.

5.2.5 Plane Wave Illumination and Reciprocity

The forms of Eqs. (5.15) and (5.20) show similarities in the form of γ_1 and γ_3 . As γ_3 describes the initial scattering before reaching the analyte particle, it is the only enhancement factor to depend on \mathbf{E}_0 . The relationship between γ_1 and γ_3 is made even clearer when \mathbf{E}_0 is a (lossless) plane wave $\mathbf{E}_0(\mathbf{r}) = \mathcal{E}_0 \hat{\boldsymbol{\xi}} \exp(i\mathbf{k}_{inc} \mathbf{r})$, with spatially uniform polarisation $\hat{\boldsymbol{\xi}}$ and amplitude \mathcal{E}_0 , so

$\mathbf{p}_{0,j} = \alpha_j A \hat{\boldsymbol{\xi}} e^{i\mathbf{k}_{inc} \cdot \mathbf{r}_j}$, With this incident field, γ_3 becomes

$$\gamma_3 = \mathbb{I} + \sum_{i,j=1}^N \alpha_{N+1} G_{N+1,i} M_{ij}^{-1} e^{i\mathbf{k}_{inc} \cdot (\mathbf{r}_j - \mathbf{r}_{N+1})} \frac{\alpha_j \hat{\boldsymbol{\xi}} (\alpha_{N+1} \hat{\boldsymbol{\xi}})^\dagger}{|\alpha_{N+1} \hat{\boldsymbol{\xi}}|^2}. \quad (5.29)$$

In comparison to Eq. (5.24), these forms of γ_1 and γ_3 share an exponential factor $\sim \exp\{i\mathbf{q} \cdot (\mathbf{r}_i - \mathbf{r}_{N+1})\}$, a Green's tensor propagating from or to the analyte scatterer and M_{ij}^{-1} . Further simplification is possible assuming isotropic scatterers so that α_i are all scalars, reducing Eq. (5.29) to

$$\gamma_3 = \mathbb{I} + \sum_{i,j=1}^N G_{N+1,i} M_{ij}^{-1} \alpha_j e^{i\mathbf{k}_{inc} \cdot (\mathbf{r}_j - \mathbf{r}_{N+1})} \hat{\boldsymbol{\xi}} \hat{\boldsymbol{\xi}}^\dagger. \quad (5.30)$$

Finally, assuming the initial scatterers are identical ($\alpha_j = \alpha$ for all $j \leq N$), γ_3 can be expressed

$$\begin{aligned} \gamma_3 &= \mathbb{I} + \sum_{i,j=1}^N \alpha G_{N+1,i} M_{ij}^{-1} e^{i\mathbf{k}_{inc} \cdot (\mathbf{r}_j - \mathbf{r}_{N+1})} \hat{\boldsymbol{\xi}} \hat{\boldsymbol{\xi}}^\dagger \\ &= \mathbb{I} + \sum_{i,j=1}^N \alpha G_{i,N+1}^T (M_{ji}^{-1})^T e^{i\mathbf{k}_{inc} \cdot (\mathbf{r}_j - \mathbf{r}_{N+1})} \hat{\boldsymbol{\xi}} \hat{\boldsymbol{\xi}}^\dagger \end{aligned}$$

From Eq. (5.18), γ_3 only acts on $\mathbf{p}_{0,N+1}$, which for isotropic scatterers is oriented in the direction $\hat{\boldsymbol{\xi}}$. When operating on vectors of this orientation, the projection operator $\hat{\boldsymbol{\xi}} \hat{\boldsymbol{\xi}}^\dagger$ is equivalent to the identity and thus, when acting on vectors in this space, γ_3 can be expressed

$$\begin{aligned} \gamma_3 &= \left[\mathbb{I} + \sum_{i,j=1}^N \alpha G_{i,N+1}^T (M_{ji}^{-1})^T e^{i\mathbf{k}_{inc} \cdot (\mathbf{r}_j - \mathbf{r}_{N+1})} \right] \hat{\boldsymbol{\xi}} \hat{\boldsymbol{\xi}}^\dagger \\ &= \left[\mathbb{I} + \sum_{i,j=1}^N e^{i\mathbf{k}_{inc} \cdot (\mathbf{r}_j - \mathbf{r}_{N+1})} M_{ij}^{-1} \alpha G_{i,N+1} \right]^T \hat{\boldsymbol{\xi}} \hat{\boldsymbol{\xi}}^\dagger. \end{aligned} \quad (5.31)$$

The term in brackets is identical to $\gamma_1(-\mathbf{k}_{inc,\parallel})$ from Eq. (5.24), for the case $R^\pm = \mathbb{I}$. Thus, up to the projection operator, γ_3 is the transpose of γ_1 in the far field. In the scalar case, the transpose and projection operators drop out and $\gamma_3 = \gamma_1(-\mathbf{k}_{inc,\parallel})$. This symmetry arises due to the reciprocal symmetry of scattering paths in γ_1 and γ_3 . Scattering paths involved in γ_1 have a reciprocal scattering path visiting the same scatterers in reverse order. The contribution of such paths to the enhancement factor is identical up to the outgoing propagation to \mathbf{r} in the

case of γ_1 as compared to the incoming propagation to \mathbf{r}_{N+1} for γ_3 . Because of the reciprocal symmetry, the scattering of the incoming plane wave to \mathbf{r}_{N+1} gives an identical enhancement to scattering from \mathbf{r}_{N+1} into an outgoing plane wave $-\mathbf{k}_{inc}$, hence the relation of Eq. (5.31). Relaxing the assumptions (identical, isotropic scatterers and lossless plane wave illumination) results in Eq. (5.31) breaking down, though the relationship between reciprocal scattering paths still has consequences with regards to the relationship between γ_1 and γ_3 , which will be discussed later.

5.2.6 Intensity Sensitivity Enhancement

The enhancement factors γ_i give the enhancement in the field perturbation, but as discussed in Chapter 2, any experimental method would rely on measurement of intensities. Just as in the single scattering analysis (see Eq. (3.85)), the intensity perturbation ΔI contains an interference between the field perturbation and the background field as follows

$$\Delta I(\mathbf{r}) = |\Delta \mathbf{E}(\mathbf{r})|^2 + 2 \operatorname{Re} [\Delta \mathbf{E}(\mathbf{r}) \cdot \mathbf{E}^*(\mathbf{r})]. \quad (5.32)$$

While similar to Eq. (3.85), this result differs in that the interference is not with the field scattered from the analyte particle, but with the full perturbation $\Delta \mathbf{E}$ including multiple scattering via the enhancement factors. The interferometric enhancement discussed in the context of single scattering applies equally here, however it is not a multiple scattering effect. Since the enhancement factors are complex, the phase of $\Delta \mathbf{E}$ is changed by multiple scattering, however, the phase difference between \mathbf{E} and $\Delta \mathbf{E}$ is random in both the single and multiple scattering regimes, meaning that the phase statistics of the interference term are essentially unchanged by the degree of multiple scattering. The primary difference in the statistics of the interference term between single and multiple scattering lies in the different amplitudes $|\Delta \mathbf{E}|$, captured by $\|\gamma_1 \gamma_2 \gamma_3\|$.

5.3 Analytic Results

In order to analytically calculate statistical properties of the enhancement factors, one must define the statistics of the scattering configuration. It is assumed that the N initial scatterers are identical and have the same orientation relative to the surface, implying $\alpha_i = \alpha$ for $i \leq N$. Note that α is still allowed to be a tensor. Furthermore, the transverse positions $\boldsymbol{\rho}_i$ of these scatterers are assumed to be independently randomly distributed with uniform probability across a 2D planar region of area L^2 ($L \gg \lambda_0$) on the surface of a substrate, with the same height $z_i = z_s$ for $i \leq N$. This model is more appropriate for a nanostructured surface with discrete scatterers randomly positioned on the surface of the multilayer stack, as opposed to a modelling scattering from continuous surface roughness. The properties (α_{N+1} and \boldsymbol{r}_{N+1}) of the added scatterer are not restricted and may be different to the background scatterers. It will also be convenient to denote the sums appearing in Eqs. (5.15), (5.19) and (5.20) such that

$$\gamma_1 = \mathbb{I} + S_1 \quad (5.33)$$

$$\gamma_2 = (\mathbb{I} - S_2)^{-1} \quad (5.34)$$

$$\gamma_3 = \mathbb{I} + S_3, \quad (5.35)$$

and therefore the sums, taking the observation position in the far field, are given by

$$S_1(\mathbf{k}_{\parallel}) = \frac{k_0^2}{\varepsilon_0} \sum_{i,j=1}^N [R^{\pm}(z_i, z_{N+1}) e^{-i\mathbf{k}_{\parallel} \cdot (\boldsymbol{\rho}_i - \boldsymbol{\rho}_{N+1})} e^{-ik_z(z_i - z_{N+1})} M_{ij}^{-1} \alpha_j G_{j,N+1}] \quad (5.36)$$

$$S_2 = \left(\frac{k_0^2}{\varepsilon_0} \right)^2 \sum_{i,j=1}^N \alpha_{N+1} G_{N+1,i} M_{ij}^{-1} \alpha_j G_{j,N+1} \quad (5.37)$$

$$S_3(\mathbf{E}_0) = \frac{k_0^2}{\varepsilon_0} \sum_{i,j=1}^N \alpha_{N+1} G_{N+1,i} M_{ij}^{-1} \frac{\boldsymbol{p}_{0,j} \boldsymbol{p}_{0,i}^{\dagger}}{|\boldsymbol{p}_{0,N+1}|^2}. \quad (5.38)$$

To calculate the mean values $\langle S_{1,2,3} \rangle$, it is helpful to introduce a Fourier decomposition of the sums. Expressing the Green's tensors using Eq. (3.29) and also expressing the incident field as

an inverse Fourier transform (of the form of Eq. (3.21)), the sums can be expressed as

$$S_1(\mathbf{k}_{\parallel}) = \frac{k_0^2}{\varepsilon_0} \int \frac{d^2 \mathbf{q}}{(2\pi)^2} R^{\pm}(z_s, z_{N+1}) A(\mathbf{k}_{\parallel}, \mathbf{q}) \alpha \tilde{G}(\mathbf{q}; z_s, z_{N+1}) e^{ik_z(\mathbf{k}_{\parallel})(z_s - z_{N+1})} \quad (5.39)$$

$$S_2 = \left(\frac{k_0^2}{\varepsilon_0} \right)^2 \int \frac{d^2 \mathbf{q}}{(2\pi)^2} \frac{d^2 \mathbf{q}'}{(2\pi)^2} \alpha_{N+1} \tilde{G}(\mathbf{q}; z_{N+1}, z_s) A(\mathbf{q}, \mathbf{q}') \alpha \tilde{G}(\mathbf{q}'; z_s, z_{N+1}) \quad (5.40)$$

$$S_3(\mathbf{E}_0) = \frac{k_0^2}{\varepsilon_0} \int \frac{d^2 \mathbf{q}}{(2\pi)^2} \frac{d^2 \mathbf{q}'}{(2\pi)^2} \alpha_{N+1} \tilde{G}(\mathbf{q}; z_{N+1}, z_s) A(\mathbf{q}, \mathbf{q}') \alpha \tilde{\mathbf{E}}_0(\mathbf{q}_2; z_s) e^{i\mathbf{q}' \cdot \boldsymbol{\rho}_{N+1}} \frac{\mathbf{p}_{0,N+1}^{\dagger}}{|\mathbf{p}_{0,N+1}|^2}, \quad (5.41)$$

where the function $A(\mathbf{q}, \mathbf{q}')$ has been introduced, defined by

$$A(\mathbf{q}, \mathbf{q}') = \sum_{i,j=1}^N e^{-i\mathbf{q} \cdot (\boldsymbol{\rho}_i - \boldsymbol{\rho}_{N+1})} M_{ij}^{-1} e^{i\mathbf{q}' \cdot (\boldsymbol{\rho}_j - \boldsymbol{\rho}_{N+1})}. \quad (5.42)$$

The dependence of Eqs. (5.39)–(5.41) on the random transverse positions $\boldsymbol{\rho}_i$, and therefore the statistics, is entirely contained within A . Substituting the Born expansion from Eq. (5.21) into Eq. (5.42) results in

$$A(\mathbf{q}, \mathbf{q}') = \sum_{p=0}^{\infty} A^{(p)}(\mathbf{q}, \mathbf{q}') \quad (5.43)$$

$$A^{(p)}(\mathbf{q}, \mathbf{q}') = \left(\frac{k_0^2}{\varepsilon_0} \right)^p \sum_{\substack{i,j,l_1,l_2, \\ \dots, l_{p-1}}} e^{-i\mathbf{q} \cdot (\boldsymbol{\rho}_i - \boldsymbol{\rho}_{N+1})} \alpha G_{il_1} \alpha G_{l_1 l_2} \alpha G_{l_2 l_3} \dots \alpha G_{l_{p-1} l_p} \alpha G_{l_p j} e^{i\mathbf{q}' \cdot (\boldsymbol{\rho}_j - \boldsymbol{\rho}_{N+1})}, \quad (5.44)$$

where $A^{(p)}$ is the term containing the p th order term from the Born expansion and the limits and exclusions (i.e. that the scattering path cannot visit the same scatterer consecutively) from the sums in Eq. (5.44) are left implicit. Substituting the Green's tensors in Eq. (5.44) with its Fourier decomposition from Eq. (3.29) allows the dependence on the scatterer positions to be included within an exponential factor as follows

$$A^{(p)}(\mathbf{q}, \mathbf{q}') = \int \prod_{b=1}^p \frac{d^2 \mathbf{q}_b}{(2\pi)^2} \frac{k_0^2}{\varepsilon_0} \alpha \tilde{G}(\mathbf{q}_b; z_s, z_s) \times \sum_{i,j,l_1,\dots,l_{p-1}} e^{-i\mathbf{q} \cdot (\boldsymbol{\rho}_i - \boldsymbol{\rho}_{N+1})} e^{i\mathbf{q}_1 \cdot (\boldsymbol{\rho}_i - \boldsymbol{\rho}_{l_1})} e^{i\mathbf{q}_2 \cdot (\boldsymbol{\rho}_{l_1} - \boldsymbol{\rho}_{l_2})} \dots e^{i\mathbf{q}_p \cdot (\boldsymbol{\rho}_{l_{p-1}} - \boldsymbol{\rho}_j)} e^{i\mathbf{q}' \cdot (\boldsymbol{\rho}_j - \boldsymbol{\rho}_{N+1})}. \quad (5.45)$$

The dependence on the random $\boldsymbol{\rho}_i$ in Eq. (5.45) is contained only in the exponential factors. Regrouping the exponents so that each \boldsymbol{r}_{l_i} term is in the same exponent allows the sum to be rewritten as

$$\sum_{i,j,l_1,\dots,l_{p-1}} e^{i(\mathbf{q}_1-\mathbf{q})\cdot\boldsymbol{\rho}_i} e^{i(\mathbf{q}_2-\mathbf{q}_1)\cdot\boldsymbol{\rho}_{l_1}} \dots e^{i(\mathbf{q}_{c+1}-\mathbf{q}_c)\cdot\boldsymbol{\rho}_{l_c}} \dots e^{i(\mathbf{q}'-\mathbf{q}_p)\cdot\boldsymbol{\rho}_j} e^{i(\mathbf{q}-\mathbf{q}')\cdot\boldsymbol{\rho}_{N+1}}. \quad (5.46)$$

In order to find $\langle A^{(p)} \rangle$ (angled brackets here denoting averaging over the positions $\boldsymbol{\rho}_j$ of the N initial scatterers), one only has to find the average of this sum of exponential factors. Since the probability distribution of $\boldsymbol{\rho}_j$ is uniform across the 2D planar region \mathcal{A} , any average is found simply by integration over this region, $\langle f(\boldsymbol{\rho}_j) \rangle = \int_{\boldsymbol{\rho}_j \in \mathcal{A}} f(\boldsymbol{\rho}_j) d^2 \boldsymbol{\rho}_j / L^2$. The mean of a complex exponential function can be calculated to be

$$\langle e^{i\mathbf{q}\cdot\boldsymbol{\rho}_j} \rangle = \frac{4}{L^2} \frac{\sin(q_x L_x)}{q_x} \frac{\sin(q_y L_y)}{q_y} \quad (5.47)$$

where the \mathcal{A} has been assumed to be rectangular of sides L_x and L_y and q_x and q_y are the components of \mathbf{q} . Since Eq. (5.46) involves the sum over scatterers, it is important to consider the average of a sum over scatterers

$$\begin{aligned} \left\langle \sum_{j=1}^N e^{i\mathbf{q}\cdot\boldsymbol{\rho}_j} \right\rangle &= \frac{4N}{L^2} \frac{\sin(q_x L_x)}{q_x} \frac{\sin(q_y L_y)}{q_y} \\ &= 4n \frac{\sin(q_x L_x)}{q_x} \frac{\sin(q_y L_y)}{q_y}. \end{aligned} \quad (5.48)$$

The 2D scatterer density $n = N/L^2$ has been defined. In the limit $N \rightarrow \infty$ and $L_{x,y} \rightarrow \infty$ with n fixed, Eq. (5.48) becomes

$$\left\langle \sum_{j=1}^N e^{i\mathbf{q}\cdot\boldsymbol{\rho}_j} \right\rangle = n(2\pi)^2 \delta(q_x) \delta(q_y) = n(2\pi)^2 \delta(\mathbf{q}). \quad (5.49)$$

This limit is the one that shall be used in this analysis, owing in part to its analytic simplicity but also it is wide applicability, since generally strong multiple scattering requires a large number of scatterers, and the illuminated region is typically significantly larger than the wavelength,

so that $L \rightarrow \infty$ is valid. While Eq. (5.46) is a sum over exponential factors, one cannot directly apply Eq. (5.49) to each exponential factor, since some of the terms in the sum have the indices taking the same value, i.e. $l_r = l_c$, so that the two exponential factors are not statistically independent and need to be averaged together. Recalling the interpretation of the Born expansion as a sum over scattering paths, terms where a scatterer position appears in two different exponents correspond to loop paths visiting the same scatterer twice. Scattering paths can have multiple loops or visit the same scatterer any number of times. Accounting for all the different possible paths with repeated scatterers is difficult, and, following an approach similar to Ref. [349], it is convenient to group the different terms into how many repeated scatterer positions occur. Firstly, consider only terms with no repeated scatterers, meaning each l_i is distinct. For such terms, each exponential $e^{i(\mathbf{q}_{c+1}-\mathbf{q}_c)\cdot\boldsymbol{\rho}_{l_c}}$ can be averaged independently from the rest as $\boldsymbol{\rho}_{l_c} \neq \boldsymbol{\rho}_{l_r}$ for $l_c \neq l_r$. Since there are $p+1$ distinct scatterers appearing in such terms, the sum in Eq. (5.46) contains $N(N-1)(N-2)\dots(N-p) \approx N^{p+1}$ such terms with no repeated scatterers. Averaging over the $p+1$ different scatterer positions gives a factor of $(L^2)^{-(p+1)}$, so that the contribution of these distinct scatterer terms to the mean of Eq. (5.46) scales as n^{p+1} . Taking the next class of scattering paths with a single repeated scatterer (corresponding to scattering paths with a single loop to revisit the scatterer), these paths visit p distinct scatterers (as the order p term contains $p+1$ scattering events, but two are from the same scatterer). Choosing p scatterers out of N options gives $N(N-1)\dots(N-(p-1)) \sim N^p$ such terms in the mean of Eq. (5.46), while averaging over the p scatterer positions give $(L^2)^{-p}$. Thus, the one loop terms scale as n^p . Applying the same logic to any number of repeated scatterers, it can be deduced that the contribution of terms with r repeated scatterers to the mean of Eq. (5.46) scales as n^{p+1-r} . Ignoring loops and taking only the highest order contribution in n (i.e. the no repeated scatterers terms) allows the mean of Eq. (5.46) to be analytically calculated by applying Eq. (5.49) to each $\boldsymbol{\rho}_{l_c}$ separately, resulting in

$$n^{p+1}(2\pi)^2\delta(\mathbf{q}_1-\mathbf{q})(2\pi)^2\delta(\mathbf{q}_2-\mathbf{q}_1)(2\pi)^2\delta(\mathbf{q}_3-\mathbf{q}_2)\dots(2\pi)^2\delta(\mathbf{q}_{c+1}-\mathbf{q}_c)\dots(2\pi)^2\delta(\mathbf{q}'-\mathbf{q}_p)e^{i(\mathbf{q}-\mathbf{q}')\cdot\boldsymbol{\rho}_{N+1}}. \quad (5.50)$$

Substituting this averaged result into the mean of Eq. (5.45) then gives

$$\langle A^{(p)}(\mathbf{q}, \mathbf{q}') \rangle \approx n^{p+1} \left[\frac{k_0^2}{\varepsilon_0} \alpha \tilde{G}(\mathbf{q}; z_s, z_s) \right]^p (2\pi)^2 \delta(\mathbf{q} - \mathbf{q}'). \quad (5.51)$$

where the delta functions restricts $\mathbf{q}_b = \mathbf{q} = \mathbf{q}'$ for every integration variable \mathbf{q}_b . Using Eq. (5.43), summing over p gives the no loop approximation result for $\langle A \rangle$ as

$$\langle A(\mathbf{q}, \mathbf{q}') \rangle \approx n(2\pi)^2 \delta(\mathbf{q} - \mathbf{q}') \left[\mathbb{I} - n \frac{k_0^2}{\varepsilon_0} \alpha \tilde{G}(\mathbf{q}; z_s, z_s) \right]^{-1}. \quad (5.52)$$

The means of Eqs. (5.39)–(5.41) follow from substitution of Eq. (5.52) and are

$$\langle S_1(\mathbf{k}_{\parallel}) \rangle = R^{\pm}(z_s, z_{N+1}) \left[\mathbb{I} - n \frac{k_0^2}{\varepsilon_0} \alpha \tilde{G}(\mathbf{k}_{\parallel}; z_s, z_s) \right]^{-1} n \alpha \frac{k_0^2}{\varepsilon_0} \tilde{G}(\mathbf{k}_{\parallel}; z_s, z_{N+1}) e^{ik_z(\mathbf{k}_{\parallel})(z_s - z_{N+1})} \quad (5.53)$$

$$\langle S_2 \rangle = n \left(\frac{k_0^2}{\varepsilon_0} \right)^2 \int \frac{d^2 \mathbf{q}}{(2\pi)^2} \alpha_{N+1} \tilde{G}(\mathbf{q}; z_{N+1}, z_s) \left[\mathbb{I} - n \frac{k_0^2}{\varepsilon_0} \alpha \tilde{G}(\mathbf{q}; z_s, z_s) \right]^{-1} \alpha \tilde{G}(\mathbf{q}; z_s, z_{N+1}) \quad (5.54)$$

$$\begin{aligned} \langle S_3(\mathbf{E}_0) \rangle &= n \frac{k_0^2}{\varepsilon_0} \int \frac{d^2 \mathbf{q}}{(2\pi)^2} \alpha_{N+1} \tilde{G}(\mathbf{q}; z_{N+1}, z_s) \left[\mathbb{I} - n \frac{k_0^2}{\varepsilon_0} \alpha \tilde{G}(\mathbf{q}; z_s, z_s) \right]^{-1} \\ &\times \alpha \tilde{\mathbf{E}}_0(\mathbf{q}; z_s) e^{i\mathbf{q} \cdot \boldsymbol{\rho}_{N+1}} \frac{\boldsymbol{\rho}_{0,N+1}^{\dagger}}{|\boldsymbol{\rho}_{0,N+1}|^2}. \end{aligned} \quad (5.55)$$

While $\langle \gamma_1 \rangle$ and $\langle \gamma_3 \rangle$ follow simply from addition of the identity matrix, $\langle \gamma_2 \rangle$ does not follow from $\langle S_2 \rangle$.

The derivation of Eqs. (5.53)–(5.55) made no assumptions on \mathbf{E}_0 and analyte scatterer position \mathbf{r}_{N+1} , but it is informative to consider a couple of special cases. Firstly, the case when $\mathbf{E}_0(\mathbf{r}) = \mathcal{E}_0(z) \hat{\boldsymbol{\xi}} \exp(i\mathbf{k}_{inc,\parallel} \boldsymbol{\rho})$ is a lossless plane wave simplifies $\langle S_3 \rangle$. Note that, by allowing a z -dependent amplitude $\mathcal{E}_0(z)$, one can account for a single transmitted plane wave ($\mathcal{E}_0(z) = \mathcal{E}_0 \exp(ik_{inc,z}z)$) or the sum of a plane wave and a reflected wave ($\mathcal{E}_0(z) = \mathcal{E}_0[\exp(-ik_{inc,z}z) + r(k_{inc,\parallel}) \exp(ik_{inc,z}z)]$ with appropriate choice of reflection coefficient). The transverse Fourier transform of this plane wave form of \mathbf{E}_0 is $\tilde{\mathbf{E}}_0(\mathbf{q}; z_s) = \mathcal{E}_0(z) \hat{\boldsymbol{\xi}} (2\pi)^2 \delta(\mathbf{q} - \mathbf{k}_{inc,\parallel})$, which reduces

Eq. (5.55) to

$$\langle S_3(\mathbf{k}_{inc,\parallel}) \rangle = n \frac{k_0^2}{\varepsilon_0} \alpha_{N+1} \tilde{G}(\mathbf{q}; z_{N+1}, z_s) \left[\mathbb{I} - n \frac{k_0^2}{\varepsilon_0} \alpha \tilde{G}(\mathbf{k}_{inc,\parallel}; z_s, z_s) \right]^{-1} \frac{\mathcal{E}_0(z_s)}{\mathcal{E}_0(z_{N+1})} \frac{(\alpha \hat{\boldsymbol{\xi}})(\alpha_{N+1} \hat{\boldsymbol{\xi}})^\dagger}{|\alpha_{N+1} \hat{\boldsymbol{\xi}}|^2}. \quad (5.56)$$

This form bears close resemblance to $\langle S_1 \rangle$, due to the reciprocal symmetry discussed in Section 5.2.5. Another informative case is when the analyte particle is at the same height as the initial scatterers $z_{N+1} = z_s$. In the context of the sensing setup, since the SPP field decays rapidly away from the surface, scattering only occurs close to the surface and thus $z_{N+1} = z_s$ can approximately describe most situations where the analyte particle scatters light sufficiently. With $z_{N+1} = z_s$, Eq. (5.53) becomes

$$\langle S_1(\mathbf{k}_{\parallel}) \rangle = \left[\mathbb{I} - n \frac{k_0^2}{\varepsilon_0} \alpha \tilde{G}(\mathbf{k}_{\parallel}; z_s, z_s) \right]^{-1} n \alpha \frac{k_0^2}{\varepsilon_0} \tilde{G}(\mathbf{k}_{\parallel}; z_s, z_s) \quad (5.57)$$

Using the matrix push through identity [352], $(\mathbb{I} - X)^{-1} = \mathbb{I} + (\mathbb{I} - X)^{-1}X$, $\langle \gamma_1 \rangle$ is therefore

$$\langle \gamma_1 \rangle = \left[\mathbb{I} - n \frac{k_0^2}{\varepsilon_0} \alpha \tilde{G}(\mathbf{k}_{\parallel}; z_s, z_s) \right]^{-1}. \quad (5.58)$$

This simple form of $\langle \gamma_1 \rangle$ has some significant consequences. Specifically, $\langle \gamma_1 \rangle$ diverges when $\mathbb{I} - n(k_0^2/\varepsilon_0)\alpha\tilde{G}(\mathbf{k}_{\parallel}; z_s, z_s)$ is singular, or in the scalar case, when $n(k_0^2/\varepsilon_0)\alpha\tilde{G}(\mathbf{k}_{\parallel}; z_s, z_s) = 1$. Furthermore, when the divergence condition is almost satisfied (i.e. $\det \left[\mathbb{I} - n(k_0^2/\varepsilon_0)\alpha\tilde{G}(\mathbf{k}_{\parallel}; z_s, z_s) \right]$ is close to zero), $\langle \gamma_1 \rangle$ will become very large, suggesting large multiple scattering enhancements. As a result, under these conditions, it can be concluded that the multiple scattering regime is significantly more sensitive to the addition of a scatterer than the single scattering regime. In the scalar case, the density (for a given α) at which this occurs can be calculated analytically. It can be shown $|\langle \gamma_1 \rangle|$ has a maximum value $|\langle \gamma_1 \rangle|_{\max} > 1$ provided $\text{Re}[\alpha\tilde{G}(\mathbf{k}_{\parallel}; z_s, z_s)] > 0$, occurring at

a density $n_{\text{opt},1}$, where

$$n_{\text{opt},1} = \frac{\text{Re} \left[\alpha \tilde{G}(\mathbf{k}_{\parallel}; z_s, z_s) \right]}{\frac{k_0^2}{\varepsilon_0} \left| \alpha \tilde{G}(\mathbf{k}_{\parallel}; z_s, z_s) \right|^2} \quad (5.59)$$

$$|\langle \gamma_1 \rangle|_{\text{max}} = \frac{\left| \alpha \tilde{G}(\mathbf{k}_{\parallel}; z_s, z_s) \right|}{\text{Im} \left[\alpha \tilde{G}(\mathbf{k}_{\parallel}; z_s, z_s) \right]}. \quad (5.60)$$

Already, this allows us to say a few things about how the scattering parameters affect the sensitivity enhancements. Firstly, the presence of an optimum scatterer density suggests that the enhancement is not monotonic, and that increasing the level of multiple scattering can reduce sensitivity beyond a certain point. Furthermore, the properties of the individual scatterers (i.e. α) also affect the statistics, as can be seen through the α dependence. In particular, depending on the value of α , the behaviour can range from having no optimum density and $|\langle \gamma_1 \rangle|_{\text{max}} < 1$ regardless of density (when $\text{Re}[\alpha \tilde{G}(\mathbf{k}_{\parallel}; z_s, z_s)] < 0$) right through to the mean diverging to ∞ (under the approximations made) when $\text{Im} \left[\alpha \tilde{G}(\mathbf{k}_{\parallel}; z_s, z_s) \right] = 0$. It is important to understand these features in designing a sensing environment where it is desirable to maximise the sensitivity, since it implies one can tune the scatterer density and composition to achieve this increased sensitivity.

5.3.1 Mean Amplitude of Enhancement Factors

While the complex means capture some of the statistics, it does not give an idea of the typical magnitude of the enhancement factors. This is because cancellation of large realisations of γ_i with differing phases can lead to a comparatively small value of $\langle \gamma_i \rangle$. A more useful indication of the typical magnitude of the enhancement is $\langle |\gamma_i| \rangle$. An approximate analytic result for $\gamma_1(\mathbf{k}_{\parallel})$ can be calculated in the scalar case, for which $\gamma_1 = 1 + \sum_{i=1}^N \mathcal{R}_i$ with $\mathcal{R}_i = (k_0^2/\varepsilon_0) \exp[-i\mathbf{k}_{\text{out}} \cdot (\mathbf{r}_i - \mathbf{r}_{N+1})] \sum_{j=1}^N M_{ij}^{-1} \alpha G_{j,N+1}$. Under the assumption that the phase of \mathcal{R}_i is uniformly distributed over the full 2π range and independent of the amplitudes $|\mathcal{R}_i|$, along with the further assumption that N is large, such that the central limit theorem can be applied, it then follows that the phasor sum $S_1 = \sum_{i=1}^N \mathcal{R}_i$ is a zero mean complex circular

Gaussian random variable with variance $\sigma_1^2 = N\langle|\mathcal{R}_i|^2\rangle/2$ as discussed in Chapter 3 [21]. It is clear from Eq. (5.53) that $\langle S_1 \rangle \neq 0$ and therefore the assumptions on the phase and amplitude statistics do not hold exactly (as they would imply $\langle S_1 \rangle = 0$). The consequences of approximating the phase and amplitude statistics in this manner will be assessed later. Given the assumption of a zero mean circular Gaussian distribution on S_1 , it follows that the PDF of γ_1 is a Gaussian with mean of 1 (rotationally symmetric about the mean) and the same standard deviation. The amplitude $|\gamma_1|$ thus follows a Rician distribution, from which the mean amplitude can be calculated using [21]

$$\langle |\gamma_1| \rangle = \sigma_{1,3}(\pi/2)^{1/2} L_{1/2}(-1/(2\sigma_{1,3}^2)) \quad (5.61)$$

where $L_{1/2}(x)$ is a generalised Laguerre polynomial. All that remains to calculate $\langle |\gamma_1| \rangle$ is to calculate σ_1 , which can be done using the framework of diagrammatic methods.

Averaged Perturbed Green's Function

Before calculating σ_1 explicitly, a useful intermediate quantity to calculate is the mean exciting Green's function $\langle G^{sc}(\mathbf{r}, \mathbf{r}') \rangle$, where the quantity G^{sc} has been defined as

$$G^{sc}(\mathbf{r}_i, \mathbf{r}') = \sum_{j=1}^N M_{ij}^{-1} \alpha G(\mathbf{r}_j, \mathbf{r}'). \quad (5.62)$$

Physically, this corresponds to the solution for \mathbf{p}_i when the incident field is taken as the unperturbed Green's tensor $\mathbf{E}_0(\mathbf{r}) = G(\mathbf{r}, \mathbf{r}')$, and therefore is closely related to the Green's tensor for the scattering system including the initial N scatterers, $G^{(N)}$, which would be given by putting $\mathbf{p}_i = G^{sc}(\mathbf{r}_i, \mathbf{r}')$ into Eq. (5.2), resulting in

$$G^{(N)}(\mathbf{r}, \mathbf{r}') = G(\mathbf{r}, \mathbf{r}') + \sum_{i=1}^N G(\mathbf{r}, \mathbf{r}_i) G^{sc}(\mathbf{r}_i, \mathbf{r}'). \quad (5.63)$$

The utility of G^{sc} in the context of calculating $\sigma_1^2 = N\langle |\mathcal{R}_i|^2 \rangle / 2$ is that the sum elements \mathcal{R}_i can be written in terms of the perturbed exciting Green's function as

$$\mathcal{R}_i = e^{-i\mathbf{k}_{\text{out}} \cdot (\mathbf{r}_i - \mathbf{r}_{N+1})} G^{sc}(\mathbf{r}_i, \mathbf{r}_{N+1}). \quad (5.64)$$

Expanding M_{ij}^{-1} with Eq. (5.21) as was done for the enhancement factor gives

$$G^{sc}(\mathbf{r}, \mathbf{r}') = \sum_{k=0}^{\infty} \left(\frac{k_0^2}{\varepsilon_0} \right)^{k+1} \sum_{l_1, \dots, l_k} \alpha G(\mathbf{r}, \mathbf{r}_{l_1}) \alpha G_{l_1 l_2} \alpha G_{l_2 l_3} \dots \alpha G_{l_{k-2} l_{k-1}} \alpha G_{l_{k-1} l_k} \alpha G(\mathbf{r}_{l_k}, \mathbf{r}'). \quad (5.65)$$

This form of G^{sc} bears close resemblance to Eq. (5.44), and the same approach as was used to calculate $\langle A(\mathbf{q}, \mathbf{q}') \rangle$ can be applied here. Using the inverse Fourier transform of G as before results in

$$G^{sc}(\mathbf{r}, \mathbf{r}') = \sum_{k=0}^{\infty} \left(\frac{k_0^2}{\varepsilon_0} \right)^{k+1} \int \prod_{b=1}^{k+1} \frac{d^2 \mathbf{q}_b}{(2\pi)^2} \alpha \tilde{G}(\mathbf{q}_b; z_s, z_s) e^{i(\mathbf{q}_1 \cdot \boldsymbol{\rho} - \mathbf{q}_{k+1} \cdot \boldsymbol{\rho}')} \quad (5.66)$$

$$\times \sum_{l_1, \dots, l_k} e^{i(\mathbf{q}_2 - \mathbf{q}_1) \cdot \boldsymbol{\rho}_{l_1}} e^{i(\mathbf{q}_3 - \mathbf{q}_2) \cdot \boldsymbol{\rho}_{l_2}} e^{i(\mathbf{q}_4 - \mathbf{q}_3) \cdot \boldsymbol{\rho}_{l_3}} \dots e^{i(\mathbf{q}_k - \mathbf{q}_{k-1}) \cdot \boldsymbol{\rho}_{l_{k-1}}} e^{i(\mathbf{q}_{k+1} - \mathbf{q}_k) \cdot \boldsymbol{\rho}_{l_k}}. \quad (5.67)$$

Applying the same no loop approximation as used in deriving Eq. (5.50) to the exponential sum, the mean of Eq. (5.66) is

$$\langle G^{sc}(\mathbf{r}, \mathbf{r}') \rangle = \int \left[1 - n\alpha \frac{k_0^2}{\varepsilon_0} \tilde{G}(\mathbf{q}; z_s, z_s) \right]^{-1} \alpha \frac{k_0^2}{\varepsilon_0} \tilde{G}(\mathbf{q}) e^{i\mathbf{q} \cdot (\boldsymbol{\rho} - \boldsymbol{\rho}')} \frac{d^2 \mathbf{q}}{(2\pi)^2} \quad (5.68)$$

$$\langle \tilde{G}_{sc}(\mathbf{q}; z_s, z_s) \rangle = \left[1 - n\alpha \frac{k_0^2}{\varepsilon_0} \tilde{G}(\mathbf{q}; z_s, z_s) \right]^{-1} \alpha \frac{k_0^2}{\varepsilon_0} \tilde{G}(\mathbf{q}; z_s, z_s). \quad (5.69)$$

Physically, $\langle G^{sc}(\mathbf{r}, \mathbf{r}') \rangle$ describes propagation via all possible scattering paths from \mathbf{r}' to \mathbf{r} averaged over the scatterer positions. This turns out to be a valuable quantity in calculating σ_1 .

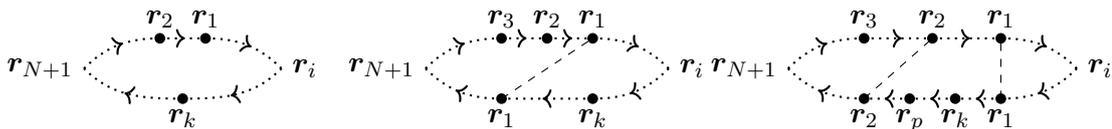


Figure 5.4: Diagrammatic representations of some examples of the interference of scattering paths $\mathcal{L}_i \mathcal{L}_j^*$ contributing to σ_1^2 . Dotted lines with a forward arrow represent a factor of G (freely) propagating the field between two scatterers, while those with a backward arrow represent G^* in the conjugated path. Intermediate scatterer positions are integrated over upon averaging. Dashed lines show shared scatterers between the two paths. In the left example, the paths share no scatterers and are therefore statistically independent. For these paths $\langle \mathcal{L}_i \mathcal{L}_j \rangle = \langle \mathcal{L}_i \rangle \langle \mathcal{L}_j \rangle$ and their contribution is included in $|\langle G^{sc} \rangle|^2$. The other two paths share scatterers and therefore are not accounted for in the Drude-Boltzmann approximation.

Ladder Approximation

The calculation of σ_1 involves considering the average of the interference between all scattering paths from the $N + 1$ th scatterer to the i th scatterer, given by

$$\begin{aligned} \sigma_1^2 &= \frac{N}{2} \langle \mathcal{R}_i \mathcal{R}_i^* \rangle \\ &= \frac{n}{2} \int \langle G^{sc}(\mathbf{r}_i, \mathbf{r}_{N+1}) G^{sc*}(\mathbf{r}_i, \mathbf{r}_{N+1}) \rangle d^2 \boldsymbol{\rho}_i \end{aligned} \quad (5.70)$$

where the integral over $\boldsymbol{\rho}_i$ comes from averaging over the i th scatterer's position. One cannot directly insert Eq. (5.68) here, since $\langle |G^{sc}(\mathbf{r}_i, \mathbf{r}_{N+1})|^2 \rangle \neq |\langle G^{sc}(\mathbf{r}_i, \mathbf{r}_{N+1}) \rangle|^2$ (though this approximation is used in some contexts and is sometimes referred to as the Drude-Boltzmann approximation [24]). Since G^{sc} contains all scattering paths between the start and end point, the averaged product can be thought of as $\sum_{i,j \in \text{paths}} \langle \mathcal{L}_i \mathcal{L}_j^* \rangle$ where \mathcal{L}_i is the contribution to G^{sc} from one scattering path. A given scattering path visits a certain subset of scatterers. In the product $\langle \mathcal{L}_i \mathcal{L}_j^* \rangle$, any scatterers visited in \mathcal{L}_i but not in \mathcal{L}_j can be averaged over independently from \mathcal{L}_j^* and vice versa for scatterers appearing in \mathcal{L}_j^* but not in \mathcal{L}_i . Fig. 5.4 demonstrates diagrammatic representations of some examples of these paths with shared and non-shared scatterers. In this coupled dipole model, nodes corresponding to scattering events introduce a factor of $\alpha k_0^2 / \epsilon_0$, and intermediate scatterer positions are summed over (since any of the N scatterers can be responsible for the scattering event) and integrated over (from averaging over position) so a node introduces $\sum_i \int \alpha_i k_0^2 / \epsilon_0 d^2 \boldsymbol{\rho}_i / L^2$. The connecting lines are Green's functions as usual in diagrammatic scattering theory. The contribution to $\langle G^{sc}(\mathbf{r}_i, \mathbf{r}_{N+1}) G^{sc*}(\mathbf{r}_i, \mathbf{r}_{N+1}) \rangle$ from interference of

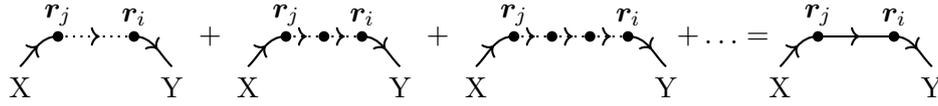


Figure 5.5: Diagrammatic representation of how the use of G^{sc} , represented by a solid line with a forward arrow (backward arrow denotes G^{sc*}) to propagate between scatterers accounts for averaging over all scatterer positions not appearing in both \mathcal{L}_i and \mathcal{L}_j . Intermediate scatterer positions are not labelled, but are integrated over. Replacing a dotted unperturbed Green's function line that connects one part of a diagram (represented by X in the figure) with the rest of the diagram (represented by Y) with a solid line representing G^{sc} in a diagram accounts for the whole set of diagrams visiting intermediate scatterers between X and Y, provided those intermediate scatterers do not appear in the rest of the diagram (i.e. in X and Y). As such, connecting scattering events with solid lines and using G^{sc} , one only needs to consider shared scatterers between two paths.

paths with no shared scatterers is contained within $\langle G^{sc}(\mathbf{r}_i, \mathbf{r}_{N+1}) \rangle \langle G^{sc*}(\mathbf{r}_i, \mathbf{r}_{N+1}) \rangle$. By using $G^{sc}(\mathbf{r}_a, \mathbf{r}_b)$ to propagate between the shared scatterers a and b in both \mathcal{L}_i and \mathcal{L}_j rather than $G(\mathbf{r}_a, \mathbf{r}_b)$, one accounts for all possible scattering between the shared scatterers. This principle is illustrated by the diagrams of Fig. 5.5. This allows the average $\langle G^{sc}(\mathbf{r}_i, \mathbf{r}_{N+1}) G^{sc*}(\mathbf{r}_i, \mathbf{r}_{N+1}) \rangle$ to be represented by a sum over all interferences between paths where every scatterer is shared, with diagrammatic representation of some examples of these terms shown in Fig. 5.6. There still remain a large variety of different diagrams and summing all of them remains challenging. In the case when the scattering mean free path is sufficiently large that the phase is randomised between scattering events, i.e. $k_{\text{SPP}} l_s \gg 1$, it is possible to ignore certain diagrams. Two paths which visit the shared scatterers in a different order will have a random phase difference owing the phase randomisation between scattering events and thus the net contribution from these paths will tend to average out. This leaves only the contribution from the interference of paths visiting the shared scatterers in the same order, and those visiting the same scatterers in reverse order. Taking only the first class of diagrams that visit the scatterers in the same order,

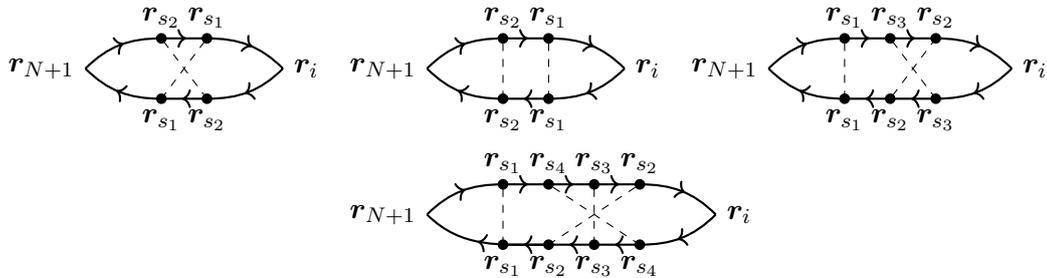


Figure 5.6: Examples of different kinds of path interferences contributing to σ_1^2 , where every scatterer appears in the both paths.

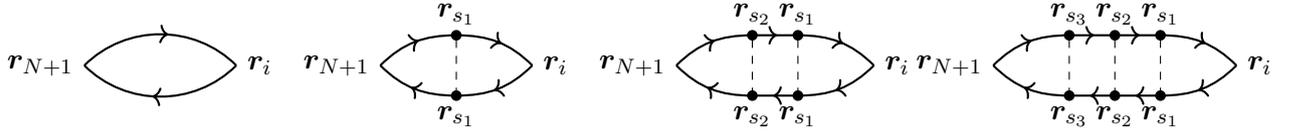


Figure 5.7: The first four ladder diagrams which give rise to Eq. (5.71). Solid lines with a forward arrow represent G^{sc} while backwards arrows denote G^{sc*} , intermediate scatterer positions are integrated over. Dashed lines connect shared scatterers.

represented diagrammatically in Fig 5.7, is known as the ladder approximation [24]. Within this approximation, one finds

$$\begin{aligned} \sigma_1^2 &= \frac{n}{2} \int \langle G^{sc}(\mathbf{r}_i, \mathbf{r}_{N+1}) \rangle \langle G^{sc*}(\mathbf{r}_i, \mathbf{r}_{N+1}) \rangle \frac{d^2 \boldsymbol{\rho}_i}{L^2} \\ &+ \frac{n}{2} \sum_{k=1}^{\infty} \sum_{s_1, \dots, s_k=1}^N \int |\langle G^{sc}(\mathbf{r}_i, \mathbf{r}_{s_1}) \rangle \langle G^{sc}(\mathbf{r}_{s_1}, \mathbf{r}_{s_2}) \rangle \langle G^{sc}(\mathbf{r}_{s_2}, \mathbf{r}_{s_3}) \rangle \dots \langle G^{sc}(\mathbf{r}_{s_k}, \mathbf{r}_{N+1}) \rangle|^2 \\ &\quad \times \frac{d^2 \boldsymbol{\rho}_i}{L^2} \frac{d^2 \boldsymbol{\rho}_{s_1}}{L^2} \frac{d^2 \boldsymbol{\rho}_{s_2}}{L^2} \dots \frac{d^2 \boldsymbol{\rho}_{s_k}}{L^2}, \end{aligned} \quad (5.71)$$

where \mathbf{r}_{s_j} are the shared scatterer positions being averaged over. The first term corresponds to the interference of all paths with no shared scatterers (i.e. the Drude-Boltzmann contribution) while each term in the sum over k is the contribution from the interference between paths \mathcal{L}_i from G^{sc} visiting k scatterers s_1, s_2, \dots, s_k and a conjugated path \mathcal{L}_j^* from G^{sc*} visiting the same k scatterers in the same order. Introducing the Fourier space averaged Green's function \tilde{G}^{sc} gives

$$\begin{aligned} \sigma_1^2 &= \frac{n}{2} \int \left| \langle \tilde{G}_{sc}(\mathbf{q}) \rangle \right|^2 \frac{d^2 \mathbf{q}}{(2\pi)^2} \\ &+ \frac{n}{2} \sum_{k=1}^{\infty} \sum_{s_1, \dots, s_k=1}^N \int \langle \tilde{G}_{sc}(\mathbf{q}_1) \rangle \langle \tilde{G}_{sc}(\mathbf{q}_2) \rangle \dots \langle \tilde{G}_{sc}(\mathbf{q}_k) \rangle \langle \tilde{G}_{sc}^*(\mathbf{q}'_1) \rangle \langle \tilde{G}_{sc}^*(\mathbf{q}'_2) \rangle \dots \langle \tilde{G}_{sc}^*(\mathbf{q}'_k) \rangle \\ &\quad \times e^{i(\mathbf{q}_1 - \mathbf{q}'_1) \cdot (\boldsymbol{\rho}_i - \boldsymbol{\rho}_{s_1})} e^{i(\mathbf{q}_2 - \mathbf{q}'_2) \cdot (\boldsymbol{\rho}_{s_1} - \boldsymbol{\rho}_{s_2})} \dots e^{i(\mathbf{q}_k - \mathbf{q}'_k) \cdot (\boldsymbol{\rho}_{s_k} - \boldsymbol{\rho}_{N+1})} d^2 \boldsymbol{\rho}_i \frac{d^2 \boldsymbol{\rho}_{s_1}}{L^2} \frac{d^2 \boldsymbol{\rho}_{s_2}}{L^2} \dots \frac{d^2 \boldsymbol{\rho}_{s_k}}{L^2} \\ &\quad \times \frac{d^2 \mathbf{q}_1}{(2\pi)^2} \frac{d^2 \mathbf{q}'_1}{(2\pi)^2} \frac{d^2 \mathbf{q}_2}{(2\pi)^2} \frac{d^2 \mathbf{q}'_2}{(2\pi)^2} \dots \frac{d^2 \mathbf{q}_k}{(2\pi)^2} \frac{d^2 \mathbf{q}'_k}{(2\pi)^2}. \end{aligned} \quad (5.72)$$

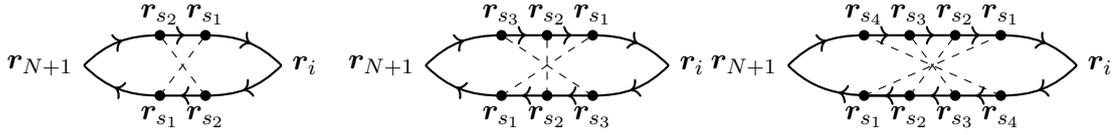


Figure 5.8: The three lowest order examples of maximally crossed diagrams, which are responsible for coherent back scattering and weak localisation effects.

Integrating over each position gives a factor $(2\pi)^2\delta(\mathbf{q}_i - \mathbf{q}'_i)/L^2$, while the sum over scatterers \sum_{s_1, \dots, s_k} reduces to a factor of N^k . Therefore, Eq. (5.71) becomes

$$\begin{aligned} \sigma_1^2 &= \frac{n}{2} \int \left| \langle \tilde{G}_{sc}(\mathbf{q}) \rangle \right|^2 \frac{d^2\mathbf{q}}{(2\pi)^2} + \frac{n}{2} \sum_{k=1}^{\infty} \frac{N^k}{L^{2k}} \left(\int \left| \langle \tilde{G}_{sc}(\mathbf{q}) \rangle \right|^2 \frac{d^2\mathbf{q}}{(2\pi)^2} \right)^{k+1} \\ &= \frac{1}{2} \frac{n\mathcal{J}}{1 - n\mathcal{J}}, \end{aligned} \quad (5.73)$$

where the integral $\mathcal{J} = \int \left| \langle \tilde{G}_{sc}(\mathbf{q}) \rangle \right|^2 d^2\mathbf{q}/(2\pi)^2$ has been defined and the geometric series in $n\mathcal{J}$ has been assumed to converge. The divergence when $n\mathcal{J} = 1$ is not physical, but arises due to the approximations, and the ladder approximation breaks down as the scatterer density approaches this divergence. The ladder approximation result ignores the interference between exactly reversed paths, represented diagrammatically in Fig. 5.8. These maximally crossed diagrams (also referred to as Cooperons) do not separate into separate integrals which can all be summed like the ladder diagram, and no analytic contribution to σ_1 from these diagrams has been calculated (or any other class of diagrams) in this thesis. It is, however, worth considering a few features of them to aid the discussion and physical understanding of results. Firstly, while the reciprocity means the phase and amplitude acquired by the field is the same in the reversed path $s_k \rightarrow s_{k-1} \rightarrow \dots \rightarrow s_2 \rightarrow s_1$ as in the path $s_1 \rightarrow s_2 \rightarrow \dots \rightarrow s_{k-1} \rightarrow s_k$, there is a phase difference due to the difference in propagating from \mathbf{r}_{N+1} to the first scatterer (at \mathbf{r}_{s_1} and \mathbf{r}_{s_k}) and from the last scatterer (at \mathbf{r}_{s_k} and \mathbf{r}_{s_1}) to \mathbf{r}_i . If the $|\mathbf{r}_i - \mathbf{r}_{N+1}|$ is large, the dephasing between the reciprocal paths in the incoming and outgoing legs means this contribution averages out and can be ignored. When $\mathbf{r}_i = \mathbf{r}_{N+1}$, the contribution from the interference of a path with its reciprocal path is exactly equal to the contribution of the interference of the path with itself (i.e. the ladder diagram contribution). This means that $\langle |G^{sc}(\mathbf{r}, \mathbf{r})|^2 \rangle$ is twice that of the ladder approximation result. This result is analogous to coherent backscattering and is

the simplest set of diagrams contributing to weak localisation. In calculating σ_1 , however, the quantity required is $\int \langle |G^{sc}(\mathbf{r}, \mathbf{r}_{N+1})|^2 \rangle d^2 \boldsymbol{\rho}_i$, and thus, provided this coherent backscattering peak around $\boldsymbol{\rho}_i \approx \boldsymbol{\rho}_{N+1}$ is sufficiently narrow, the maximally crossed diagrams' contribution to σ_1 is negligible compared to the ladder diagram contribution. Beyond the maximally crossed diagrams, there are yet more contributions involving partially reversed scattering paths that, at short mean free paths, can give rise to Anderson localisation. Anderson localisation would have significant consequences for $\langle |\gamma_1| \rangle$, as $\langle |G^{sc}|^2 \rangle$ is confined to decay over the localisation length ξ , and thus the integral in Eq. (5.70) is effectively over a smaller area $\sim \xi^2$. While the contribution of diagrams beyond the ladder approximation to $\langle |\gamma_1| \rangle$ are not calculated analytically in this thesis, they still play a role in the enhancement factor statistics.

A notable feature of Eq. (5.70) (and thus any analytic approximation to it such as the ladder approximation) is that σ_1 is independent of the observation wavevector \mathbf{k}_\parallel . This should be contrasted with Eq. (5.53), which does depend on \mathbf{k}_\parallel . As a result, while the statistics of γ_1 do depend on the observation point, the mean amplitude of this enhancement factor, when the random phasor sum model applies, is independent of observation position.

Mean Amplitude of γ_3

The analysis of the mean amplitude was applied to $\langle |\gamma_1| \rangle$, however, it can also be applied to γ_3 in the scalar case, when the illuminating field is a plane wave. In this case, $S_3 = \sum_{j=1}^N \mathcal{U}_j$ with $\mathcal{U}_j = (k_0^2/\varepsilon_0) \exp[i\mathbf{k}_{in} \cdot (\mathbf{r}_j - \mathbf{r}_{N+1})] \sum_{i=1}^N \alpha G_{N+1,i} M_{ij}^{-1}$. Applying the same phasor sum logic as in the γ_1 case, the Rician mean of Eq. (5.61) applies but with $\sigma_3^2 = N \langle |\mathcal{U}_i|^2 \rangle / 2$ instead of σ_1 . Under the assumptions already made (i.e. identical scatterers at $z_i = z_s$) $M_{ij}^{-1} \alpha G_{j,N+1} = \alpha G_{N+1,j} M_{ji}^{-1}$ and therefore, analogous to Eq. (5.64), one has

$$\mathcal{U}_i = e^{i\mathbf{k}_{in} \cdot (\mathbf{r}_i - \mathbf{r}_{N+1})} G^{sc}(\mathbf{r}_{N+1}, \mathbf{r}_i). \quad (5.74)$$

Unlike Eq. (5.64) however, the presence of loss means $\mathbf{k}_{in} \in \mathbb{C}$ and therefore σ_3 is given by

$$\sigma_3^2 = \frac{n}{2} \int \langle G^{sc}(\mathbf{r}_{N+1}, \mathbf{r}_i) G^{sc*}(\mathbf{r}_{N+1}, \mathbf{r}_i) \rangle e^{-2\text{Im}[k_{in}]x_i} d^2 \boldsymbol{\rho}_i, \quad (5.75)$$

where, without loss of generality, the x_i -axis has been aligned parallel to \mathbf{k}_{in} . In the lossless case, $\text{Im}[k_{\text{in}}] = 0$ and therefore $\sigma_3 = \sigma_1$ and $\langle |\gamma_3| \rangle = \langle |\gamma_1| \rangle$. More generally, when the incident field has a position dependent amplitude $A(\mathbf{r})$ and phase $\phi_0(\mathbf{r})$, so $E_{0,z} = A(\mathbf{r})e^{i\phi_0(\mathbf{r})}$, the distribution width parameter is

$$\sigma_3^2 = \frac{n}{2} \int \langle G^{sc}(\mathbf{r}_{N+1}, \mathbf{r}_i) G^{sc*}(\mathbf{r}_{N+1}, \mathbf{r}_i) \rangle \left| \frac{A(\boldsymbol{\rho}_i, z_s)}{A(\boldsymbol{\rho}_{N+1}, z_s)} \right|^2 d^2 \boldsymbol{\rho}_i. \quad (5.76)$$

5.4 Scalar SPP Scattering Results

The analytic results so far have been derived with relatively few assumptions about the background. In order to study the multiple scattering enhancements in more depth within an SPP scattering context, it is important to introduce an explicit form for G . The Green's tensor can be split into contributions from different modes [213], including SPPs as

$$G(\mathbf{r}, \mathbf{r}') = G_{\text{SPP}}(\mathbf{r}, \mathbf{r}') + G_s(\mathbf{r}, \mathbf{r}') + G_p(\mathbf{r}, \mathbf{r}') \quad (5.77)$$

where G_{SPP} is the contribution from SPP modes and $G_{s,p}$ is the contribution from s - and p - polarised plane wave modes. For a single metal-dielectric interface, the SPP contribution has been calculated analytically above the surface ($z, z' > 0$) in the limit of large transverse separations $|\boldsymbol{\rho} - \boldsymbol{\rho}'| \gg \lambda_0$ and small heights $z, z' \ll \lambda_0$ to be [213, 353, 354]

$$G_{\text{SPP}}(\mathbf{r}, \mathbf{r}') = iA_0 H_0^{(1)}(k_{\text{SPP}}|\boldsymbol{\rho} - \boldsymbol{\rho}'|) e^{-\kappa_d(z+z')} (\hat{\mathbf{z}} - ia\hat{\boldsymbol{\rho}})(\hat{\mathbf{z}} + ia\hat{\boldsymbol{\rho}})^T \quad (5.78)$$

where $A_0 = ak_{\text{SPP}}/[2(1 - a^4)(1 - a^2)]$, $a = (\varepsilon_d/(-\varepsilon_m))^{1/2}$ and $H_n^{(1)}(x)$ is the n th order Hankel function of the first kind. In addition, for points close to the surface, SPP modes are dominant and $G \approx G_{\text{SPP}}$ [353]. Furthermore, while this form of G_{SPP} has been derived for a single metal-dielectric interface, the SPP modes take the same functional form in the $z > 0$ halfspace even for more general multilayer structures (see Eq. (3.5)), and therefore G_{SPP} takes the same form as Eq. (5.78) for a more general layered structure, though the dependence of the parameters A_0 , a , k_{SPP} and κ_d on the physical parameters differs from the single interface case.

Since the off diagonal $\hat{\mathbf{z}}\hat{\boldsymbol{\rho}}^T$ and $\hat{\boldsymbol{\rho}}\hat{\mathbf{z}}^T$ terms in Eq. (5.78) are proportional to a (and the $\hat{\boldsymbol{\rho}}\hat{\boldsymbol{\rho}}^T$ component scales as a^2), in the limit $|a| \ll 1$, $G_{\text{SPP}} \propto \hat{\mathbf{z}}\hat{\mathbf{z}}^T$ projects all sources onto their z component. As a result, one only needs to consider the z components E_z and p_z of the field and dipole moments in the coupled dipole model. This allows the treatment of SPP scattering within a scalar model [354, 355], and thus the scalar forms of γ_i can be used. The scalar Green's function follows from the $\hat{\mathbf{z}}\hat{\mathbf{z}}^T$ component of Eq. (5.78) as

$$G_{\text{SPP,scal}}(\mathbf{r}, \mathbf{r}') = iA_0 H_0^{(1)}(k_{\text{SPP}}|\boldsymbol{\rho} - \boldsymbol{\rho}'|)e^{-\kappa_d(z+z')}. \quad (5.79)$$

Within this scalar model, the scalar value of α appearing in the coupled dipole equations and hence enhancement factors corresponds to α_{zz} , such that $p_z = \alpha_{zz}E_z$. The model allows for calculation of an elastic scattering cross-section $\sigma_{\text{SPP}} = 4|\mu|^2/k_{\text{SPP}}'$, where $\mu = \alpha_{zz}(k_0^2/\varepsilon_0)A_0 \exp[-2\kappa_d z_s]$ [354, 355]. From now on within the scalar model, α_{zz} shall be referred to simply as α . The transverse Fourier space scalar Green's function follows from the Fourier transform of Eq. (5.79) and is

$$\tilde{G}_{\text{SPP,scal}}(\mathbf{q}; z, z') = \frac{-4A_0 e^{-\kappa_d(z+z')}}{k_{\text{SPP}}^2 - q^2}. \quad (5.80)$$

With the scalar Green's function, assuming $z_{N+1} = z_s$ and an incident (lossy) plane wave $E_{0,z}(x, z) \propto \Theta(x) \exp(ik_{\text{SPP}}x - \kappa_d z)$, Eqs. (5.53)–(5.55) result in

$$\langle \gamma_1(k_{\parallel}) \rangle = \frac{k_{\text{SPP}}^2 - k_{\parallel}^2}{k_{\text{SPP}}^2 - k_{\parallel}^2 + 4n\mu} \quad (5.81)$$

$$\langle \gamma_3(x_{N+1}) \rangle = -\frac{2n\mu \exp\left[i(\tilde{k}(n) - k_{\text{SPP}})x_{N+1}\right]}{\left[k_{\text{SPP}} - \tilde{k}(n)\right] \tilde{k}(n)} \quad (5.82)$$

$$\langle S_2 \rangle = -\frac{\mu_{N+1}}{\pi} \log\left(1 + \frac{4n\mu}{k_{\text{SPP}}^2}\right) \quad (5.83)$$

where $\tilde{k}(n) = (k_{\text{SPP}}^2 + 4n\mu)^{1/2}$ has been defined and μ_{N+1} is defined analogously with α_{N+1} replacing α . While Eq. (5.81) follows directly from the substitution of Eq. (5.80) into Eq. (5.58), evaluation of the integrals in Eqs. (5.54) and (5.55) to arrive at Eqs (5.82) and (5.83) is

slightly more involved and is presented in Appendix A. In addition, the scalar model allows use of the scalar theory for $\langle |\gamma_{1,3}| \rangle$. The perturbed (Fourier space) Green's function is

$$\tilde{G}^{sc}(\mathbf{q}; z_s, z_s) = \frac{-4\mu}{k_{\text{SPP}}^2 - q^2 + 4n\mu}. \quad (5.84)$$

This result demonstrates that one effect of multiple scattering is to shift the effective wavevector $k_{\text{SPP}} \rightarrow \tilde{k}(n)$, as was discussed in Section 5.1.1. From Eq. (5.84), the width parameter σ_1 can be calculated by evaluating the integral

$$\begin{aligned} \mathcal{J} &= \int \frac{16|\mu|^2}{(k_{\text{SPP}}^2 - q^2 + 4n\mu)(k_{\text{SPP}}^{*2} - q^2 + 4n\mu^*)} \frac{d^2\mathbf{q}}{(2\pi)^2} \\ &= \frac{4|\mu|^2}{\text{Im}[k_{\text{SPP}}^2 + 4n\mu]} \\ &= \frac{\sigma_{\text{SPP}}}{L_{\text{SPP}}^{-1} + 4n \text{Im}[\mu]/k'_{\text{SPP}}}, \end{aligned} \quad (5.85)$$

resulting in

$$\sigma_1^2 = \frac{1}{2} \frac{l_s^{-1}}{L_{\text{SPP}}^{-1} + 4n \text{Im}[\mu]/k'_{\text{SPP}} - l_s^{-1}}. \quad (5.86)$$

From Eq. (5.86), $\langle |\gamma_1| \rangle$ can be calculated via Eq. (5.61). Note that the divergence of this result occurs when $l_s = L_{\text{abs}}$ where L_{abs} is an effective absorption length defined by $L_{\text{abs}}^{-1} = L_{\text{SPP}}^{-1} + 4n \text{Im}[\mu]/k'_{\text{SPP}}$, which accounts for absorption of the SPP in the metal substrate (described by L_{SPP}) and also absorption by the scatterers.

5.4.1 Numerical Results

In order to further understand the enhancement factor statistics, and to assess the validity of the derived analytic results, numerical Monte Carlo simulations were performed using the scalar SPP scattering model. A realisation of randomly distributed scatterers was generated and the corresponding scattered field calculated by solving Eq. (5.6) with G_{ij} calculated using Eq. (5.79). The simulation was repeated with an additional particle from which the field perturbation and individual enhancement factors could be calculated. The scatterer positions were randomly

generated with a uniform distribution in a 2D box of sides L , while all scatterers (including the additional one) were at the same height z_s and had the same polarisability α . While the simulations were performed with $\alpha_{N+1} = \alpha$ for speed of computation, in a sensing context generally α_{N+1} will be different and typically smaller due to being a small biological particle. As described in Section 5.2.2, in the scalar case γ_1 and γ_3 are independent of the value of α_{N+1} while γ_3 for any value of α_{N+1} for any given realisation can be calculated from the $\alpha_{N+1} = \alpha$ case using the scaling $S_2 \propto \alpha_{N+1}$. Thus, there is no need to perform simulations at different values of α_{N+1} . In order to vary the scatterer density n (and therefore the mean free path l_s and level of multiple scattering), L was varied. Unless otherwise stated, averages were calculated from 50,000 realisations.

Low Loss Single Interface

The first set of simulations applies to a single gold-water interface at a free space wavelength of 650nm. In this case $\varepsilon_d = 1.77$ and $\varepsilon_m = -13.68 + 1.04i$ [280], meaning that $k_{\text{SPP}} = (1.42 + 0.008i)k_0$ and $L_{\text{SPP}} = 9.9\lambda_0$. The number of scatterers was fixed at $N = 700$, while the box size was varied between $L = 9.3\lambda_0$ and $L = 118\lambda_0$. Fig. 5.9 shows the complex means and mean amplitudes of the enhancement factors as a function of density for an observation point at 70° to the surface normal in the backward direction ($\mathbf{k}_{\parallel} = -\varepsilon_d^{1/2}k_0 \sin(70^\circ)\hat{\mathbf{x}}$) with an incident plane wave SPP $E_{0,z}(x, z) = \Theta(x) \exp(ik_{\text{SPP}}x - \kappa_d z)$ and assuming a polarisability α_{g1} corresponding to a gold sphere of radius $R_{g1} = 40\text{nm}$ sitting on the gold surface (i.e. $z_s = 40\text{nm}$). This polarisability was calculated using a quasistatic result with surface dressing included using the electrostatic approximation to the Green's tensor [354, 353] to give

$$\alpha_{zz} = \frac{\alpha_0}{1 + 2v\beta} \quad (5.87)$$

where $v = (\varepsilon_d - \varepsilon_m)(\varepsilon_p - \varepsilon_d)/[(\varepsilon_d + \varepsilon_m)(\varepsilon_p + 2\varepsilon_d)]$, $\beta = (R_{g1}/(2z_s))^3$ and α_0 is the homogeneous space quasistatic result from Eq. (2.6). The additional scatterer was added at a position $\mathbf{r}_{N+1} = (0, 0, z_s)$. The analytic theory results of Eqs. (5.81) to (5.83) are also shown. These results are a good description at lower scatterer density, before breaking down at shorter mean

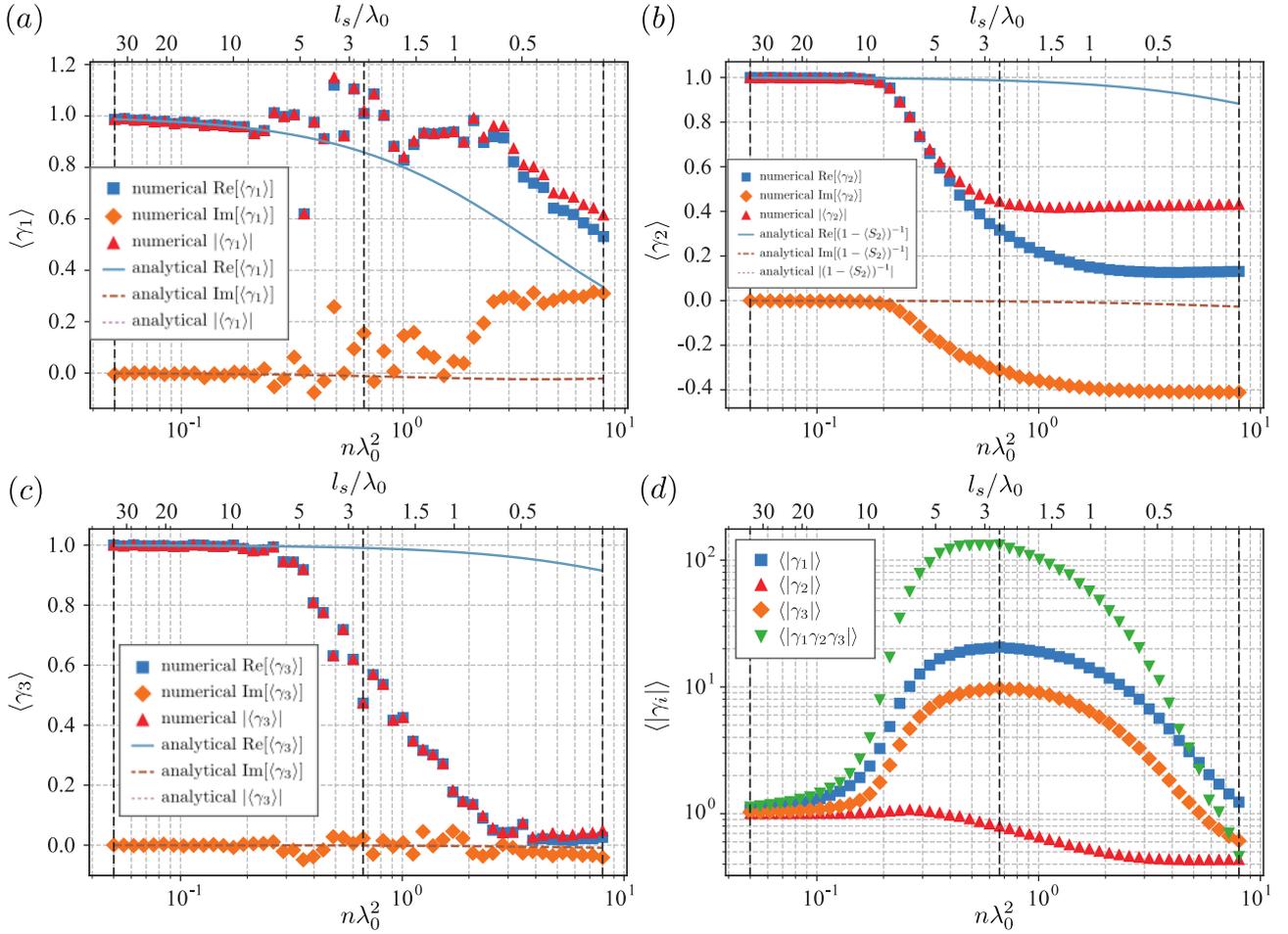


Figure 5.9: Complex means (a) $\langle \gamma_1 \rangle$, (b) $\langle \gamma_2 \rangle$, (c) $\langle \gamma_3 \rangle$ and (d) the mean amplitudes of the individual enhancement factor $\langle |\gamma_i| \rangle$ and the total enhancement factor $\langle |\gamma_1 \gamma_2 \gamma_3| \rangle$ for a gold-water interface at $\lambda_0 = 650\text{nm}$ with 40nm gold sphere scatterers. Vertical dashed lines indicate densities used in Fig. 5.10. *Figure reprinted (with alteration) with permission from Ref.[288] © 2021 American Physical Society.*

free paths $l_s \sim 5\lambda_0$, at which point the loop paths ignored in the derivation become significant. Note the mean free path shown in the plots is calculated via $l_s = (n\sigma_{\text{SPP}})^{-1}$ and therefore strictly only accurate at lower densities. The value of l_s shown on the plots is intended only as an alternative parameterisation to n .

Importantly, Fig. 5.9(d) shows the total amplitude enhancement can be as large as $\sim 10^2$, and that there is an optimum density at which this field amplitude enhancement is maximised. More detailed understanding of the enhancement factor statistics is revealed in looking at the probability distribution over the complex plane, shown in Fig. 5.10. While, as indicated by both the analytical and numerical complex means, the centre of distributions for γ_1 and γ_3 migrate from 1 to 0 with increasing density, the width of the distributions increases initially with density

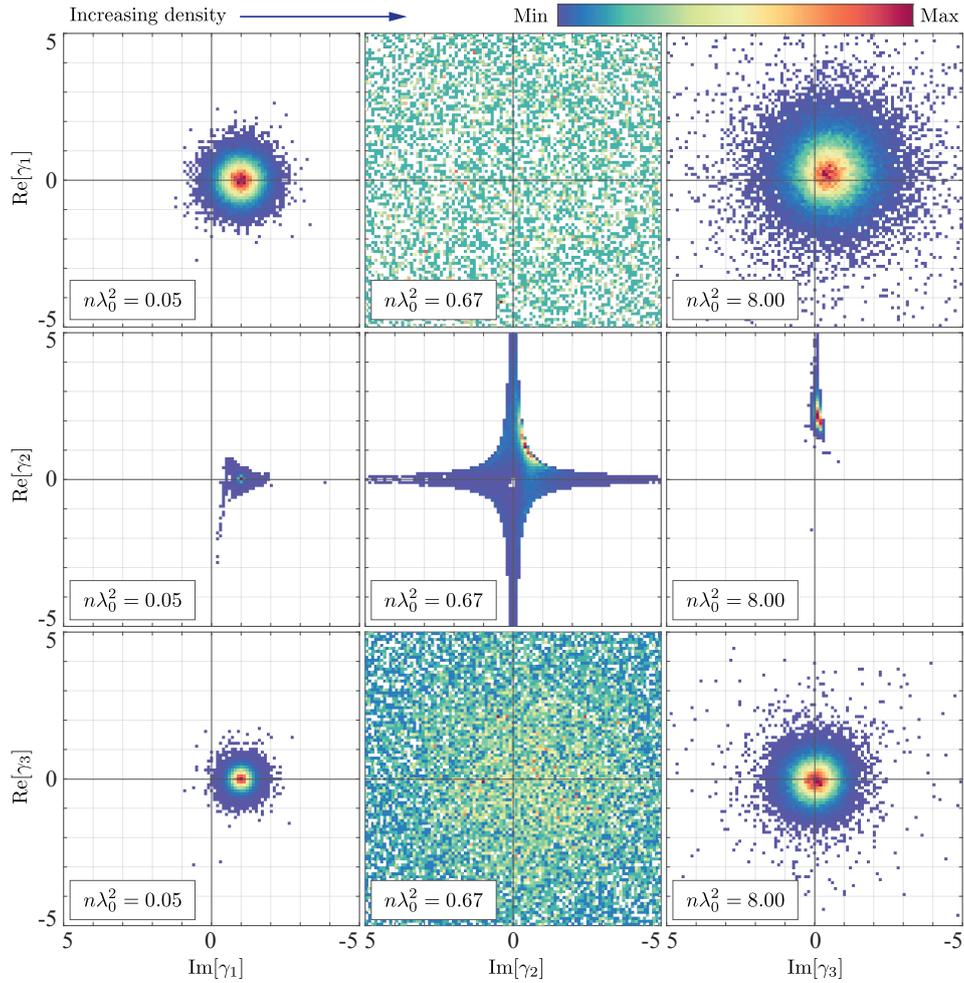


Figure 5.10: The relative frequency of γ_1 (top row), γ_2 (middle row) and γ_3 (bottom row) over the complex plane for scatterer densities $n\lambda_0^2 = 0.05$ (left column), 0.67 (middle column) and 8.0 (right column). *Figure reprinted with permission from Ref. [288] © 2021 American Physical Society.*

before shrinking beyond the optimum density. The width of the distribution is the primary factor controlling $\langle |\gamma_{1,3}| \rangle$, with the wide distribution at intermediate density giving rise to the peak with large $\langle |\gamma_1 \gamma_2 \gamma_3| \rangle$, while the relatively tight distributions at high and low density result in small total amplitude enhancements $\langle |\gamma_1 \gamma_2 \gamma_3| \rangle \sim 1$ and even < 1 at the highest simulated densities, indicating that multiple scattering can reduce sensitivity at very high density. The density range for which the probability distributions for $\gamma_{1,3}$ have a large variance and mean amplitude give rise to larger statistical fluctuations in the complex mean. As such, the results plotted between $n\lambda_0^2 = 0.21$ and 3.87 in Fig. 5.9 are averaged over 150,000 realisations in order to improve convergence. The shape of the $\gamma_{1,3}$ distributions appear to be close to rotationally symmetric about their centre, demonstrated by Fig.5.11(a) and (b), which show the standard

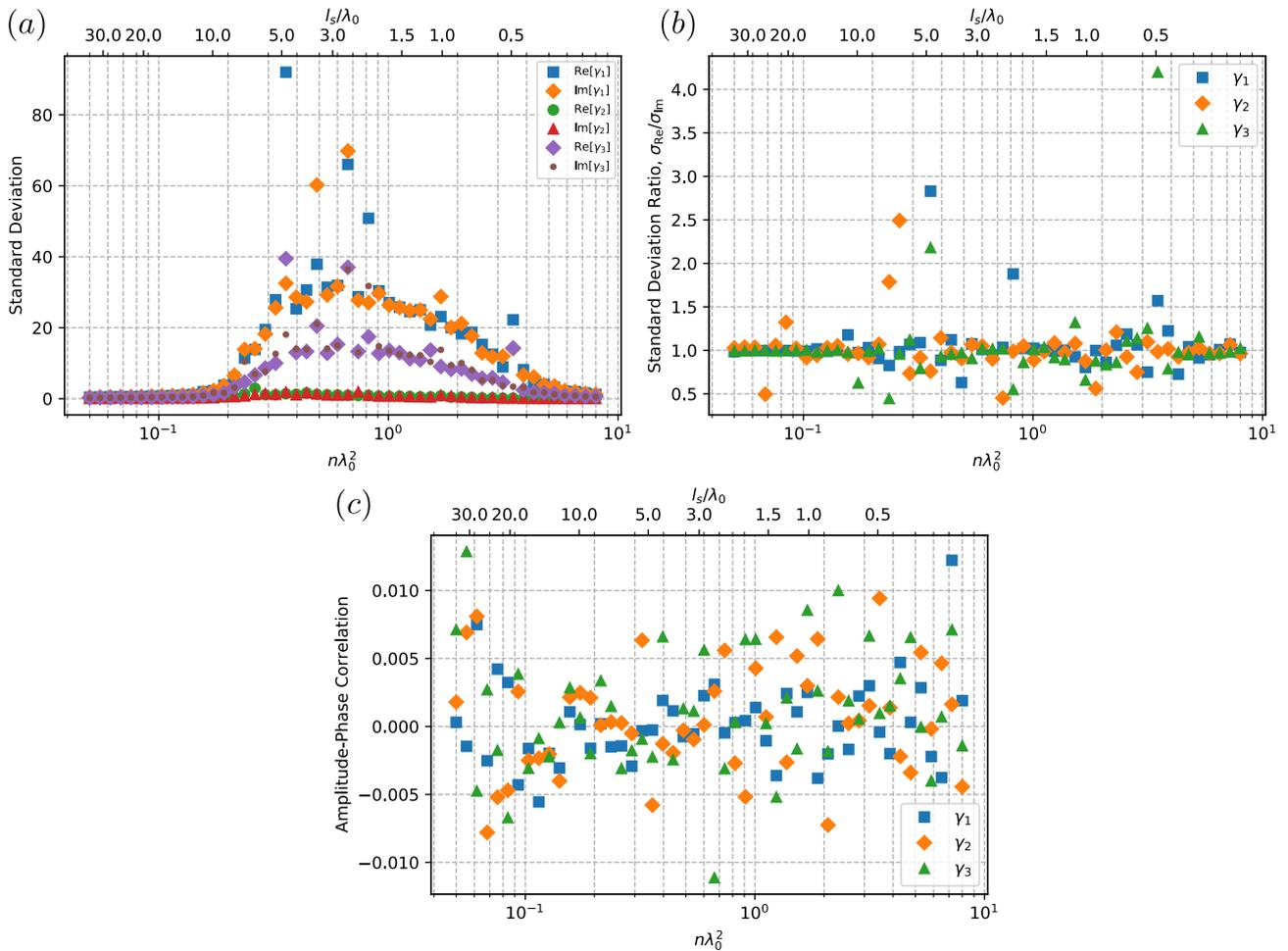


Figure 5.11: Plots showing, as a function of n , (a) the standard deviations σ_{Re} and σ_{Im} of the real and imaginary parts of $\gamma_{1,2,3}$, (b) the ratio $\sigma_{\text{Re}}/\sigma_{\text{Im}}$ and (c) the Pearson correlation coefficient of $|\gamma_{1,2,3} - \langle \gamma_{1,2,3} \rangle|$ and $\arg(\gamma_{1,2,3} - \langle \gamma_{1,2,3} \rangle)$.

deviations of the real and imaginary parts of each enhancement factor and their ratios. The real and imaginary standard deviations are generally very similar, with large discrepancies attributable to outliers, since the convergence of the standard deviation is slower than that of the mean. In addition, the correlation coefficients between $|\gamma_i - \langle \gamma_i \rangle|$ and $\arg(\gamma_i - \langle \gamma_i \rangle)$, shown in Fig. 5.11, are close to zero, which would be expected for a rotationally symmetric distribution. Rotational symmetry is consistent with the phasor sum model used to derive Eq. (5.70), which predicts a rotationally symmetric Gaussian distribution on the complex plane. Conversely, γ_2 has a more complicated distribution over the complex plane and is confined to close to the real/imaginary axes, bearing some resemblance to the previously studied eigenvalue distributions of Euclidean matrices arising in similar scattering studies [348]. Most significantly, from a sensitivity viewpoint, the very large values of $|\gamma_2|$ occur with low probability so that

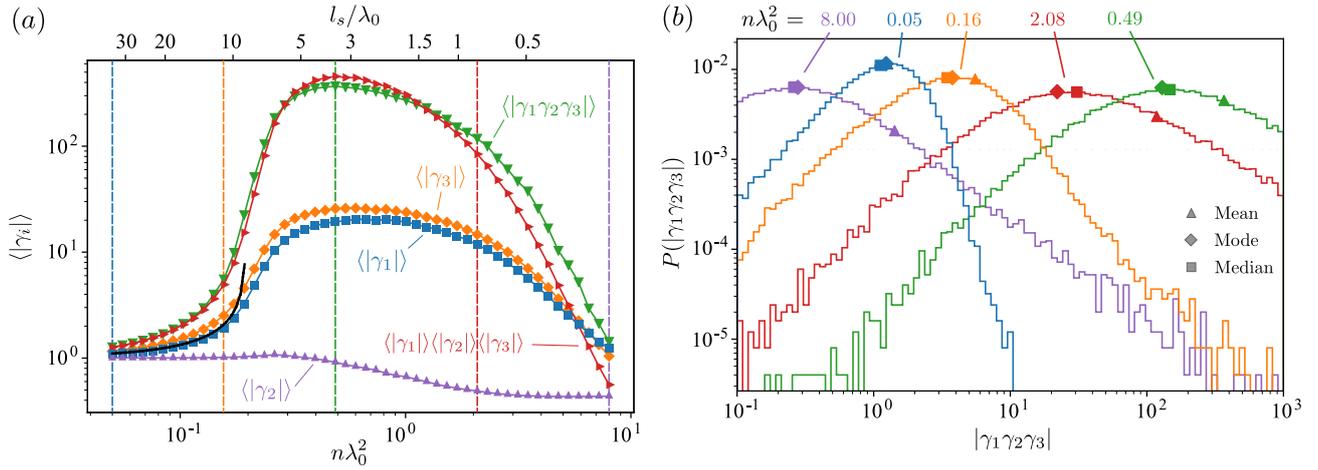


Figure 5.12: Enhancement factor amplitude statistics for an incident field $E_{0,z} = \exp(ik_{\text{SPP}}x - \kappa_d z)$, (a) the density dependence of $\langle |\gamma_i| \rangle$ and $\langle |\gamma_1\gamma_2\gamma_3| \rangle$. The black line indicates the analytic ladder approximation result for $\langle |\gamma_1| \rangle$ from Eqs. (5.61) and (5.86), plotted up to the point it diverges. (b) The relative frequency of a given total amplitude enhancement for a different scatterer densities. The vertical dashed coloured lines in (a) indicate the densities plotted in (b). *Figure reprinted (with alteration) with permission from Ref. [287] © 2021 American Chemical Society.*

$\langle |\gamma_2| \rangle$ remains close to 1 (as evidenced in Fig. 5.9(d)).

The density dependences of $\langle |\gamma_1| \rangle$ and $\langle |\gamma_3| \rangle$ are very similar, but they differ in that $\langle |\gamma_3| \rangle$ is notably lower. As indicated by Eq. (5.75), it is expected that in the lossless limit $k_{\text{SPP}}'' \rightarrow 0$, $\langle |\gamma_1| \rangle = \langle |\gamma_3| \rangle$, but here, in addition to the exponential decay from loss, there is also the fact the SPP is launched from $x = 0$ (captured in the presence of the Heaviside step function in the incident field), and thus scatterers located at $x_i < 0$ do not scatter the incident field. As a result, in the limit $k_{\text{SPP}}'' \rightarrow 0$, $\sigma_3 = \sigma_1/2$ and $\langle |\gamma_3| \rangle$ is reduced relative to $\langle |\gamma_1| \rangle$. Fig. 5.12(a) shows the equivalent of Fig. 5.9(d) in the absence of the step function (i.e. $E_{0,z} = \exp(ik_{\text{SPP}}x - \kappa_d z)$). In this case, $\langle |\gamma_3| \rangle$ and $\langle |\gamma_1| \rangle$ can be seen to be close, though the presence of loss does mean they differ slightly, with $\langle |\gamma_3| \rangle$ slightly larger. The maximum value of $\langle |\gamma_1\gamma_2\gamma_3| \rangle$ is ~ 367 at an optimal density of $n_{\text{opt,tot}} = 0.49/\lambda_0^2$ ($l_{\text{opt,tot}} = 3.51\lambda_0$).

The ladder approximation provides a good description of $\langle |\gamma_1| \rangle$ up to the point it diverges. Beyond this point, the role of the interference of reciprocal paths, captured in more complicated diagrams such as the ones shown in Fig. 5.8 becomes significant. While the breakdown of the ladder approximation means there is no analytic result describing any of the mean amplitudes $\langle |\gamma_i| \rangle$ beyond the density where $n\mathcal{J} = 1$, an understanding of the nature of the ladder approximation and the diagrams ignored in making it helps provide an explanation of

the density dependence of $\langle |\gamma_{1,3}| \rangle$ and $\langle |\gamma_1 \gamma_2 \gamma_3| \rangle$. At low densities, the ladder approximation is valid, and $\langle |\gamma_{1,3}| \rangle$ grows with density as the scatterers become closer together on average and thus the amplitude of individual scattering paths increase due to the larger values of G_{ij} (which decays with scatterer separation as $\sim |\boldsymbol{\rho}_i - \boldsymbol{\rho}_j|^{-1/2}$ for large separations). At higher densities, the interference of reciprocal paths becomes significant. As noted in the discussion of Fig. 5.8, the maximally crossed diagram contribution to $\langle |G^{sc}(\mathbf{r}, \mathbf{r}_{N+1})|^2 \rangle$ is strongly peaked around \mathbf{r}_{N+1} , and more generally these diagrams and others give rise to Anderson localisation such that $\langle |G^{sc}(\mathbf{r}, \mathbf{r}_{N+1})|^2 \rangle$ decays over a length scale given by the Anderson localisation length ξ . As a result, light scattered from the analyte particle is confined to a region within $\sim \xi$ of \mathbf{r}_{N+1} , and thus only scatterers within this region contribute to $|\gamma_1|$. Similarly, only scatterers within $\sim \xi$ can scatter light onto the analyte particle and contribute to γ_3 . The reduced number of scatterers around the analyte particle contributing to the enhancement factor results in a smaller enhancement factor with increasing density. As a result, an optimum density exists at which the average amplitude enhancement is maximised, before localisation effects begin to reduce the enhancement. Using the result for the localisation length of a 2D system of $\xi = l_s \exp(\pi \text{Re}[k_{\text{SPP}}] l_s / 2)$ [356], ξ becomes smaller than the system size for $l_s \approx 0.73\lambda_0$ in these simulations (remembering that l_s is calculated using a low density result that may not be fully accurate at higher densities). At the largest density, $l_s = 0.21\lambda_0$ which give $\xi/L = 0.42$, so that only a fraction $\sim (\xi/L)^2 \sim 0.17$ of scatterers receive intensity scattered from the additional scatterer and similarly scatter intensity onto the additional scatterer.

The density histograms shown in Fig. 5.12(b) show that low densities give a distribution of total amplitude enhancements tightly centred around ~ 1 . On the other hand, for n close to the optimum value, the probability distribution exhibits a long tail. A given scattering configuration at the optimum density $n_{\text{opt,tot}}$ consequently has a high probability of producing a significant sensitivity enhancement, but it is important to point out that the total enhancement for a given realisation with $n \approx n_{\text{opt,tot}}$ will likely be smaller than the mean total enhancement (mode \approx median \ll mean), though still typically of the order ~ 100 . Importantly, there is a small but non-negligible probability of a very large enhancement even as high as $\sim 10^3$. Such long tailed distributions are seen in the context of LDOS enhancements by random media [357, 358], which

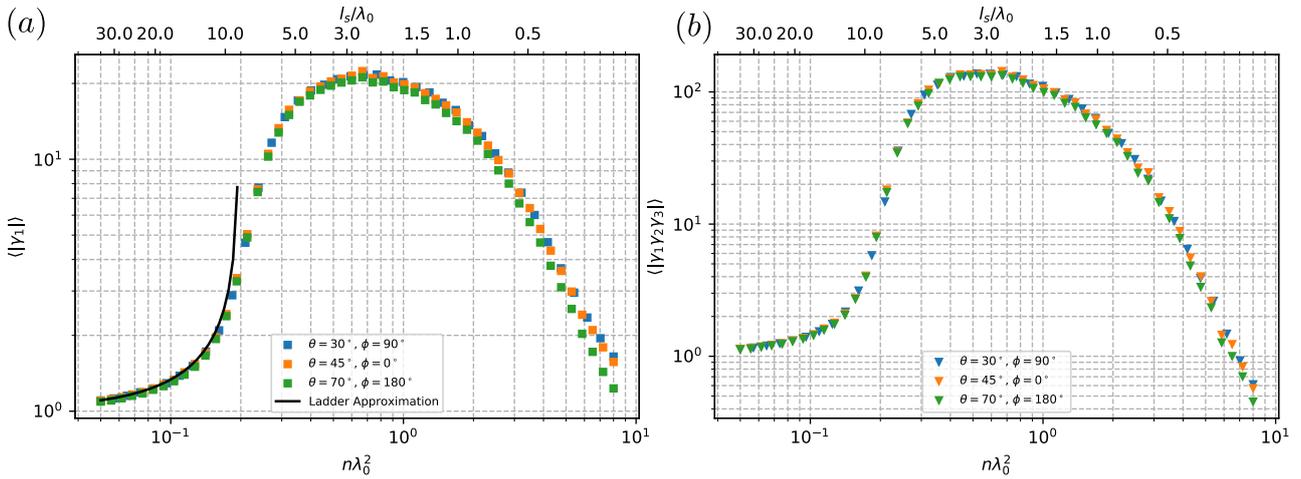


Figure 5.13: The mean amplitudes (a) $\langle |\gamma_1| \rangle$ and (b) $\langle |\gamma_1 \gamma_2 \gamma_3| \rangle$ for various different observation wavevectors $\mathbf{k}_{\parallel} = k_d(\sin \theta \cos \phi, \sin \theta \sin \phi)$. Note that $\gamma_{2,3}$ do not depend on observation position since they only deal with scattering onto the analyte particle.

describe a similar concept of enhancement of power radiated by a dipole in inhomogeneous environments. At the highest densities, the majority of realisations give $|\gamma_1 \gamma_2 \gamma_3| < 1$, reducing sensitivity, though the tail is still longer compared to the lower densities with comparable means. Consequently, even though the mean enhancements for the two limiting cases are both of order unity, for high scatterer density there exist a small number of configurations that produce an appreciable sensitivity enhancement. In contrast, at low densities, different configurations do not differ greatly in their effect on sensitivity.

The results discussed so far are for the arbitrarily chosen observation position at $\mathbf{k}_{\parallel} = -\varepsilon_d^{1/2} k_0 \sin(70^\circ) \hat{\mathbf{x}}$. Fig. 5.13 shows mean amplitude $\langle |\gamma_1| \rangle$ (the only observation position dependent enhancement factor) and the total amplitude enhancement factor $\langle |\gamma_1 \gamma_2 \gamma_3| \rangle$ for a few different observation wavevectors \mathbf{k}_{\parallel} . Importantly, the mean amplitude shows only a weak dependence, and generally only when the enhancement is small, i.e. when the width of the distributions in the complex plane are small and thus the position of the centre of the distribution (i.e. the complex mean) has a more important role in the mean amplitude. This matches with the theoretical predictions of Section 5.3.1, with Eq. (5.70) giving no \mathbf{k}_{\parallel} dependence of $\langle |\gamma_1| \rangle$. Fig. 5.14 shows the density dependence of the complex mean $\langle \gamma_1 \rangle$ for the same set of observation positions. Here, there is some differences, most notably at high density. The movement of $\langle \gamma_1 \rangle$ from 1 at $n = 0$ towards 0 as $n \rightarrow \infty$ occurs regardless of \mathbf{k}_{\parallel} .

Additionally, the effect of α on the enhancement factor statistics should be considered.

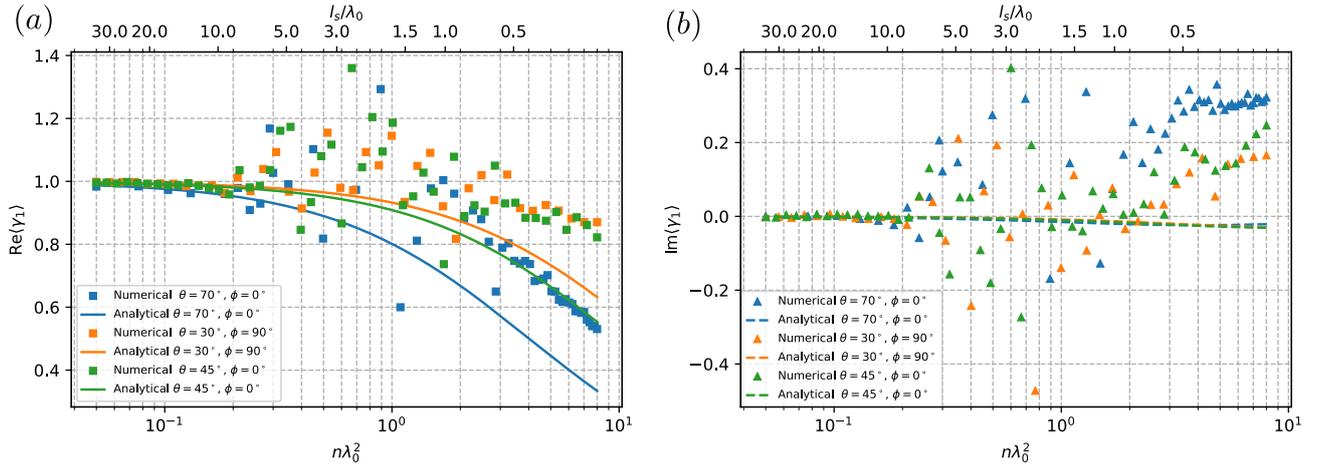


Figure 5.14: The complex mean $\langle \gamma_1 \rangle$ as a function of n for different observation positions (a) $\text{Re}\langle \gamma_1 \rangle$ and (b) $\text{Im}\langle \gamma_1 \rangle$.

The analytic results do predict a dependence on α , for both the complex mean and the mean amplitude. Fig. 5.15 shows the equivalent of Fig. 5.9 with $\alpha = i\alpha_{g1}$ phase shifted by $\pi/2$, while similarly, Fig. 5.16 shows the equivalent of Fig. 5.12 for this phase shifted value of α . The analytic results from Eqs. (5.81)–(5.82) still describe the complex means well at low densities, while Eq. (5.86) describes $\langle |\gamma_1| \rangle$ accurately up to the divergence. With this phase shifted polarisability $i\alpha_{g1}$, the general behaviour is the same as with α_{g1} , but the transition to large mean amplitudes and wide distributions occurs at a higher density. As a result, the mean amplitude enhancement is reduced at low densities compared to the results for α_{g1} in Fig. 5.12, which can be attributed to the fact the new scatterer properties give rise to more absorption, and therefore reduce both the amount of light incident on the analyte particle and the amount of light reaching other scatterers after scattering from the analyte particle. At larger values of n , however, the mean amplitudes in Fig. 5.16 behave similarly to those in Fig. 5.12, showcasing an optimum density where the amplitude enhancement is maximised and a decay with increasing density. The maximum sensitivity enhancement is smaller, though of similar magnitude (~ 119) and additionally it occurs at a higher density ($n_{\text{opt,tot}} = 2.31/\lambda_0^2$, $l_s = 0.74\lambda_0$) compared to the $\alpha = \alpha_{g1}$ case. The decay at higher densities, which has been attributed to localisation effects, starts at a similar density and therefore large enhancements close to the peak value occur over a narrower density range than the $\alpha = \alpha_{g1}$. The analytic results for $\langle \gamma_{1,3} \rangle$ provide an accurate description up to much higher densities $n\lambda_0^2 \sim 1$ ($l_s \sim 1.5\lambda_0$) compared to the α_{g1} case. While

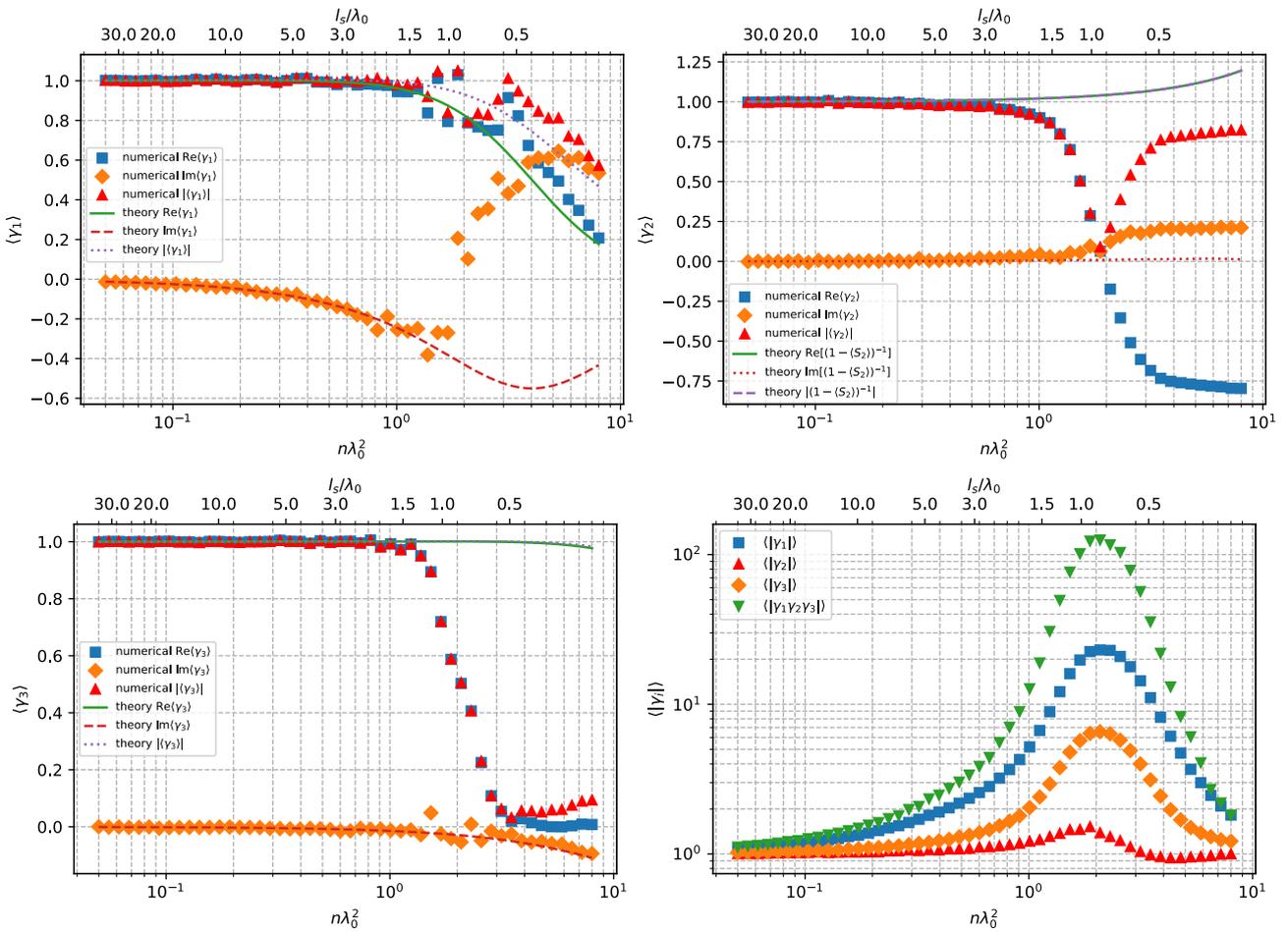


Figure 5.15: Equivalent of Fig. 5.9 but with $\alpha = i\alpha_{g1}$.

the ladder approximation provides a good low density description of $\langle |\gamma_1| \rangle$, $\langle |\gamma_3| \rangle$ is significantly reduced, due to the increased absorption from scatterers reducing the amount of scattered light arriving at the analyte particle. The probability distributions (both the amplitude distribution in Fig. 5.16(b) and the complex plane distributions shown in Fig. 5.17) behave as in the $\alpha = \alpha_{g1}$ case, though the transitions between different regimes occur at different density, in particular the transition to long tailed/high variance distribution requires a higher scatterer density. This set of results is revealing in terms of sensor design. The existence of an optimum scatterer density which maximises the mean amplitude implies the sensitivity of a sensor can be maximised by tuning the density of scattering elements on a random nanostructured surface. Furthermore, the properties of the individual scatterers (i.e. their polarisability α) affect the enhancement, offering another parameter that can be tuned. Physically, α is not freely tunable, but can be altered by tuning different properties of the scatterers, for example, variation of the material

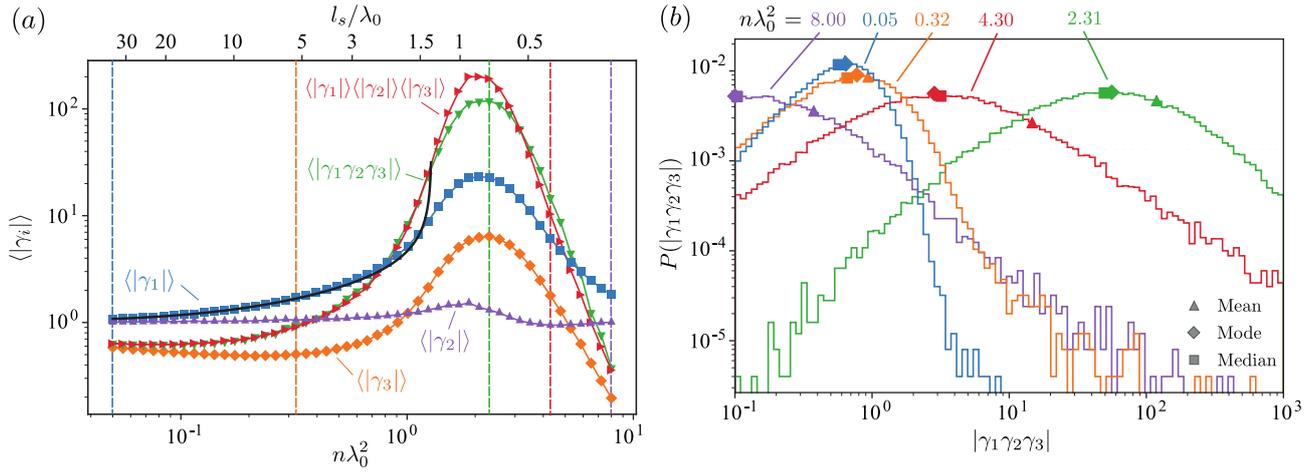


Figure 5.16: Equivalent of Fig. 5.12 showing but with $\alpha = i\alpha_{g1}$. Histograms are plotted at different densities to Fig. 5.12 since the transitions in distribution occur at different points. *Figure reprinted (with alteration) with permission from Ref. [287] © 2021 American Chemical Society.*

composition or geometrical properties. In addition, the polarisability of resonant scatterers, such as plasmonic nanoparticles exhibits a strong frequency dependence, so that tuning λ_0 can alter α . Additionally, introduction of an index matched spacer layer between the substrate and background scatterers furthermore allows the height z_s to be adjusted, which would alter the value of μ and therefore tune the statistics.

High Loss Single Interface

The use of lossy media such as metals in plasmonic structures means absorption inevitably plays an important role in SPP propagation, including in multiple scattering. The results discussed already demonstrate that absorption by scatterers can significantly affect the enhancement factor statistics. Additionally, absorption by the substrate plays a role, since Eqs. (5.81)–(5.83) and (5.86) depend on L_{SPP} . In order to further explore the role of absorption in the enhancement, simulations were performed with increased absorption within the metal. Specifically, the gold-water interface with $\lambda_0 = 600\text{nm}$ was considered, for which $\varepsilon_m = -8.0 + 2.1i$ while ε_d is unchanged. As a result $k_{\text{SPP}} = (1.49 + 0.05i)k_0$ and $L_{\text{SPP}} = 1.6\lambda_0$ is significantly shorter. The complex means and mean amplitudes for $E_{0,z} = \Theta(x) \exp(ik_{\text{SPP}}x - \kappa_d z)$ are shown in Fig. 5.18, for which all scatterers were modelled as 21.5nm gold spheres sitting on the surface, giving a polarisability α_{g2} calculated from Eq. (5.87). The observation position is $\mathbf{k}_{\parallel} = -\varepsilon_d^{1/2} k_0 \sin(70^\circ) \hat{\mathbf{x}}$ as in the ‘low loss’ case shown in Fig. 5.9. In this case, the box size L was varied between $L = 8\lambda_0$

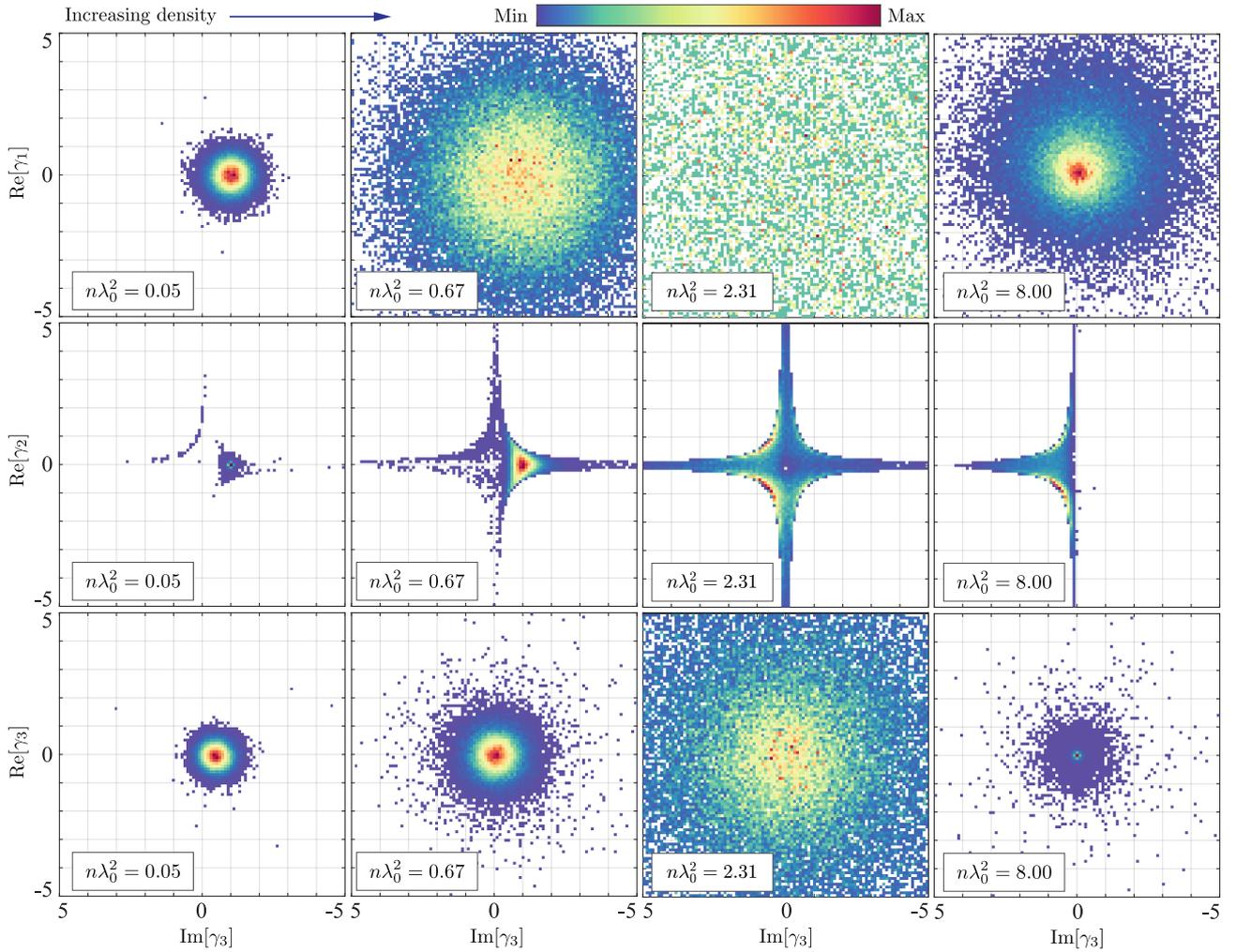


Figure 5.17: Equivalent of Fig. 5.10 showing the histograms of the individual enhancement factors on the complex plane but with $\alpha = i\alpha_{g1}$. Densities plotted are (left) $n\lambda_0^2 = 0.05$, (centre left) $n\lambda_0^2 = 0.67$, (centre right) $n\lambda_0^2 = 2.31$ and (right) $n\lambda_0^2 = 8$ which correspond to the same ones as in Fig. 5.10, with the addition of $n = n_{\text{opt,tot}} = 2.31\lambda_0^{-2}$ to show the optimum density distribution is analogous to that of Fig. 5.10.

and $L = 30\lambda_0$. In contrast to the ‘low loss’ case, the analytic results (Eqs. (5.81) and (5.82)) match the numerically calculated $\langle \gamma_{1,3} \rangle$ for the full density range. In addition, a significant difference from the ‘low loss’ results is the absence of the peak in the mean amplitudes, both for the individual enhancement factors and the total mean amplitude enhancement factor $\langle |\gamma_1 \gamma_2 \gamma_3| \rangle$. Instead, the means remain close to unity. This occurs due to the fact that the distribution of the enhancement factors in the complex plane remains narrow, as shown in Fig. 5.19, rather than the broad distributions of $\gamma_{1,3}$ seen in Fig. 5.10. The fact that absorption quenches the effect of multiple scattering, such that the enhancements are ~ 1 , is to be expected, since when $l_s > L_{\text{SPP}}$, scattered light is absorbed before it can be rescattered. The tight distributions

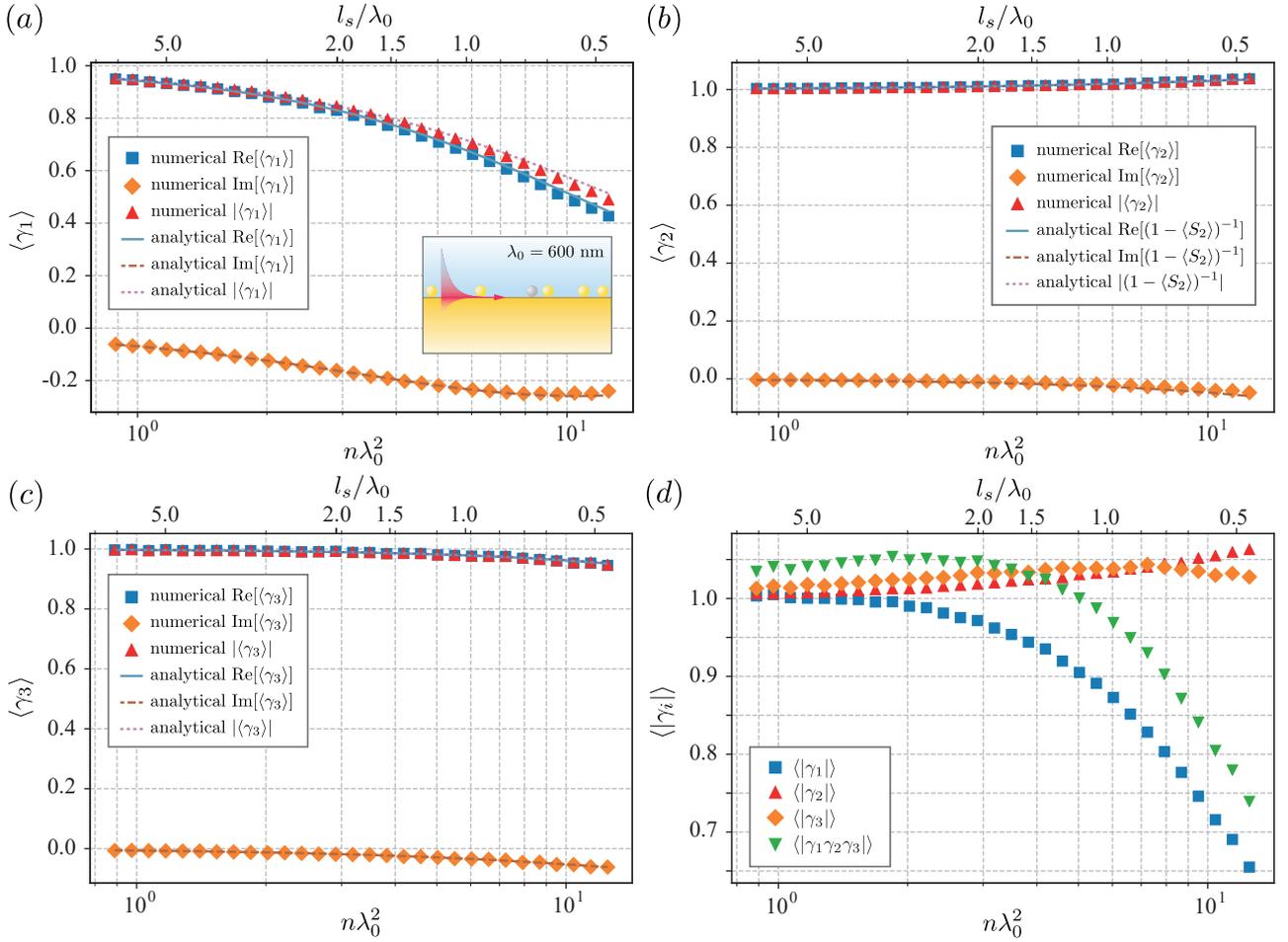


Figure 5.18: Equivalent of Fig. 5.9 for the ‘high loss’ case of $\lambda_0 = 600\text{nm}$ and $\alpha = \alpha_{g2}$. *Figure reprinted with permission from Ref. [288] © 2021 American Physical Society.*

mean that $\langle |\gamma_i| \rangle \approx |\langle \gamma_i \rangle|$, while the ladder approximation does not provide a good description of $\langle |\gamma_{1,3}| \rangle$. To explain the significant difference in the widths of the probability distributions between the high and low loss cases, it is helpful to compare the effect of a scattering event within a multiple scattering trajectory with the role propagation between scattering events plays in the same trajectory. The phase and amplitude an individual scattering path contributes to an enhancement factor can be divided into contributions from scattering events ($A_{\text{scat}} e^{i\Phi_{\text{scat}}}$) and from propagation between scattering events ($A_{\text{prop}} e^{i\Phi_{\text{prop}}}$), and the total enhancement factors are determined from the sum over all possible paths $\sim \sum_{\text{paths}} A_{\text{scat}} e^{i\Phi_{\text{scat}}} A_{\text{prop}} e^{i\Phi_{\text{prop}}}$. Different realisations give rise to different propagation factors as the relative position of scatterers changes, however the scattering contribution for a given sequence of scatterers is unchanged, since the individual scatterer properties are unchanged. Averaging over realisations leads to cancellation of the propagation contribution due to the random Φ_{prop} and thus the complex mean simplifies to

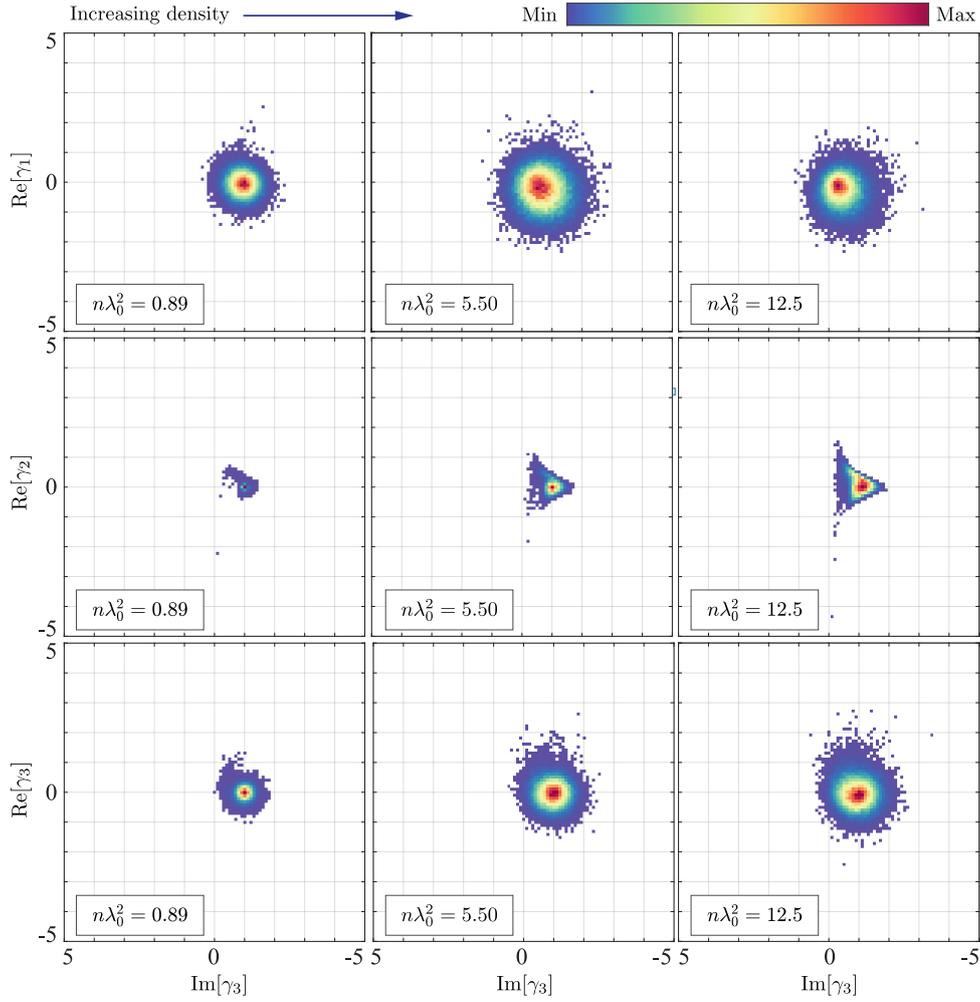


Figure 5.19: Equivalent of Fig. 5.10 for the ‘high loss’ $\lambda_0 = 600\text{nm}$ case, with $\alpha = \alpha_{g_2}$, showing the probability distribution of the enhancement factors over the complex plane.

the sum of the deterministic $A_{\text{scat}} e^{i\Phi_{\text{scat}}}$ factors arising from scattering events. In the presence of absorption, scattering paths longer than L_{SPP} have a small amplitude A_{scat} and hence contribute negligibly to the enhancement factors for that particular realisation. In the low loss case (with $L_{\text{SPP}} = 9.9\lambda_0$), a large number of scattering paths several wavelengths long contribute. As the paths extend over multiple wavelengths, the phases Φ_{prop} are essentially uniform and random and thus the sum over scattering paths can give a significantly different result to the complex mean. On the other hand, only a relatively small number of scattering paths shorter than $L_{\text{SPP}} = 1.6\lambda_0$ contribute significantly to the enhancement factor in the high loss case. Furthermore, the amplitude decay due to absorption over the wavelength scale is significant (the amplitude decays by $\sim 20\%$ over one SPP wavelength in the high loss case compared to $\sim 2\%$ in the low loss case), very short sub-wavelength scattering paths for which Φ_{prop} is close to zero will have significantly

higher amplitude and contribute more to the total enhancement factors. As a result, the high loss case is close to the complex mean since the propagation has little effect. The behaviour of $\gamma_{1,3}$ in the high loss case is therefore dominated by the scattering phase shift. The ladder approximation result is no longer valid in the high loss case since, in deriving Eq. (5.70), the phase and amplitude of each term in the sum was assumed to be independent. This assumption does not hold when $L_{\text{SPP}} \sim \lambda_{\text{SPP}}$ as the short paths with $\Phi_{\text{prop}} = 0$ (and therefore having a phase of $\sim \Phi_{\text{scat}}$) have larger amplitudes.

5.4.2 Polarisability Dependence of Peak Enhancement

The fact that the high loss case gives rise to a tighter distribution of γ_i centered close to $\langle \gamma_i \rangle$ means the amplitude enhancements inherit the dependence on $\arg(\alpha)$ and \mathbf{k}_{\parallel} , and thus are more sensitive to observation position and scatterer properties, compared to the low loss case where the random phasor sum model predicts no \mathbf{k}_{\parallel} dependence and the dependence on $\arg(\alpha)$ only arises in the modified absorption length within the ladder approximation. A stronger dependence on α is significant in the context of sensor design, since it would imply optimising the choice of the individual scatterers making up the multiple scattering environment is important. In particular, choosing α and \mathbf{k}_{\parallel} such that the divergence condition of Eq. (5.60) is satisfied could allow large enhancement factors to be achieved even in the presence of strong absorption. Since $\tilde{G}(\mathbf{k}_{\parallel}) \propto 1/(k_{\text{SPP}}^2 - k_{\parallel}^2)$, the choice of observation point such that \mathbf{k}_{\parallel} is close to the pole at k_{SPP} , i.e. $k_{\parallel} = k'_{\text{SPP}}$, ensures a large $|\tilde{G}|$ and thus the optimum density predicted by Eq. (5.59) is reduced. In addition, it ensures the optimum density condition for γ_1 and the lossless version of γ_3 coincide (since the input and output wavevectors are equal). While such an observation point is not possible for a setup consisting of a single metal-dielectric interface since $k'_{\text{SPP}} > \varepsilon_d^{1/2} k_0$, it is possible for SPPs in multilayer structures and corresponds to the leakage radiation ring. Some of the benefits of measurements in the leakage radiation ring have already been discussed within the single scattering analysis, and the multiple scattering enhancement applies in addition to these effects such as the confinement of scattered light giving large intensities. While such a multilayer configuration alters the Green's function and surface dressing, the functional form of G_{SPP} remains the same for points in the lower index dielectric near the surface of the gold

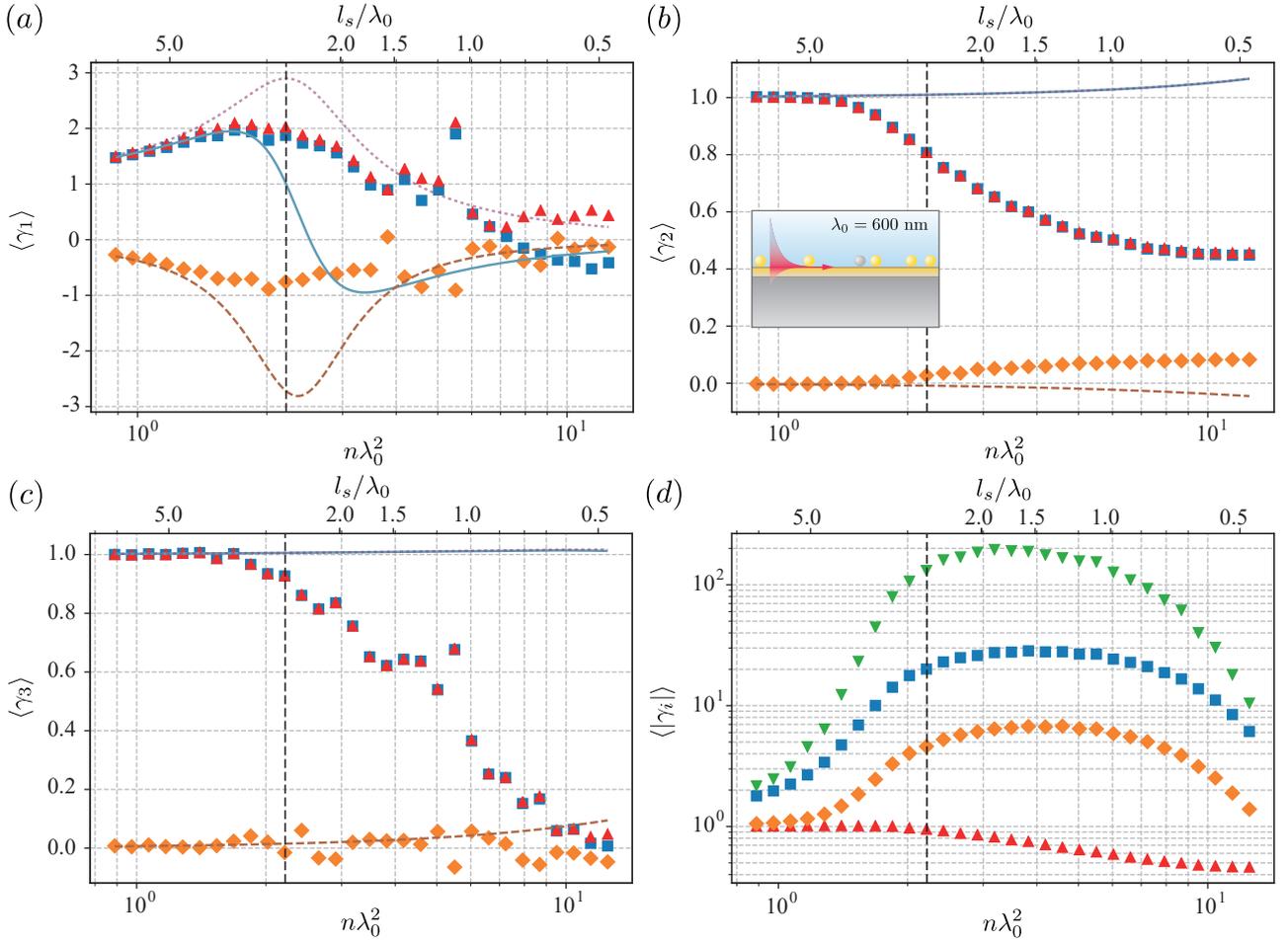


Figure 5.20: Equivalent to Fig. 5.9 for $\lambda_0 = 600\text{nm}$, $\alpha = -\alpha_{g2}$ and $\mathbf{k}_{\parallel} = -k'_{\text{SPP}}\hat{\mathbf{x}}$. Figure reprinted with permission from Ref. [288] © 2021 American Physical Society.

film, with only the parameter values changed (i.e. A_0 , k_{SPP} , a and α). Therefore, one can now consider such an observation position, keeping in mind that the parameters in the model will no longer correspond to the same physical properties of scatterers. The high loss case results equivalent to Fig. 5.18 except for with $\mathbf{k}_{\parallel} = -k'_{\text{SPP}}\hat{\mathbf{x}}$ and the polarisability phase shifted by π so that $\alpha = -\alpha_{g2}$ are shown in Fig. 5.20.

Note that since the amplitude of the polarisability is unchanged, the cross-section and mean free path are also unaltered. The phase shift to α , in addition to changing the absorption loss from a single scatterer, also affects the phase difference between the scattered and incident field. The chosen phase means the divergence condition of Eq. (5.60) is nearly satisfied. This means the phase difference between the SPP incident on a scatterer and the SPPs radiated by the scatterer is small and hence the scattered field can add in phase with the incident field to give a larger enhancement. Unlike the observation point in the upper half space, the analytic

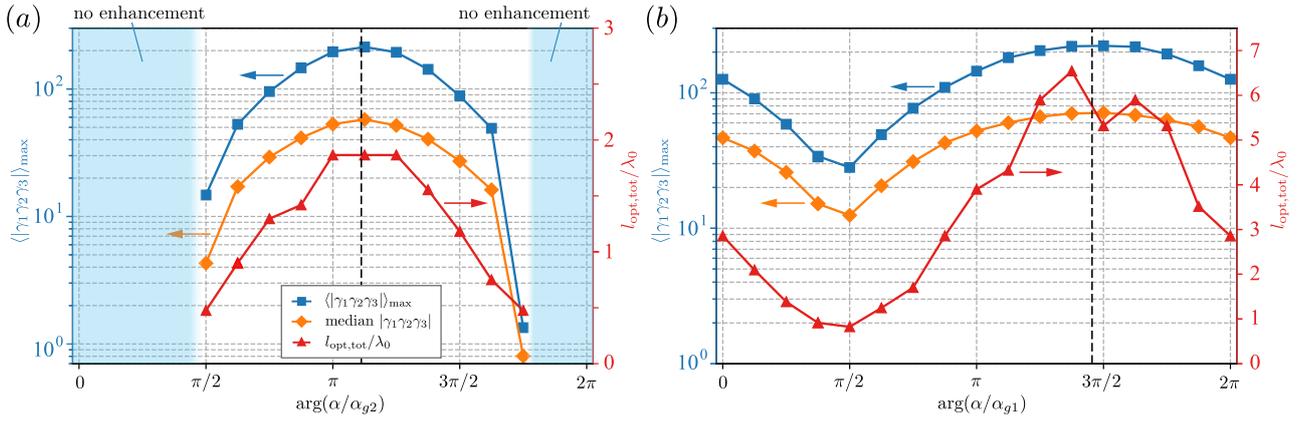


Figure 5.21: The dependence of $\langle |\gamma_1\gamma_2\gamma_3| \rangle_{\max}$ and $l_{\text{opt,tot}}$ on the phase of α (or equivalently μ) for (a) the high loss $\lambda_0 = 600\text{nm}$ case and (b) the low loss $\lambda_0 = 650\text{nm}$ case. The median value of $|\gamma_1\gamma_2\gamma_3|$ at $l_{\text{opt,tot}}$ is also shown. The vertical dashed line indicates the phase of α at which the divergence condition $\text{Im}[\alpha\tilde{G}(\mathbf{k}_{\parallel})] = 0$ from Eq. (5.60) is satisfied. Points for which $\langle |\gamma_1\gamma_2\gamma_3| \rangle_{\max} \leq 1$ are not plotted but denoted by the shaded blue region, since this indicates no amount of multiple scattering improves sensitivity on average and single scattering (i.e. $n = 0$) provides the best mean sensitivity. *Figure reprinted with permission from Ref. [288] © 2021 American Physical Society.*

results for the complex mean are only accurate at lower densities. More importantly, the mean amplitudes now exhibit the density dependence seen in Fig. 5.9, with an optimum density at which amplitude enhancements $\sim 10^2$ are achieved. In the low loss case, the leakage radiation ring observation position has little effect on the mean amplitude density dependence. The maximum mean total absolute enhancement $\langle |\gamma_1\gamma_2\gamma_3| \rangle_{\max}$ and density $n_{\text{opt,tot}}$ (and corresponding mean free path $l_{\text{opt,tot}}$) at which it occurs were calculated numerically as the phase of μ (or equivalently α) was varied with $|\mu|$ held constant and $\mathbf{k}_{\parallel} = -\text{Re}[k_{\text{SPP}}]\hat{\mathbf{x}}$. The dependence of $\langle |\gamma_1\gamma_2\gamma_3| \rangle_{\max}$ and $l_{\text{opt,tot}}$ on $\arg(\alpha)$ is shown in Fig. 5.21.

For the low loss case, Fig. 5.21(b), mean enhancements of at least one order of magnitude $\langle |\gamma_1\gamma_2\gamma_3| \rangle_{\max} > 10$ are always achievable regardless of $\arg(\alpha)$, with the value varying slightly with $\arg(\mu)$, albeit remaining $\sim 10^2$ for a broad range of phases. The optimum phase condition predicted from Eq. (5.60) coincides with the region where the numerically calculated maximum mean enhancement is largest and is also achieved at the lowest scatterer density (and therefore longest scattering mean free path). In contrast, for the high loss case, Fig. 5.21(a), there exists a range of $\arg(\alpha)$ for which no enhancement is observed on average and $\langle |\gamma_1\gamma_2\gamma_3| \rangle_{\max} \leq 1$, since absorption prevents multiple scattering enhancements. Tuning of $\arg(\alpha)$ does nevertheless allow a similar level of enhancement to the low loss case to be achieved, with the divergence condition

introduced by Eq. (5.60) providing a good predictor of the optimum phase. On the other hand, Eq. (5.59) does not provide an accurate prediction for $n_{\text{opt,tot}}$ (or indeed $n_{\text{opt},1}$ calculated from the Monte Carlo simulations). The fact the optimum phase prediction of Eq. (5.60) remains accurate, while the optimum density prediction of Eq. (5.59) does not, can be understood by considering the role of loop paths, ignored in deriving Eqs. (5.59) and (5.60). As the density increases, loop paths become significant and alter the form of Eqs. (5.53)–(5.55) so that the maximum $|\langle\gamma_i\rangle|$ occurs at a different density. In addition, the difference between $|\langle\gamma_i\rangle|$ and $\langle|\gamma_1\gamma_2\gamma_2|\rangle$ can also lead to a difference between $n_{\text{opt,tot}}$ which maximises $\langle|\gamma_1\gamma_2\gamma_2|\rangle$ and $n_{\text{opt},i}$ which maximises $|\langle\gamma_i\rangle|$. On the other hand, the phase shift from each scattering event in a trajectory is the same whether the trajectory contains loops or not. Therefore, the condition on α for the scattered field to be in phase with the incident field is unchanged by the inclusion of the loop paths and thus the no loop phase condition remains accurate.

5.4.3 Enhancement Factor Correlations

While much of this chapter has studied the statistics of individual enhancement factors separately, they are not statistically independent since they depend on the same underlying random variables (i.e. the scatterer positions \mathbf{r}_i). Indeed, Figs. 5.12 and 5.16 demonstrate this clearly through the fact $\langle|\gamma_1\gamma_2\gamma_3|\rangle \neq \langle|\gamma_1|\rangle\langle|\gamma_2|\rangle\langle|\gamma_3|\rangle$. At low densities, the fact that the enhancement factors are correlated has little effect and the two quantities are approximately equal. Close to the optimum density, the correlations between enhancement factors serve to reduce the total amplitude enhancement such that $\langle|\gamma_1\gamma_2\gamma_3|\rangle < \langle|\gamma_1|\rangle\langle|\gamma_2|\rangle\langle|\gamma_3|\rangle$, while in the localisation regime with decreasing enhancement factors, the correlations have the opposite effect as $\langle|\gamma_1\gamma_2\gamma_3|\rangle > \langle|\gamma_1|\rangle\langle|\gamma_2|\rangle\langle|\gamma_3|\rangle$. Analysis of the Pearson's correlation coefficients P_{ij} between $|\gamma_i|$ and $|\gamma_j|$ ($i \neq j$), plotted in Fig. 5.22 reveals that $|\gamma_2|$ shows weak correlation with the other enhancement factors, with $|P_{23}| < 0.2$ across the full density range and $|P_{12}| < 0.2$ except at the very highest density for the $\alpha = i\alpha_{g1}$ case, where $P_{12} \sim -0.4$. This lack of strong correlation can be understood to arise from the fact that the loop paths associated with γ_2 are distinct from the scattering paths in $\gamma_{1,3}$ that start and end at distinct scatterers. In contrast, scattering trajectories contributing to γ_1 and γ_3 are partially related by reciprocity as discussed in Section

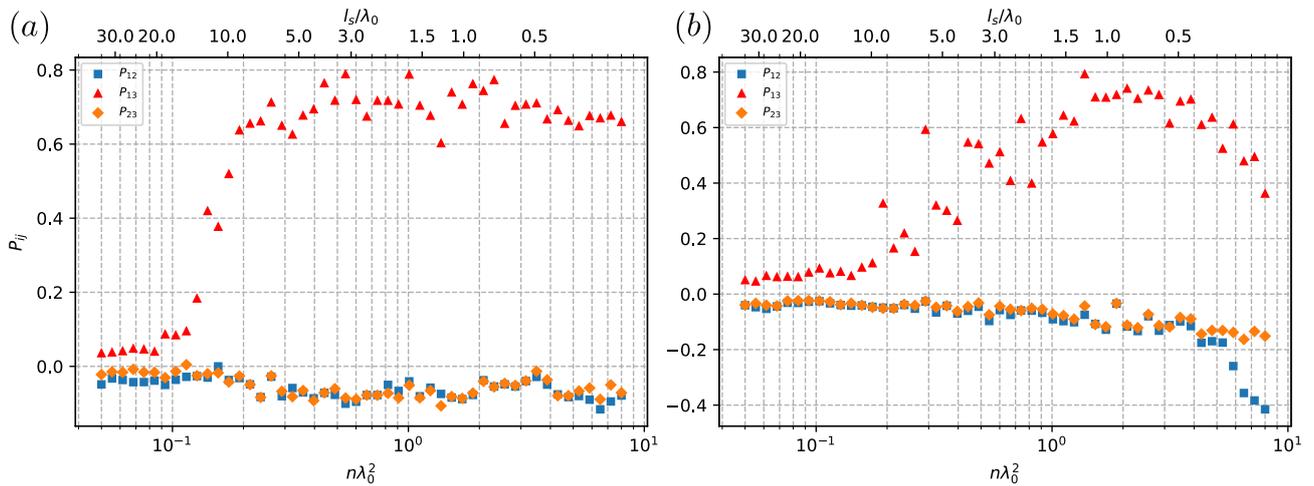


Figure 5.22: The Pearson correlation coefficients P_{ij} of the enhancement factor amplitudes in the low loss case for (a) $\alpha = \alpha_{g1}$ and (b) $\alpha = i\alpha_{g1}$.

5.2.5, with any given multiple scattering trajectory from \mathbf{r}_{N+1} to \mathbf{r}_i (associated with γ_1) having the same phase and amplitude as its reciprocal path going from \mathbf{r}_i to \mathbf{r}_{N+1} (associated with γ_3). Much like in coherent backscattering and weak localisation, the outgoing/incoming legs, i.e. propagation of the scattered field from \mathbf{r}_i to \mathbf{r} for γ_1 and propagation of the incident field to \mathbf{r}_i , can still cause the reciprocal paths to contribute different phases and amplitudes to the respective enhancement factors, allowing for decorrelation of $|\gamma_1|$ and $|\gamma_3|$. At low densities, the propagation phases, and hence enhancement factors, remain uncorrelated ($|P_{13}| \lesssim 0.1$) as the long propagation between scatterers randomise the phase, however, at higher densities the typically shorter distances between scattering sites and the analyte particle mean the phase difference of the incident and outgoing fields are smaller resulting in increased correlation ($0.6 \lesssim P_{13} \lesssim 0.8$). The phase shifted polarisability case $\alpha = i\alpha_{g1}$ transitions to this correlated regime at a higher density, coinciding with the onset of the optimum density regime.

5.5 Conclusion

In this chapter, the effect of multiple scattering on the field perturbation caused by the addition of an analyte particle was encapsulated in three enhancement factors, each arising from distinct classes of multiple scattering paths. The statistics of the enhancement factors were studied in the context of SPP scattering, both analytically and numerically with Monte Carlo simulations.

It was found that multiple scattering could increase the field sensitivity by a factor of $\sim 10^2$ relative to the single scattering sensitivity. Additionally, the enhancement showed sensitivity to scattering parameters such as scatterer density, polarisability and the amount of absorption. In particular, it was seen that, for long absorption lengths, there existed an optimum scatterer density which maximised the mean enhancement, and above this optimum density, localisation effects served to reduce the multiple scattering enhancement, and even reducing sensitivity at sufficiently high densities. For very short absorption lengths $L_{\text{SPP}} \sim \lambda_{\text{SPP}}$, the enhancement factor statistics showed greater sensitivity to the individual scatterer properties, with there also existing an optimum polarisability phase to maximise the mean amplitude of the enhancement factors.

These results can help inform the design of random nanostructured surfaces for sensing, indicating that by choosing the correct density of nano-scatterers on the surface and optimising their optical or geometric properties, one can improve the sensitivity by up to two orders of magnitude over a flat surface.

Chapter 6

Neural Network Based Tracking

The tracking method developed in Chapter 4 relies on knowledge of the relationship between the shift in analyte particle position and the resulting change in the phase and amplitude of the light scattered from the analyte particle, which in the multiple scattering regime depends on the scattering configuration, differs from realisation to realisation and is not known in general. The results in Chapter 5, however, suggest that there may be gains in sensitivity from working in the multiple scattering regime, while multiple scattering may also be unavoidable in some complex biological media. As such, it would be useful to develop a tracking algorithm that could extract the trajectory from speckle patterns formed in the multiple scattering regime. This chapter investigates achieving this aim through the use of a neural network based tracking algorithm

6.1 Neural Network Background

Machine learning has seen much research interest in recent years, enabled by rapid advancements in computing technology. It has found many applications in science and in optics in particular. An artificial neural networks is a prominent example of a machine learning tool which is widely used in many machine learning applications. The fundamental building block of a neural network is the node or Perceptron [4, 359, 360]. A single node takes in a set of M input variables a_i and has associated with it a series of M weights w_i . The node linearly combines this set of inputs with the appropriate weight to give a single number $T = w_0 + \sum_{i=1}^M w_i a_i$. The constant offset w_0 is often referred to as the bias of a node. The node output b is then given by an activation

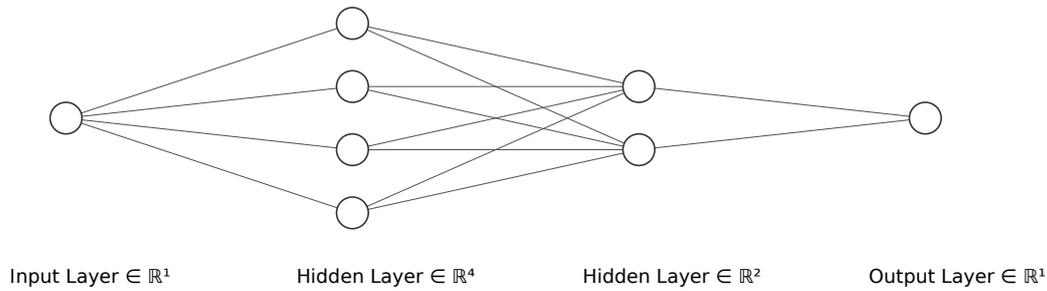


Figure 6.1: A simple example of a fully connected feed forward neural network, with each node in a layer connected to every node in the subsequent layer. The neural network shown consists of a single element input, a hidden layer of 4 nodes, a second hidden layer of 2 nodes and a single element output. Lines indicate the left node output is fed as input to the right node.

function $g(x)$ via $b = g(T)$. As such, a single node takes in M inputs and gives out a single number. A neural network is formed by a connection of these nodes [361], so that the output from a set of nodes is fed as input into other nodes. There are many ways to connect nodes in a network, giving rise to a wide range of different architectures. Typically, nodes are arranged in layers, with nodes in each layer connected to nodes in adjacent layers [362]. Layers which are not the input or output are referred to as hidden layers, since the output of nodes in such layers is not seen, and they only act as an intermediate step towards the final output. A simple example of network architecture is a fully connected network, in which a node is connected to every node in the subsequent layer, as illustrated in Fig. 6.1. Other architectures with different connections between nodes are also used, including some such as liquid state machines where nodes cannot be simply grouped into layers [363].

A neural network is determined by the architecture and the choice of weights w_i and activation function $g(x)$. Together, these define a function from the input vector \mathbf{a} to the output vector \mathbf{y} , where the vectors consist of an ordered list of all the input variables a_i or output variables y_i . The size of these vectors depend on the network architecture. A key feature of artificial neural networks is that they can act as a universal function approximator [364, 365, 366]. This means that a sufficiently wide (i.e. a large enough number of nodes in a single layer) [364, 365] or deep (i.e. a large enough number of layers) [367, 368, 369] neural network with a non-linear activation function can get arbitrarily close to any continuous function with domain and codomain matching the input and output sets respectively through the correct choice of weights. As a result, neural networks can be used to approximate a functional relationship

between real world data, even when the exact relationship is unknown.

While a neural network is capable of approximating a given real world function, the set of weights required to approximate the particular function is unknown. In order to achieve the correct weights, a machine learning element must be incorporated. This means that a subset of the data where the input and desired output are known is used to train the neural network towards getting the correct weights. Firstly, a loss function $L(\mathbf{y}_{\text{true}}, \mathbf{y}_{\text{pred}})$ is defined which quantifies the error when the neural network gives an output \mathbf{y}_{pred} for an input with a true output \mathbf{y}_{true} . The choice of loss function depends on the context [370], for example in categorisation problems, a logarithmic cross-entropy loss function is often used, while when fitting the neural network to a real valued function (the context relevant to this chapter), a mean square deviation loss function is common [371]. In general, the loss function decreases as \mathbf{y}_{pred} approaches \mathbf{y}_{true} and is minimised when $\mathbf{y}_{\text{true}} = \mathbf{y}_{\text{pred}}$. While the loss function is defined as a function of \mathbf{y}_{pred} and \mathbf{y}_{true} , it is also implicitly a function of the weights of each node, since changing the weights changes \mathbf{y}_{pred} and therefore L . The training of a neural network on a set of training data, i.e. a set of N_{train} inputs $\mathbf{a}^{(i)}$ and known true outputs $\mathbf{y}_{\text{true}}^{(i)}$, involves finding the set of weights which minimises the average value of the loss function $\sum_{j=1}^{N_{\text{train}}} L(\mathbf{y}_{\text{true}}^{(j)}, \mathbf{y}_{\text{pred}}^{(j)})/N_{\text{train}}$, where $\mathbf{y}_{\text{pred}}^{(j)}$ is the neural network output corresponding to the input $\mathbf{a}^{(j)}$. Provided the training data is representative of the underlying sample space from which it is drawn, this trained network should then give an output value close to the true value even when encountering an input not in the training set. In quantifying the performance of the neural network on unseen data, the appropriate metric is not always the value of the loss function. In a categorisation problem, a success rate (i.e. fraction of a test data set put in to the right category) is a more appropriate measure of the effectiveness of the neural network, but differs from the loss function. A quantity measuring neural network performance for a given dataset is referred to as a metric, while loss specifically refers to the value of the loss function, which in some contexts can also be an appropriate metric.

An important caveat to the idea of minimising the loss of a dataset is that it is also possible to overtrain the network. This means that the neural network learns features of the training data that do not apply to the wider sample space, and therefore does not perform well when

applied to unseen data not in the training set [372, 373]. This effect is analogous to overfitting in curve fitting, in which, with sufficient free parameters in the fitted model, one can perfectly fit a curve to pass through all the data points, while clearly not providing an accurate description of the real underlying curve between the data points. A widely used approach to avoid overfitting is early stopping [373, 374, 375]. In addition to the training data set, some data is used as a validation data set, which is not used to train the network, but instead used to calculate a performance metric (this could be the average loss of the validation dataset, or some other measure). This provides a measure of how well the network generalises, and the validation metric beginning to increase (i.e. performance gets worse) with iterations of the optimisation algorithm indicates that overfitting is starting to occur. At this point, the training is stopped to prevent overfitting. In reality, the stochastic nature of the optimisation algorithm means the validation metric can increase for a few iterations before decreasing again after more. Such behaviour is not indicative of overfitting, for which one would see an increase (or at least no decrease) in the validation metric as more optimisation iterations were performed. Thus, an early stopping algorithm typically has a patience hyperparameter, defining how many iterations to be performed showing no decrease in validation metric before stopping the training and restoring the weights to the values which minimised the validation metric. Note that a hyperparameter refers to some parameter of the training algorithm set by the user, to distinguish it from the neural network parameters (i.e. the weights) [376, 377]

6.1.1 Optimisation Algorithm

So far, the specifics of the optimisation algorithm used in training the neural network have not been discussed, and the theory behind training the network does not depend on the minimisation procedure. The nature of the optimisation problems associated with training neural networks, in particular the potentially large number of weights and training data, mean certain algorithms tend to be used more commonly for training. Most of the commonly used optimisers are based on gradient descent [378]. In this approach, starting from an initialised vector of all weights \mathbf{W}_0 , the weights are updated iteratively, with the k th iteration \mathbf{W}_k calculated from the previous

iteration via

$$\mathbf{W}_k = \mathbf{W}_{k-1} - \alpha_{\text{lr}} \nabla_{\mathbf{W}} F(\mathbf{W}_{k-1}), \quad (6.1)$$

where α_{lr} determines the size of the steps taken and is referred to as the learning rate in the context of neural network training, $\nabla_{\mathbf{W}}$ denotes the gradient in the vector space of the weights and $F(\mathbf{x})$ is the function being minimised, which here would be $F(\mathbf{W}) = \sum_{j=1}^{N_{\text{train}}} L(\mathbf{y}_{\text{true}}^{(j)}, \mathbf{y}_{\text{pred}}^{(j)}) / N_{\text{train}}$. The idea behind gradient descent is that F decreases fastest in the direction $-\nabla_{\mathbf{W}} F$, and thus, provided α_{lr} is not too big, stepping in this direction from \mathbf{W}_{k-1} to \mathbf{W}_k decreases the objective function, $F(\mathbf{W}_k) < F(\mathbf{W}_{k-1})$. As a result, the iterations give lower and lower values of F , heading towards a (local) minimum. The form of F presents a problem when there is a large amount of data, as one has

$$\nabla_{\mathbf{W}} F = \frac{1}{N_{\text{train}}} \sum_{j=1}^{N_{\text{train}}} \nabla_{\mathbf{W}} L(\mathbf{y}_{\text{true}}^{(j)}, \mathbf{y}_{\text{pred}}^{(j)}), \quad (6.2)$$

meaning that one would have to calculate the loss function gradient for each member of the training set. For a large dataset, calculating so many gradients can be prohibitively time consuming, especially if there is also a large number of weights. As a result, variants of a stochastic gradient descent (SGD) algorithm are used [378, 379]. In SGD, rather than use the full training dataset to calculate the gradient, a random subset is used to estimate the gradient. Thus, the iterative relation is

$$\mathbf{W}_k = \mathbf{W}_{k-1} - \frac{\alpha_{\text{lr}}}{N_{\text{batch}}} \sum_{j \in \text{batch } k} \nabla_{\mathbf{W}} L(\mathbf{y}_{\text{true}}^{(j)}, \mathbf{y}_{\text{pred}}^{(j)}), \quad (6.3)$$

where the training data has been randomly divided into $\text{ceil}(N_{\text{train}}/N_{\text{batch}})$ batches of size N_{batch} (though the last batch will be smaller if N_{batch} does not divide N_{train}). Once all the batches have been used, the algorithm has completed an epoch. The training data is divided into new random batches and the process continued in a new epoch. Sometimes, SGD specifically refers to the case where $N_{\text{batch}} = 1$, while the case $1 < N_{\text{batch}} < N_{\text{train}}$ is called mini-batch gradient descent. Here, both cases shall be referred to as SGD and the batch size treated as a parameter

of the learning process. There exist many adaptations of SGD based on slightly modifying the estimated step between weight iterations calculated from a batch from a simple gradient. One of the most widely used versions is the ‘Adam’ SGD [380], which is the optimiser used to train all the neural networks presented in this chapter. The name ‘Adam’ comes from adaptive moment estimation, and it makes two key alterations to the basic SGD algorithm. Firstly, rather than using the gradient estimate calculated from a single batch to generate a step, the gradient estimate is calculated as a weighted average of the gradient from the previous step and the gradient calculated from the current batch. Additionally, a running second moment estimate (i.e. uncentered variance) is also used to calculate the step size, ensuring that batches with high uncertainty as to how well the gradient estimate matches the true gradient give comparatively smaller steps. More detailed description of the algorithm can be found in Ref. [380]. In addition to α_{lr} , ‘Adam’ requires two additional learning hyperparameters, β_1 and β_2 between 0 and 1 which describe the weight given to the estimate from the previous iteration of the first and second moment of the gradient respectively in calculating the updated estimate of the moments.

6.1.2 Neural Networks in Optics and Imaging

Neural networks have found many uses within optics, microscopy and imaging. Applying neural networks in image processing has a clear logic, in that one wishes to find a function from a collection of spatially ordered pixel values onto some information that is known to be contained within the image, but such a function is not known. Due to the spatially ordered nature of images, convolutional neural networks (CNNs) have found much use in image processing [381]. In a CNN, rather than linearly combining all the input pixels in a node, the image is convolved with a kernel, to give a feature map which retains the spatial ordering information. Convolution is a linear operation, and thus computationally, the process is not significantly demanding compared to linearly combining all the pixels as in the case of a perceptron discussed above. The weights (i.e. parameters tuned in training) in a CNN are the elements of the kernel, while the kernel shape is a manually set hyperparameter. The use of a non-linear activation function is still present in CNNs. One image processing application of neural networks is in categorising what is shown in the image [382, 383]. For example, a human can recognise if an image shows a

cat, so some link exists between the properties of the image (i.e. the spatially ordered pixel values) and the categorisation as a cat. A neural network learns to approximate a function making this link from image to cat. Another image processing application of neural networks, relevant to the context of this thesis, is in the field of localisation microscopy. As discussed in Chapter 2, point emitters give rise to a PSF in the microscope image. As a result, there is a function linking the image to the emitter position, and a neural network can be applied to learn this function [4, 384]. In this case, there exists non-machine learning based approaches in which one fits a known PSF, for example through maximum likelihood estimation (MLE) with a PSF calculated through diffraction theory [385]. Though it is not required, using a neural network approach does bring advantages over the standard PSF fitting based approaches. Firstly, the PSF fitting approaches such as MLE can be extremely computationally demanding, especially when multiple emitters must be localised in many different frames [386]. In contrast, once trained, a neural network only needs to evaluate a series of linear transforms and function evaluations, meaning a neural network can perform higher throughput localisation compared to conventional PSF fitting approaches. In addition, a neural network can be trained to extract additional information on top of the emitter position, for example labelling individual objects or isolating them [177]. Finally, in the context of PSF engineering, the optimisation of the system PSF can be incorporated in the learning process, so that, in addition to the neural network learning how to localise emitters, the PSF is being optimised to improve localisation [387, 388]. Generally, such neural network based localisation microscopy approaches follow the methods described in Chapter 2, except with the image processing stage to extract object's position (or any other information) is performed by the neural network [389, 390]. Thus, techniques such as TIRF, STORM and PALM have had neural network based processing incorporated [391, 392, 393]. Highly accurate numerical simulations of microscope image formation allow a large amount of training data to be generated numerically rather than experimentally, making comprehensive training of the localisation neural networks possible without time consuming experiments.

Another area of optics where neural networks and machine learning in general has found applications is imaging through scattering media, and extracting information from speckle

patterns. Neural networks can enable imaging through scattering media by learning the non-linear inverse scattering function linking the intensity in the object plane to the speckle image formed after propagation through the scattering medium [394, 395, 396]. This allows for imaging through fixed random configurations such as multi-mode fibres or glass diffusers, but since the neural network learns the particular inverse scattering function for the realisation of disorder it is trained on, it does not generalise to different realisations [397, 398]. Via training on multiple different realisations of disorder, a CNN has, however, been shown to be capable of imaging through realisations of disorder different to those in the training data, which would allow imaging through dynamic scattering media where the realisation changes between imaging frames [399]. As well as imaging, neural networks have been trained to extract other information from speckle patterns, such as identifying the scatterer composition [400] or labelling objects [401]. In addition to imaging, neural networks have been used in other light control applications in scattering media, such as focusing, beam and wavefront shaping and learning the relationship between transmitted and reflected speckle patterns [396, 402, 403].

This is by no means an exhaustive review of the numerous applications of neural networks in optics, for example it also finds use in holography and phase extraction [404, 405]. The wide range of applications arises in part due to the universal function approximation capability. The use of neural networks in both localisation microscopy and the study of random light scattering media is emphasised here as they are most relevant to the work in this thesis. In both these applications, there is known theoretically to be a functional dependence of the measured image or speckle pattern on the desired quantity (e.g. particle position or object plane intensity distribution), but it is not feasible to find this link using the detailed physics, so a neural network learns the link from data.

6.2 Theory of Neural Network Localisation from Speckle

As established in Chapter 3, Maxwell's equations provide a link between a given scattering dielectric distribution and the electric field. Indeed, for a fixed incident field, Eq. (3.16) can be considered to define a higher order function (i.e. a function that maps a set of functions

onto another set of functions) from a dielectric distribution $\varepsilon_s(\mathbf{r})$ onto an electric field $\mathbf{E}(\mathbf{r})$, $\mathbf{E}(\mathbf{r}) = F_{\mathbf{E}}(\varepsilon_s(\mathbf{r}))$. As a result, there is also a function mapping ε_s onto the intensity, $I(\mathbf{r}) = F_I(\varepsilon_s(\mathbf{r}))$. Note that no single scattering approximations have been made, as the full form of Eq. (3.16), rather than the Born approximation, defines the mapping functions $F_{\mathbf{E}}$ and F_I . If F_I is invertible, then it can also be deduced that there is a function from the intensity distribution onto the scattering configuration, namely $\varepsilon_s(\mathbf{r}) = F_I^{-1}(I(\mathbf{r}))$. Requiring F_I to be invertible means that two scattering configurations cannot give rise to the same intensity distribution. This condition has been shown to hold in certain cases [406, 407, 408, 409, 410, 411], depending on the properties of ε_s and the domain over which $I(\mathbf{r})$ is defined (i.e. where the intensity is measured in an experimental context). For scalar waves in 3D, for example, it has been shown that the values of $I(\mathbf{r})$ on the surface of a sphere uniquely determine $\varepsilon_s(\mathbf{r})$, provided ε_s is real and has compact support (along with some more technical conditions on the differentiability of ε_s) [406]. It shall be assumed F_I^{-1} exists. In the case where ε_s is a random scattering potential, $I(\mathbf{r})$ is a speckle pattern, with the implication being there exists a function from a measured speckle pattern to a scattering potential. Thus, while the speckle pattern is often modelled statistically, the properties are still a deterministic function of the given scattering configuration. Furthermore, if ε_s can be parameterised by some other function or set of numbers, then there exists a function relating the speckle pattern to these parameters. For example, for a randomly rough surface, Eq. (3.35) parameterises ε_s in terms of the surface height profile $\zeta(\boldsymbol{\rho})$, and thus there is a function from the speckle pattern to the rough surface profile, $\zeta(\boldsymbol{\rho}) = F_{\zeta}^{-1}(I(\mathbf{r}))$. Indeed, a neural network based on learning F_{ζ} to map random nanostructured surfaces supporting SPPs onto leakage ring speckle patterns has been developed [278]. Finally, considering the case where the random scattering configuration is fixed except the analyte particle, there is a function from the set of speckle patterns generated by this arrangement onto the analyte particle state, parameterised by the position and orientation as discussed in Chapter 4, i.e. $[\mathbf{r}_a, \boldsymbol{\xi}_a] = F_{\mathbf{r}}^{-1}(I(\mathbf{r}))$. Finally, if one discretises the measured speckle intensity at N_I points \mathbf{r}_i with $I_i = I(\mathbf{r}_i)$, the function maps a N_I element vector onto the analyte particle parameters. In discretising, N_I must be sufficiently large and the pixels distributed over a wide enough region to retain the uniqueness of the speckle pattern to scattering configuration mapping, since even if the speckle

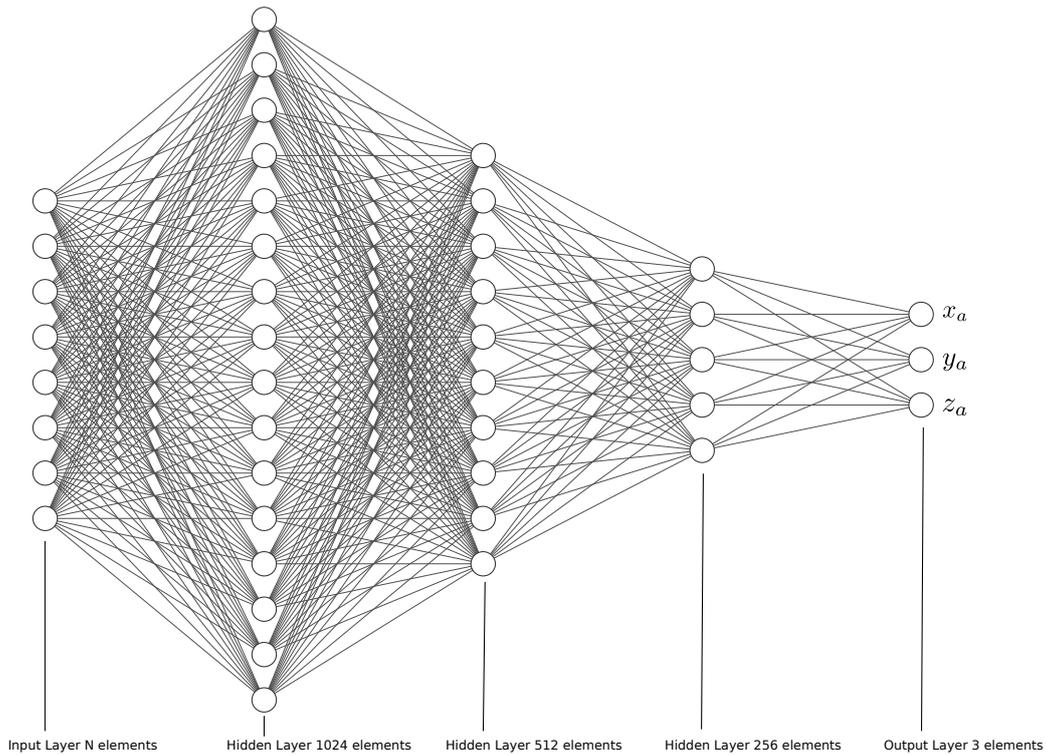


Figure 6.2: Schematic of the structure of the dense neural network used to localise the particle position from the leakage ring speckle pattern intensity profile. The number of nodes shown does not represent the true size of each layer (except for the final output layer).

patterns for two different scattering configurations are different, the intensities can be equal at a finite set of points. With the existence of this speckle pattern to analyte particle parameters established, it can now be seen how to apply neural network based tracking algorithm. The neural network learns to approximate the mapping from a training dataset of measured speckle intensity profiles where the position of the analyte particle is known. Importantly, the algorithm, unlike the one developed in Chapter 4, makes no assumptions about the scattering regime and thus should work in the presence of strong multiple scattering.

6.3 Simulation of Neural Network Tracker

To verify the effectiveness of the neural network tracking algorithm, a dense feed-forward neural network was designed to extract the analyte particle position from the speckle intensity profile around the leakage radiation ring in the random SPP scattering system introduced in Chapter 3 and tested on simulated data. The structure of the neural network is shown in Fig. 6.2, consisting of an input layer, the vector of size N_ϕ consisting of the discretised speckle intensity

around the leakage ring $I_i = I(\phi_i)$, three hidden layers of size 1024, 512 and 256 and an output layer corresponding to the position of the analyte particle (either 2 or 3 outputs, depending on whether the algorithm was performing 2D or 3D tracking). A rectified linear unit (ReLU) activation function is used, defined by $g(x) = \max(0, x)$ [412, 413].

6.3.1 Normalisation of the Input

It is standard practice in neural networks to normalise the input so that the values are generally distributed between -1 and 1 [414, 415]. This can be either a hard limit (i.e. use $a_i / \max(|a_i|)$) or a z-score type normalisation (i.e. $(a_i - \mu) / \sigma$ where μ and σ are the mean and standard deviation of the input values a_i) where input values of magnitude much larger than 1 become increasingly improbable. There are several good reasons for this from a machine learning viewpoint. One of the main motivations for normalising is that it often improves generalisation, removing the scale of the training data from the neural network. For example, in classifying objects in images, doubling all the pixel values in a given image clearly does not change what is shown in the image, but without normalisation, a neural network could give a very different output for this image, even though the doubled intensity image should be classified in the same way as the original image. It could be argued that this point is less relevant to the speckle based tracking neural network being developed in this chapter, since different binding positions \mathbf{r}_a will give different scale of intensities, for example if the particle binds in a hotspot or darkspot, and thus the scale information is relevant. There does, however, remain a generalisation advantage to normalising. If one trains the network on data from a specific analyte particle, the network only learns speckle patterns for that analyte particle, and would not work for different size or refractive index particles. The intensity perturbation $\Delta I = I - I_b$ (I_b being the speckle intensity in the absence of the analyte particle) contains the particle scattered intensity and the interference with the background speckle, just for now considering the single scattering regime. The interference term scales as $\sim |\alpha| \propto R_a^3$, while the direct term scales according to $\sim |\alpha|^2 \propto R_a^6$. If the interference term dominates, as it does for small biological particles, the intensity perturbation has a simple proportional scaling with $|\alpha|$ (or R_a^3) which can be normalised out to allow the neural network tracker to generalise to different size analyte particles.

Allowing for the effects of multiple scattering, rather than the direct intensity and interference of the light scattered from the particle, the intensity perturbation is made up of a direct term and interference term of $\delta\mathbf{E}(\mathbf{r}) = G(\mathbf{r}, \mathbf{r}_a)\gamma_1\gamma_2\gamma_3\mathbf{p}_0$, assuming the point scattering model from Chapter 5 is valid. In Chapter 5, it was established that γ_1 was independent of the analyte particle polarisability, α_{N+1} , as was γ_3 for scalar α_{N+1} , (i.e. a rotationally symmetric analyte particle). Similarly, for sufficiently small analyte particles $\gamma_2 = (1 - \text{const.} \times \alpha_{N+1})^{-1} \approx 1$ and the effect of γ_2 is negligible. Under these conditions, $|\delta\mathbf{E}| \propto |\alpha_{N+1}|$ and the same scaling of the interference term with $|\alpha_{N+1}|$ as the single scattering regime is found. As such, the generalisation of the neural network tracker to different analyte particles to the training analyte particle via appropriate normalisation also applies in the multiple scattering regime.

In addition to the improved generalisation, there are other reasons for normalising data. By ensuring the scale of the data is comparable to that of other previously studied neural networks, the same heuristic rules can be applied, for example the values of the hyperparameters such as α_{lr} . Furthermore, the most commonly used activation functions have a natural scale of unity. Machine learning programming libraries are often optimised for input and output values in the range -1 to 1 , so normalising is easier than trying to adapt the various functions to match the scale of the unnormalised data. For the tracking neural network, the normalised input $\widetilde{\Delta I}_i = \widetilde{\Delta I}(\phi_i)$ is taken to be the intensity differences normalised by the maximum magnitude intensity difference so that all elements lie between -1 and 1 , i.e. $\widetilde{\Delta I}(\phi_i) = \Delta I(\phi_i) / \max[|\Delta I(\phi_i)|]$. Similarly, the output is normalised, with a normalised analyte particle position $\tilde{\mathbf{r}}_a = 2\mathbf{r}_a / \max(L_x, L_y)$. Here, $L_{x,y}$ are the widths in the x and y directions of the box bounding the training data. This normalisation means all training data has a true output vector with elements between -1 and 1 , though it is possible for the neural network output to be outside this range.

6.3.2 Simulation Methods

Simulated data was generated via two methods, both discussed in Chapters 4 and 5. Both simulations assumed a rotationally symmetric analyte particle, so that there was no ξ_a dependence. The first method, described in Chapter 4, involved generating a Gaussian distributed background random speckle field and then adding the dipolar field for a dipole located at \mathbf{r}_a .

This method fails to account for any multiple scattering, but can rapidly generate a lot of data and therefore is useful for assessing the neural network tracker in the single scattering regime. Since the motivation in developing the neural network tracker was to enable tracking in the multiple scattering regime, a simulation method accounting for multiple scattering is required. This is achieved using the coupled dipole model discussed in Chapter 5. While this accounts for all scattering orders, it is significantly slower owing to the requirement to invert M , which is of size $3(N_{sc} + 1) \times 3(N_{sc} + 1)$ for N_{sc} background dipole scatterers, plus the analyte particle. In order to facilitate faster computation, the speckle profiles I_i were calculated on a 2D (x, y) grid with spacing $\lambda_0/30$, and then ΔI_i for any random point was calculated via linear interpolation from the grid at each observation angle ϕ_i . Furthermore, for coupling between dipoles, the scalar SPP Green's function defined in Eq. (5.79) was used, while the far field Green's tensor is used in propagating to the leakage ring. The use of these Green's tensors reduces M to an $(N_{sc} + 1) \times (N_{sc} + 1)$ matrix by virtue of the scalar coupling, while also ensuring that all Green's functions can be calculated analytically, rather than requiring numerical integration of Eq. (3.29) as would be required to find the full Green's tensor between different dipoles. While these approximations may not be fully justified, it should be emphasised that the purpose of these simulations is not to accurately describe a specific speckle pattern generated by a specific instance of disorder, but rather to verify that the neural network based tracking algorithm works in the multiple scattering regime and understand some of the important physical parameters affecting its performance. Thus, the fact that this model incorporates multiple scattering in a physically consistent manner is more important than its accuracy describing a specific physical arrangement. A final caveat for the coupled dipole simulations is that the analyte particle was restricted to a fixed plane $z = R_a$ above the surface, and only 2D localisation was performed.

Most simulations were performed at $\lambda_0 = 721\text{nm}$, for a 50nm thick gold film ($n_m = 0.13 + 4.24i$) on glass ($n_g = 1.51$) with water above the film ($n_d = 1.33$). The analyte particle was modelled as a homogeneous sphere of refractive index $n_p = 1.4$, while in the coupled dipole simulations, the background scatterers were modelled as 50nm radius gold spheres sat on the metal surface. The N_{sc} background scatterers were uniformly randomly distributed on the surface across a rectangle of dimensions $4L_{\text{SPP}} \times 4L_{\text{SPP}}$ (noting that the random configuration is fixed for a

given dataset).

6.3.3 Training

The neural network, implemented in a TensorFlow python framework [416], was trained on a set of N_{train} simulated speckle profiles with the analyte particle at uniformly distributed random positions \mathbf{r}_a in a box of dimensions $L_x \times L_y \times L_z$, or $L_x \times L_y$ for the 2D coupled dipole case. The coordinate system was chosen such that the (x, y) origin lay at the centre of the box (and also coinciding with the centre of the box within which the background scatterers are distributed in the coupled dipole simulations), so that $-L_x/2 \leq x_a \leq L_x/2$, $-L_y/2 \leq y_a \leq L_y/2$ and $R_a \leq z_a \leq L_z$. The loss function $L(\hat{\mathbf{r}}_a) = |\hat{\mathbf{r}}_a - \tilde{\mathbf{r}}_a|$ corresponded to the Euclidean distance between the (normalised) estimated and real position. Note that this is also an appropriate metric to measure the localisation performance, so in this case the loss is also used as a metric. The training used the ‘Adam’ SGD method for weight optimisation, with hyperparameters $\alpha_{\text{lr}} = 0.001$, $\beta_1 = 0.9$ and $\beta_2 = 0.999$ (from a combination of numerical experimentation and previous literature [380]). Early stopping (with the validation loss used as the validation metric) was used with a patience of 50 epochs. The training data was split so that 80% was used for training and 20% used for validation.

6.3.4 Neural Network Performance

The primary metric used to assess the performance of the trained neural networks is the mean total error $\langle |\mathbf{r}_a - \hat{\mathbf{r}}_a| \rangle$, where the average is taken over many different true values of \mathbf{r}_a . Unless otherwise stated, all mean error values are calculated from 2×10^5 random position realisations uniformly distributed with equal probability throughout the training box. Note that the test data is unseen, meaning it is not from the training data. An important parameter affecting the performance is the training set size, N_{train} (i.e. the number of different I_i speckle profiles for different \mathbf{r}_a provided to the neural network to learn from). Intuitively, the more speckle examples provided, the better the localisation precision as there is more likely to be an analyte particle position in the training data near a particular point. The training set size can be converted into an effective training sampling density $\rho_{\text{train}}^{3D} = N_{\text{train}}/L_x L_y L_z$ (in 2D $\rho_{\text{train}}^{2D} = N_{\text{train}}/L_x L_y$),

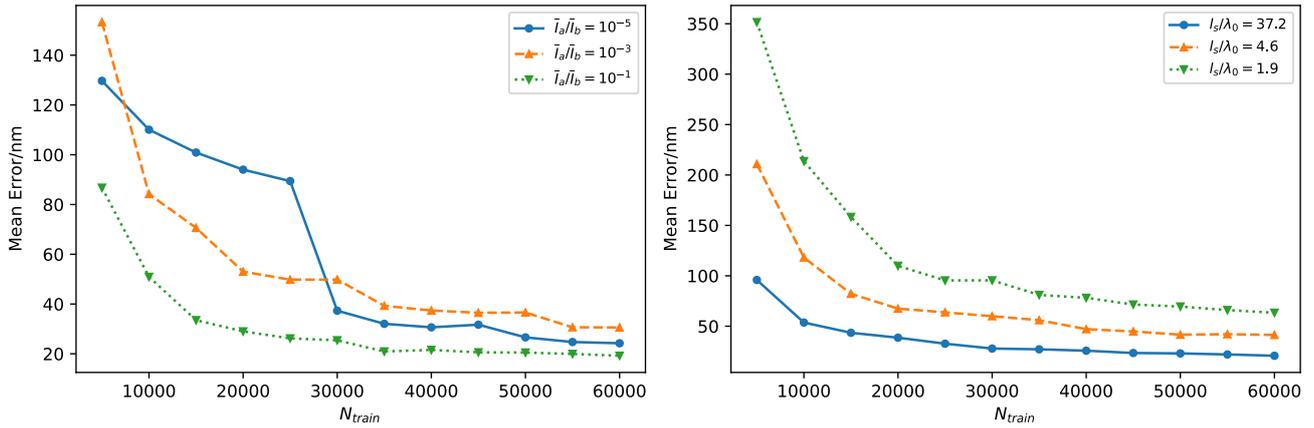


Figure 6.3: Mean error as a function of training set size for (left) Gaussian random vector simulated speckle at different ratios of particle scattered intensity to background speckle intensity \bar{I}_a/\bar{I}_b and (right) coupled dipole simulated speckle at different scattering mean free paths l_s (or equivalently scatterer densities).

and mean separation of training points $\Delta r_{\text{train}}^{3D} = (4\pi\rho_{\text{train}}^{3D}/3)^{-1/3}$ ($\Delta r_{\text{train}}^{2D} = (\pi\rho_{\text{train}}^{2D})^{-1/2}$). The dependence of the mean error on training set size is shown in Fig. 6.3 for both Gaussian simulated single scattering speckle and coupled dipole simulations with $L_x = L_y = L_{\text{SPP}}$ and $L_z = 2L_d$. Clearly, more training points does indeed improve the performance, with the random learning points covering the entire range of possible analyte particle positions, ensuring there are not patches of sensor surface that the neural network is unfamiliar with. However, beyond a sampling density corresponding to $N_{\text{train}} \sim 3 \times 10^4$ (equivalent to $\Delta r_{\text{train}}^{2D} \approx 18.7\text{nm}$ and $\Delta r_{\text{train}}^{3D} \approx 40.7\text{nm}$), adding additional learning points has little effect as the neural network has already learned the region. Once above this threshold of N_{train} , the single scattering results show slight improvement with the intensity ratio, defined as in Chapter 4 as the ratio of the spatially averaged analyte particle scattered and background speckle intensities around the ring \bar{I}_a/\bar{I}_b . Similarly, the mean error of the coupled dipole data increases with increased multiple scattering (smaller l_s), which can in part be attributed to multiple scattering effects, but the fact that the the average background intensity increases as more background scatterers are scattering the incident field, and thus \bar{I}_b/\bar{I}_a decreases as the density of scatterers increases, could also affect performance, which is not multiple scattering effect..

The sampling density can also be changed by altering the training region size, and the dependence on this is shown in Fig. 6.4, in which $L_x = L_y$ is varied with fixed $N_{\text{train}} = 5 \times 10^4$ (and fixed $L_z = 2L_d$ for 3D). Since the mean separation is linearly proportional to $(L_x L_y)^{1/2} = L_x$,

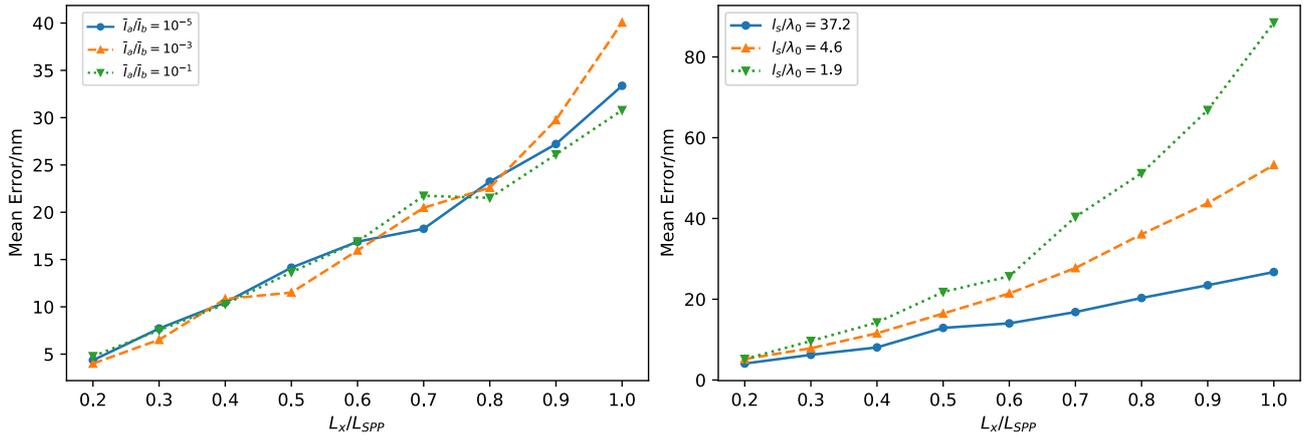


Figure 6.4: Mean error as a function of training region size L_x for (left) Gaussian random vector simulated speckle and (right) coupled dipole simulated speckle.

the mean error in the single scattering case (i.e. Gaussian speckle simulations and long mean free path coupled dipole simulations) decreases approximately linearly with L_x . In the multiple scattering case, the local field at the particle includes a near field speckle component scattered from the background scatterers, which the neural net must learn. As the training region decreases, there is less of this near field speckle to learn, and thus the mean error decreases more rapidly with decreasing L_x as compared to the linear decrease in the single scattering regime. This indicates, that while in the single scattering regime, the size of the training region has little effect on the mean error provided the ρ_{train} is kept constant, for multiple scattering, reducing the training region size is advantageous even if ρ_{train} is unchanged. This suggests a hierarchy of neural network localisation algorithms may be effective in the multiple scattering regime, in which an initial neural network coarsely localises the analyte particle, before the data is fed in to a different neural network trained only in the subregion the analyte has been localised to, allowing a more precise estimate of the position. The process could then be iterated with ever smaller subregions. Such approach does not require more training data, since the subset of data within the relevant region can be used to train the subregion neural networks, though it would require more training time to train multiple neural networks. There is little benefit to this approach in the single scattering regime as training density of the subregion is essentially the same as that of the entire region. The results presented in Figs. 6.3 and 6.4 arose from a neural network trained and tested on noiseless data, but it is important to understand the effect of noise on the performance. Fig. 6.5 shows the mean error as a function of SNR_{tot}

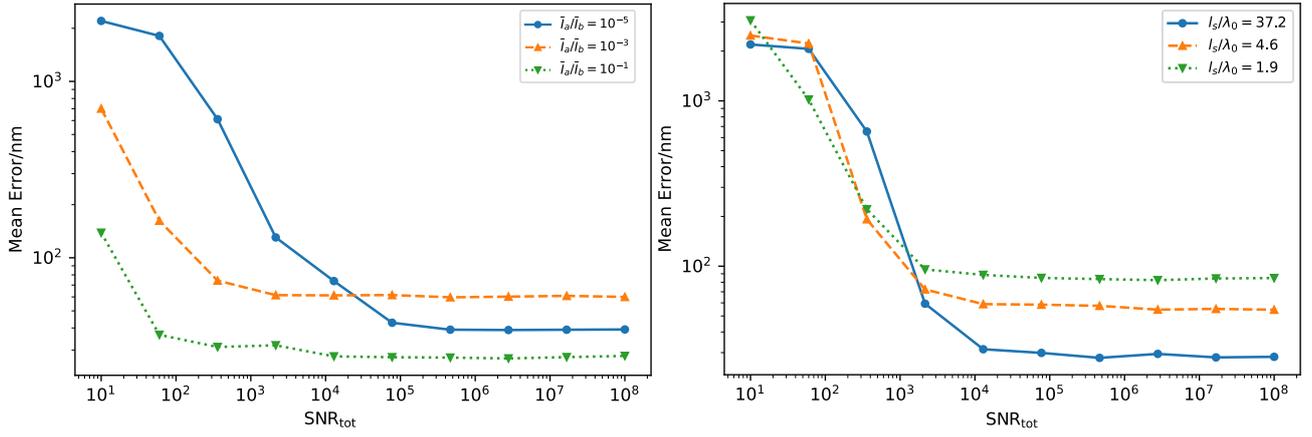


Figure 6.5: Mean error as a function of SNR_{tot} for (left) Gaussian random vector simulated speckle and (right) coupled dipole simulated speckle.

for different \bar{I}_a/\bar{I}_b and l_s with $N_{\text{train}} = 3 \times 10^4$ and $L_x = L_y = L_{\text{SPP}}$ fixed. The noise in each pixel is simulated via adding a random error to the simulated value of I_i drawn from a zero mean Gaussian distribution with standard deviation σ_i , with the SNR_{tot} defined as I_i/σ_i . From the Gaussian random speckle simulated data results, at high noise levels, larger \bar{I}_a/\bar{I}_b give a reduced mean error, as the intensity perturbation is larger compared to the noise, and therefore less obscured by noise. Note that the definition of SNR_{tot} differs from the SNR as defined in Chapter 2, in that it includes the background speckle as part of the signal. Following the Chapter 2 definition as the ratio of the perturbation ΔI_i relative to the noise, the SNR is defined by $\text{SNR} = \overline{\Delta I}_i/\sigma_i$, and the SNR dependence of coupled dipole simulations is shown in Fig. 6.6, for different values of l_s and N_{train} . Noticeably, while increasing the training set size improves peak performance at a given SNR as one would expect, it does not improve the robustness against noise in the sense that the neural network shows little dependence on SNR once $\text{SNR} \gtrsim 10$ regardless of the amount of training. Unsurprisingly, when the noise amplitude is larger or comparable to the size of the perturbation to the speckle pattern from scattering from the analyte particle, the error is large as the perturbation is not discernible from noise. Once the noise is less than about 10% of the signal perturbation, the neural network performance is essentially equivalent to the noiseless case.

Coupled dipole results presented so far consistently show, the mean error is smaller in the single scattering regime. For example, in Fig. 6.3, the case $l_s = 37.2\lambda_0 \gg L_{\text{SPP}} = 8.0\lambda_0$, firmly in the single scattering regimes, achieves localisation precision $\lesssim 50\text{nm}$, while the strongly

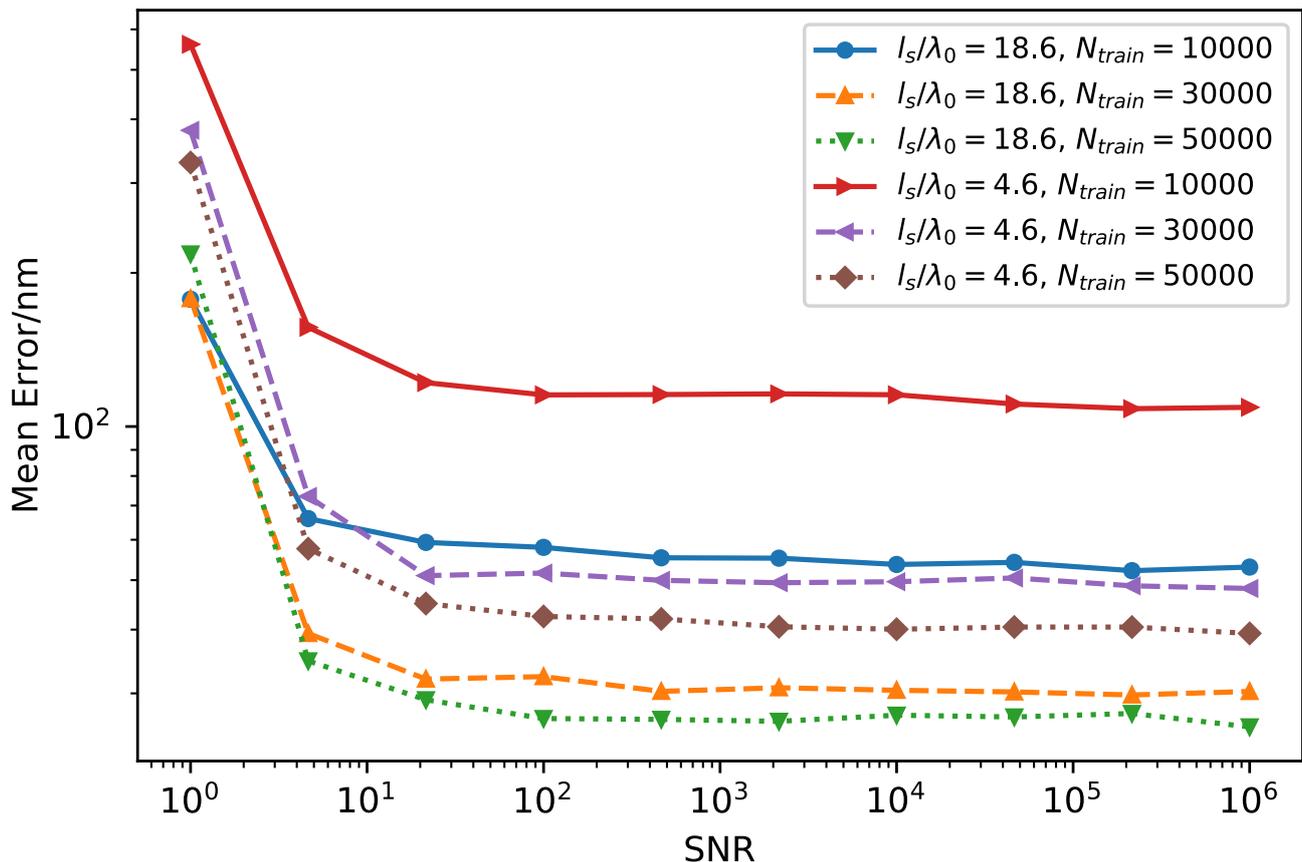


Figure 6.6: Mean error as a function of SNR for coupled dipole simulated speckle for different l_s and N_{train} .

multiple scattering case $l_s = 1.9\lambda_0 < L_{SPP}$ can only achieve a mean error $\sim 100\text{nm}$. This can be partly attributed to the fact that, in the single scattering case, the network is essentially learning the smoothly varying fringe pattern of Fig. 4.1, whereas in the multiple scattering case, as discussed in relation to Fig. 6.4, the network must learn the effects of the near field speckle pattern at \mathbf{r}_a . The effect of multiple scattering on the mean error can be seen in more detail in Fig. 6.7, showing the mean error as a function of l_s , for training sets of 50,000 and 100,000 and over training regions $L_x = L_y = L_{SPP}$ and $0.4L_{SPP}$. The mean error increases with increased multiple scattering, but not drastically, barring a spike near $l_s \sim 3\lambda_0$, which occurs for a few different random configurations of the background scatterers and thus does not appear to be arising due to the particular realisation the neural network was trained on. For the smaller training region, the mean error shows a weaker dependence on l_s , consistent with the notion that the reduced performance with multiple scattering is due to the neural network having to learn the near field speckle, and thus with less of the speckled SPP field near the sensor surface

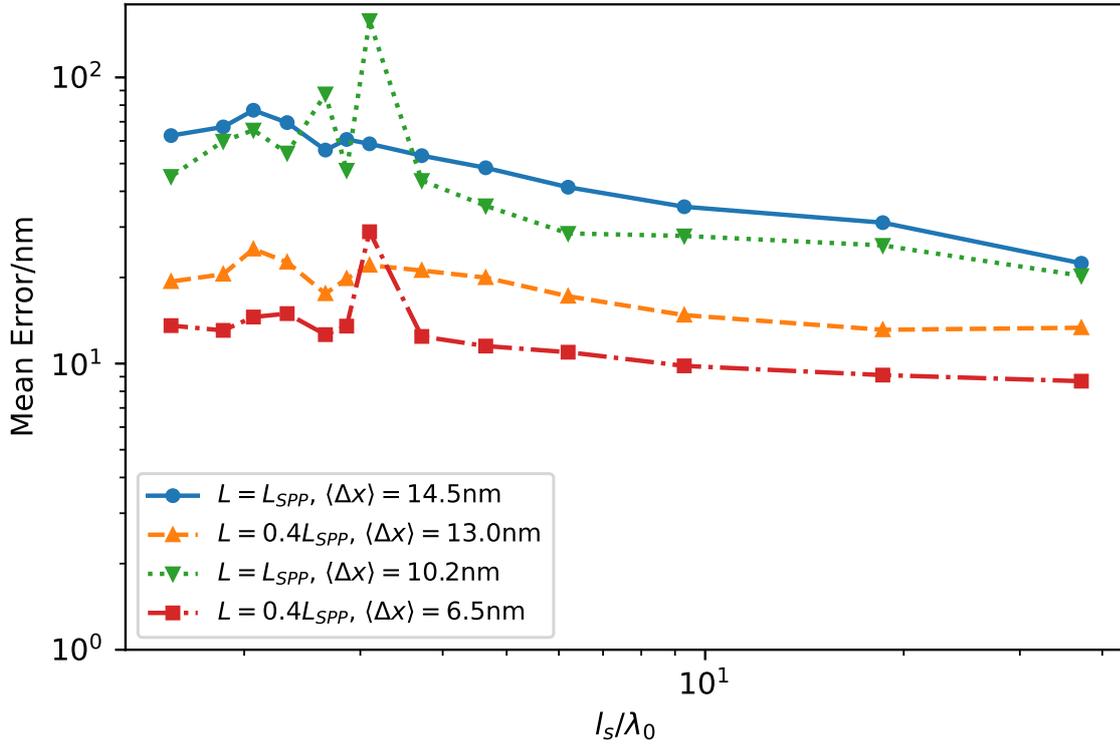


Figure 6.7: Mean error as a function of scattering mean free path for coupled dipole simulated speckle. The plots shown are for $N_{\text{train}} = 50,000$ (blue and green lines) and $N_{\text{train}} = 100,000$ (yellow and red lines), with corresponding mean separation of training points $\langle \Delta x \rangle$ shown.

to learn, the mean error does not suffer as much.

In Section 6.3.1, it was shown that, for sufficiently small scatterers where the interference term dominated and the effect of loop paths was negligible (i.e. $\gamma_2 \approx 1$), the normalisation of the neural network input should allow the results to generalise to analyte particles. This is an important property from a practical viewpoint, since to train the neural network on experimental data, one must generate data with known particle positions. Therefore, one must already have a method to localise the analyte particle. Additionally, it is desirable to apply a single trained neural network to any analyte particle, rather than having to train a neural network for each analyte particle, which would be impractically time consuming. The results shown in Fig. 6.8 show the error as a function of particle radius R_a for coupled dipole simulations for different mean free paths and training radii r_{tr} , defined as the radius of analyte particle from which training data was generated. Even in the presence of significant multiple scattering, a neural network trained for $r_{tr} \lesssim 100\text{nm}$ performs almost identically on different analyte particles in the

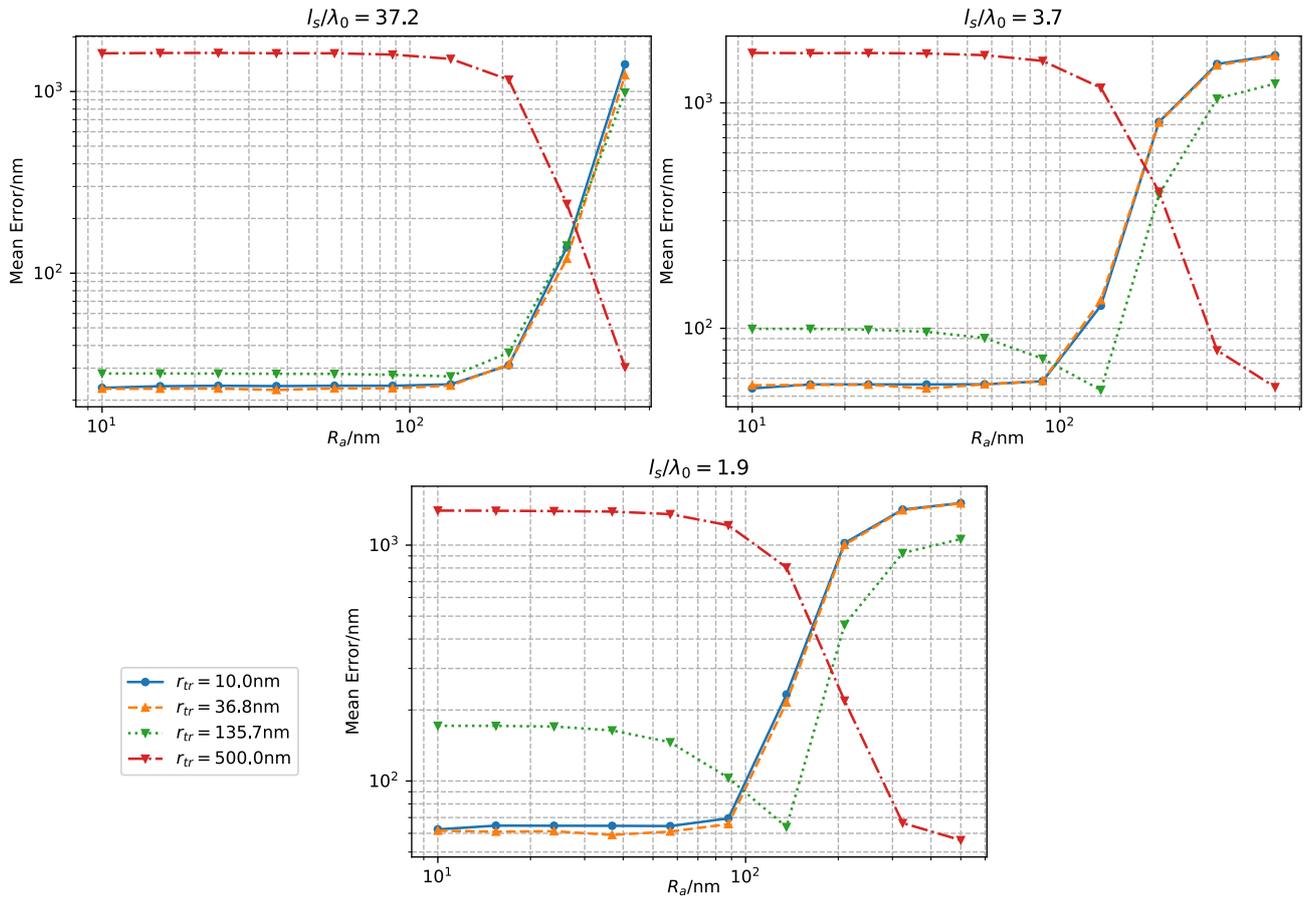


Figure 6.8: Mean error as a function of analyte particle radius, for different training radii r_{tr} and mean free path l_s .

same size range, while a neural network trained on a large particle (e.g. the $r_{tr} = 500nm$ case) does not generalise well to smaller analyte particles. It does appear that the intermediate case $r_{tr} = 135.7nm$, while generalising very well to smaller particles in the single scattering regime, generalises significantly worse in the presence of multiple scattering

As a result of this feature, provided the training analyte particle is not too large, the trained neural network can be applied to track a range of different small analyte particle. Thus, one can choose an easily localised particle to generate training data, such as one with a fluorescent label attached or one larger than the analyte particle one wishes to track (though not too much larger), before applying the trained neural network to track the desired analyte particle (e.g. the smaller, label-free particle), which cannot be tracked in an alternative manner. In addition, provided the phase of the polarisability, ϕ_α , does not change, the same principle applied to changing the analyte particle refractive index. This can be seen from the fact that

the interference terms is given (in the single scattering regime) by

$$\text{interference} = 2 \frac{k_0^2}{\varepsilon_0} \text{Re}[\mathbf{E}_b^*(\mathbf{r}) \cdot G(\mathbf{r}, \mathbf{r}_a) \alpha \mathbf{E}_0(\mathbf{r}_a)] \quad (6.4)$$

$$= 2 \frac{k_0^2}{\varepsilon_0} |\alpha| \text{Re}[\mathbf{E}_b^*(\mathbf{r}) \cdot G(\mathbf{r}, \mathbf{r}_a) e^{i\phi_\alpha} \mathbf{E}_0(\mathbf{r}_a)]. \quad (6.5)$$

For dielectric particles in a dielectric background medium where all permittivities are real, Eq. (2.6) gives a $\phi_\alpha = 0$ regardless of the magnitude of the particle permittivity. Provided there is not a large surface dressing effect, changing ε_a , provided it remains real, only changes $|\alpha|$, which has been normalised out of the neural network.

6.3.5 Comparison to Single Scattering Algorithm

Considering the results from Chapter 4, it can be seen that the mean error of the neural network localisation algorithm never approaches the best case (i.e. noiseless or low noise case) performance of the single scattering algorithm, which achieves errors as low as $\sim 1\text{nm}$. The single scattering algorithm benefits from the input of the additional information of the phase and amplitude shift functions, i.e. Eqs. (4.7) and (4.8). The neural network approach does, however bring in several advantages. Most obviously, and the main motivation behind developing the neural network approach, is the capability to work in the multiple scattering regime. Another benefit is that, while the single scattering algorithm only gives relative position shifts between frame, the neural network gives an absolute position in each frame. This can be advantageous if, for example, one was interested in the analyte particle position relative to some object on the surface, in order to study the interaction between them. As such, the neural network based approach remains of interest, especially if further performance improvements can enable it to achieve localisation precision $\lesssim 10\text{nm}$, comparable to state of the art localisation techniques. One possible extension is to work in the prior knowledge of the interference effects and changes in propagation phase into a neural network localisation algorithm in the manner of physics based machine learning [417, 418, 419], in which physics is used to provide constraints on the neural network outputs. For example, one approach could be to provide the theoretical phase and amplitude shifts of Eqs. (4.7) and (4.8) with an additional phase and amplitude shift term

arising from multiple scattering, which is learned by the neural network.

6.4 Conclusion

This chapter has demonstrated how the use of a neural network as a universal function approximator can be used to learn the link between a speckle pattern and the position of an analyte particle scattering light that contributes to the speckle pattern in order to localise the analyte particle. This idea, while applicable more generally to other (static) random scattering environments, has been explored in the context of the plasmonic random scattering system introduced and studied in earlier chapters, using numerical simulations. The minimum achievable errors, for a training region covering an area $L_{\text{SPP}} \times L_{\text{SPP}}$ of the surface, were found to be of the order $\sim 30\text{nm}$. In addition, the dependence of the neural network's performance on various parameters was quantified, and the difference between the single and multiple scattering regimes was studied. Importantly, while the algorithm did show a slight drop in performance in the multiple scattering regime, it could still localise to below 100nm accuracy, with the increasing mean error mitigated by reducing the training region. Furthermore, it was demonstrated that the normalisation of the input data allowed the generalisation of the neural network to different sized analyte particles, rather than requiring training on an identical analyte particle, a crucial feature in making any experimental realisation of this algorithm feasible.

Chapter 7

Conclusion

The primary aim of this thesis was to theoretically investigate techniques for sensing and tracking (or localisation) of single nanoscale biological molecules via the use of random scattering of SPPs. In particular, the sensing setup proposed in Chapter 4 (see Fig. 3.2) was investigated. This was motivated by the discussion in Chapter 2, in which it was seen that achieving label free sensing and tracking/localisation with single particle sensitivity in a relatively simple SPR-like experimental setup would be a step towards low cost, commercially available single particle sensors and trackers. Analysis of the sensing properties of the proposed system in the single scattering regime was performed in Chapter 3. The results of this analysis elucidated some of the benefits of the proposed sensing setup. In particular, the enhanced interaction between the electric field and the analyte particle due to plasmonic enhancement, the interferometric signal enhancement arising from interference between the light scattered from the analyte particle and the random surface scattered light and the strong confinement of light scattered from near the surface to the leakage radiation ring all give rise to larger scattered intensities which are beneficial to sensing. In addition, the randomness of the scattering from the surface means there is no stringent requirement on the quality of metal surfaces used. The results of Chapter 3 predict a fractional intensity change to a speckle pattern on the order of 10^{-4} when a 10nm analyte ‘biological’ particle approaches the surface and enters the sensing volume.

The fact that the measured intensity in the single scattering regime corresponded to the interference between a random speckle pattern and the field scattered from the analyte particle

suggested that the intensity would be sensitive to phase changes to the analyte particle scattered field, such as those that occur for small (sub-wavelength) shifts in the particle's position. In Chapter 4, an algorithm to track the analyte particle was derived, extracting the trajectory based only on intensity measurements in the leakage radiation ring. This algorithm was based on the fact that, ignoring multiple scattering effects, the phase and amplitude changes as the particle moved could be derived within a simple Born approximation (see Eqs. (4.7) and (4.8)). As such, the algorithm was only valid in the single scattering regime when multiple scattering effects are negligible. The use of three frames to simultaneously extract two shifts in particle position allowed for the elimination of the unknown speckle field phase and amplitude, while enabling a consistency check to be built in to the algorithm. When tested on simulated data, the algorithm was found to be capable of consistently achieving errors on the order of 1nm ($\sim 1\%$ error in estimating a 90nm shift in particle position) under optimal noise conditions. There was also a discussion of how the algorithm could be extended to more general situations via modifying the phase and amplitude shift functions, including potentially being applicable to other experimental setups with (static) random scattering, for example disordered systems supporting other surface waves, or even electromagnetic waves in a homogeneous medium.

The effect of multiple scattering on the sensing setup was investigated in Chapter 5. In order to facilitate this, three enhancement factors were derived within a coupled dipole model describing how multiple scattering alters the field perturbation arising from addition of an analyte particle, as compared to the single scattering perturbation. Each enhancement factor was given a physical interpretation in terms of an associated set of multiple scattering. The enhancement factors as presented in Eqs. (5.15), (5.19) and (5.20) apply generally to scenarios modelled by a coupled dipole model (or even more generally, any point scatterer wave scattering model). The statistics of these enhancement factors was studied for SPP scattering, both analytically and numerically through Monte Carlo simulations. The most significant feature was that, due to competition between stronger coupling with reduced scatterer separation and Anderson localisation effects at high scatterer density preventing coupling of dipoles separated by much more than a localisation length, there exists an optimum scatterer density at which the enhancement to the field perturbation amplitude is maximised. At this optimum density, the

mean total amplitude enhancement was found to be $\sim 10^2$, indicating the sensitivity of the SPP random scatterer sensor could be increased by up to two orders of magnitude by fabricating a random nanostructure of scatterers on the surface with a optimised number density. Furthermore, the statistics of the enhancement factors were found to have a dependence on the composition of the individual scatterers (through their polarisability), in particular in the high loss case where the suppression of long scattering paths results in greater sensitivity to the properties of a single scattering event. A notable feature of this dependence on background scatterer polarisability is that the optimum scatterer density and the maximum mean amplitude enhancement achieved display a dependence on the phase of the background scatterer polarisability. While present in the low loss case, this dependence is most prominent in the high loss case, where one can go from no optimum density (i.e. the mean amplitude enhancement is always below 1 and the multiple scattering regime is always less sensitive on average than the single scattering regime) to having an optimum density at which peak mean enhancements of $\sim 10^2$ are achieved. The optimum polarisability phase is well described by the divergence condition of Eq. (5.60), which corresponds to requiring the phase shift acquired in a single scattering event to vanish. The dependence on scatterer polarisability indicates that the random SPP scattering nanostructure based sensor could be further optimised by the tuning the parameters of the individual scatterers (in both composition and geometry).

Motivated both by the potential benefits of multiple scattering discovered in Chapter 5 and the fact that some degree of multiple scattering may be unavoidable in practical systems, Chapter 6 developed a neural network based algorithm to localise an analyte particle based on the speckle pattern measured around the leakage radiation ring, which did not rely on the assumption of negligible multiple scattering like the tracking algorithm described in Chapter 4. The neural network learns the link between the perturbation to the (fixed) background speckle and the analyte particle position based on training data of speckle pattern perturbations with known analyte particle positions. The algorithm was verified on simulations, with the best case performance giving a mean error of $\sim 30\text{nm}$ for a large training region (i.e. covering an area $L_{\text{SPP}} \times L_{\text{SPP}}$ of the surface). Reducing the size of the training region further improves performance due to the increased density of training points along with reduced region of SPP

speckle to learn, with a mean localisation error below 10nm possible for a region smaller than about $0.4L_{\text{SPP}} \times 0.4L_{\text{SPP}}$ ($0.4L_{\text{SPP}} \approx 2.30\mu\text{m}$). The mean error was found to only to increase slightly as multiple scattering increases, with the increase in mean error with decreasing scattering mean free path being reduced for smaller training regions. Furthermore, the normalisation of the input data was shown to allow the generalisation of a trained neural network to particles of different size and refractive index, provided the phase of the analyte particle polarisability is unchanged and the average background speckle intensity is significantly larger than the intensity scattered from the analyte particle used in training the network, such that the interference term dominates the intensity perturbation. As a result, the training data can be generated using an analyte particle which can be localised with some independent method (e.g. a fluorescent particle), but still applied to a wide range of different analyte particles. Importantly, this analyte particle size independence feature was shown to still hold in the presence of multiple scattering since the loop scattering paths, the effect of which is not normalised out in the neural network input, have negligible effect for small biological particles. The neural network used for the algorithm in Chapter 6 had a very simple dense, fully connected architecture with only 2 hidden layers. Given the wealth of different neural network architectures, it is plausible that further experimentation with different architectures, such as the use of CNNs, could improve performance, reducing the mean error or the amount of training data required. The amount of training data required for the neural network in Chapter 6 to achieve reasonable mean error $\lesssim 100\text{nm}$ ($N_{\text{train}} \gtrsim 10^5$) could be challenging to experimentally realise, and thus reducing this threshold would be desirable.

In light of the results of subsequent chapters, it is worth revisiting the summary of benefits of the proposed sensing/tracking setup discussed in Section 3.7 and adding a few additional benefits as follows

- simple SPR-like experimental setup
- strong interaction with analyte particle due to plasmonic confinement
- interferometric signal enhancement as discussed in Section 2.2.3
- no requirement for interferometric stability or external reference field

- stronger signal in the leakage radiation ring due to highly directional scattering compared to diffuse dipolar scattering in the absence of leaky modes
- no need for high quality surfaces due to random roughness
- dependence on phase difference gives rise to dependence of scattered intensity on \mathbf{r}_a , enabling tracking in the single scattering regime (Chapter 4) and multiple scattering regime (Chapter 6)
- speckle acts as the reference field, improving robustness against (static) random scatterers compared to techniques such as iSCAT that use a known reference field, which can be degraded by speckle
- surface scatterers can give rise to multiple scattering, which can possibly increase the size of the perturbation to the speckle pattern by up to two orders of magnitude from the single scattering case. This effect can be maximised by optimising the parameters of a random nanostructure of scatterers deliberately fabricated on the surface (Chapter 5).

Experimental verification and realisation of the theoretical results presented in this thesis are a work in progress. Recent research has been able to detect single proteins based on the interference of the light scattered from a protein bound to a receptor on a rough metal surface with the light scattered from the roughness, when SPPs are excited on the metal surface [204]. In this case, the scattered light was measured above the metal surface, as opposed to in the leakage ring. Similarly, a proof of principle experiment has shown the use of the leakage radiation speckle arising from random gold nanoislands on a gold film supporting SPPs to detect the binding of dielectric and gold nanoparticles, with the speckle Pearson correlation coefficient introduced in Section 3.6.1 acting as a sensing signal [420]. In addition to the development of experimental realisations of the proposed sensing setup investigated theoretically in this thesis, it is hoped that the work in this thesis can be applied to similar studies of random scattering systems for sensing or tracking applications, enabling further studies into extracting information from speckle patterns and how scattering parameters affect how well such information can be extracted. For example, the enhancement factors can be applied quite generally to many

wave scattering problems, and different types of waves may exhibit quite different behaviour to what was seen for SPP scattering in Chapter 5, especially for waves with different Anderson localisation properties.

References

- [1] M. Lelek, M. T. Gyparaki, G. Beliu, F. Schueder, J. Griffié, S. Manley, R. Jungmann, M. Sauer, M. Lakadamyali, and C. Zimmer, “Single-molecule localization microscopy,” *Nature Reviews Methods Primers*, vol. 1, no. 1, 2021.
- [2] S. Pujals, N. Feiner-Gracia, P. Delcanale, I. Voets, and L. Albertazzi, “Super-resolution microscopy as a powerful tool to study complex synthetic materials,” *Nat. Rev. Chem.*, vol. 3, no. 2, pp. 68–84, 2019.
- [3] E. J. Rees, M. Erdelyi, G. S. K. Schierle, A. Knight, and C. F. Kaminski, “Elements of image processing in localization microscopy,” *J. Opt.*, vol. 15, no. 9, p. 094012, 2013.
- [4] L. Möckl, A. R. Roy, and W. E. Moerner, “Deep learning in single-molecule microscopy: fundamentals, caveats, and recent developments [Invited],” *Biomedical Opt. Express*, vol. 11, no. 3, p. 1633, 2020.
- [5] M. Saxena, G. Eluru, and S. S. Gorthi, “Structured illumination microscopy,” *Adv. Opt. Photonics*, vol. 7, no. 2, pp. 241–275, 2015.
- [6] I. Moreno, J. A. Davis, T. M. Hernandez, D. M. Cottrell, and D. Sand, “Complete polarization control of light from a liquid crystal spatial light modulator,” *Opt. Express*, vol. 20, no. 1, pp. 364–376, 2012.
- [7] A. Von Diezmann, Y. Shechtman, and W. E. Moerner, “Three-Dimensional Localization of Single Molecules for Super-Resolution Imaging and Single-Particle Tracking,” *Chem. Rev.*, vol. 117, no. 11, pp. 7244–7275, 2017.

- [8] C. P. Toseland, “Fluorescent labeling and modification of proteins,” *J. Chem. Biol.*, vol. 6, no. 3, pp. 85–95, 2013.
- [9] E. C. Jensen, “Use of Fluorescent Probes: Their Effect on Cell Biology and Limitations,” *The Anatomical Record*, vol. 295, no. 12, pp. 2031–2036, 2012.
- [10] R. W. Taylor and V. Sandoghdar, “Interferometric Scattering Microscopy: Seeing Single Nanoparticles and Molecules via Rayleigh Scattering,” *Nano Lett.*, vol. 19, no. 8, pp. 4827–4835, 2019.
- [11] F. Vollmer, L. Yang, and S. Fainman, “Label-free detection with high-Q microcavities: A review of biosensing mechanisms for integrated devices,” *Nanophotonics*, vol. 1, no. 3-4, pp. 267–291, 2012.
- [12] D. Pines and D. Bohm, “A Collective Description of Electron Interactions: II. Collective vs Individual Particle Aspects of the Interactions,” *Phys. Rev.*, vol. 85, no. 2, pp. 338–353, 1952.
- [13] S. A. Maier, *Plasmonics: Fundamentals and applications*. Springer, 1 ed., 2007.
- [14] A. V. Zayats, I. I. Smolyaninov, and A. A. Maradudin, “Nano-optics of surface plasmon polaritons,” *Phys. Reports*, vol. 408, pp. 131–314, 2005.
- [15] H. Yu, Y. Peng, Y. Yang, and Z. Y. Li, “Plasmon-enhanced light–matter interactions and applications,” *npj Computational Materials*, vol. 5, no. 1, 2019.
- [16] J. Homola, S. S. Yee, and G. Gauglitz, “Surface plasmon resonance sensors: review,” *Sens. Actuators B Chem.*, vol. 54, pp. 3–15, 1999.
- [17] J. N. Anker, W. Paige Hall, O. Lyandres, N. C. Shah, J. Zhao, and R. P. Van Duyne, “Biosensing with plasmonic nanosensors,” *Nat. Mater.*, vol. 7, p. 442–453, 2008.
- [18] J. R. Mejía-Salazar and O. N. Oliveira, “Plasmonic Biosensing,” *Chem. Rev.*, vol. 118, no. 20, pp. 10617–10625, 2018.

- [19] J. Homola, “Present and future of surface plasmon resonance biosensors,” *Analytical and Bioanalytical Chemistry*, vol. 377, pp. 528–539, 2003.
- [20] P. Zijlstra, P. M. Paulo, and M. Orrit, “Optical detection of single non-absorbing molecules using the surface plasmon resonance of a gold nanorod,” *Nat. Nanotechnol.*, vol. 7, no. 6, pp. 379–382, 2012.
- [21] J. W. Goodman, *Speckle phenomena in optics: theory and applications*. Roberts and Company Publishers, 2005.
- [22] J. Stetefeld, S. A. Mckenna, and T. R. Patel, “Dynamic light scattering: a practical guide and applications in biomedical sciences,” *Biophys. Rev. s*, vol. 8, pp. 409–427, 2016.
- [23] D. J. Pine, D. A. Weitz, P. M. Chaikin, and E. Herbolzheimer, “Diffusing wave spectroscopy,” *Phys. Rev. Lett.*, 1988.
- [24] E. Akkermans and G. Montambaux, *Mesoscopic Physics of Electrons and Photons*. Cambridge University Press, 1 ed., 2007.
- [25] A. Dogariu and R. Carminati, “Electromagnetic field correlations in three-dimensional speckles,” *Phys. Rep.*, 2015.
- [26] S. Rotter and S. Gigan, “Light fields in complex media : mesoscopic scattering meets wave control,” *Rev. Mod. Phys.*, vol. 89, no. 1, p. 015005, 2017.
- [27] M. C. W. van Rossum and T. M. Nieuwenhuizen, “Multiple scattering of classical waves: microscopy, mesoscopy, and diffusion,” *Rev. Mod. Phys.*, 1999.
- [28] R. Berkovits and S. Feng, “Correlations in coherent multiple scattering,” *Phys. Rep.*, vol. 238, no. 3, pp. 135–172, 1994.
- [29] K. Belkebir, P. C. Chaumet, and A. Sentenac, “Influence of multiple scattering on three-dimensional imaging with optical diffraction tomography,” *J. Opt. Soc. Am. A*, vol. 23, no. 3, pp. 586–595, 2006.

- [30] Y. Choi, T. D. Yang, C. Fang-Yen, P. Kang, K. J. Lee, R. R. Dasari, M. S. Feld, and W. Choi, “Overcoming the diffraction limit using multiple light scattering in a highly disordered medium,” *Phys. Rev. Lett.*, vol. 107, no. 2, pp. 023902–, 2011.
- [31] P. Alonso-González, P. Albella, M. Schnell, J. Chen, F. Huth, A. García-Etxarri, F. Casanova, F. Golmar, L. Arzubiaga, L. E. Hueso, J. Aizpurua, and R. Hillenbrand, “Resolving the electromagnetic mechanism of surface-enhanced light scattering at single hot spots,” *Nat. Commun.*, vol. 3, no. 1, p. 684, 2012.
- [32] P. Mehrotra, “Biosensors and their applications – a review,” *J. Oral Biol. Craniofac. Res.*, vol. 6, no. 2, pp. 153–159, 2016.
- [33] D. Thevenot, K. Toth, R. Durst, G. Wilson, and G. W. Electrochemical, “Electrochemical biosensors: recommended definitions and classification,” *Biosens. Bioelectron.*, vol. 16, pp. 121–131, 2001.
- [34] D. A. Armbruster and T. Pry, “Limit of Blank, Limit of Detection and Limit of Quantitation,” *Clin. Biochem. Rev.*, vol. 29, pp. S49–S52, 2008.
- [35] J. Homola, “Surface plasmon resonance sensors for detection of chemical and biological species,” *Chem. Rev.*, vol. 108, no. 2, pp. 462–493, 2008.
- [36] F. Wang, W. B. Tan, Y. Zhang, X. Fan, and M. Wang, “Luminescent nanomaterials for biological labelling,” *Nanotechnology*, vol. 17, no. 1, pp. R1–R13, 2006.
- [37] L. Cagnet, C. Tardin, D. Boyer, D. Choquet, P. Tamarat, and B. Lounis, “Single metallic nanoparticle imaging for protein detection in cells,” *PNAS*, vol. 100, no. 20, pp. 11350–11355, 2003.
- [38] M. Zhang, L. Vojtech, Z. Ye, F. Hladik, and E. Nance, “Quantum Dot Labeling and Visualization of Extracellular Vesicles,” *ACS Appl. Nano Mater.*, vol. 3, no. 7, pp. 7211–7222, 2020.
- [39] M. S. T. Gonçalves, “Fluorescent Labeling of Biomolecules with Organic Probes,” *Chem. Rev.*, vol. 109, no. 1, pp. 190–212, 2009.

- [40] C. Martin, T. Yuan, E. Ghia, W. W. W, and P. D. C, “Green Fluorescent Protein as a Marker for Gene Expression,” *Science*, vol. 263, no. 5148, pp. 802–805, 1994.
- [41] L. Yin, W. Wang, S. Wang, F. Zhang, S. Zhang, and N. Tao, “How does fluorescent labeling affect the binding kinetics of proteins with intact cells?,” *Biosens. Bioelectron.*, vol. 66, pp. 412–416, 2015.
- [42] L. Ma and S. L. Cockroft, “Biological nanopores for single-molecule biophysics,” *Chem-BioChem*, vol. 11, no. 1, pp. 25–34, 2010.
- [43] C. Dekker, “Solid-state nanopores,” *Nat. Nanotechnol.*, vol. 2, no. 4, pp. 209–215, 2007.
- [44] M. Ayub and H. Bayley, “Engineered transmembrane pores,” *Current Opinion in Chemical Biology*, vol. 34, pp. 117–126, 2016.
- [45] A. Patel, E. Belykh, E. Miller, L. George, N. Martirosyan, V. Byvaltsev, and M. Preul, “MinION rapid sequencing: Review of potential applications in neurosurgery,” *Surgical Neurology International*, vol. 9, no. 1, p. 157, 2018.
- [46] P. Euskirchen, F. Bielle, K. Labreche, W. P. Kloosterman, S. Rosenberg, M. Daniau, C. Schmitt, J. Masliah-Planchon, F. Bourdeaut, C. Dehais, Y. Marie, J. Y. Delattre, and A. Idbaih, “Same-day genomic and epigenomic diagnosis of brain tumors using real-time nanopore sequencing,” *Acta Neuropathologica*, vol. 134, no. 5, pp. 691–703, 2017.
- [47] S. Li, C. Cao, J. Yang, and Y. T. Long, “Detection of Peptides with Different Charges and Lengths by Using the Aerolysin Nanopore,” *ChemElectroChem*, vol. 6, pp. 126–129, 2019.
- [48] S. Zhou, P. Tang, Y. J. Wang, L. Wang, and D. Q. Wang, “Applications of Nanopore Sensing in Detection of Toxic Molecules,” *Chinese J. Anal. Chem.*, 2018.
- [49] S. Rauf, L. Zhang, A. Ali, Y. Liu, and J. Li, “Label-Free Nanopore Biosensor for Rapid and Highly Sensitive Cocaine Detection in Complex Biological Fluids,” *ACS Sens.*, vol. 2, no. 2, pp. 227–234, 2017.
- [50] R. M. M. Smeets, U. F. Keyser, M. Y. Wu, N. H. Dekker, and C. Dekker, “Nanobubbles in Solid-State Nanopores,” *Phys. Rev. Lett.*, vol. 97, no. 8, p. 088101, 2006.

- [51] R. M. M. Smeets, U. F. Keyser, N. H. Dekker, and C. Dekker, “Noise in solid-state nanopores,” *PNAS*, vol. 105, no. 2, pp. 417–421, 2008.
- [52] D. Branton, D. W. Deamer, A. Marziali, H. Bayley, S. A. Benner, T. Butler, M. Di Ventra, S. Garaj, A. Hibbs, X. Huang, S. B. Jovanovich, P. S. Krstic, S. Lindsay, X. S. Ling, C. H. Mastrangelo, A. Meller, J. S. Oliver, Y. V. Pershin, J. M. Ramsey, R. Riehn, G. V. Soni, V. Tabard-Cossa, M. Wanunu, M. Wiggin, and J. A. Schloss, “The potential and challenges of nanopore sequencing,” *Nat. Biotechnol.*, vol. 26, no. 10, pp. 1146–1153, 2008.
- [53] F. Patolsky, G. Zheng, and C. M. Lieber, “Nanowire sensors for medicine and the life sciences,” *Nanomedicine*, vol. 1, pp. 51–65, 2006.
- [54] P. Xie, Q. Xiong, Y. Fang, Q. Qing, and C. M. Lieber, “Local electrical potential detection of DNA by nanowire-nanopore sensors HHS Public Access,” *Nat Nanotechnol.*, vol. 7, no. 2, pp. 119–125, 2011.
- [55] F. Patolsky and C. M. Lieber, “Nanowire nanosensors,” *Mater. Today*, vol. 8, no. 4, pp. 20–28, 2005.
- [56] A. Gao, N. Zou, P. Dai, N. Lu, T. Li, Y. Wang, J. Zhao, and H. Mao, “Signal-to-noise ratio enhancement of silicon nanowires biosensor with rolling circle amplification,” *Nano Lett.*, vol. 13, no. 9, pp. 4123–4130, 2013.
- [57] G. Zheng, F. Patolsky, Y. Cui, W. U. Wang, and C. M. Lieber, “Multiplexed electrical detection of cancer markers with nanowire sensor arrays,” *Nat. Biotechnol.*, vol. 23, no. 10, pp. 1294–1301, 2005.
- [58] M. J. Kim, M. Wanunu, D. C. Bell, and A. Meller, “Rapid fabrication of uniformly sized nanopores and nanopore arrays for parallel DNA analysis,” *Advanced Materials*, vol. 18, no. 23, pp. 3149–3153, 2006.
- [59] R. W. P. Drever, J. L. Hall, F. V. Kowalski, J. Hough, G. M. Ford, A. J. Munley, and H. Ward, “Laser Phase and Frequency Stabilization Using an Optical Resonator,” *Appl. Phys. B*, vol. 31, pp. 97–105, 1983.

- [60] M. R. Foreman, J. D. Swaim, and F. Vollmer, “Whispering gallery mode sensors,” *Adv. Opt. Photonics*, vol. 7, no. 3, p. 632, 2015.
- [61] M. R. Foreman, W.-L. Jin, and F. Vollmer, “Optimizing detection limits in whispering gallery mode biosensing,” *Opt. Express*, vol. 22, no. 5, p. 5491, 2014.
- [62] V. S. Ilchenko and L. Maleki, “Novel whispering-gallery resonators for lasers, modulators, and sensors,” in *Laser Resonators IV* (A. V. Kudryashov and A. H. Paxton, eds.), vol. 4270, p. 120 – 130, SPIE, 2001.
- [63] L. Shao, X. F. Jiang, X. C. Yu, B. B. Li, W. R. Clements, F. Vollmer, W. Wang, Y. F. Xiao, and Q. Gong, “Detection of single nanoparticles and lentiviruses using microcavity resonance broadening,” *Advanced Materials*, vol. 25, no. 39, pp. 5616–5620, 2013.
- [64] J. Zhu, S. K. Ozdemir, Y. F. Xiao, L. Li, L. He, D. R. Chen, and L. Yang, “On-chip single nanoparticle detection and sizing by mode splitting in an ultrahigh-Q microresonator,” *Nat. Photonics*, vol. 4, no. 1, pp. 46–49, 2010.
- [65] S. Rosenblum, Y. Lovsky, L. Arazi, F. Vollmer, and B. Dayan, “Cavity ring-up spectroscopy for ultrafast sensing with optical microresonators,” *Nat. Commun.*, vol. 6, no. 1, p. 6788, 2015.
- [66] H. R. Gwon and S. H. Lee, “Spectral and angular responses of surface plasmon resonance based on the kretschmann prism configuration,” *Mater. Trans.*, vol. 51, no. 6, pp. 1150–1155, 2010.
- [67] M. Puiu and C. Bala, “SPR and SPR imaging: Recent trends in developing nanodevices for detection and real-time monitoring of biomolecular events,” *Sensors (Switzerland)*, vol. 16, no. 6, 2016.
- [68] B. Sepúlveda, A. Calle, L. M. Lechuga, and G. Armelles, “Highly sensitive detection of biomolecules with the magneto-optic surface-plasmon-resonance sensor,” *Opt. Lett.*, vol. 31, no. 8, pp. 1085–1087, 2006.

- [69] C. T. Campbell and G. Kim, “Spr microscopy and its applications to high-throughput analyses of biomolecular binding events and their kinetics,” *Biomaterials*, vol. 28, no. 15, pp. 2380–2392, 2007. *Imaging Techniques for Biomaterials Characterization*.
- [70] J. Homola, *Surface plasmon resonance based sensors*, vol. 4. Springer Science & Business Media, 2006.
- [71] T. P. Burg, M. Godin, S. M. Knudsen, W. Shen, G. Carlson, J. S. Foster, K. Babcock, and S. R. Manalis, “Weighing of biomolecules, single cells and single nanoparticles in fluid,” *Nature*, vol. 446, no. 7139, pp. 1066–1069, 2007.
- [72] T. Thundat, E. A. Wachter, S. L. Sharp, and R. J. Warmack, “Detection of mercury vapor using resonating microcantilevers,” *Appl. Phys. Lett.*, vol. 66, no. 13, pp. 1695–1697, 2002.
- [73] J. L. Arlett, E. B. Myers, and M. L. Roukes, “Comparative advantages of mechanical biosensors,” *Nat. Nanotechnol.*, vol. 6, no. 4, pp. 203–215, 2011.
- [74] T. Braun, M. K. Ghatkesar, N. Backmann, W. Grange, P. Boulanger, L. Letellier, H. P. Lang, A. Bietsch, C. Gerber, and M. Hegner, “Quantitative time-resolved measurement of membrane protein-ligand interactions using microcantilever array sensors,” *Nat. Nanotechnol.*, vol. 4, no. 3, pp. 179–185, 2009.
- [75] A. Webster, F. Vollmer, and Y. Sato, “Probing biomechanical properties with a centrifugal force quartz crystal microbalance,” *Nat. Comm.*, vol. 5, no. 1, 2014.
- [76] K. Ekinici, “Electromechanical transducers at the nanoscale: Actuation and sensing of motion in nanoelectromechanical systems (nems),” *Small*, vol. 1, pp. 786–797, 2005.
- [77] E. V. Alieva and V. N. Konopsky, “Biosensor based on surface plasmon interferometry independent on variations of liquid’s refraction index,” *Sens. Actuators B Chem.*, vol. 99, no. 1, pp. 90–97, 2004.
- [78] M. Piliarik and J. Homola, “Surface plasmon resonance (SPR) sensors: approaching their limits?,” *Opt. Express*, vol. 17, no. 19, pp. 16505–16517, 2009.

- [79] H. H. Nguyen, J. Park, S. Kang, and M. Kim, “Surface plasmon resonance: A versatile technique for biosensor applications,” *Sensors*, vol. 15, no. 5, pp. 10481–10510, 2015.
- [80] J. S. Shumaker-Parry and C. T. Campbell, “Quantitative Methods for Spatially Resolved Adsorption/Desorption Measurements in Real Time by Surface Plasmon Resonance Microscopy,” *Anal. Chem.*, vol. 76, no. 4, pp. 907–917, 2004.
- [81] R. L. Rich and D. G. Myszka, “Higher-throughput, label-free, real-time molecular interaction analysis,” *Analytical Biochemistry*, vol. 361, no. 1, pp. 1–6, 2007.
- [82] M. Canovi, J. Lucchetti, M. Stravalaci, F. Re, D. Moscatelli, P. Bigini, M. Salmona, and M. Gobbi, “Applications of Surface Plasmon Resonance (SPR) for the Characterization of Nanoparticles Developed for Biomedical Purposes,” *Sensors*, vol. 12, pp. 16420–16432, 2012.
- [83] G. Colombo, B. Margosio, L. Ragona, M. Neves, S. Bonifacio, D. S. Annis, M. Stravalaci, S. Tomaselli, R. Giavazzi, M. Rusnati, M. Presta, Z. Lucia, D. F. Mosher, D. Ribatti, M. Gobbi, and G. Taraboletti, “Non-peptidic thrombospondin-1 mimics as fibroblast growth factor-2 inhibitors: an integrated strategy for the development of new antiangiogenic compounds,” *J. Biol. Chem.*, vol. 285, pp. 8733–8742, 2010.
- [84] V. Hodnik and G. Anderluh, “Toxin Detection by Surface Plasmon Resonance,” *Sensors*, vol. 9, pp. 1339–1354, 2009.
- [85] J. Jackson, *Classical Electrodynamics*. John Wiley & Sons, 1998.
- [86] X. Fan, W. Zheng, and D. J. Singh, “Light scattering and surface plasmons on small spherical particles,” *Light Sci. Appl.*, vol. 3, 2014.
- [87] H. Oka and Y. Ohdaira, “Simple model of saturable localised surface plasmon,” *Sci. Rep.*, vol. 8, no. 1, 2018.
- [88] R. Colom, A. Devilez, S. Enoch, B. Stout, and N. Bonod, “Polarizability expressions for predicting resonances in plasmonic and Mie scatterers,” *Phys. Rev. A*, vol. 95, no. 6, 2017.

- [89] K. M. Mayer and J. H. Hafner, “Localized surface plasmon resonance sensors,” *Chem. Rev.*, vol. 111, no. 6, pp. 3828–3857, 2011.
- [90] W. J. Galush, S. A. Shelby, M. J. Mulvihill, A. Tao, P. Yang, and J. T. Groves, “A Nanocube Plasmonic Sensor for Molecular Binding on Membrane Surfaces,” *Nano Lett.*, vol. 9, no. 5, pp. 2077–2082, 2009.
- [91] H. Bin Jeon, P. Vuka Tsalu, and J. Won Ha, “Shape Effect on the Refractive Index Sensitivity at Localized Surface Plasmon Resonance Inflection Points of Single Gold Nanocubes with Vertices,” *Sci. Rep.*, vol. 9, 2019.
- [92] L. Zhang, X. Li, Y. Wang, K. Sun, X. Chen, H. Chen, and J. Zhou, “Reproducible Plasmonic Nanopyramid Array of Various Metals for Highly Sensitive Refractometric and Surface-Enhanced Raman Biosensing,” *ACS Omega*, vol. 3, pp. 14181–14187, 2018.
- [93] J. Yang, H. Giessen, and P. Lalanne, “Simple Analytical Expression for the Peak-Frequency Shifts of Plasmonic Resonances for Sensing,” *Nano Lett.*, vol. 15, no. 5, pp. 3439–3444, 2015.
- [94] M. D. Baaske, N. Asgari, D. Punj, and M. Orrit, “Transient Optoplasmonic Detection of Single Proteins with Sub-Microsecond Resolution,” 2021. arXiv:2106.01881 [physics.optics].
- [95] A. Wax and K. Sokolov, “Molecular imaging and darkfield microspectroscopy of live cells using gold plasmonic nanoparticles,” *Laser Photonics Rev.*, vol. 3, no. 2, pp. 146–158, 2009.
- [96] K. Sokolov, M. Follen, J. Aaron, I. Pavlova, A. Malpica, R. Lotan, and R. Richards-Kortum, “Real-Time Vital Optical Imaging of Precancer Using Anti-Epidermal Growth Factor Receptor Antibodies Conjugated to Gold Nanoparticles 1,” *Cancer Research*, vol. 63, pp. 1999–2004, 2003.
- [97] M. Hentschel, M. Schäferling, X. Duan, H. Giessen, and N. Liu, “Chiral plasmonics,” *Science Advances*, vol. 3, no. 5, 2017.

- [98] S. Laing, L. E. Jamieson, K. Faulds, and D. Graham, “Surface-enhanced Raman spectroscopy for in vivo biosensing,” *Nat. Rev. Chem.*, vol. 1, p. 0060, 2017.
- [99] K. H. Rieder, S. Nie, and S. R. Emory, “Probing Single Molecules and Single Nanoparticles by Surface-Enhanced Raman Scattering,” *Science*, vol. 275, no. 1, pp. 1102–1106, 1997.
- [100] Y. Wang and J. Irudayaraj, “Surface-enhanced Raman spectroscopy at single-molecule scale and its implications in biology,” *Philos. Trans. R. Soc. B*, vol. 368, no. 1611, 2013.
- [101] I. Ament, J. Prasad, A. Henkel, S. Schmachtel, and C. S\“{o}nnichsen, “Single Unlabeled Protein Detection on Individual Plasmonic Nanoparticles,” *Nano Lett.*, vol. 12, pp. 1092–1095, 2012.
- [102] M. P. Raphael, J. A. Christodoulides, J. B. Delehanty, J. P. Long, P. E. Pehrsson, and J. M. Byers, “Quantitative LSPR imaging for biosensing with single nanostructure resolution,” *Biophysical Journal*, vol. 104, no. 1, pp. 30–36, 2013.
- [103] M. E. Stewart, C. R. Anderton, L. B. Thompson, J. Maria, S. K. Gray, J. A. Rogers, and R. G. Nuzzo, “Nanostructured Plasmonic Sensors,” *Chem. Rev.*, vol. 108, no. 2, pp. 494–521, 2008.
- [104] C. L. Haynes, A. D. McFarland, M. T. Smith, J. C. Hulteen, and R. P. Van Duyne, “Angle-Resolved Nanosphere Lithography: Manipulation of Nanoparticle Size, Shape, and Interparticle Spacing,” *J. Phys. Chem. B*, vol. 106, no. 8, pp. 1898–1902, 2002.
- [105] P. Gao, M. Pu, X. Ma, X. Li, Y. Guo, C. Wang, Z. Zhao, and X. Luo, “Plasmonic lithography for the fabrication of surface nanostructures with a feature size down to 9 nm,” *Nanoscale*, vol. 12, no. 4, pp. 2415–2421, 2020.
- [106] L. Gunnarsson, T. Rindzevicius, J. Prikulis, B. Kasemo, M. Käll, S. Zou, and G. C. Schatz, “Confined Plasmons in Nanofabricated Single Silver Particle Pairs: Experimental Observations of Strong Interparticle Interactions,” *J. Phys. Chem. B*, vol. 109, no. 3, pp. 1079–1087, 2005.

- [107] N. Nath and A. Chilkoti, “A Colorimetric Gold Nanoparticle Sensor To Interrogate Biomolecular Interactions in Real Time on a Surface,” *Anal. Chem.*, vol. 74, no. 3, pp. 504–509, 2002.
- [108] I. Ruach-Nir, T. A. Bendikov, I. Doron-Mor, Z. Barkay, A. Vaskevich, and I. Rubinstein, “Silica-Stabilized Gold Island Films for Transmission Localized Surface Plasmon Sensing,” *J. Am. Chem. Soc.*, vol. 129, no. 1, pp. 84–92, 2007.
- [109] T. Son, G. Moon, H. Lee, and D. Kim, “Metallic 3D Random Nanocomposite Islands For Near-Field Spatial Light Switching,” *Adv. Opt. Mater.*, vol. 6, no. 10, p. 1701219, 2018.
- [110] A. G. Brolo, R. Gordon, B. Leathem, and K. L. Kavanagh, “Surface Plasmon Sensor Based on the Enhanced Light Transmission through Arrays of Nanoholes in Gold Films,” *Langmuir*, vol. 20, no. 12, pp. 4813–4815, 2004.
- [111] A. De Leebeeck, L. K. S. Kumar, V. de Lange, D. Sinton, R. Gordon, and A. G. Brolo, “On-Chip Surface-Based Detection with Nanohole Arrays,” *Anal. Chem.*, vol. 79, no. 11, pp. 4094–4100, 2007.
- [112] A. I. Aristov, M. Manousidaki, A. Danilov, K. Terzaki, C. Fotakis, M. Farsari, and A. V. Kabashin, “3D plasmonic crystal metamaterials for ultra-sensitive biosensing,” *Sci. Rep.*, vol. 6, no. 1, p. 25380, 2016.
- [113] M. E. Stewart, N. H. Mack, V. Malyarchuk, J. A. N. T. Soares, T.-W. Lee, S. K. Gray, R. G. Nuzzo, and J. A. Rogers, “Quantitative multispectral biosensing and 1D imaging using quasi-3D plasmonic crystals,” *PNAS*, vol. 103, no. 46, pp. 17143–17148, 2006.
- [114] X. He, H. Yi, J. Long, X. Zhou, J. Yang, and T. Yang, “Plasmonic crystal cavity on single-mode optical fiber end facet for label-free biosensing,” *Appl. Phys. Lett.*, vol. 108, no. 23, p. 231105, 2016.
- [115] V. Malyarchuk, M. E. Stewart, R. G. Nuzzo, and J. A. Rogers, “Spatially resolved biosensing with a molded plasmonic crystal,” *Appl. Phys. Lett.*, vol. 90, no. 20, p. 203113, 2007.

- [116] A. B. Dahlin, J. O. Tegenfeldt, and F. Höök, “Improving the Instrumental Resolution of Sensors Based on Localized Surface Plasmon Resonance,” *Anal. Chem.*, vol. 78, no. 13, pp. 4416–4423, 2006.
- [117] A. Dahlin, M. Zäch, T. Rindzevicius, M. Käll, D. S. Sutherland, and F. Höök, “Localized Surface Plasmon Resonance Sensing of Lipid-Membrane-Mediated Biorecognition Events,” *J. Am. Chem. Soc.*, vol. 127, no. 14, pp. 5043–5048, 2005.
- [118] J. H. Wade and R. C. Bailey, “Applications of Optical Microcavity Resonators in Analytical Chemistry,” *Annual Rev. Anal. Chem.*, vol. 9, pp. 1–25, 2016.
- [119] W. Bogaerts, P. de Heyn, T. van Vaerenbergh, K. de Vos, S. Kumar Selvaraja, T. Claes, P. Dumon, P. Bienstman, D. van Thourhout, and R. Baets, “Silicon microring resonators,” *Laser Photonics Rev.*, vol. 6, pp. 47–73, 2012.
- [120] K. De Vos, I. Bartolozzi, E. Schacht, P. Bienstman, and R. Baets, “Silicon-on-Insulator microring resonator for sensitive and label-free biosensing,” *Opt. Express*, vol. 15, no. 12, p. 7610, 2007.
- [121] S. Lee, S. C. Eom, J. S. Chang, C. Huh, G. Y. Sung, and J. H. Shin, “A silicon nitride microdisk resonator with a 40-nm-thin horizontal air slot,” *Opt. Express*, vol. 18, no. 11, p. 11209, 2010.
- [122] D. K. Armani, T. J. Kippenberg, S. M. Spillane, and K. Vahala, “Ultra-high-Q toroid microcavity on a chip,” *Nature*, vol. 421, pp. 925–928, 2003.
- [123] P. Wang, J. Ward, Y. Yang, X. Feng, G. Brambilla, G. Farrell, and N. Chormaic, “Lead-silicate glass optical microbubble resonator,” *Appl. Phys. Lett.*, vol. 106, p. 61101, 2015.
- [124] V. Huet, A. Rasoloniaina, P. Guillemé, P. Rochard, P. Féron, M. Mortier, A. Levenson, K. Bencheikh, A. Yacomotti, and Y. Dumeige, “Millisecond Photon Lifetime in a Slow-Light Microcavity,” *Phys. Rev. Lett.*, vol. 116, no. 13, 2016.
- [125] A. A. Savchenkov, A. B. Matsko, V. S. Ilchenko, and L. Maleki, “Optical resonators with ten million finesse,” *Opt. Express*, vol. 15, no. 11, p. 6768, 2007.

- [126] S. Frustaci and F. Vollmer, “Whispering-gallery mode (WGM) sensors: review of established and WGM-based techniques to study protein conformational dynamics,” *Current Opinion in Chemical Biology*, vol. 51, pp. 66–73, 2019.
- [127] M. D. Baaske, M. R. Foreman, and F. Vollmer, “Single-molecule nucleic acid interactions monitored on a label-free microcavity biosensor platform,” *Nat. Nanotechnol.*, vol. 9, no. 11, pp. 933–939, 2014.
- [128] E. Kim, M. D. Baaske, and F. Vollmer, “In Situ Observation of Single-Molecule Surface Reactions from Low to High Affinities,” *Adv. Mater.*, vol. 28, pp. 9941–9948, 2016.
- [129] L. He, S. K. Ozdemir, J. Zhu, W. Kim, and L. Yang, “Detecting single viruses and nanoparticles using whispering gallery microlasers,” *Nat. Nanotechnol.*, vol. 6, pp. 428–432, 2011.
- [130] F. Sedlmeir, R. Zeltner, G. Leuchs, and H. G. Schwefel, “High-Q MgF₂ whispering gallery mode resonators for refractometric sensing in aqueous environment,” *Opt. Express*, vol. 22, no. 25, p. 30934, 2014.
- [131] V. N. Konopsky and E. V. Alieva, “Photonic crystal surface mode imaging biosensor based on wavelength interrogation of resonance peak,” *Sens. Actuators B Chem.*, vol. 276, pp. 271–278, 2018.
- [132] G. Pitruzzello and T. F. Krauss, “Photonic crystal resonances for sensing and imaging,” *J. Opt.*, vol. 20, no. 7, p. 073004, 2018.
- [133] Q. Huang and B. T. Cunningham, “Microcavity-Mediated Spectrally Tunable Amplification of Absorption in Plasmonic Nanoantennas,” *Nano Lett.*, vol. 19, no. 8, pp. 5297–5303, 2019.
- [134] J. N. Liu, Q. Huang, K. K. Liu, S. Singamaneni, and B. T. Cunningham, “Nanoantenna-Microcavity Hybrids with Highly Cooperative Plasmonic-Photonic Coupling,” *Nano Lett.*, vol. 17, no. 12, pp. 7569–7577, 2017.

- [135] F. De Angelis, M. Patrini, G. Das, I. Maksymov, M. Galli, L. Businaro, L. C. Andreani, and E. Di Fabrizio, "A hybrid plasmonic-photonic nanodevice for label-free detection of a few molecules," *Nano Lett.*, vol. 8, no. 8, pp. 2321–2327, 2008.
- [136] S. Ghosh, N. Li, Y. Xiong, Y.-G. Ju, M. P. Rathslag, E. G. Onal, E. Falkiewicz, M. Kohli, B. T. Cunningham, E. G. Onal, E. Falkiewicz, M. Kohli, B. T. Cunningham, and B. T. Cunningham, "A compact photonic resonator absorption microscope for point of care digital resolution nucleic acid molecular diagnostics," *Biomedical Opt. Express*, vol. 12, no. 8, p. 4637, 2021.
- [137] K. Takemura, "Surface plasmon resonance (SPR)-and localized SPR (LSPR)-based virus sensing systems: Optical vibration of nano-and micro-metallic materials for the development of next-generation virus detection technology," *Biosensors*, vol. 11, no. 250, 2021.
- [138] Y.-F. Chang, W.-H. Wang, Y.-W. Hong, R.-Y. Yuan, K.-H. Chen, Y.-W. Huang, P.-L. Lu, Y.-H. Chen, Y.-M. Arthur Chen, L.-C. Su, and S.-F. Wang, "Simple Strategy for Rapid and Sensitive Detection of Avian Influenza A H7N9 Virus Based on Intensity-Modulated SPR Biosensor and New Generated Antibody," *Anal. Chem.*, vol. 90, pp. 1861–1869, 2018.
- [139] Y. Wang, H. Wang, Y. Chen, Y. Wang, H.-Y. Chen, X. Shan, and N. Tao, "Fast Electrochemical and Plasmonic Detection Reveals Multitime Scale Conformational Gating of Electron Transfer in Cytochrome c," *J. Am. Chem. Soc.*, vol. 139, 2017.
- [140] X.-L. Zhou, Y. Yang, S. Wang, and X.-W. Liu, "Surface Plasmon Resonance Microscopy: From Single-Molecule Sensing to Single-Cell Imaging," *Angew Chem Int Ed Engl.*, vol. 59, no. 5, pp. 1776–1785, 2020.
- [141] S. S. Aćimović, H. Šípová-Jungová, G. Emilsson, L. Shao, A. B. Dahlin, M. Käll, and T. J. Antosiewicz, "Antibody-antigen interaction dynamics revealed by analysis of single-molecule equilibrium fluctuations on individual plasmonic nanoparticle biosensors," *ACS Nano*, vol. 12, no. 10, pp. 9958–9965, 2018.

- [142] M. R. Foreman, D. Keng, E. Treasurer, J. R. Lopez, and S. Arnold, “Whispering gallery mode single nanoparticle detection and sizing: the validity of the dipole approximation,” *Opt. Lett.*, vol. 42, no. 5, pp. 963–966, 2017.
- [143] T. J. Kippenberg, J. Kalkman, A. Polman, and K. J. Vahala, “Demonstration of an erbium-doped microdisk laser on a silicon chip,” *Phys. Rev. A*, vol. 74, no. 5, p. 051802, 2006.
- [144] Y. F. Chen, X. Serey, R. Sarkar, P. Chen, and D. Erickson, “Controlled photonic manipulation of proteins and other nanomaterials,” *Nano Lett.*, vol. 12, no. 3, pp. 1633–1637, 2012.
- [145] F. Wang, R. E. Christiansen, Y. Yu, J. Mørk, and O. Sigmund, “Maximizing the quality factor to mode volume ratio for ultra-small photonic crystal cavities,” *Appl. Phys. Lett.*, vol. 113, no. 24, p. 241101, 2018.
- [146] J. Chaste, A. Eichler, J. Moser, G. Ceballos, R. Rurali, and A. Bachtold, “A nanomechanical mass sensor with yoctogram resolution,” *Nat. Nanotechnol.*, vol. 7, pp. 301–304, 2012.
- [147] C. Jiang, Q. Li, J. Huang, S. Bi, R. Ji, and Q. Guo, “Single-Layer MoS₂ Mechanical Resonant Piezo-Sensors with High Mass Sensitivity,” *ACS Appl. Mater. Interfaces*, vol. 12, no. 37, pp. 41991–41998, 2020.
- [148] P. S. Waggoner and H. G. Craighead, “Micro- and nanomechanical sensors for environmental, chemical, and biological detection,” *Lab on a Chip*, vol. 7, no. 10, pp. 1238–1255, 2007.
- [149] S. Sorgenfrei, C. Y. Chiu, R. L. Gonzalez, Y. J. Yu, P. Kim, C. Nuckolls, and K. L. Shepard, “Label-free single-molecule detection of DNA-hybridization kinetics with a carbon nanotube field-effect transistor,” *Nat. Nanotechnol.*, vol. 6, no. 2, pp. 126–132, 2011.

- [150] Y. Li, C. Yang, and X. Guo, “Single-Molecule Electrical Detection: A Promising Route toward the Fundamental Limits of Chemistry and Life Science,” *Acc. Chem. Res.*, vol. 53, pp. 159–169, 2020.
- [151] B. Tian, J. Liu, T. Dvir, L. Jin, J. H. Tsui, Q. Qing, Z. Suo, R. Langer, D. S. Kohane, and C. M. Lieber, “Macroporous nanowire nanoelectronic scaffolds for synthetic tissues,” *Nat. Mater.*, vol. 11, pp. 986–994, 2012.
- [152] W. Zhou, X. Dai, and C. M. Lieber, “Advances in nanowire bioelectronics,” *Rep. Prog. Phys.*, vol. 80, p. 016701, 2017.
- [153] J. H. Ahn, J. Y. Kim, K. Choi, D. I. Moon, C. H. Kim, M. L. Seol, T. J. Park, S. Y. Lee, and Y. K. Choi, “Nanowire FET biosensors on a bulk silicon substrate,” *IEEE Trans. Electron Devices*, vol. 59, no. 8, pp. 2243–2249, 2012.
- [154] S. Zafar, C. D’Emic, A. Jagtiani, E. Kratschmer, X. Miao, Y. Zhu, R. Mo, N. Sosa, H. Hamann, G. Shahidi, and H. Riel, “Silicon Nanowire Field Effect Transistor Sensors with Minimal Sensor-to-Sensor Variations and Enhanced Sensing Characteristics,” *ACS Nano*, vol. 12, no. 7, pp. 6577–6587, 2018.
- [155] K. J. Freedman, L. M. Otto, A. P. Ivanov, A. Barik, S. H. Oh, and J. B. Edel, “Nanopore sensing at ultra-low concentrations using single-molecule dielectrophoretic trapping,” *Nat. Commun.*, vol. 7, 2016.
- [156] C. G. Brown and J. Clarke, “Nanopore development at Oxford Nanopore,” *Nat. Biotechnol.*, vol. 34, no. 8, p. 810, 2016.
- [157] R. W. Taylor, R. G. Mahmoodabadi, V. Rauschenberger, A. Giessel, A. Schambony, and V. Sandoghdar, “Interferometric scattering microscopy reveals microsecond nanoscopic protein motion on a live cell membrane,” *Nat. Photonics*, vol. 13, no. 7, pp. 480–487, 2019.
- [158] C.-L. Hsieh, S. Spindler, J. Ehrig, and V. Sandoghdar, “Tracking Single Particles on Supported Lipid Membranes: Multimobility Diffusion and Nanoscopic Confinement,” *J. Phys. Chem. B*, vol. 118, no. 6, pp. 1545–1554, 2014.

- [159] U. Endesfelder, S. van de Linde, S. Wolter, M. Sauer, and M. Heilemann, “Subdiffraction-Resolution Fluorescence Microscopy of Myosin–Actin Motility,” *ChemPhysChem*, vol. 11, no. 4, pp. 836–840, 2010.
- [160] K. Sun, C. Zhao, X. Zeng, Y. Chen, X. Jiang, X. Ding, L. Gou, H. Xie, X. Li, X. Zhang, S. Lin, L. Dou, L. Wei, H. Niu, M. Zhang, R. Tian, E. Sawyer, Q. Yuan, Y. Huang, P. Chen, C. Zhao, C. Zhou, B. Ying, B. Shi, X. Wei, R. Jiang, L. Zhang, G. Lu, and J. Geng, “Active DNA unwinding and transport by a membrane-adapted helicase nanopore,” *Nat. Commun.*, vol. 10, p. 5083, 2019.
- [161] K. S. Karunatilaka, D. J. Filman, M. Strauss, J. J. Loparo, and J. M. Hogle, “Real-time imaging of polioviral rna translocation across a membrane,” *mBio*, vol. 12, no. 1, pp. 1–16, 2021.
- [162] H. Shen, L. J. Tauzin, R. Baiyasi, W. Wang, N. Moringo, B. Shuang, and C. F. Landes, “Single Particle Tracking: From Theory to Biophysical Applications,” *Chem. Rev.*, vol. 117, no. 11, pp. 7331–7376, 2017.
- [163] J. Goodman, *Introduction to Fourier Optics*. Roberts & Company Publishers, 3 ed., 2005.
- [164] S. Zhang, B. Hu, P. Sebbah, and A. Z. Genack, “Speckle evolution of diffusive and localized waves,” *Phys. Rev. Lett.*, vol. 99, no. 6, p. 63902, 2007.
- [165] Y. Shechtman, L. E. Weiss, A. S. Backer, M. Y. Lee, and W. E. Moerner, “Multicolour localization microscopy by point-spread-function engineering,” *Nat. Photonics*, 2016.
- [166] K. I. Mortensen, L. S. Churchman, J. A. Spudich, and H. Flyvbjerg, “Optimized localization analysis for single-molecule tracking and super-resolution microscopy,” *Nat. Methods*, vol. 7, no. 5, pp. 377–381, 2010.
- [167] U. Kubitscheck, “Fluorescence Microscopy: Single Particle Tracking,” in *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*, pp. 579–583, Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.

- [168] H. Ueno, S. Nishikawa, R. Iino, K. V. Tabata, S. Sakakihara, T. Yanagida, and H. Noji, “Simple Dark-Field Microscopy with Nanometer Spatial Precision and Microsecond Temporal Resolution,” *Biophysical Journal*, vol. 98, no. 9, pp. 2014–2023, 2010.
- [169] S. Ram, J. Chao, P. Prabhat, E. S. Ward, and R. J. Ober, “A novel approach to determining the three-dimensional location of microscopic objects with applications to 3D particle tracking,” in *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XIV*, vol. 6443, p. 64430D, SPIE, 2007.
- [170] S. R. P. Pavani and R. Piestun, “Three dimensional tracking of fluorescent microparticles using a photon-limited double-helix response system,” *Opt. Express*, vol. 16, no. 26, p. 22048, 2008.
- [171] S. R. P. Pavani, M. A. Thompson, J. S. Biteen, S. J. Lord, N. Liu, R. J. Twieg, R. Piestun, and W. E. Moerner, “Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function,” *PNAS*, vol. 106, no. 9, pp. 2995–2999, 2009.
- [172] Y. Shechtman, S. J. Sahl, A. S. Backer, and W. Moerner, “Optimal Point Spread Function Design for 3D Imaging,” *Phys. Rev. Lett.*, vol. 113, no. 13, p. 133902, 2014.
- [173] Y. Shechtman, L. E. Weiss, A. S. Backer, S. J. Sahl, and W. E. Moerner, “Precise Three-Dimensional Scan-Free Multiple-Particle Tracking over Large Axial Ranges with Tetrapod Point Spread Functions,” *Nano Lett.*, vol. 15, no. 6, pp. 4194—4199, 2015.
- [174] N. Opatovski, Y. Shalev Ezra, L. E. Weiss, B. Ferdman, R. Orange-Kedem, and Y. Shechtman, “Multiplexed PSF Engineering for Three-Dimensional Multicolor Particle Tracking,” *Nano Lett.*, vol. 21, no. 13, pp. 5888–5895, 2021.
- [175] M. Liebel, J. O. Arroyo, V. S. Beltrán, J. Osmond, A. Jo, H. Lee, R. Quidant, and N. F. van Hulst, “3d tracking of extracellular vesicles by holographic fluorescence imaging,” *Sci. Adv.*, vol. 6, no. 45, p. eabc2508, 2020.

- [176] Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, “Deep learning microscopy,” *Optica*, vol. 4, no. 11, p. 1437, 2017.
- [177] J. A. Grant-Jacob, B. S. Mackay, J. A. G. Baker, Y. Xie, D. J. Heath, M. Loxham, R. W. Eason, and B. Mills, “A neural lens for super-resolution biological imaging,” *J. Phys. Commun.*, vol. 3, no. 6, p. 065004, 2019.
- [178] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, “Imaging Intracellular Fluorescent Proteins at Nanometer Resolution,” *Science*, vol. 313, no. 5793, pp. 1642–1645, 2006.
- [179] M. J. Rust, M. Bates, and X. Zhuang, “Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM),” *Nat. Methods*, vol. 3, no. 10, pp. 793–795, 2006.
- [180] S. W. Hell and J. Wichmann, “Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy,” *Opt. Lett.*, vol. 19, no. 11, p. 780, 1994.
- [181] T. A. Klar and S. W. Hell, “Subdiffraction resolution in far-field fluorescence microscopy,” *Opt. Lett.*, vol. 24, no. 14, p. 954, 1999.
- [182] T. Müller, C. Schumann, and A. Kraegeloh, “STED microscopy and its applications: New insights into cellular processes on the nanoscale,” *ChemPhysChem*, vol. 13, no. 8, pp. 1986–2000, 2012.
- [183] H. Blom and H. Brismar, “STED microscopy: Increased resolution for medical research?,” *J. Intern. Med.*, vol. 276, no. 6, pp. 560–578, 2014.
- [184] J. R. Moffitt, C. Osseforth, and J. Michaelis, “Time-gating improves the spatial resolution of STED microscopy,” *Opt. Express*, vol. 19, no. 5, p. 4242, 2011.
- [185] G. Vicidomini, A. Schönl, H. Ta, K. Y. Han, G. Moneron, C. Eggeling, and S. W. Hell, “STED Nanoscopy with Time-Gated Detection: Theoretical and Experimental Aspects,” *PLoS One*, vol. 8, no. 1, p. e54421, 2013.

- [186] N. T. Urban, M. R. Foreman, S. W. Hell, and Y. Sivan, “Nanoparticle-Assisted STED Nanoscopy with Gold Nanospheres,” *ACS Photonics*, vol. 5, no. 7, pp. 2574–2583, 2018.
- [187] E. Cortés, P. A. Huidobro, H. G. Sinclair, S. Guldbbrand, W. J. Peveler, T. Davies, S. Parrinello, F. Görlitz, C. Dunsby, M. A. A. Neil, Y. Sivan, I. P. Parkin, P. M. W. French, and S. A. Maier, “Plasmonic Nanoprobes for Stimulated Emission Depletion Nanoscopy,” *ACS Nano*, vol. 10, no. 11, pp. 10454–10461, 2016.
- [188] Y. Sonnefraud, H. G. Sinclair, Y. Sivan, M. R. Foreman, C. W. Dunsby, M. A. A. Neil, P. M. French, and S. A. Maier, “Experimental Proof of Concept of Nanoparticle-Assisted STED,” *Nano Lett.*, vol. 14, no. 8, pp. 4449–4453, 2014.
- [189] D. Wildanger, B. R. Patton, H. Schill, L. Marseglia, J. P. Hadden, S. Knauer, A. Schönle, J. G. Rarity, J. L. O’Brien, S. W. Hell, and J. M. Smith, “Solid immersion facilitates fluorescence microscopy with nanometer resolution and sub-Ångström emitter localization,” *Advanced Materials*, vol. 24, no. 44, 2012.
- [190] F. Bergermann, L. Alber, S. J. Sahl, J. Engelhardt, and S. W. Hell, “2000-fold parallelized dual-color STED fluorescence nanoscopy,” *Opt. Express*, vol. 23, no. 1, p. 211, 2015.
- [191] K. C. Gwosch, J. K. Pape, and F. Balzarotti, “MINFLUX nanoscopy delivers 3D multicolor nanometer resolution in cells,” *Nat. Methods*, vol. 17, pp. 217–224, 2020.
- [192] F. Balzarotti, Y. Eilers, K. C. Gwosch, A. H. Gynnå, V. Westphal, F. D. Stefani, J. Elf, and S. W. Hell, “Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes,” *Science*, vol. 355, pp. 606–612, 2017.
- [193] S. A. Jones, S.-H. Shim, J. He, and X. Zhuang, “Fast, three-dimensional super-resolution imaging of live cells,” *Nat. Methods*, vol. 8, no. 6, pp. 499–508, 2011.
- [194] R. W. Taylor and V. Sandoghdar, *Interferometric Scattering (iSCAT) Microscopy and Related Techniques*, pp. 25–65. Springer International Publishing, 2019.

- [195] K. Lindfors, T. Kalkbrenner, P. Stoller, and V. Sandoghdar, “Detection and Spectroscopy of Gold Nanoparticles Using Supercontinuum White Light Confocal Microscopy,” *Phys. Rev. Lett.*, vol. 93, no. 3, 2004.
- [196] P. Kukura, H. Ewers, C. Müller, A. Renn, A. Helenius, and V. Sandoghdar, “High-speed nanoscopic tracking of the position and orientation of a single virus,” *Nat. Methods*, vol. 6, no. 12, pp. 923–927, 2009.
- [197] V. Jacobsen, P. Stoller, C. Brunner, V. Vogel, and V. Sandoghdar, “Interferometric optical detection and tracking of very small gold nanoparticles at a water-glass interface,” *Opt. Express*, vol. 14, no. 1, p. 405, 2006.
- [198] M. Xiao, L.-A. Wu, and H. J. Kimble, “Precision measurement beyond the shot-noise limit,” *Phys. Rev. Lett.*, vol. 59, pp. 278–281, 1987.
- [199] M. Konnik and J. Welsh, “High-level numerical simulations of noise in CCD and CMOS photosensors: review and tutorial,” 2014. arXiv:1412.4031 [astro-ph.IM].
- [200] J. C. Mullikin, L. J. v. Vliet, H. Netten, F. R. Boddeke, G. v. d. Feltz, and I. T. Young, “Methods for CCD camera characterization,” in *Proc.SPIE* (H. C. Titus and A. Waks, eds.), vol. 2173, pp. 73–84, 1994.
- [201] M. Liebel, J. T. Hugall, and N. F. Van Hulst, “Ultrasensitive Label-Free Nanosensing and High-Speed Tracking of Single Proteins,” *Nano Lett.*, vol. 17, no. 2, pp. 1277–1281, 2017.
- [202] D. Cole, G. Young, A. Weigel, A. Sebesta, and P. Kukura, “Label-Free Single-Molecule Imaging with Numerical-Aperture-Shaped Interferometric Scattering Microscopy,” *ACS Photonics*, vol. 4, no. 2, pp. 211–216, 2017.
- [203] Y. Yang, G. Shen, H. Wang, H. Li, T. Zhang, N. Tao, X. Ding, and H. Yu, “Interferometric plasmonic imaging and detection of single exosomes,” *PNAS*, vol. 115, no. 41, pp. 10275–10280, 2018.

- [204] P. Zhang, G. Ma, W. Dong, Z. Wan, S. Wang, and N. Tao, “Plasmonic scattering imaging of single proteins and binding kinetics,” *Nat. Methods*, vol. 17, no. 10, pp. 1010–1017, 2020.
- [205] N. Li, T. D. Canady, Q. Huang, X. Wang, G. A. Fried, and B. T. Cunningham, “Photonic resonator interferometric scattering microscopy,” *Nat. Commun.*, vol. 12, no. 1, pp. 1–9, 2021.
- [206] M. A. Doucey and S. Carrara, “Nanowire Sensors in Cancer,” *Trends Biotechnol.*, vol. 37, no. 1, pp. 86–99, 2019.
- [207] X. Shi, Z. Zhang, L. Ge, j. Xiang, and S. Yi, “Dispersion manipulation of multilayer dielectric plasmonic metasurfaces,” *Phys. Lett. A*, vol. 395, p. 127225, 2021.
- [208] D. Sarid, “Long-Range Surface-Plasma Waves on Very Thin Metal Films,” *Phys. Rev. Lett.*, vol. 47, no. 26, pp. 1927–1930, 1981.
- [209] K. Welford and J. Sambles, “Coupled Surface Plasmons in a Symmetric System,” *J. Mod. Opt.*, vol. 35, no. 9, pp. 1467–1483, 1988.
- [210] A. P. Vinogradov, A. V. Dorofeenko, A. A. Pukhov, and A. A. Lisyansky, “Exciting surface plasmon polaritons in the Kretschmann configuration by a light beam,” *Phys. Rev. B*, 2018.
- [211] A. Drezet, A. Hohenau, D. Koller, A. Stepanov, H. Ditlbacher, B. Steinberger, F. R. Aussenegg, A. Leitner, and J. R. Krenn, “Leakage radiation microscopy of surface plasmon polaritons,” *Mater. Sci. Eng. B*, vol. 149, pp. 220–229, 2008.
- [212] S. I. Bozhevolnyi, V. S. Volkov, and K. Leosson, “Localization and Waveguiding of Surface Plasmon Polaritons in Random Nanostructures,” *Phys. Rev. Lett.*, vol. 89, no. 18, p. 186801, 2002.
- [213] T. Søndergaard and S. I. Bozhevolnyi, “Surface plasmon polariton scattering by a small particle placed near a metal surface: An analytical study,” *Phys. Rev. B*, vol. 69, no. 4, p. 045422, 2004.

- [214] L. Dal Negro and S. V. Boriskina, “Deterministic aperiodic nanostructures for photonics and plasmonics applications,” *Laser Photonics Rev.*, vol. 6, no. 2, pp. 178–218, 2012.
- [215] E. Le Moal, S. Lévêque-Fort, M.-C. Potier, and E. Fort, “Nanoroughened plasmonic films for enhanced biosensing detection,” *Nanotechnology*, vol. 20, no. 22, p. 225502, 2009.
- [216] J. Cesario, M. U. Gonzalez, S. Cheylan, W. L. Barnes, S. Enoch, and R. Quidant, “Coupling localized and extended plasmons to improve the light extraction through metal films,” *Opt. Express*, vol. 15, no. 17, p. 10533, 2007.
- [217] K. Arya and R. Zeyher, “Optical response from a randomly rough surface: Theory and numerical results,” *Phys. Rev. B*, vol. 28, no. 8, pp. 4090–4105, 1983.
- [218] D. C. Prieve and N. A. Frej, “Total internal reflection microscopy: a quantitative tool for the measurement of colloidal forces,” *Langmuir*, vol. 6, no. 2, pp. 396–403, 1990.
- [219] S. Y. Buhmann, *Dispersion Forces II*, vol. 248 of *Springer Tracts in Modern Physics*. Springer-Verlag Berlin Heidelberg, 1 ed., 2012.
- [220] Z. Cheng, “Localization and the field intensity of surface-phonon polaritons,” *Phys. Rev. B*, vol. 51, no. 15, 1995.
- [221] M. I. Mishchenko, L. D. Travis, and A. A. Lacis, *Multiple Scattering of Light by Particles: Radiative Transfer and Coherent Backscattering*. Cambridge University Press, 1 ed., 2006.
- [222] J. J. Sakurai, *Modern Quantum Mechanics*. Cambridge University Press, 1986.
- [223] A. R. Holt and B. Santoso, “On the born series for scattering amplitudes,” *J. Phys. B*, vol. 7, no. 9, 1974.
- [224] P. Sheng, *Scattering and Localization of Classical Waves in Random Media*. World Scientific, 1990.
- [225] G. Berginc, “Dyson equation for electromagnetic scattering of heterogeneous media with spatially disordered particles: Properties of the effective medium,” in *Journal of Physics: Conference Series*, vol. 738, Institute of Physics Publishing, 2016.

- [226] O. J. Martin, C. Girard, and A. Dereux, “Generalized field propagator for electromagnetic scattering and light confinement,” *Phys. Rev. Lett.*, 1995.
- [227] O. Keller, “Attached and radiated electromagnetic fields of an electric point dipole,” *J. Opt. Soc. Am. B*, vol. 16, no. 5, pp. 835–847, 1999.
- [228] L. Novotny and B. Hecht, *Principles of nano-optics*. Cambridge University Press, 2 ed., 2012.
- [229] W. C. Chew, *Waves and fields in inhomogeneous media*, vol. 16. Wiley-IEEE Press, 1995.
- [230] S. V. Zhukovsky, O. Kidwai, and J. E. Sipe, “Physical nature of volume plasmon polaritons in hyperbolic metamaterials,” *Opt. Express*, vol. 21, no. 12, p. 14982, 2013.
- [231] A. Delfan, I. Degli-Eredi, and J. E. Sipe, “Long-range surface plasmons in multilayer structures,” *J. Opt. Soc. Am. B*, vol. 32, no. 8, p. 1615, 2015.
- [232] A. Pitelet, E. Mallet, E. Centeno, and A. Moreau, “Fresnel coefficients and Fabry-Perot formula for spatially dispersive metallic layers,” *Phys. Rev. B*, vol. 96, no. 4, p. 041406, 2017.
- [233] J. P. McClure and R. Wong, “Multidimensional stationary phase approximation: boundary stationary point,” *J. Comput. Appl. Math.*, 1990.
- [234] J. W. Goodman, *Statistical optics*. John Wiley & Sons, 2015.
- [235] P. N. Pusey and W. Van Megen, “Dynamic light scattering by non-ergodic media,” *Physica A*, vol. 157, no. 2, pp. 705–741, 1989.
- [236] P. Zakharov, “Ergodic and non-ergodic regimes in temporal laser speckle imaging,” *Opt. Lett.*, vol. 42, no. 12, 2017.
- [237] K. A. O’Donnell, “Speckle statistics of doubly scattered light,” *J. Opt. Soc. Am.*, vol. 72, no. 11, p. 1459, 1982.

- [238] J. W. Goodman, “Role Of Coherence Concepts In The Study Of Speckle;,” in *Proc. SPIE 0194, Applications of Optical Coherence* (W. H. Carter, ed.), vol. 0194, pp. 86–94, SPIE, 1979.
- [239] N. Shnerb and M. Kaveh, “Non-Rayleigh statistics of waves in random systems,” *Phys. Rev. B*, vol. 43, no. 1, pp. 1279–1282, 1991.
- [240] E. Kogan and M. Kaveh, “Statistics of Waves Propagating in a Random Medium,” *Found. Phys.*, vol. 26, no. 5, pp. 679–690, 1996.
- [241] A. Z. Genack and N. Garcia, “Intensity Statistics and Correlation in Absorbing Random Media,” *Europhys. Lett.*, vol. 21, no. 7, pp. 753–758, 1993.
- [242] S. Sangu, T. Okamoto, J. Uozumi, and T. Asakura, “Effect of absorption on surface speckles in random media,” *Waves Random Media*, vol. 9, no. 1, pp. 27–36, 1999.
- [243] Y. Bromberg and H. Cao, “Generating Non-Rayleigh Speckles with Tailored Intensity Statistics,” *Phys. Rev. Lett.*, vol. 112, no. 21, p. 213904, 2014.
- [244] A. A. Maradudin and D. L. Mills, “Scattering and absorption of electromagnetic radiation by a semi-infinite medium in the presence of surface roughness,” *Phys. Rev. B*, 1975.
- [245] A. Maradudin and W. Zierau, “Effects of surface roughness on the surface-polariton dispersion relation,” *Phys. Rev. B*, vol. 14, no. 2, pp. 484–499, 1976.
- [246] M. R. Foreman, “Field Correlations in Surface Plasmon Speckle,” *Sci. Rep.*, vol. 9, no. 1, p. 8359, 2019.
- [247] G. Rasigni, F. Varnier, M. Rasigni, J. P. Palmari, and A. Llebaria, “Autocovariance functions, root-mean-square-roughness height, and autocovariance length for rough deposits of copper, silver, and gold,” *Phys. Rev. B*, vol. 25, no. 4, 1982.
- [248] J. Laverdant, S. Buil, B. Bérini, and X. Quélin, “Polarization dependent near-field speckle of random gold films,” *Phys. Rev. B*, vol. 77, no. 16, 2008.

- [249] J. M. Elson and R. H. Ritchie, “Photon interactions at a rough metal surface,” *Phys. Rev. B*, 1971.
- [250] P. Bousquet, F. Flory, and P. Roche, “Scattering from multilayer thin films: theory and experiment,” *J. Opt. Soc. Am.*, vol. 71, no. 9, p. 1115, 1981.
- [251] K. Arya, R. Zeyher, and A. A. Maradudin, “Electric field enhancement near a randomly rough metal surface,” *Solid State Commun.*, vol. 42, no. 6, pp. 461–464, 1982.
- [252] K. Arya, Z. B. Su, and J. L. Birman, “Localization of the Surface Plasmon Polariton Caused by Random Roughness and its Role in Surface-Enhanced Optical Phenomena,” *Phys. Rev. Lett.*, vol. 54, no. 14, pp. 1559–1562, 1985.
- [253] P. Lehmann, “Surface-roughness measurement based on the intensity correlation function of scattered light under speckle-pattern illumination,” *Appl. Opt.*, vol. 38, no. 7, pp. 1144–1152, 1999.
- [254] A. Altland and B. Simons, *Condensed Matter Field Theory*. Cambridge University Press, 2 ed., 2010.
- [255] G. A. Farias and A. A. Maradudin, “Surface plasmons on a randomly rough surface,” *Phys. Rev. B*, vol. 28, no. 10, pp. 5675–5687, 1983.
- [256] E. Wolf, “Coherence properties of partially polarized electromagnetic radiation,” *Nuovo Cim.*, vol. 13, no. 6, pp. 1165–1181, 1959.
- [257] N. Wiener, “Coherency Matrices and Quantum,” *J. Math. and Phys.*, vol. 7, pp. 109–125, 1928.
- [258] P. Amblard, M. Gaeta, and J. Lacoume, “Statistics for complex variables and signals — part i: Variables,” *Signal Processing*, vol. 53, no. 1, pp. 1–13, 1996.
- [259] X.-B. Hu, M.-X. Dong, Z.-H. Zhu, W. Gao, and C. Rosales-Guzmán, “Does the structure of light influence the speckle size?,” *Sci. Rep.*, vol. 10, p. 199, 2020.

- [260] T. M. Habashy, R. W. Groom, and B. R. Spies, “Beyond the Born and Rytov approximations: A nonlinear approach to electromagnetic scattering,” *J. Geophys. Res.*, vol. 98, pp. 1759–1775, 1993.
- [261] K. Belkebir, P. C. Chaumet, and A. Sentenac, “Superresolution in total internal reflection tomography,” *J. Opt. Soc. Am. A*, vol. 22, no. 9, p. 1889, 2005.
- [262] B. Stout, A. Devilez, B. Rolly, and N. Bonod, “Multipole methods for nanoantennas design: applications to Yagi-Uda configurations,” *J. Opt. Soc. Am. B*, vol. 28, no. 5, pp. 1213–1223, 2011.
- [263] R. Colom, A. Devilez, N. Bonod, and B. Stout, “Optimal interactions of light with magnetic and electric resonant particles,” *Phys. Rev. B*, vol. 93, no. 4, 2016.
- [264] H. C. Van De Hulst, *Light scattering by small particles*. Courier Corporation, 2003.
- [265] P. de Vries, D. V. van Coevorden, and A. Lagendijk, “Point scatterers for classical waves,” *Rev. Mod. Phys.*, 1998.
- [266] X. J. Liang, A. Q. Liu, C. S. Lim, T. C. Ayi, and P. H. Yap, “Determining refractive index of single living cell using an integrated microchip,” *Sens. Actuators A Phys.*, 2007.
- [267] T. L. McMeekin, M. L. Groves, and N. J. Hipp, “Refractive Indices of Amino Acids, Proteins, and Related Substances,” in *Amino Acids and Serum Proteins*, vol. 44 of *Advances in Chemistry*, ch. 4, pp. 54–66, American Chemical Society, 1964.
- [268] Y. Pang, H. Song, and W. Cheng, “Using optical trap to measure the refractive index of a single animal virus in culture fluid with high precision,” *Biomed. Opt. Express*, vol. 7, no. 5, pp. 1672–1689, 2016.
- [269] A. J. Cox, A. J. DeWeerd, and J. Linden, “An experiment to measure Mie and Rayleigh total scattering cross sections,” *Am. J. Phys.*, vol. 70, no. 6, pp. 620–625, 2002.
- [270] E. Goulielmakis, M. Uiberacker, R. Kienberger, A. Baltuska, V. Yakovlev, A. Scrinzi, T. Westerwalbesloh, U. Kleineberg, U. Heinzmann, M. Drescher, and F. Krausz, “Direct Measurement of Light Waves,” *Science*, vol. 305, pp. 1267–1269, 2004.

- [271] M. Piliarik and V. Sandoghdar, “Direct optical sensing of single unlabelled proteins and super-resolution imaging of their binding sites,” *Nat Commun*, vol. 5, no. 4459, 2014.
- [272] A. Webster, *Novel Methods in Optical and Mechanical Biosensors*. PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2018.
- [273] J. Berk, C. Paterson, and M. R. Foreman, “Tracking Single Particles Using Surface Plasmon Leakage Radiation Speckle,” *J. Light. Technol.*, vol. 39, no. 12, pp. 3950–3960, 2021.
- [274] D. Dominguez, N. Alharbi, M. Alhusain, A. A. Bernussi, and L. G. D. Peralta, “Fourier plane imaging microscopy,” *J. Appl. Phys.*, vol. 116, no. 10, 2014.
- [275] J. H. Bruning, D. R. Herriott, J. E. Gallagher, D. P. Rosenfeld, A. D. White, and D. J. Brangaccio, “Digital Wavefront Measuring Interferometer for Testing Optical Surfaces and Lenses,” *J. Opt. Soc. Am. A*, vol. 13, no. 11, pp. 2693–2703, 1974.
- [276] G. Lai and T. Yatagai, “Generalized phase-shifting interferometry,” *J. Opt. Soc. Am. A*, vol. 8, no. 5, pp. 822–827, 1991.
- [277] R. Martí, “Multi-Start Methods,” in *Handbook of Metaheuristics* (F. Glover and G. A. Kochenberger, eds.), pp. 355–368, Springer US, 2003.
- [278] J. Kim, H. Lee, S. Im, S. A. Lee, D. Kim, and K.-A. Toh, “Machine learning-based leaky momentum prediction of plasmonic random nanosubstrate,” *Opt. Express*, vol. 29, no. 19, p. 30625, 2021.
- [279] G. P. Baxter, L. L. Burgess, and H. W. Daudt, “The refractive index of water,” *J. Am. Chem. Soc.*, vol. 33, no. 6, pp. 893–901, 1911.
- [280] P. B. Johnson and R. W. Christy, “Optical constants of the noble metals,” *Phys. Rev. B*, vol. 6, pp. 4370–4379, 1972.
- [281] C. Min, Z. Shen, J. Shen, Y. Zhang, H. Fang, G. Yuan, L. Du, S. Zhu, T. Lei, and X. Yuan, “Focused plasmonic trapping of metallic particles,” *Nat. Commun.*, vol. 4, 2013.

- [282] V. Garcés-Chávez, R. Quidant, P. J. Reece, G. Badenes, L. Torner, and K. Dholakia, “Extended organization of colloidal microparticles by surface plasmon polariton excitation,” *Phys. Rev. B*, vol. 73, no. 8, p. 085417, 2006.
- [283] J. Y. Wong, T. L. Kuhl, J. N. Israelachvili, N. Mullah, and S. Zalipsky, “Direct Measurement of a Tethered Ligand-Receptor Interaction Potential,” *J. Am. Chem. Soc.*, vol. 275, no. 5301, pp. 820–822, 1997.
- [284] P. J. Van Laarhoven and E. H. Aarts, “Simulated annealing,” in *Simulated annealing: Theory and applications*, pp. 7–15, Springer, 1987.
- [285] R. Jungmann, C. Steinhauer, M. Scheible, A. Kuzyk, P. Tinnefeld, and F. C. Simmel, “Single-molecule kinetics and super-resolution microscopy by fluorescence imaging of transient binding on dna origami,” *Nano Lett.*, vol. 10, no. 11, pp. 4756–4761, 2010.
- [286] T. G. Baboolal, G. I. Mashanov, T. A. Nenasheva, M. Peckham, and J. E. Molloy, “A combination of diffusion and active translocation localizes myosin 10 to the filopodial tip,” *J. Biol. Chem.*, vol. 291, no. 43, pp. 22373–22385, 2016.
- [287] J. Berk and M. R. Foreman, “Theory of Multiple Scattering Enhanced Single Particle Plasmonic Sensing,” *ACS Photonics*, vol. 8, no. 8, pp. 2227–2233, 2021.
- [288] J. Berk and M. R. Foreman, “Role of multiple scattering in single particle perturbations in absorbing random media,” *Phys. Rev. Res.*, vol. 3, no. 3, p. 033111, 2021.
- [289] J. A. DeSanto and G. G. Zipfel, “Scattering of a scalar wave from a random rough surface: A diagrammatic approach,” *J. Math. Phys.*, vol. 13, pp. 1903–1911, 1974.
- [290] D. Hornauer, H. Kapitza, and H. Raether, “The dispersion relation of surface plasmons on rough surfaces,” *J. Phys. D: Appl. Phys.*, vol. 7, pp. L100–L102, 1974.
- [291] E. Kroger and E. Kretschmann, “Surface Plasmon and Polariton Dispersion at Rough Boundaries,” *Phys. Stat. Solid. B*, vol. 76, pp. 515–523, 1976.

- [292] D. Hall and A. Braundmeier, “Dispersion relation for surface plasmons on randomly rough surfaces: A quantum-mechanical approach,” *Phys. Rev. B*, vol. 17, no. 10, pp. 3808–3813, 1978.
- [293] E. Kretschmann, T. L. Ferrel, and J. C. Ashley, “Splitting of the Dispersion Relation of Surface Plasmons on a Rough Surface,” *Phys. Rev. Lett.*, vol. 42, no. 19, pp. 1312–1314, 1979.
- [294] A. Ishimaru, *Wave propagation and scattering in random media*, vol. 2. Academic press New York, 1978.
- [295] A. Ishimaru, “Wave propagation and scattering in random media and rough surfaces,” *Proceedings of the IEEE*, vol. 79, no. 10, pp. 1359–1366, 1991.
- [296] K. M. Yoo, F. Liu, and R. R. Alfano, “When does the diffusion approximation fail to describe photon transport in random media?,” *Phys. Rev. Lett.*, vol. 64, pp. 2647–2650, 1990.
- [297] R. Elaloufi, R. Carminati, and J.-J. Greffet, “Definition of the diffusion coefficient in scattering and absorbing media,” *J. Opt. Soc. Am. A*, vol. 20, no. 4, p. 678, 2003.
- [298] H. Cang, A. Labno, C. Lu, X. Yin, M. Liu, C. Gladden, Y. Liu, and X. Zhang, “Probing the electromagnetic field of a 15-nanometre hotspot by single molecule imaging,” *Nature*, vol. 469, no. 7330, pp. 385–388, 2011.
- [299] H. Liu, Z. Yang, L. Meng, Y. Sun, J. Wang, L. Yang, J. Liu, and Z. Tian, “Three-Dimensional and Time-Ordered Surface-Enhanced Raman Scattering Hotspot Matrix,” *J. Am. Chem. Soc.*, vol. 136, no. 14, pp. 5332–5341, 2014.
- [300] T. Itoh, Y. S. Yamamoto, Y. Kitahama, and J. Balachandran, “One-dimensional plasmonic hotspots located between silver nanowire dimers evaluated by surface-enhanced resonance Raman scattering,” *Phys. Rev. B*, vol. 95, no. 11, pp. 115441–undefined, 2017.

- [301] Y. Huang, L. Ma, J. Li, and Z. Zhang, “Nanoparticle-on-mirror cavity modes for huge and/or tunable plasmonic field enhancement,” *Nanotechnology*, vol. 28, no. 10, p. 105203, 2017.
- [302] J. J. Baumberg, J. Aizpurua, M. H. Mikkelsen, and D. R. Smith, “Extreme nanophotonics from ultrathin metallic gaps,” *Nat. Mater.*, vol. 18, no. 7, pp. 668–678, 2019.
- [303] J. M. Wylie and J. E. Sipe, “Quantum electrodynamics near an interface,” *Phys. Rev. A*, vol. 30, no. 3, pp. 1185–1193, 1984.
- [304] L. S. Froufe-Pérez, R. Carminati, and J. J. Sáenz, “Fluorescence decay rate statistics of a single molecule in a disordered cluster of nanoparticles,” *Phys. Rev. A*, vol. 76, no. 1, 2007.
- [305] J. Liu, X. Jiang, S. Ishii, V. Shalaev, and J. Irudayaraj, “Quantifying the local density of optical states of nanorods by fluorescence lifetime imaging,” *New J. Phys.*, vol. 16, no. 6, p. 063069, 2014.
- [306] B. A. Van Tiggelen and S. E. Skipetrov, “Fluctuations of local density of states and C₀ speckle correlations are equal,” *Phys. Rev. E*, vol. 73, no. 4, p. 045601, 2006.
- [307] B. Shapiro, “New Type of Intensity Correlation in Random Media,” *Phys. Rev. Lett.*, vol. 83, no. 23, pp. 4733–4735, 1999.
- [308] S. Skipetrov and R. Maynard, “Nonuniversal correlations in multiple scattering,” *Phys. Rev. B*, vol. 62, no. 2, pp. 886–891, 2000.
- [309] R. Pierrat and R. Carminati, “Intensity correlations between reflected and transmitted speckle patterns,” *Phys. Rev. A*, vol. 033827, p. 33827, 2015.
- [310] S. Feng, C. Kane, P. A. Lee, and A. D. Stone, “Correlations and fluctuations of coherent wave transmission through disordered media,” *Phys. Rev. Lett.*, vol. 61, no. 7, pp. 834–837, 1988.

- [311] A. Garcia-Martin, F. Scheffold, M. Nieto-Vesperinas, and J. J. Saenz, “Finite-size effects on intensity correlations in random media,” *Phys. Rev. Lett.*, vol. 88, no. 14, p. 143901, 2002.
- [312] M. Segev, Y. Silberberg, and D. N. Christodoulides, “Anderson localization of light,” *Nat. Photonics*, vol. 7, pp. 197–204, 2013.
- [313] R. Carminati and J. C. Schotland, *Principles of Scattering and Transport of Light*. Cambridge University Press, 2021.
- [314] N. Byrnes and M. R. Foreman, “Symmetry constraints for vector scattering and transfer matrices containing evanescent components: Energy conservation, reciprocity, and time reversal,” *Phys. Rev. Res.*, vol. 3, no. 1, p. 013129, 2021.
- [315] A. A. Maradudin, A. R. McGurn, and E. R. Méndez, “Surface plasmon polariton mechanism for enhanced backscattering of light from one-dimensional randomly rough metal surfaces,” *J. Opt. Soc. Am. A*, vol. 12, no. 11, 1995.
- [316] P. Schlagheck and J. Dujardin, “Coherent backscattering in the Fock space of ultracold bosonic atoms,” *Annalen der Physik*, vol. 529, no. 7, 2017.
- [317] P. E. Wolf and G. Maret, “Weak localization and coherent backscattering of photons in disordered media,” *Phys. Rev. Lett.*, vol. 55, no. 24, 1985.
- [318] M. Boguslawski, S. Brake, D. Leykam, A. S. Desyatnikov, and C. Denz, “Observation of transverse coherent backscattering in disordered photonic structures,” *Sci. Rep.*, vol. 7, no. 1, 2017.
- [319] G. Jacucci, O. D. Onelli, A. De Luca, J. Bertolotti, R. Sapienza, and S. Vignolini, “Coherent backscattering of light by an anisotropic biological network,” *Interface Focus*, vol. 9, no. 1, 2019.
- [320] B. Fazio, A. Irrera, S. Pirotta, C. D’Andrea, S. Del Sorbo, M. Josè Lo Faro, P. G. Gucciardi, M. A. Iatì, R. Saija, M. Patrini, P. Musumeci, C. Salvatore Vasi, D. S. Wiersma, M. Galli,

- and F. Priolo, “Coherent backscattering of Raman light,” *Nat. Photonics*, vol. 11, no. 3, 2017.
- [321] D. Maystre and M. Saillard, “Localization of light by randomly rough surfaces: concept of localization,” *J. Opt. Soc. Am. A*, vol. 11, no. 2, pp. 680–690, 1994.
- [322] S. I. Bozhevolnyi, B. Vohnsen, I. I. Smolyaninov, and A. V. Zayats, “Direct observation of surface polariton localization caused by surface roughness,” *Opt. Commun.*, vol. 117, no. 5-6, pp. 417–423, 1995.
- [323] C. Gómez-Navarro, P. J. De Pablo, J. Gómez-Herrero, B. Biel, F. J. Garcia-Vidal, A. Rubio, and F. Flores, “Tuning the conductance of single-walled carbon nanotubes by ion irradiation in the Anderson localization regime,” *Nat. Mater.*, vol. 4, no. 7, 2005.
- [324] J. Chabé, G. Lemarié, B. Grémaud, D. Delande, P. Szriftgiser, and J. C. Garreau, “Experimental Observation of the Anderson Metal-Insulator Transition with Atomic Matter Waves,” *Phys. Rev. Lett.*, vol. 101, no. 25, p. 255702, 2008.
- [325] D. H. White, T. A. Haase, D. J. Brown, M. D. Hoogerland, M. S. Najafabadi, J. L. Helm, C. Gies, D. Schumayer, and D. A. Hutchinson, “Observation of two-dimensional Anderson localisation of ultracold atoms,” *Nat. Commun.*, vol. 11, no. 1, 2020.
- [326] G. Roati, C. D’Errico, L. Fallani, M. Fattori, C. Fort, M. Zaccanti, G. Modugno, M. Modugno, and M. Inguscio, “Anderson localization of a non-interacting Bose-Einstein condensate,” *Nature*, vol. 453, no. 7197, 2008.
- [327] H. Hu, A. Strybulevych, J. H. Page, S. E. Skipetrov, and B. A. Van Tiggelen, “Localization of ultrasound in a three-dimensional elastic network,” *Nature Physics*, vol. 4, no. 12, 2008.
- [328] A. Lagendijk, B. Van Tiggelen, and D. S. Wiersma, “Fifty years of Anderson localization,” *Physics Today*, vol. 62, no. 8, 2009.
- [329] A. Mafi, “Transverse Anderson localization of light: a tutorial,” *Adv. Opt. Photonics*, vol. 7, no. 3, 2015.

- [330] H. H. Sheinfux, Y. Lumer, G. Ankonina, A. Z. Genack, G. Bartal, and M. Segev, “Observation of Anderson localization in disordered nanophotonic structures,” *Science*, vol. 356, no. 6341, 2017.
- [331] T. Schwartz, G. Bartal, S. Fishman, and M. Segev, “Transport and Anderson localization in disordered two-dimensional photonic lattices,” *Nature*, vol. 446, no. 7131, 2007.
- [332] S. Karbasi, C. R. Mirr, R. J. Frazier, P. G. Yarandi, K. W. Koch, and A. Mafi, “Detailed investigation of the impact of the fiber design parameters on the transverse Anderson localization of light in disordered optical fibers,” *Opt. Express*, vol. 20, no. 17, 2012.
- [333] S. Karbasi, R. J. Frazier, C. R. Mirr, K. W. Koch, and A. Mafi, “Fabrication and characterization of disordered polymer optical fibers for transverse Anderson localization of light,” *J. Vis. Exp.*, no. 77, p. 50679, 2013.
- [334] Z. Cherpakova, F. Bleckmann, T. Vogler, and S. Linden, “Transverse Anderson localization of surface plasmon polaritons,” *Opt. Lett.*, vol. 42, no. 11, 2017.
- [335] Y. Zhu, H. Jing, R. W. Peng, C. Y. Li, J. He, B. Xiong, and M. Wang, “Realizing Anderson localization of surface plasmon polaritons and enhancing their interactions with excitons in 2D disordered nanostructures,” *Appl. Phys. Lett.*, vol. 116, no. 20, 2020.
- [336] J. Duan, S. Xiao, and J. Chen, “Anderson Localized Plasmon in Graphene with Random Tensile-Strain Distribution,” *Advanced Science*, vol. 6, no. 7, 2019.
- [337] S. Skipetrov and I. Sokolov, “Absence of Anderson Localization of Light in a Random Ensemble of Point Scatterers,” *Phys. Rev. Lett.*, vol. 112, no. 2, p. 023905, 2014.
- [338] S. E. Skipetrov and J. H. Page, “Red light for Anderson localization,” *New J. Phys.*, vol. 18, no. 2, 2016.
- [339] S. E. Skipetrov and Y. M. Beltukov, “Anderson transition for elastic waves in three dimensions,” *Phys. Rev. B*, vol. 98, p. 064206, 2018.
- [340] L. Novotny, B. Hecht, and D. W. Pohl, “Interference of locally excited surface plasmons,” *J. Appl. Phys.*, vol. 81, no. 4, pp. 1798–1806, 1997.

- [341] A. Lakhtakia, “Macroscopic theory of the coupled dipole approximation method,” *Opt. Commun.*, vol. 79, no. 1-2, 1990.
- [342] R. Berkovits, “Sensitivity of the multiple scattering speckle pattern to the motion of a single scatterer,” *Phys. Rev. B*, vol. 43, no. 10, pp. 8638–8640, 1991.
- [343] T. M. Nieuwenhuizen and M. C. W. van Rossum, “Role of a single scatterer in a multiple scattering medium,” *Phys. Lett. A*, vol. 177, no. 1, pp. 102–106, 1993.
- [344] J. Huang, “A perturbative multiple scattering theory,” *Phys. Lett. A*, vol. 322, pp. 10–18, 2004.
- [345] R. Berkovits and S. Feng, “Theory of speckle pattern tomography in multiple scattering media,” *Phys. Rev. Lett.*, vol. 65, no. 25, pp. 3120–3123, 1990.
- [346] P. N. den Outer, A. Lagendijk, and T. M. Nieuwenhuizen, “Location of objects in multiple-scattering media,” *J. Opt. Soc. Am. A*, vol. 10, no. 6, p. 1209, 1993.
- [347] A. Goetschy and S. E. Skipetrov, “Non-Hermitian Euclidean random matrix theory,” *Phys. Rev. E*, vol. 84, no. 1, p. 011150, 2011.
- [348] A. Goetschy and S. E. Skipetrov, “Euclidean random matrices and their applications in physics,” 2013. arXiv:1303.2880 [math-ph].
- [349] V. Martin-Mayor, M. Mézard, G. Parisi, and P. Verrocchio, “The dynamical structure factor in topologically disordered systems,” *J. Chem. Phys.*, vol. 114, no. 18, pp. 8068–8081, 2001.
- [350] M. Mézard, G. Parisi, and A. Zee, “Spectra of euclidean random matrices,” *Nucl. Phys. B.*, vol. 559, no. 3, pp. 689–701, 1999.
- [351] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Johns Hopkins University Press, 4 ed., 2013.
- [352] H. V. Henderson and S. R. Searle, “On deriving the inverse of a sum of matrices,” *SIAM Review*, vol. 23, no. 1, pp. 53–60, 1981.

- [353] T. Søndergaard and S. I. Bozhevolnyi, “Vectorial model for multiple scattering by surface nanoparticles via surface polariton-to-polariton interactions,” *Phys. Rev. B*, vol. 67, no. 16, p. 165405, 2003.
- [354] A. B. Evlyukhin and S. I. Bozhevolnyi, “Point-dipole approximation for surface plasmon polariton scattering: Implications and limitations,” *Phys. Rev. B*, vol. 71, no. 13, 2005.
- [355] S. I. Bozhevolnyi and V. Coello, “Elastic scattering of surface plasmon polaritons: Modeling and experiment,” *Phys. Rev. B*, vol. 58, no. 16, pp. 10899–10910, 1998.
- [356] P. Sheng, *Introduction to Wave Scattering, Localization, and Mesoscopic Phenomena*. Academic Press Inc., 1995.
- [357] V. Krachmalnicoff, E. Castanié, Y. De Wilde, and R. Carminati, “Fluctuations of the local density of states probe localized surface plasmons on disordered metal films,” *Phys. Rev. Lett.*, vol. 105, no. 18, p. 183901, 2010.
- [358] F. Riboli, F. Ucheddu, G. Monaco, N. Caselli, F. Intonti, M. Gurioli, and S. E. Skipetrov, “Tailoring Correlations of the Local Density of States in Disordered Photonic Materials,” *Phys. Rev. Lett.*, vol. 119, no. 4, p. 043902, 2017.
- [359] F. Rosenblatt, “The perceptron: a probabilistic model for information storage and organization in the brain.,” *Psychol. Rev.*, vol. 65, no. 6, p. 386, 1958.
- [360] S. Gallant, “Perceptron-Based Learning Algorithms,” *IEEE Trans. neural networks*, vol. 1, no. 2, pp. 179–191, 1990.
- [361] B. Widrow and M. A. Lehr, “30 years of adaptive neural networks: perceptron, Madaline, and backpropagation,” *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1415–1442, 1990.
- [362] R. Reed and R. J. MarksII, *Neural smithing: supervised learning in feedforward artificial neural networks*. MIT Press, 1999.
- [363] W. Maass and H. Markram, “On the computational power of circuits of spiking neurons,” *J. Comput. Syst. Sci.*, vol. 69, no. 4, pp. 593–616, 2004.

- [364] G. Cybenko, “Approximation by superpositions of a sigmoidal function,” *Math. Control Signals Syst.*, vol. 2, no. 4, pp. 303–314, 1989.
- [365] K. Hornik, “Multilayer Feedforward Networks are Universal Approximators,” *Neural Networks*, vol. 2, pp. 359–366, 1989.
- [366] K. Hornik, “Approximation capabilities of multilayer feedforward networks,” *Neural Networks*, vol. 4, no. 2, pp. 251–257, 1991.
- [367] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang, “The Expressive Power of Neural Networks: A View from the Width,” in *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), vol. 30, Curran Associates, Inc., 2017.
- [368] B. Hanin, “Universal Function Approximation by Deep Neural Nets with Bounded Width and ReLU Activations,” *Mathematics*, vol. 7, no. 10, p. 992, 2019.
- [369] P. Kidger and T. Lyons, “Universal approximation with deep narrow networks,” in *Conference on learning theory*, pp. 2306–2327, PMLR, 2020.
- [370] K. Janocha and W. M. Czarnecki, “On loss functions for deep neural networks in classification,” 2017. arXiv:1702.05659 [cs.LG].
- [371] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.
- [372] J. Sjöberg and L. Ljung, “Overtraining, regularization and searching for a minimum, with application to neural networks,” *Int. J. Control*, vol. 62, no. 6, pp. 1391–1407, 1995.
- [373] R. Caruana, S. Lawrence, and L. Giles, “Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping,” *Adv. Neural Inf. Process. Syst.*, pp. 402–408, 2001.
- [374] L. Prechelt, “Automatic early stopping using cross validation: quantifying the criteria,” *Neural Networks*, vol. 11, no. 4, pp. 761–767, 1998.

- [375] C. D. Doan and S. Y. Liong, “Generalization for multilayer neural network bayesian regularization or early stopping,” in *Proceedings of Asia Pacific association of hydrology and water resources 2nd conference*, pp. 5–8, 2004.
- [376] J. N. van Rijn and F. Hutter, “Hyperparameter importance across datasets,” in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD ’18, (New York, NY, USA), p. 2367–2376, Association for Computing Machinery, 2018.
- [377] P. Probst, B. Bischl, and A.-L. Boulesteix, “Tunability: Importance of hyperparameters of machine learning algorithms,” 2018. arXiv:1802.09596 [stat.ML].
- [378] S. Ruder, “An overview of gradient descent optimization algorithms,” 2017. arXiv:1609.04747 [cs.LG].
- [379] W. A. Gardner, “Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique,” *Signal Process.*, vol. 6, no. 2, pp. 113–133, 1984.
- [380] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014. arXiv:1412.6980 [cs.LG].
- [381] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a convolutional neural network,” in *2017 International Conference on Engineering and Technology (ICET)*, pp. 1–6, 2017.
- [382] G. Giacinto and F. Roli, “Design of effective neural network ensembles for image classification purposes,” *Image Vis. Comput.*, vol. 19, no. 9-10, pp. 699–707, 2001.
- [383] I. Kanellopoulos and G. G. Wilkinson, “Strategies and best practice for neural network image classification,” *Int. J. Remote Sens.*, vol. 18, no. 4, pp. 711–725, 1997.
- [384] C. Belthangady and L. A. Royer, “Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction,” *Nat. Methods*, vol. 16, no. 12, pp. 1215–1225, 2019.

- [385] P. N. Petrov, Y. Shechtman, and W. E. Moerner, “Measurement-based estimation of global pupil functions in 3D localization microscopy,” *Opt. Express*, vol. 25, no. 7, p. 7945, 2017.
- [386] P. Zelger, K. Kaser, B. Rossboth, L. Velas, G. J. Schütz, and A. Jesacher, “Three-dimensional localization microscopy using deep learning,” *Opt. Express*, vol. 26, no. 25, p. 33166, 2018.
- [387] E. Nehme, B. Ferdman, D. Freedman, L. E. Weiss, R. Gordon-Soffer, T. Naor, R. O. Kedem, O. Alalouf, T. Michaeli, Y. Shechtman, and Y. Shechtman, “Three-Dimensional Localization Microscopy by Deep Learning,” in *OSA Imaging and Applied Optics Congress 2021 (3D, COSI, DH, ISA, pcAOP)*, (Washington, D.C.), p. 3W3G.5, OSA, 2021.
- [388] E. Nehme, D. Freedman, R. Gordon, B. Ferdman, L. E. Weiss, O. Alalouf, T. Naor, R. Orange, T. Michaeli, and Y. Shechtman, “DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning,” *Nat. Methods*, vol. 17, no. 7, pp. 734–740, 2020.
- [389] P. Zhang, S. Liu, A. Chaurasia, D. Ma, M. J. Mlodzianoski, E. Culurciello, and F. Huang, “Analyzing complex single-molecule emission patterns with deep learning,” *Nat. Methods*, vol. 15, no. 11, pp. 913–916, 2018.
- [390] T. Kim, S. Moon, and K. Xu, “Information-rich localization microscopy through machine learning,” *Nat. Commun.*, vol. 10, no. 1, p. 1996, 2019.
- [391] E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, “Deep-STORM: super-resolution single-molecule microscopy by deep learning,” *Optica*, vol. 5, no. 4, p. 458, 2018.
- [392] W. Ouyang, A. Aristov, M. Lelek, X. Hao, and C. Zimmer, “Deep learning massively accelerates super-resolution localization microscopy,” *Nat. Biotechnol.*, vol. 36, no. 5, pp. 460–468, 2018.

- [393] H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Günaydin, L. A. Bentolila, C. Kural, and A. Ozcan, “Deep learning enables cross-modality super-resolution in fluorescence microscopy,” *Nat. Methods*, vol. 16, no. 1, pp. 103–110, 2019.
- [394] R. Horisaki, R. Takagi, and J. Tanida, “Learning-based imaging through scattering media,” *Opt. Express*, vol. 24, no. 13, p. 13738, 2016.
- [395] Q. Li, J. Zhao, Y. Zhang, X. Lai, Z. Chen, and J. Pu, “Imaging reconstruction through strongly scattering media by using convolutional neural networks,” *Opt. Commun.*, vol. 477, p. 126341, 2020.
- [396] A. Turpin, I. Vishniakou, and J. d. Seelig, “Light scattering control in transmission and reflection with neural networks,” *Opt. Express*, vol. 26, no. 23, p. 30911, 2018.
- [397] N. Borhani, E. Kakkava, C. Moser, and D. Psaltis, “Learning to see through multimode fibers,” *Optica*, vol. 5, no. 8, p. 960, 2018.
- [398] S. Li, M. Deng, J. Lee, A. Sinha, and G. Barbastathis, “Imaging through glass diffusers using densely connected convolutional networks,” *Optica*, vol. 5, no. 7, p. 803, 2018.
- [399] 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, p. 1181, 2018.
- [400] E. Valent and Y. Silberberg, “Scatterer recognition via analysis of speckle patterns,” *Optica*, vol. 5, no. 2, p. 204, 2018.
- [401] T. Ando, R. Horisaki, and J. Tanida, “Speckle-learning-based object recognition through scattering media,” *Opt. Express*, vol. 23, no. 26, p. 33902, 2015.
- [402] R. Horisaki, R. Takagi, and J. Tanida, “Learning-based focusing through scattering media,” *Appl. Opt.*, vol. 56, no. 15, p. 4358, 2017.
- [403] K. Liu, H. Zhang, B. Zhang, Q. Liu, B. Zhang, B. Zhang, Q. Liu, Q. Liu, and Q. Liu, “Hybrid optimization algorithm based on neural networks and its application in wavefront shaping,” *Opt. Express*, vol. 29, no. 10, p. 15517, 2021.

- [404] Y. Rivenson, Y. Zhang, H. Günaydın, D. Teng, and A. Ozcan, “Phase recovery and holographic image reconstruction using deep learning in neural networks,” *Light Sci. Appl.*, vol. 7, no. 2, p. 17141, 2018.
- [405] A. Sinha, J. Lee, S. Li, and G. Barbastathis, “Lensless computational imaging through deep learning,” *Optica*, vol. 4, no. 9, p. 1117, 2017.
- [406] A. G. Ramm, “Uniqueness of the solution to inverse scattering problem with scattering data at a fixed direction of the incident wave,” *J. Math. Phys.*, vol. 52, no. 12, p. 123506, 2011.
- [407] X. Xu, B. Zhang, and H. Zhang, “Uniqueness in inverse scattering problems with phaseless far-field data at a fixed frequency,” *SIAM J. Appl. Math.*, vol. 78, no. 3, pp. 1737–1753, 2018.
- [408] E. Wolf and T. Habashy, “Invisible Bodies and Uniqueness of the Inverse Scattering Problem,” *J. Mod. Opt.*, vol. 40, no. 5, pp. 785–792, 1993.
- [409] R. G. Novikov, “The inverse scattering problem at fixed energy for the three-dimensional Schrödinger equation with an exponentially decreasing potential,” *Commun. Math. Phys.*, vol. 161, no. 3, pp. 569–595, 1994.
- [410] A. G. Ramm, “Recovery of the potential from fixed-energy scattering data,” *Inverse Probl.*, vol. 5, no. 2, p. 510, 1989.
- [411] Y. Saitō, “An approximation formula in the inverse scattering problem,” *J. Math. Phys.*, vol. 27, no. 4, pp. 1145–1153, 1986.
- [412] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
- [413] A. F. Agarap, “Deep Learning using Rectified Linear Units (ReLU),” 2018. arXiv:1803.08375 [cs.NE].
- [414] S. Bhanja and A. Das, “Impact of Data Normalization on Deep Neural Network for Time Series Forecasting,” 2018. arXiv:1812.05519 [cs.LG].

- [415] J. Shao, K. Hu, C. Wang, X. Xue, and B. Raj, “Is normalization indispensable for training deep neural network? ,” in *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, eds.), vol. 33, pp. 13434–13444, Curran Associates, Inc., 2020.
- [416] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015. Software available from [tensorflow.org](https://www.tensorflow.org).
- [417] J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar, “Integrating physics-based modeling with machine learning: A survey,” 2020. arXiv:2003.04919 [physics.comp-ph].
- [418] V. Gavrishchaka, O. Senyukova, and M. Koepke, “Synergy of physics-based reasoning and machine learning in biomedical applications: towards unlimited deep learning with limited data,” *Advances in Physics: X*, vol. 4, no. 1, p. 1582361, 2019.
- [419] W. E, J. Han, and L. Zhang, “Integrating Machine Learning with Physics-Based Modeling,” 2020. arXiv:2006.02619 [physics.comp-ph].
- [420] H. Lee, J. Berk, A. Webster, D. Kim, and M. R. Foreman, “Label-free detection of single nanoparticles with disordered nanoisland surface plasmon sensor,” *Nanotechnology*, vol. 33, no. 16, p. 165502, 2022.

Appendix A

Evaluation of Enhancement Factor Integrals

This appendix details the integration of Eqs. (5.54) and (5.55) within the scalar SPP scattering model to arrive at Eqs (5.82) and (5.83). Starting with Eq. (5.55), in the scalar case, the α_{N+1} factors cancel, while Eq. (5.80) can be used for \tilde{G} . The incident field $E_{0,z}(x, z_s) = \Theta(x) \exp(ik_{\text{SPP}}x - \kappa_d z_s)$ (recalling that only the z -component is considered in the scalar model, and unit amplitude at $x = 0$ is used without loss of generality since the amplitude cancels with the $p_{N+1,z}$ factors) has a transverse Fourier transform $\tilde{E}_{0,z}(\mathbf{q}, z_s) = -2\pi\delta(q_y) \exp(-\kappa_d z_s)/(i(k_{\text{SPP}} - q_x))$. Substituting these factors into Eq. (5.55) results in

$$\begin{aligned} \langle S_3 \rangle &= n\alpha \frac{k_0^2}{\varepsilon_0} \int \frac{d^2\mathbf{q}}{(2\pi)^2} \frac{-4A_0\alpha e^{-2\kappa_d z_s}}{k_{\text{SPP}}^2 - q^2} \frac{1}{1 + 4n\alpha \frac{k_0^2}{\varepsilon_0} A_0 e^{-2\kappa_d z_s} \frac{1}{k_{\text{SPP}}^2 - q^2}} \cdot \frac{-(2\pi)\delta(q_y) e^{i(q_x - k_{\text{SPP}})x_{N+1}}}{i(k_{\text{SPP}} - q_x)} \\ &= e^{-ik_{\text{SPP}}x_{N+1}} \int_{-\infty}^{\infty} \frac{dq_x}{2\pi} \frac{4n\mu}{i(k_{\text{SPP}} - q_x)(k_{\text{SPP}}^2 - q_x^2 + 4n\mu)} e^{iq_x x_{N+1}}, \end{aligned} \quad (\text{A.1})$$

where $q^2 = |\mathbf{q}|^2 = q_x^2 + q_y^2$. Evaluation of the q_x integral in Eq. (A.1) can be done using complex analysis by considering the contour integral of the integrand in the complex plane around a closed loop consisting of a line along the real axis from $-R$ to R and a semi-circular arc in the upper half plane of radius R going from R to $-R$ (see Fig. A.1). By Jordan's lemma, for $x_{N+1} > 0$, the contribution from the arc vanishes as the radius R goes to ∞ , while the integral along the real line corresponds to the integral in Eq. (A.1). As a result the integral is calculated

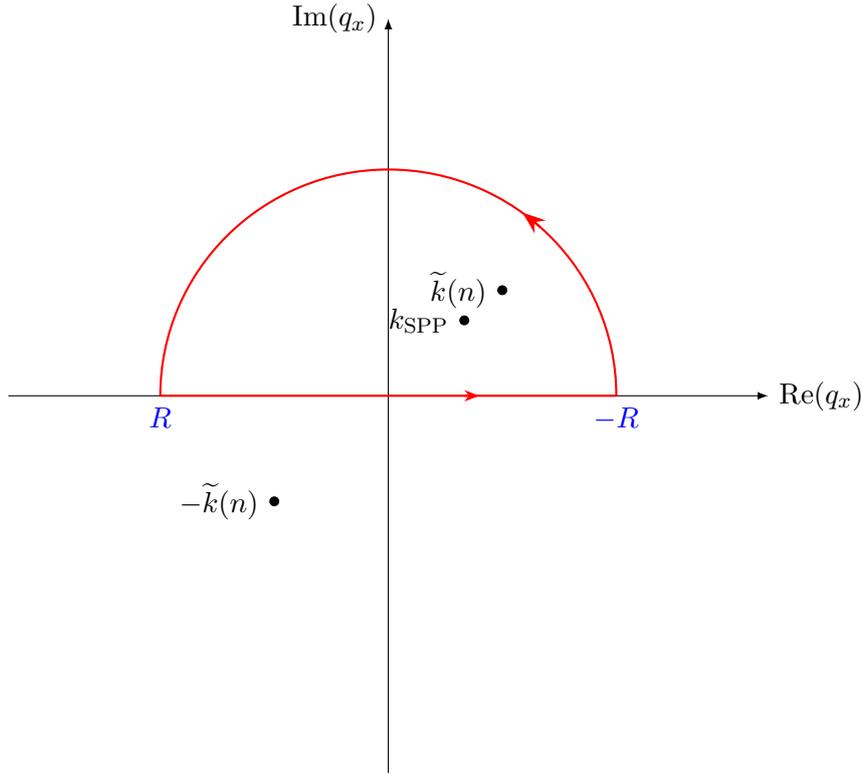


Figure A.1: Diagram showing the contour used to evaluate the integral in Eq. (A.1) in the limit $R \rightarrow \infty$ and the location of the poles of the integrand.

from the sum of the residues in the upper half plane via Cauchy's theorem. The integrand has simple poles at k_{SPP} and $\tilde{k}(n) = (k_{\text{SPP}}^2 + 4n\mu)^{1/2}$ (shown in Fig. A.1), with the pole at $-\tilde{k}(n)$ being in the lower half plane and therefore not contributing. The residues are found to be

$$\text{Res}_{q_x=k_{\text{SPP}}}[h(q_x)] = -\frac{e^{ik_{\text{SPP}}}}{2\pi i} \quad (\text{A.2})$$

$$\text{Res}_{q_x=\tilde{k}(n)}[h(q_x)] = -\frac{4n\mu e^{i\tilde{k}(n)x_{N+1}}}{4\pi i [k_{\text{SPP}} - \tilde{k}(n)] \tilde{k}(n)}, \quad (\text{A.3})$$

where $h(q_x)$ is the integrand in Eq. (A.1) (excluding the $e^{-ik_{\text{SPP}}x_{N+1}}$ factor taken outside the integration). Applying Cauchy's theorem to Eq. (A.1) results in

$$\begin{aligned} \langle S_3 \rangle &= 2\pi i e^{-ik_{\text{SPP}}x_{N+1}} \left(\text{Res}_{q_x=k_{\text{SPP}}}[h(q_x)] + \text{Res}_{q_x=\tilde{k}(n)}[h(q_x)] \right) \\ &= -1 - \frac{2n\mu \exp\left(i(\tilde{k}(n) - k_{\text{SPP}})x_{N+1}\right)}{[k_{\text{SPP}} - \tilde{k}(n)] \tilde{k}(n)}. \end{aligned} \quad (\text{A.4})$$

Using the fact $\langle \gamma_3 \rangle = 1 + \langle S_3 \rangle$ then results in Eq. (5.82). Note that the $x_{N+1} < 0$ case does not need to be considered since $\mathbf{E}_0(x_{N+1} < 0) = \mathbf{0}$ and therefore the single scattering contribution is zero and one cannot define an enhancement factor.

Moving on to calculating the scalar model result for $\langle S_2 \rangle$, given in Eq. (5.83), substitution of Eq. (5.80) for \tilde{G} in Eq. (5.54) results in

$$\langle S_2 \rangle = 16n\mu\mu_{N+1} \int \frac{d^2\mathbf{q}}{(2\pi)^2} \frac{1}{(k_{\text{SPP}}^2 - q^2)(k_{\text{SPP}}^2 - q^2 + 4n\mu)}. \quad (\text{A.5})$$

Doing the angular integration results in

$$\langle S_2 \rangle = 16n\mu\mu_{N+1} \int_{q=0}^{\infty} \frac{qdq}{2\pi} \frac{1}{(k_{\text{SPP}}^2 - q^2)(\tilde{k}(n)^2 - q^2)} \quad (\text{A.6})$$

Letting $u = q^2$, then splitting the integrand into partial fractions allows the integral over q to be found as follows

$$\begin{aligned} \langle S_2 \rangle &= 16n\mu\mu_{N+1} \int_{u=0}^{\infty} \frac{du}{4\pi} \frac{1}{(k_{\text{SPP}}^2 - u)(\tilde{k}(n)^2 - u)} \\ &= \frac{\mu_{N+1}}{\pi} \int_{u=0}^{\infty} \left(\frac{1}{k_{\text{SPP}}^2 - u} - \frac{1}{\tilde{k}(n)^2 - u} \right) \\ &= -\frac{\mu_{N+1}}{\pi} \log \left(\frac{\tilde{k}(n)^2}{k_{\text{SPP}}^2} \right). \end{aligned} \quad (\text{A.7})$$

From Eq. (A.7), the main text result of Eq. (5.83) follows from using $\tilde{k}(n)^2/k_{\text{SPP}}^2 = 1 + 4n\mu/k_{\text{SPP}}^2$.