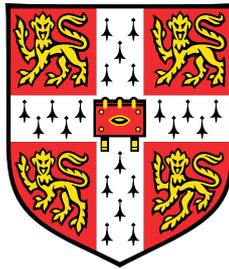


# Light scattering

## From ensembles to single particles



**Ilya Manyakin**

Department of Physics  
University of Cambridge

This dissertation is submitted for the degree of  
*Doctor of Philosophy*

Hughes Hall

December 2021

## **Declaration**

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Ilya Manyakin  
December 2021

# **Light scattering: From ensembles to single particles**

Ilya Manyakin

## **Abstract**

Light scattering enables the interrogation of nanoscale systems with minimal perturbation to their dynamics. However, this advantage of optical techniques is partially offset by requiring complex inference procedures to accurately estimate physical quantities of interest from features extracted by the optical measurement. In this thesis, I investigate optical techniques for the characterization of nanoparticle properties in solutions, the associated statistical inference problems and consider how nanosensors can be used to extend optical methods to probe nanoscale systems.

In the first part, we consider how multiwavelength light scattering combined with statistical inference can be applied to Photon Correlation Spectroscopy. By viewing the inverse problem of size distribution estimation within the Bayesian framework, a method for extracting an uncertainty quantified (UQ) estimate of the size distribution is presented. The technique is further generalized from a static inverse problem to a dynamic one, allowing sequences of temporally spaced measurements to be inverted simultaneously. Next, a novel single-particle tracking configuration is presented for simultaneous dual colour scattering inside hollow-core anti-resonant (HC-ARF) fibres at microsecond timescales. Simultaneous monitoring of multiple scattering signals allows observation of transient signatures linked to the reorientation of particles as they undergo rotational diffusion. The second part of the thesis considers transmission modulation in nanoapertures caused by local refractive index changes due to nanoparticles flowing through them. Protocols for the fabrication of nanoapertures are presented for gold films deposited on thin silicon nitride membranes. Effects of aperture parameters on optical response investigated using Finite-Difference Time-Domain (FDTD) simulations. Optical transmission measurements are performed using a constructed transmission microscope, with the nanoapertures integrated into microfluidic chip, enabling both optical interrogation and electrical flow control. These measurements allow detection of nanoparticle translocation through the nanoapertures as well as docking and ejection of larger nanoparticle aggregates, sterically prevented from translocating. Evidence of nanoaperture geometry and surface charge variation between the metal and silicon nitride layers manifests

as asymmetry in the response of the optical signal to the applied potential. Finally, to extend the optical measurement modalities available in the constructed microscope, a photon counting system based on time-to-digital converter is developed using Field-Programmable Gate-Arrays (FPGA). The system has a time resolution in the 30 ps range with continuous photon readout rates of up to 3 million counts per second, providing comparable performance to commercial instrumentation at a fraction of the cost, enabling many optical measurement techniques.

Dedicated to my grandfather, Konstantin Pilichev.

## **Acknowledgements**

Firstly, I would like to thank my supervisor Jeremy Baumberg for the opportunity to be part of Nanophotonics. His feedback and insight have been instrumental during the years of my PhD studies. I want to thank everyone in Nanophotonics for making my time in the group very special. In particular, thanks to Angelos Xomalis, Matthew Cheetham, and Omid Siddiqui for their friendship and discussions of science. A special thanks also go to Evangelos Logaras, Alun Hunt, and Ryan Smith for their advice and feedback on FPGAs and electronics. My sincere thanks go to William Deacon for teaching me experimental optics. Last but not least, I would like to express my deep gratitude to Ermanno Miele. Throughout these years, his guidance and friendship have made the PhD a wholly different experience from what it would have otherwise been. Finally, I would like to thank my family for their unwavering support.

# Table of contents

<b>List of figures</b>	<b>iv</b>
<b>List of tables</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Theoretical background</b>	<b>4</b>
2.1 Scattering theory . . . . .	4
2.1.1 Rayleigh Scattering . . . . .	5
2.1.2 Mie scattering . . . . .	6
2.1.3 Nanoaperture optics . . . . .	7
2.2 Optical detection . . . . .	9
2.3 Particle motion in solution . . . . .	11
2.3.1 Diffusion . . . . .	11
2.3.2 Electrokinetic phenomena . . . . .	12
<b>3 Photon correlation spectroscopy</b>	<b>15</b>
3.1 Theoretical background . . . . .	18
3.1.1 Random Phasor sums . . . . .	19
3.1.2 Stochastic processes . . . . .	19
3.1.3 Correlation functions . . . . .	21
3.2 Inverse problems . . . . .	24
3.2.1 Linear inverse problems . . . . .	25
3.2.2 Tikhonov regularization . . . . .	29
3.2.3 Posterior sampling . . . . .	31
3.2.4 Optimization-based sampling . . . . .	32
3.3 Experimental results . . . . .	35
3.3.1 System overview . . . . .	35

## Table of contents

---

3.3.2	Spectral filtering . . . . .	36
3.3.3	Inference engine . . . . .	40
3.3.4	Nonlinear inversion and extensions . . . . .	50
3.4	Summary . . . . .	53
<b>4</b>	<b>Single particle tracking</b>	<b>54</b>
4.1	Introduction . . . . .	54
4.2	Nanoaggregate scattering and diffusion . . . . .	55
4.2.1	Generalized diffusion . . . . .	56
4.3	Multiparticle Mie Scattering . . . . .	58
4.4	Experimental setup . . . . .	60
4.5	Particle tracking . . . . .	65
4.6	Summary . . . . .	68
<b>5</b>	<b>Time correlated single photon counting</b>	<b>69</b>
5.1	Introduction . . . . .	69
5.1.1	Detector technologies . . . . .	70
5.1.2	Time interval measurement . . . . .	74
5.2	Field-Programmable Gate Arrays . . . . .	80
5.2.1	Zynq System-on-Chip . . . . .	81
5.2.2	Time in digital systems . . . . .	82
5.2.3	Advanced eXtensible Interface (AXI) . . . . .	84
5.2.4	HDL/Verilog . . . . .	86
5.2.5	Linux . . . . .	87
5.3	Nanosecond TDC . . . . .	89
5.3.1	Analog-to-digital converter . . . . .	89
5.3.2	High speed counter . . . . .	90
5.4	Picosecond TDC . . . . .	91
5.4.1	Signal conditioning . . . . .	92
5.4.2	Carry chain . . . . .	92
5.4.3	Taps and synchronizers . . . . .	94
5.4.4	Delay line decoding . . . . .	95
5.4.5	Trigger and Hit signals . . . . .	97
5.4.6	Clock domain crossing . . . . .	98
5.4.7	System architecture . . . . .	99
5.4.8	Performance evaluation . . . . .	100
5.5	Signal conditioning board . . . . .	104

5.6	Future work . . . . .	106
5.7	Summary . . . . .	106
<b>6</b>	<b>Nanoaperture fabrication</b>	<b>107</b>
6.1	Introduction . . . . .	107
6.1.1	Nanoapertures . . . . .	109
6.1.2	Translocation control . . . . .	110
6.2	Device Fabrication . . . . .	111
6.2.1	Silicon Nitride Membranes . . . . .	112
6.2.2	Metal layers . . . . .	113
6.2.3	Microfabrication . . . . .	115
6.2.4	Surface treatment . . . . .	119
6.3	Fluidics . . . . .	123
6.4	Electrical model . . . . .	127
6.4.1	Electrodes . . . . .	127
6.4.2	Nanoaperture conductance . . . . .	128
6.5	Summary . . . . .	130
<b>7</b>	<b>Nanoaperture scattering</b>	<b>131</b>
7.1	Microscope . . . . .	131
7.1.1	Diode laser . . . . .	135
7.1.2	Data acquisition and automation . . . . .	137
7.2	Experimental results . . . . .	140
7.3	Summary . . . . .	145
<b>8</b>	<b>Conclusion and outlook</b>	<b>146</b>
	<b>References</b>	<b>153</b>

# List of figures

2.1	Mie angular scattering of AuNPs . . . . .	6
2.2	Nanoaperture schematic . . . . .	8
2.3	Randles equivalent circuit model and Debye layer . . . . .	13
3.1	Photon correlation spectroscopy overview . . . . .	17
3.2	Simulated particle size distribution and resulting correlation functions . . . . .	23
3.3	Field correlation function quadrature comparison . . . . .	27
3.4	Autocorrelation singular value decomposition and Picard plot . . . . .	28
3.5	Multiangle photon correlation spectroscopy system . . . . .	36
3.6	Autocorrelation function variation with laser bandwidth . . . . .	37
3.7	Tuneable spectral filtering instrumentation . . . . .	38
3.8	Spectral filtering performance . . . . .	39
3.9	Linear inverse problem comparison . . . . .	44
3.10	Stacked linear inverse problems . . . . .	45
3.11	Linear inverse problem solutions with uncertainty quantification . . . . .	46
3.12	Nonlinear inverse problem inversion for AuNP scatterers . . . . .	49
3.13	Polarization variation and gold nanorod scattering autocorrelation . . . . .	51
3.14	Relative Mie scattering . . . . .	52
4.1	Gold nanoaggregate SEM and angular scattering schematic . . . . .	55
4.2	Gold nanoaggregate schematic, optical cross-sections and simulated scattering intensity timeseries . . . . .	59
4.3	Dual colour high speed imaging microscope . . . . .	61
4.4	Example dual colour scattering timeseries . . . . .	63
4.5	Intensity joint distribution . . . . .	64
4.6	Reconstructed single particle tracking using stacked traces . . . . .	65
4.7	Intensity timeseries autocorrelation functions . . . . .	66

4.8 Dual colour AuNP aggregate tracks exhibiting correlated and anti-correlated properties . . . . .	67
5.1 Time interval measurement using Nutt method . . . . .	76
5.2 Tapped delay line . . . . .	77
5.3 Code density test . . . . .	78
5.4 Differential and integral nonlinearity . . . . .	79
5.5 Carry chain cascading with single CLB . . . . .	81
5.6 Analog-digital converter TDC . . . . .	89
5.7 SPAD dead-time measurement . . . . .	91
5.8 Pulse conditioner . . . . .	92
5.9 Ripple carry adder . . . . .	93
5.10 Single carry chain delay line post place-and-route . . . . .	94
5.11 Dual stage synchronizer . . . . .	95
5.12 Bubble errors and LUT decoding . . . . .	96
5.13 LUT decoder and Wallace tree adder . . . . .	97
5.14 Timetagger system diagram . . . . .	99
5.15 Time-interval calibration configuration . . . . .	101
5.16 Streaming data rate . . . . .	102
5.17 Time-to-digital converter deadtime . . . . .	103
5.18 Pulse conditioner board diagram and assembled system image . . . . .	104
5.19 Comparator evaluation configuration . . . . .	105
6.1 Nanopore and nanoaperture . . . . .	108
6.2 Analyte capture rate and nanopore capture radius . . . . .	110
6.3 Fabrication of free standing $Si_3N_4$ membranes . . . . .	112
6.4 Metal layer deposition using Thermal Physical Vapour Deposition . . . . .	113
6.5 Dual beam FIB/SEM and ion range distributions . . . . .	115
6.6 Metal coated SiNx membrane schematic . . . . .	116
6.7 Nanoaperture diameter evolution . . . . .	118
6.8 SEM images of nanoapertures and SiNx membrane . . . . .	119
6.9 Nanoaperture z-scan and spectra . . . . .	121
6.10 Darkfield imaging configuration . . . . .	122
6.11 Nanoaperture transmission spectra . . . . .	123
6.12 Fluidic chip . . . . .	124
6.13 Milling . . . . .	125
6.14 Fluidic chip sealing and assembly . . . . .	127

## List of figures

---

6.15	Nanoaperture equivalent circuit . . . . .	128
6.16	Bulk solution conductivity and Debye length with salt concentration and nanoaperture radius . . . . .	130
7.1	Transmission microscope system . . . . .	132
7.2	Autocollimator diagram . . . . .	133
7.3	Diode laser assemble and gain chip . . . . .	134
7.4	Diode laser relative intensity noise and characterization . . . . .	136
7.5	Measurement system control and acquisition schematic . . . . .	138
7.6	Nanoaperture 3D transmission intensity distribution . . . . .	139
7.7	Nanoaperture transmission modulation from nanoparticles . . . . .	140
7.8	Nanoaperture transmission as a function of effective refractive index inside aperture . . . . .	141
7.9	Transmission intensity and noise variation with applied electrical bias across electrodes . . . . .	142
7.10	Nanoaggregate docking and release from nanoaperture . . . . .	143
7.11	Optical charging and discharging from white light illumination . . . . .	144
7.12	Permanent pore enlargement and electrical detection of clogging . . . . .	145

# List of tables

3.1	PCS detection configuration coherence factor $\beta$ . . . . .	24
5.1	Zynq-7000 CLB logic . . . . .	81
5.2	Programmable logic parameters . . . . .	82
5.3	Full-adder for adding bits $A$ and $B$ . . . . .	93
5.4	TDC time resolution results . . . . .	103
5.5	NIM pulse standard . . . . .	104
6.1	Properties of polymers utilized in microfluidics . . . . .	125

# Chapter 1

## Introduction

Scattering encompasses a broad range of phenomena associated with a deviation in the trajectory of a probe particle or wave during propagation through a medium, arising from non-uniformities in the medium. Properties of the transmitted and scattered probe enable one to make inferences about features and parameters of the medium under investigation, such as presence or absence of structural periodicity, the motion of the medium relative to the probe source, size of the individual scattering centers, and chemical composition of the medium.

‘Electromagnetic’ scattering methods use an electromagnetic wave as the probe and form a diverse range of techniques capable of measuring structures from the galactic scale down to the atomic. A key feature of electromagnetic waves is their frequency (or wavelength) which determines how and at what scale the field interacts with a target under investigation.

Optical methods utilize frequencies in the 750-430 THz range, corresponding to 400-700nm wavelengths, forming the visible part of the spectrum detectable by the human eye. Small structures, down to the micron length scale, are probed using the optical microscope that focuses the probe light to the length scale of a single wavelength and projects the scattering signal back up to the macro scale, detectable by the human eye or a camera.

The ability to visualize structures on the micrometer length scale using an optical microscope has been key to enabling studies of biological systems with a characteristic length scale of a single cell falling in the micrometer range. The non-contact nature of optical methods combined with their ability to probe structures in-situ has made the optical microscope an indispensable tool in biology and spurred the continued need to improve its resolution, enabling observation of finer details revealing new biological mechanisms and structures.

## Introduction

---

With continuing advancement in understanding comes the need to affect, control and interface with biological processes. Nanoparticle-based technologies are one possible route to achieving this goal, both at the single-cell level and on the tissue and organ scales. Some examples are lipid vesicles used for drug delivery or noble metal nanoparticles used for localized chemical sensing. Single nanoparticle sizes are typically in the tens to hundreds of nanometer range, with a large surface area to volume ratio. The surface properties of nanoparticles largely dictate how they interact with the surrounding environment, commonly a heterogeneous mixture containing various chemicals and structures. Thus, to robustly design such systems, one must accurately characterize and predict the behavior of the nanoparticles in-situ. This brings the need to develop optical methods. They are one of the few techniques capable of non-invasive, in-situ, real-time monitoring of the nanoparticles at various scales, from a single particle to a large ensemble of millions of particles.

The challenge in applying optical techniques to measuring nanoparticles is the diffraction limit bounding the maximum spatial resolution achievable with optical systems to  $\sim \lambda/2$ . Aside from increasing resolution through reducing wavelength, indirect measurement techniques can be used. Such methods attempt to relate the ‘hidden’ variables of interest to the experimentally measured quantities through a mathematical model, derived analytically or estimated empirically. Estimates of the hidden variables can then be obtained through an ‘inference’ process that maps from the measured quantity to the hidden parameters of interest.

Being ‘indirect’ allows these measurements to exploit information encoded in features of the optical signal other than the two-dimensional intensity distribution recorded by a conventional imaging camera. Such attributes include spectral information, light polarization, or temporal intensity variations in the optical signal, each requiring often different instruments capable of suitably recording these signals and storing them for analysis. Being less intuitive than imaging techniques has limited the uptake of indirect methods and consequently limited the availability of instrumentation capable of performing such indirect measurements. Many such systems are inaccessible even to research laboratories due to the cost and their complexity in operation. Thus nanoparticle characterization can benefit not only from improvements in optical techniques but also from developments in instrumentation for recording the optical signals themselves.

This thesis describes work aimed at improving upon current techniques for characterizing nanoparticle properties using optical scattering techniques. In it, I describe three optical systems constructed for performing optical scattering measurements for characterizing single nanoparticles and ensembles of nanoparticles, a single photon counting readout system

---

for recording single photons' arrival times, and an inference engine developed to perform statistical inference mapping between experimentally measured quantities and hidden variable of interest, through the solution of either linear or nonlinear inverse problems. The systems are developed with the aim of improving the resolution and reliability of nanoparticle characterization techniques, as well as lowering the entry barriers to techniques utilizing single-photon sensitivity detection requiring high temporal resolutions.

The **next chapter** provides an introduction to the theory of scattering by small particles and transmission through plasmonic nanoapertures, as well as the basic theory of diffusion and electrokinetic phenomena in solutions, a medium where nanoparticle techniques have found many applications.

**Chapter 3** describes the construction of a dynamic light scattering apparatus for measuring sub-micron size nanoparticles, introduces the theory of inverse problems required for reconstruction of the particle size distribution, and proposes a new approach to performing estimation and uncertainty quantification of the particle size distribution through optimization-based sampling.

**Chapter 4** describes a new experimental method for single-particle tracking at high temporal resolutions and demonstrates the capability of the constructed system in measuring transient intensity fluctuations linked to the reorientation of plasmonic nanoparticle aggregates undergoing rotational diffusion.

**Chapter 5** details techniques for high-speed single-photon detection and presents a developed time-to-digital converter capable of determining individual photon arrival time with a 30 picosecond resolution, capable of continuous high data rate readout along with a designed high bandwidth comparator used to condition different detector pulse standards - enabling the interfacing of the implemented time-to-digital converter to single-photon detectors.

**Chapter 6** describes the fabrication of plasmonic nanoapertures and microfluidic chips with integrated electrical interfaces used to perform simultaneous electrical and optical measurements of nano aperture transmission modulation upon translocation of a nanoparticle through the nanochannel formed by the nano aperture on a thin silicon nitride membrane.

**Chapter 7** presents the constructed optical transmission microscope designed for performing high-speed optical and electrical measurements of nanoparticle translocation through the plasmonic nanoaperture along with preliminary experimental results verifying the operation of the system and outlines future directions for this work.

# Chapter 2

## Theoretical background

In this chapter I introduce the basic concepts of light scattering from small particles to define key terms and quantities involved in the formulation of a scattering problem. Two regimes, Rayleigh and Mie scattering, are outlined. A related problem of transmission through a small hole in a thin metallic film, a nanoaperture, is also briefly discussed and the transmission properties related to the geometry and size of the nanoaperture.

Following the scattering, we will introduce Brownian motion of small particles as well as review the effects of an applied electric field on motion of charged particles and ions in fluids where we will relate the applied field to their steady-state velocities.

Finally, we will introduce key equations modelling single-photon statistics that govern optical detection at low light levels showing the relation between the count per unit time interval and the statistics of inter-arrival times of the photons.

### 2.1 Scattering theory

Consider light of incident free-space wavelength  $\lambda_0$  to scatter at an angle  $\theta$  from a sample solution with bulk refractive index  $n$ . If we consider a point scatterer and work under the assumption that  $|k_{in}| \approx |k_{out}|$ , the elastic scattering assumption, the scattering vector  $q$  is the difference between the incident wavevector  $k_{in}$  and the outgoing wavevector  $k_{out}$ .

$$q = |\mathbf{q}| = |\mathbf{k}_{out} - \mathbf{k}_{in}| = \frac{4\pi n}{\lambda_0} \sin\left(\frac{\theta}{2}\right) \quad (2.1)$$

Modelling the electric field as a scalar quantity, ignoring polarization effects, the wave relation determines the amplitude of the incident and outgoing fields given by Eq. 2.2 [1].

$$u_{sca} = S(\theta, \phi) \frac{e^{-ikr+i\omega t}}{ikr} u_{inc} \quad (2.2)$$

Here  $u_{inc} = e^{-ikz+i\omega t}$  is the infinite-extent plane monochromatic wave oscillating at angular frequency  $\omega$  and with wavevector  $k = 2\pi/\lambda$  and  $r$  being the radial distance from the particle to the detector. The ‘amplitude function’  $S(\theta, \phi) = s \cdot e^{i\sigma}$  acts as the transformation relating the incident and scattered fields, with the amplitude  $s \in \mathbb{R}_+$  and phase shift  $\sigma$ . For the case of light polarized parallel and perpendicular to the scattering plane, defined as the plane containing the incident and scattering wavevectors, the amplitude scattering matrix  $\mathbf{S}(\theta, \phi)$  generalizes to [1] Eq. 2.3.

$$\begin{pmatrix} E_{\parallel, sca} \\ E_{\perp, sca} \end{pmatrix} = \begin{pmatrix} S_2 & S_3 \\ S_4 & S_1 \end{pmatrix} \cdot \frac{e^{-ikr+ikz}}{ikr} \begin{pmatrix} E_{\parallel, inc} \\ E_{\perp, inc} \end{pmatrix} \quad (2.3)$$

The amplitude functions  $S_1, S_2, S_3, S_4$  now relate the two pairs of fields. Under the assumption that no depolarization occurs ( $S_3 = S_4 = 0$ ) the ratio of incident and scattered intensity for each of the two polarizations is given by Eq. 2.4.

$$\begin{aligned} \frac{I_{\parallel, sca}}{I_{\parallel, inc}} &= \frac{|S_2|^2}{k^2 r^2} \\ \frac{I_{\perp, sca}}{I_{\perp, inc}} &= \frac{|S_1|^2}{k^2 r^2} \end{aligned} \quad (2.4)$$

### 2.1.1 Rayleigh Scattering

When the particle radius  $R$  is much less than the incident wavelength  $\lambda$ ,  $R \ll \lambda$  the induced dipole moment  $p$  is related linearly to the incident field  $E_{inc}$  by the polarizability  $\alpha$

$$p = \alpha E_{inc} \quad (2.5)$$

This represents the isotropic case, while in the general case, the polarizability is a tensor and the scattering properties of the particle are determined both by oscillation direction of the incident field and the orientation of the particles relative to it.

## Theoretical background

In this simplified model we can write  $\alpha = \alpha'V$ , where  $\alpha'$  is a dimensionless mean volume polarizability while  $V$  is the volume of the particle. As the dipole moment is parallel to the field, we neglect the orientation of the particle, giving rise to the Rayleigh scattering formulation of the amplitude matrix  $S(\theta, \phi)$  as

$$\begin{pmatrix} S_2 & S_3 \\ S_4 & S_1 \end{pmatrix} = ik^3 \alpha \begin{pmatrix} \cos\theta & 0 \\ 0 & 1 \end{pmatrix} \quad (2.6)$$

Given that the above describes relations for the field, and that the intensity  $I = |E|^2$  the scattered intensity scales as  $\propto k^4 V^2$ , assuming that there is no additional  $\lambda$  dependence in  $\alpha$ ,  $\alpha(\lambda) = \alpha$ . For particles with complex polarizability, a similar scaling relation for absorption [1] is found to be  $\propto kV$ . The comparison between the two scaling relations suggests that absorption measurements may be more sensitive to smaller particles than scattering measurements and act as motivation for photothermal techniques [2].

### 2.1.2 Mie scattering

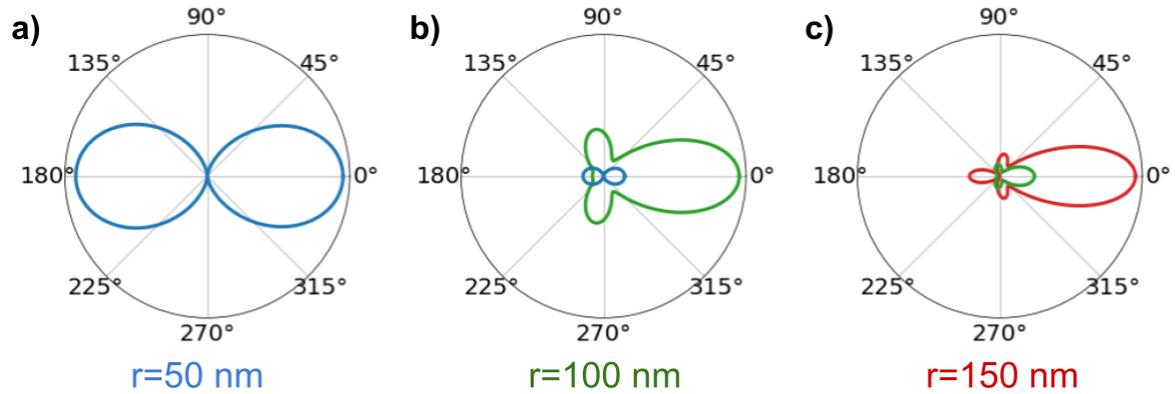


Fig. 2.1 Angular scattering intensities  $|S_2|^2$  of gold nanoparticles (AuNPs) for wavelength  $\lambda = 633$  nm in water,  $n_0 = 1.33$  for a) Particle radius  $r = 50$  nm. b) Particle radius  $r = 100$  nm. c) Particle radius  $r = 150$  nm. Plots contain smaller size particle to illustrate scale.

Mie theory can predict the amplitude scattering functions, scattering, and absorption cross-sections for light incident on a homogeneous spherical particle. The theory can be derived by expanding the incident plane wave and the scattered fields external and internal to the particle in terms of partial spherical waves. Through matching boundary conditions, the forms of the Mie coefficients  $a_m, b_m$  can be obtained. These are then used to express amplitude functions and cross-sections [3, 1]:

$$S_1(\theta) = \sum_{m=1}^{\infty} \frac{2m+1}{m(m+1)} [a_m \pi_m(\cos\theta) + b_m \tau_m(\cos\theta)] \quad (2.7)$$

$$S_2(\theta) = \sum_{m=1}^{\infty} \frac{2m+1}{m(m+1)} (b_m \pi_m(\cos\theta) + a_m \tau_m(\cos\theta)) \quad (2.8)$$

$$\sigma_{sca} = \frac{\lambda_0^2}{2\pi n_0^2} \sum_{m=1}^{\infty} (2m+1) (|a_m|^2 + |b_m|^2) \quad (2.9)$$

$$\sigma_{abs} = \frac{\lambda_0^2}{2\pi n_0^2} \sum_{m=1}^{\infty} (2m+1) [Re(a_m + b_m) - (|a_m|^2 + |b_m|^2)] \quad (2.10)$$

Here  $n_0$  denotes the refractive index of the medium,  $\pi_m, \tau_m$  are functions of scattering angle which one computes using recurrence relations. Implicit in the above formulae are the wavelength and particle refractive index dependence, encoded in the Mie coefficients  $a_m = a_m(\lambda), b_m = b_m(\lambda)$ . As particle sizes increase they begin to exhibit angular scattering patterns such as those depicted for gold spherical nanoparticles (AuNPs) in Fig.2.1.

### 2.1.3 Nanoaperture optics

Similar to scattering by spherical particles, the study of apertures dates back to the foundations of optics and remains a crucial component of many introductory courses - the single slit is an instance of a one-dimensional aperture. For small apertures in metallic films, the first theoretical treatment was given by Bethe [4] in the 1940s, deriving transmission intensity of an aperture for an infinitely thin, perfectly conducting (PEC, permittivity  $\epsilon = -\infty$ ) film exhibiting a  $(r/\lambda)^4$  scaling of transmission (Eq. 2.11). Subsequently, the treatment was extended to include higher-order expansion terms by Bouwkamp [5], with other theoretical results continuing to be derived even today.

Bethe's theory predicts a finite 'cut-off' wavelength  $\lambda_c$  below which the aperture does not transmit (Fig.2.2b). Extension of this formulation to a film of finite thickness for the PEC was developed by Roberts [6] through the coupled-mode method (CMM), based on considering a metallic plane of finite thickness between two semi-infinite dielectric half-spaces. The incident and transmitted fields are expanded in the plane wave basis while the field on the PEC is expanded using aperture waveguide modes, obeying the Helmholtz equation for which analytic solutions exist for some geometries, including the cylindrical aperture [5].

$$T_{Bethe} = 64(kr)^4 / 27\pi^2 \quad (2.11)$$

## Theoretical background

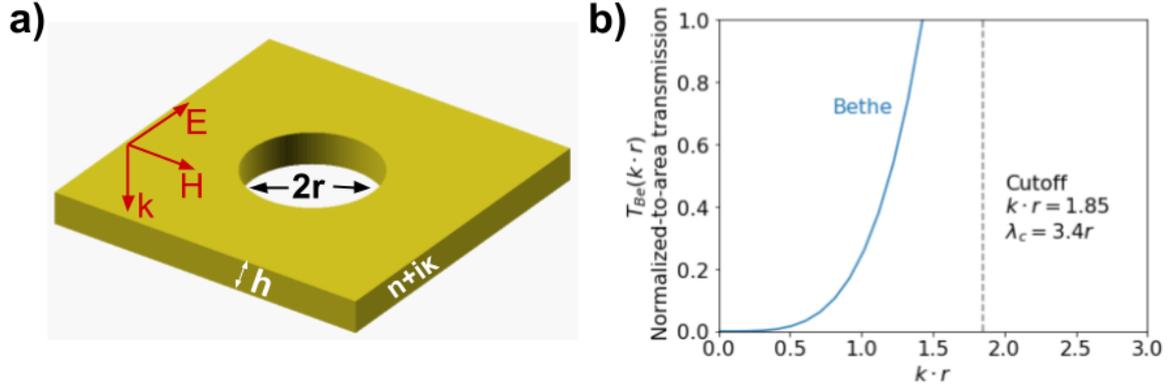


Fig. 2.2 a) Circular nanoaperture in metal film of finite thickness b) Bethe theory transmission through infinitely thin PEC film.

The model is completed by introducing boundary conditions (BCs) on the electric (E) and magnetic (H) fields (Eq. 2.12) at the reflection (I) and transmission (III) boundaries (Fig.2.2) of the PEC. Solving this allows calculation of system response to an incident field. The incorporation of ‘surface impedance boundary conditions’ (SIBCs) further extends this framework, allowing generalization to real metals like silver (Ag) or gold (Au) through incorporating effects like a finite skin depth of the metals  $\delta$  into the formulation. Even for the ideal PEC, the above formulation is found to exhibit features such as polarization sensitivity [6] induced when the incident wave vector is not normal to the film surface.

$$n \times E|_{interface} = 0; n \cdot B|_{interface} = 0 \quad (2.12)$$

Experimentally a key parameter of interest is the transmission coefficient  $T$ , normalized to aperture area, measuring the ratio of transmitted flux to flux incident on the area of the physical aperture (given by Eq. 2.11 for transmission through an infinitely thin PEC film [4]). The discovery, in 1998 by Ebbesen [7], of configurations where this can exceed unity ( $T > 1$ ) termed ‘Extraordinary optical transmission’ (EOT) has renewed interest in such nanoaperture systems in the optical regime. While initially discovered for periodic nanoaperture arrays, this has also been studied experimentally and theoretically for single apertures.

Modelling effects of different parameters on transmission properties is typically performed through in-silico simulation of the nanoaperture systems, with different approaches including rigorous coupled-mode methods, finite-difference time-domain (FDTD) simulations, and transfer matrix techniques. Such results are however limited to a given set of simulation parameters and there are few results providing empirical equations that can predict optical

properties of a particular nanoaperture geometry given features like size and shape, metal layer thickness, and surrounding medium refractive index. A model for a finite thickness PEC is however provided for different geometries by Nikitin et al [8] which allows computation of transmission properties for finite thickness PEC structures of different shapes and sizes through Eq. 2.13. The formulation extends the previous results by incorporating thickness and geometry correction model  $\psi(h)$  including effects of metal film thickness  $h$ . This model also introduces the exponential decay of transmission with linear thickness increase through  $e^{-2|q_{z0}|h}$ , where  $q_{z0}$  is the fundamental TE waveguide mode propagation constant of the aperture, acting as the waveguide. The term  $C(h)$  represents the area aspect ratio dependent polynomial correction factor.

$$T = (kr)^4 \underbrace{e^{-2|q_{z0}|h} C(h)}_{\psi(h)} \quad (2.13)$$

In the context of nanoparticle characterization, the nanoaperture systems introduced above can be used as sensors for detecting individual analyte nanoparticles, through monitoring the changes in transmission [9–11] occurring when a nanoparticle with an refractive index, different to that of the surrounding medium, enters the nanoaperture. The change of refractive index modifies the propagation constant of the waveguide mode, producing a resonance shift. This modulation can be recorded by monitoring intensity transmitted by the nanoaperture when illuminated by a resonance detuned laser. In the case when the particle refractive index exceeds that of the solution, the resonance shift occurs towards the longer wavelengths. Using a probe beam with a wavelength that is red-shifted relative to the nanoaperture resonance produces an increase in transmission intensity upon nanoparticle translocation. The ability to tailor properties of the nanoaperture, such as metal layer thickness, aperture geometry [12, 13] and size of such structures may in future enable designers the necessary parametric freedom to optimize such systems for a target application where the use of the transmission signal, as opposed to a pure scattering signal would allow improved detection limits as compared to more traditional scattering techniques.

## 2.2 Optical detection

Detection of the scattered field is achieved by transducing the optical signal to an electrical one, noting that optical detectors are of the ‘square-law’ type, measuring intensity  $I = |E|^2$  not electric field  $E$ . For the detection process, the field can be considered semi-classically, with the incident field being continuous but with the individual photon absorptions by

## Theoretical background

---

the detector material being discrete. Using this semi-classical formulation the process of detecting photons can be viewed as following a discrete Poisson distribution (Eq. 2.14) with  $n \in \mathbb{N}$  representing the number of detection events in a fixed time interval. The parameter controlling the ‘rate’ (or intensity), a continuous quantity, is represented by  $\lambda \in \mathbb{R}_+$ .

$$\mathbb{P}(n|\lambda) = \frac{\lambda^n e^{-\lambda}}{n!} \quad (2.14)$$

This distribution exhibits the characteristic of having a coupling between the mean  $\mathbb{E}(n) = \lambda$  and variance  $Var(n) = \lambda$  as well as non-negativity constraints on both the number of events in the time interval  $n$  and rate  $\lambda$ . The relation between the mean and variance result in a  $\sqrt{\lambda}$  scaling of the signal-to-noise ratio  $\mathbb{E}(n)/\sqrt{Var(n)} = \mu/\sigma$  for Poisson processes, often termed the ‘shot-noise limit’. While the above rate  $\lambda$  is assumed constant, this assumption can be relaxed to incorporate a temporal dependence  $\lambda = \lambda(t)$ . Cases where the intensity variation  $\lambda(t)$  is stochastic are termed ‘doubly stochastic Poisson processes’ [14].

This idealized count process rarely manifests as output from real detectors as the process of detection is itself non-ideal, introducing further nonlinearities and additional sources of noise. Some nonlinearities are related to the design of the detection circuits and can be more easily understood by considering the distribution of time intervals between detection events. This can be formulated as the time interval  $\tau$  between a ‘start’ and ‘stop’ photon pair. In general, these need not be consecutive, that is the stop photon need not follow exactly after the start photon. Denoting the number of intermediate photons between the start and stop as  $n$ , the distribution of time intervals is then given by Eq. 2.15 [15, 16]

$$\mathbb{P}(\tau|n, \lambda) = \frac{\lambda^{n+1} \tau^n e^{-\lambda \tau}}{n!} \quad (2.15)$$

This distribution is closely related to an ‘autocorrelation’ function introduced in the next chapter for analysis of self-similarity of intensity fluctuations, detected as count rate variation of the detected photons. For the special case where the stop follows consecutively after the start,  $n = 0$ , the resulting time interval distribution is given by

$$\mathbb{P}(\tau|\lambda) = \lambda e^{-\lambda \tau} \quad (2.16)$$

Thus the probability density of the inter-arrival times is high at short times and decays exponentially towards long time periods. Non-ideal effects such as ‘deadtime’  $\tau_d$ , where a detector is briefly deactivated after detecting a photon, manifests as a nulling of the probability

density for short times below  $\tau_d$ ,  $\mathbb{P}(\tau < \tau_d) = 0$ . Other non-ideal effects such as ‘afterpulsing’ caused by release of trapped carriers in the detectors active region, producing artificial output pulses when no photons are incident can similarly be considered using the above, and involve biasing the distribution to having greater probability at short timescales.

While by no means exhaustive these basic concepts are useful for understanding the differences between various detectors as well as how these differences can propagate through to experimental measurements.

## 2.3 Particle motion in solution

The motion of a particle in a medium is determined by both the properties of the solution and the properties of the particle or the structure it is attached to. In the case where the target of investigation is the medium, tracer particles with known parameters may be introduced into the medium to act as environmental probes allowing the sensing of flow fields (particle imaging velocimetry), viscosity (nano-rheology) and temperature, among others. Alternatively, when properties of the medium are known, the motion of particles can give information about the physical properties of particles themselves, such as their size and shape, density or charge.

When considering nanoparticles, with sizes below the diffraction limit, we are unable to directly measure their size optically. Instead the random motion of the particle can allow us to estimate its effective ‘hydrodynamic’ size if certain simplifying assumptions are made. This is what we discuss next.

### 2.3.1 Diffusion

In a medium, a particle will experience collisions with solvent molecules. These collisions will transfer momentum to the particle, modelled as an instantaneous stochastic force  $\xi(t)$  in the Langevin equation (Eq. 2.17), for a particle of mass  $m$  and velocity  $v$ . This stochastic differential equation describes the motion of the particle in the solution. The effects of hydrodynamic friction in this equation are incorporated through the  $-\gamma v$  term with  $\gamma$  being the hydrodynamic friction coefficient. This can be related to particle radius  $r$  through  $\gamma = 6\pi\eta r$ , under the simplifying assumption of a spherical particle, in a fluid with dynamic viscosity  $\eta$ .

$$m \frac{dv}{dt} = -\gamma v + \xi(t) \quad (2.17)$$

## Theoretical background

---

By considering the evolution of position for a particle moving subject to the Langevin equation, we note that the expected mean remains zero with time. In contrast, the second moment relative to an initial position evolves according to Eq. 2.18.

$$\begin{aligned}\langle x(t)^2 \rangle &= 2D_t t \\ D_t &= \frac{k_B T^{-1}}{\gamma} = \frac{k_B T}{6\pi\eta r}\end{aligned}\tag{2.18}$$

This type of motion occurs when the flow of surrounding fluid is laminar - the low Reynold regime ( $R_e \lesssim 0.3$  [17]). For a single sphere  $R_e = 2rv_s\rho/\eta$  where  $v_s$  is the particles ‘settling velocity’,  $\rho$  is fluid density. Thus even for particle velocities as high as  $v_s = 1.5 \text{ ms}^{-1}$ , radius  $r = 100 \text{ nm} = 10^{-7} \text{ m}$  in water ( $\eta = 10^{-3} \text{ Pa s}$ ,  $\rho = 10^3 \text{ kg/m}^3$ ) the low Reynolds regime holds (with  $R_e = 0.3$ ) and thus the description of diffusive motion, under assumption of no flow is valid. This shows that through tracking the trajectory of a particle it is possible to compute an effective ‘hydrodynamic’ radius. This core principle forms the basis of both nanoparticle tracking [18, 19] analysis as well as dynamic light scattering [20].

### 2.3.2 Electrokinetic phenomena

While flow fields may complicate measurements of Brownian motion, they may be beneficial in methods where the velocity of the induced flow can be related to, for example, particle charge. When a solution contains charge carrier ions, applying an electrical potential across the solution can generate a flow. This arises from the resulting field exerting a force on charged particles, both to individual ions and as the nanoparticles. A time-varying electric potential may then be used to induce oscillatory flows and motion that augment the information obtained from Brownian diffusion alone. Alternatively, the electrical potential can be applied to actively control the motion of particles and bring them to a desired sensing region. The ability to achieve either of these is however determined by the experimental setup, comprising of the solution, electrode type and geometry, and particle properties.

Begin by considering a single particle in an electrolyte where the application of an electric field produces a force and under steady-state conditions inducing electrophoretic flow, in the linear regime given by Eq. 2.19 where  $\mu_{ep}$  is termed ‘electrophoretic mobility’ [21].

$$v_{ep} = \mu_{ep} E \tag{2.19}$$

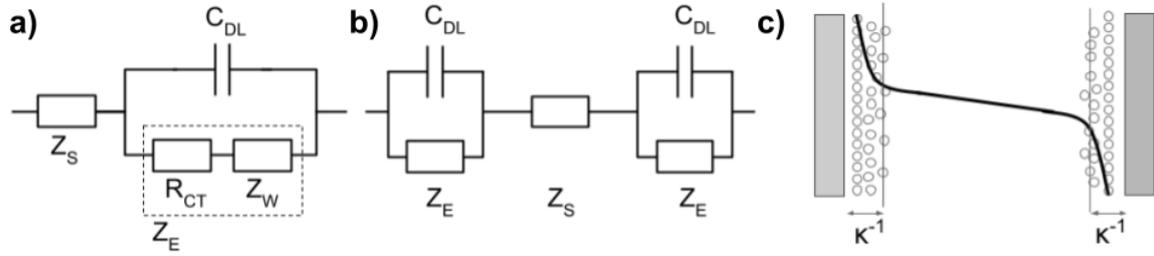


Fig. 2.3 a) Randles equivalent circuit for electrochemical half-cell. b) Randles equivalent circuit for full electrochemical cell. c) Debye layer near electrode surface and potential variation in solution.

Drag experienced by the ions produces a net motion of the surrounding solution in the channel under the applied field - ‘electrosmotic’ flow  $v_{eo}$  (Eq. 2.20) with its associated ‘electrosmotic mobility’  $\mu_{eo}$ . The resulting velocity of the charged particle through the solution the sum of the electrosmotic and electrophoretic velocities  $v_{total} = v_{ep} + v_{eo}$  and assumed proportional to the applied field [21].

$$v_{eo} = \mu_{eo}E \quad (2.20)$$

To determine how applying a potential difference between two electrodes affects the resulting field and by proxy, the flow velocity consider a pair of electrodes separated by a fixed distance. The electrical circuit equivalent is given by Fig.2.3b. Here the  $Z_S$  component models the impedance of the system under study, such as a flow channel or membrane, while the parallel resistor-capacitor  $C_{DL} \parallel Z_E$  pair provide simplified models of electrode interfacial impedance. A ‘Randles’ equivalent circuit, Fig.2.3a can be used to model electrochemical mechanisms occurring at the each of the two interfaces, known as the ‘Electrical Double Layer’ (DL), between the solid conductor and electrolyte solution.

The DL forms as the ions in the electrolyte rearrange in response to the charge on the electrode surfaces, depicted in Fig.2.3c. When curvature of the electrode is small relative to the length scale of the double layer, we can approximate the electrode interface in a planar geometry. The spatial extent of the DL, its characteristic length scale, is the ‘Debye length’  $\kappa^{-1}$  extending  $\sim 3\kappa^{-1}$  into the solution. Its dependence is determined by the concentration of the ionic species, and their valence, in the electrolyte (Eq. 2.21) as well as temperature.

## Theoretical background

---

$$\kappa = \left( \frac{e^2}{\epsilon_r \epsilon_0 k_B T} \sum_{i=1}^N z_i^2 n_i^0 \right)^{1/2} = \left( \frac{2000 F^2}{\epsilon_r \epsilon_0 R T} \right)^{1/2} \sqrt{I} \quad (2.21)$$

When ion species are present in solution, the Debye length can be expressed through Eq. 2.21, with bulk number density of species  $i$  is denoted as  $n_i^0$  and charge  $z_i$  is given in the general case and for the case of a monovalent ( $\forall i, z_i = \pm 1$ ) electrolyte [22]. Outside of the few fundamental constants,  $F = eN_A$  (Faraday's constant),  $R = k_B N_A$  (Molar gas constant), the Debye length depends on  $n_i^0$  (unit:  $\text{m}^{-3}$ ) bulk ion density (related to volume charge density via  $\rho = \sum_i n_i z_i e$ ), ionic concentration  $c_i$  (unit:  $\text{mol} \cdot \text{L}^{-1}$ ) with  $c_i \times 10^3 = n_i^0 / N_A$  and ionic strength  $I = \frac{1}{2} \sum_i c_i z_i^2$ .

In the Randles cell, the circuit equivalent of the Debye layer is the capacitive element  $C_{DL}$  in Fig.2.3a, approximated by a Guoy-Chapman model as  $C_{DL} = \epsilon_r \epsilon_0 A_{electrode} \kappa$ , scaling linearly with electrode area,  $A_{electrode}$ . In parallel with  $C_{DL}$  the effective resistive impedance  $Z_E$  encodes effects associated with a drop of electrical potential across the double layer, incorporating effects [23] including 'charge-transfer resistance'  $R_{CT}$  and Warburg impedance  $Z_W$  [24]. The latter impedance models effects of diffusion processes that bring(/remove) reacting species to(/from) the bulk of the electrolyte and electrode surfaces, limiting the frequency response of the electrodes. The charge-transfer  $R_{CT}$  on the other hand models processes limiting charge transfer rates between the electrolyte and electrodes, such as finite rates of reduction and oxidation (redox) reactions occurring at the surfaces. One should note the distinction between polarizable and non-polarizable electrodes and how this relates to the charge transfer processes - the former type of electrodes (such as platinum, Pt) do not react directly with ions in the solution, while the latter (such as silver/silver-chloride, Ag/AgCl) do [23, 25]. This results in smaller potential drops across the interfacial resistances, increasing the dynamic range flow velocities that can be created by applying a potential.

## Summary

In this chapter we have introduced some key concepts describing how light is scattered by small particles and transmitted through small holes in metallic films. We have also shown how motion of the small particles is governed by stochastic dynamics, caused by collisions with solvent molecules, and presented how an applied electric field gives rise to a range of electrokinetic phenomena, inducing flows in the solution. These concepts are key to understanding the experimental techniques for characterization of particles presented in subsequent chapters of this thesis.

# Chapter 3

## Photon correlation spectroscopy

The first of the techniques investigated is ‘Photon Correlation Spectroscopy’ (PCS) also known as ‘Quasi-Elastic’ (QELS) or ‘Dynamic’ Light Scattering (DLS). We use the former naming. The method is routinely applied for ensemble measurements of small particles down to 1 nm in radius. Central to the technique is the recording of intensity fluctuations arising from the relative motion of scattering particles diffusing in a solution, while illuminated by a laser light source. Each of the particles in the ensemble contributes to the total field measured at one or more detectors placed around the sample. Interference between the scattered fields gives rise to a ‘speckle’ pattern, that evolves with time as motion of the particles produces phase shifts between individual electric field contributions.

With detectors measuring the intensity, not the electrical field, a PCS measurement records time varying speckle intensity fluctuations. These intensity fluctuations exhibit a temporal self-similarity due to their origin in the relative motion of the scattering centres. We quantify this self-similarity through an ‘autocorrelation’ function (ACF) measuring the degree to which, two temporally separate points, are correlated.

A qualitative connection between diffusion of a particle, with radius  $R$  and diffusion constant  $D_t$  and speckle fluctuations, can be drawn by considering the time taken for the particle to diffuse by a wavelength  $\Delta x \approx \lambda$ . This path length difference in turn induces a phase difference  $\Delta\phi = k\Delta x$ . With many particles, this characteristic diffusion time can in turn be related to the decay of the intensity autocorrelation, as the system forgets its initial state.

When generalized to an ensemble of particles, the above reasoning can be incorporated into a ‘forward’ model, which, for given a particle size distribution computes an expected

autocorrelation function. In an experimental measurement we are however forced to work backwards from the measured autocorrelation to estimate particle size distribution, the ‘inverse’ problem. The properties of the PCS inverse problem give rise to high sensitivity of the solution to noise in the measured ACF. This sensitivity, as shall be illustrated in the chapter, is one of the primary limitations of the technique, degrading the resolution of the recovered size distributions and limiting the method to simple mono-modal size distributions. Practical applications, however, involve the need to characterize heterogeneous samples with different particle sub-populations and, consequently, multi-modal particle size distributions (PSD). This, combined with the widespread adoption, ease of use and reproducibility of PCS makes the continued improvement of the techniques resolution a worthwhile endeavour. While traditional PCS recorded the ACF at single scattering angle, recent developments have generalized this to multi-angle measurements. The multi-angle techniques exploit the non-isotropic Mie scattering of larger (>100 nm) particles in order to resolve sub-populations in the 100-500 nm size region. There are however other parameters available experimentally including wavelength and polarization, which, when combined with multi-angle measurements may provide further improvements in the resolution of PCS. A key challenge to performing such experiments is the proper ‘fusion’ of the additional information, in form of modified correlation functions, into the recovered particle size distribution. Doing so requires new approaches to solving the inverse problem, mentioned above.

Unlike many common applications of machine learning (ML) including image recognition and speech analysis, the case of PCS is not easily addressable through generation of large annotated data sets using hand labelled training data. In part this is because fitting correlation functions has never been an area where human performance exceeded that of machines. Instead the forward model is directly formulated and inverted using a statistical inference engine. Similarly to other ML techniques, developments in this area has benefited from recent improvements in computing power, algorithm development and programming languages.

In this chapter I will discuss the construction of a PCS system capable of multi-angle and multi-wavelength measurements along with the development of an inference procedure capable of inverting the full nonlinear model of PCS data to recover particle size distributions. I begin by introducing the theory of PCS and Inverse problems. This will be followed by a discussion of light source required for PCS measurements. Finally I will present results of the measurements with the PCS system along and discuss the innovations of the inference approach compared to traditional PCS inversion.

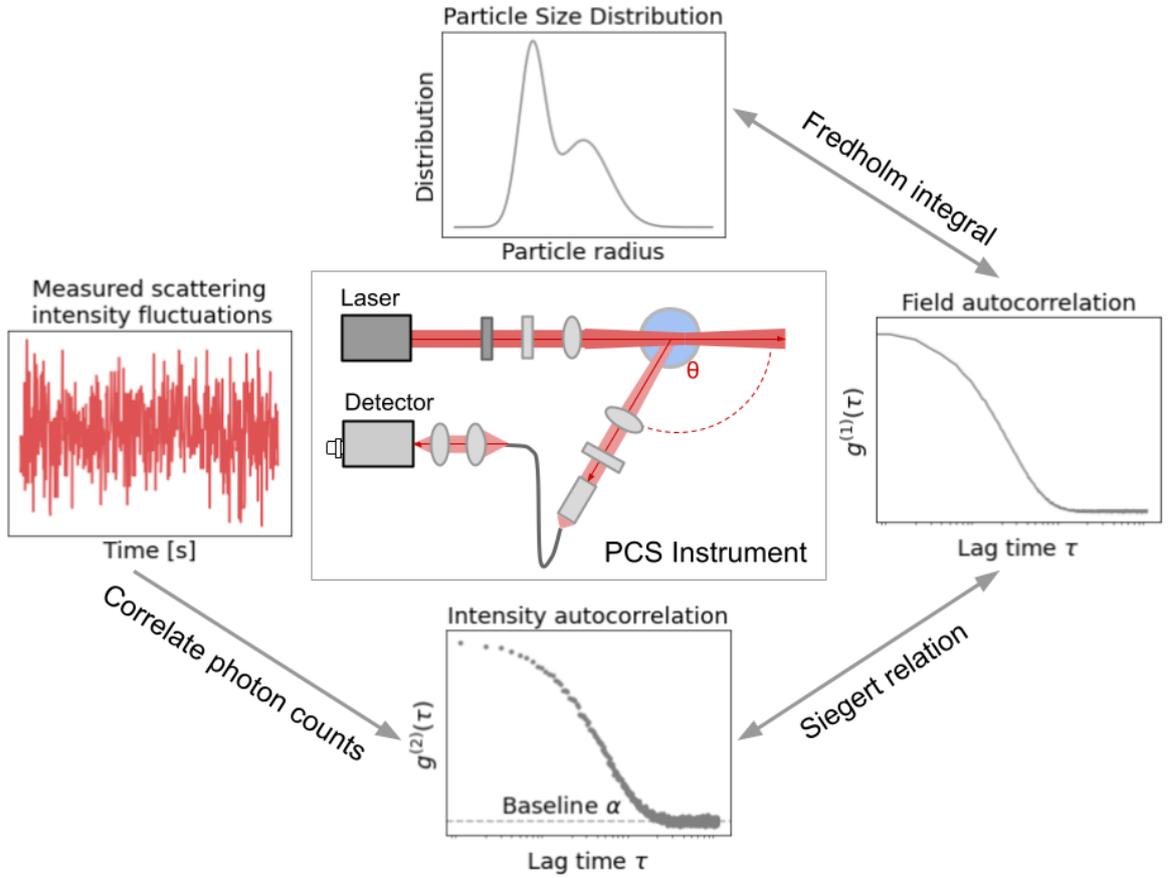


Fig. 3.1 Visual overview of a PCS measurement. A scattering signals fluctuations, produced by diffusion of particles is recorded. An intensity autocorrelation function  $g^2(\tau)$  is computed from the measured fluctuations. The intensity autocorrelation is related to the field autocorrelation function  $g^{(1)}(\tau)$ , through the Siegert relation. Inverting the field autocorrelation, by solving the Fredholm integral equation, recovers an estimate of the size distribution of scattering particles in the liquid.

### Contribution

Key contributions of this work are the development of an inference engine capable of solving both linear and nonlinear formulations of the PCS inverse problem. Along with recovered PSD estimates, the engine automatically performs uncertainty quantification (UQ) of the recovered PSD and provides automatic estimates of measurement noise levels. These provide an interpretable connection between the autocorrelation measurement noise and resolution of the recovered PSD missing in other works. For the linear case I show how the multi-angle approach can be generalized to perform simultaneous inversion of a measurement timeseries, applicable to time-varying particle size distributions by introducing temporal

priors. In the case of the nonlinear inverse problem, I show how the inverse problem can be solved efficiently by combining gradient-based and Markov-chain Monte-Carlo (MCMC) methods, allowing efficient sampling of the posterior distribution. The ability to cast the nonlinear PCS inversion as an optimization enables problems of higher dimension to be solved more efficiently, allowing for multiple measurements both in angle and time to reconstruct time-evolving particle size distributions. A vital aspect of this contribution is a clear presentation of the key steps, methods and limitations for solving inverse problems, and PCS in particular, owing to the fact that solvers incorporated into commercial instruments are seldom publically available. This ‘black-box’ nature of many PCS instruments acts to complicate interpretation of their outputs, and has the potential to be alleviated through an open source and well documented library can be developed from methods described herein. A secondary contribution of this work is the experimental investigation of different laser source suitability for PCS applications and involved construction of a multiangle and multiwavelength PCS instrument for nanoparticle characterization.

### 3.1 Theoretical background

To understand the origin of the effect enabling PCS for scattering from a solution of particles consider the interference pattern created from two narrow slits. When illuminated by a coherent light source an interference pattern on a distant screen arises from the relative path length difference between light from each slit. The path difference creates a optical phase difference of  $\phi = \frac{2\pi d}{\lambda} \sin(\theta)$ . For complete constructive interference, corresponding to an intensity maximum, we require that  $\phi = 2\pi n$  with  $n \in \mathbb{N}$  while destructive interference occurs for  $\phi = (2n + 1)\pi$ . Fixing the scattering angle  $\theta$ , and detection point, and varying the slit separation  $d$  instead produces a linearly varying phase offset between the two fields. This in turn results in a variation of the detected intensity.

Replacing the slits with particles and allowing stochastic motion in three instead of two dimensions will create stochastic fluctuations in intensity on a distant screen, the detector. Generalizing to many particles we then recover the PCS experiment. In this case the resulting sum of the field contributions at a detector varies on a timescale related to the characteristic diffusion time of the particles giving rise to intensity fluctuations that exhibit a temporal self-similarity. Quantification of the degree of temporal self-similarity and rate of its decay is achieved through calculation of an intensity correlation function. The process of ‘inversion’ of PCS data is then maps the measured intensity correlation back to a particle size distribution.

### 3.1.1 Random Phasor sums

First, consider the isolated contribution of each particle to the field at the detector. For a single particle this contribution is represented as a ‘phasor’ having an amplitude  $\alpha_i$  and phase  $\phi(r, t)$ .

$$E(r, t) = (\alpha_i / \sqrt{N}) e^{i\phi(r, t)} \quad (3.1)$$

Generalizing to many particles the resulting field on the detector is the vector sum of such phasors, having a real and imaginary component. In order to derive a form for the probability distributions of each field component consider the phasors to have amplitudes and phases be random, independent and identically distributed (i.i.d) with the amplitudes  $\alpha_k$  having first and second moments equal to  $\mathbb{E}[\alpha]$  and  $\mathbb{E}[\alpha^2]$ . The phases are similarly i.i.d but uniformly distributed on the interval  $\phi \in [-\pi, \pi]$ . The resulting sum is then written as [14]:

$$E_{total}(r, t) = \sum_{i=0}^N E_i(r, t) = \frac{1}{\sqrt{N}} \sum_{i=0}^N \alpha_i e^{i\phi(r, t)} = \frac{1}{\sqrt{N}} \sum_{i=0}^N \alpha_i [\cos(\phi) + i \sin(\phi)]$$

Appealing to the central limit theorem as  $N \rightarrow \infty$  both components of the resulting field  $E_{total}$  each follow a Gaussian distribution, with zero mean and variance given by  $\sigma^2 = \mathbb{E}(\alpha^2)/2$ . A second key result is the lack of correlation (and hence independence) between the two field components, forming a ‘Circular Gaussian random variable’ [14].

### 3.1.2 Stochastic processes

Moving from the static picture to the dynamic case, requires introducing particle motion. Here we consider the stochastic Brownian motion of particles diffusing freely in the solution. We ignore effects like convection due to heating or other flows that may be present in the solution noting that presence of such flows for PCS should be minimized, as it gives rise of subtle modelling errors that propagate through and increase biases in reconstructed particle size distributions.

From the previous section, the electric field follows a circular gaussian random variable (or a complex gaussian distribution). As the particles undergo diffusion this will evolve in time. Indexing the electric field random variables by time,  $E_t$  we treat the field at each time instance as a separate random variables. This representation forms the ‘random process’, comprised of the set of electric field random variables  $E_T$ . Similar to the case of a single random variable, it is worth making the distinction between properties of a particular instantiation(sample) of

the random process and properties of the process itself. The connection between properties of the sample and the underlying random process is dependent of the category of random process. One key feature of random processes is their stationarity properties, determining which parameters of the random process remain fixed in time. Two forms of stationarity are defined below.

**Definition 3.1.1** (Strictly Stationary [26]). The time series  $(X_t)_{t \in \mathbb{Z}}$  is strictly stationary if  $(X_{t_1}, \dots, X_{t_n}) = (X_{t_1+k}, \dots, X_{t_n+k})$

**Definition 3.1.2** (Covariance/Weakly Stationary [26]). The time series  $(X_t)_{t \in \mathbb{Z}}$  is second order stationary if the first two moments exist are invariant with shifts ( $\tau$ ) in time:

$$\begin{aligned} \mathbb{E}[X_t] &= \mu \\ \mathbb{E}[(X_{t_1} - \mu)(X_{t_2} - \mu)] &= E[(X_{t_1+\tau} - \mu)(X_{t_2+\tau} - \mu)] \end{aligned} \quad (3.2)$$

Assumptions of stationarity are common to many models, and experimental data is often transformed to satisfy them prior to model fitting. Predictions of the stationary models can then be inverse transformed into the original domain to allow predictions of the nonstationary process. In PCS the key assumption is weak stationarity. This enables the correlation function (Eq. 3.4), to depend only on time difference between two sample points, for two points  $t_1$  and  $t_2 = t_1 + \tau$  the dependence is only on the ‘lag-time’  $\tau$ , not absolute times  $t_1, t_2$ . In practice this assumption can manifest as requiring the experimental system to be either fully invariant in time, or vary much slower than the duration of the measurement.

### Gaussian processes (GP)

For the case of the electric field, with each element  $E_t$  draw from a Gaussian distribution, the random process can also be considered as being drawn from a single multi-dimensional Gaussian distribution. This leads to considering the electric field as a Gaussian random process [27] indexed by time.

**Definition 3.1.3** (Gaussian Process). A Gaussian process is a set of random variables any finite number of which have a joint distribution that is Gaussian.

The crucial property of such processes,  $f(x)$ , is that they are completely specified by their mean  $m(x)$  and covariance  $k(x, x')$  functions,  $f(x) \sim GP(m(x), k(x, x'))$ .

$$\begin{aligned} m(x) &= \mathbb{E}[f(x)] \\ k(x, x') &= \mathbb{E}[(f(x) - m(x))(f(x') - m(x')))] \end{aligned} \quad (3.3)$$

Being fully described by the mean and covariance functions, the higher order moments of the GP are also expressible in terms of these same quantities. An instance of such a higher order moment, for the E-field GP, is the intensity autocorrelation function, that can be computed from the experimentally recorded intensity fluctuations or photon counts. The relation between field and intensity correlations functions, both being moments of the E-field random process is then given by ‘moment’ theorems for Gaussian processes [14, 27].

Qualitatively the origin of the correlation between the electric field values at different time points, as specified by the covariance function can be understood by considering positions of particles at two times  $x(t)$  and  $x(t + \Delta t)$ . When  $\Delta t$  is small relative to the characteristic timescale of diffusive motion the particle position  $x(t) \approx x(t + \Delta t)$ . Correspondingly the electric field contributions will have similar phase and thus the  $E(t)$  and  $E(t + \Delta t)$  will be highly correlated, while as  $\Delta t$  increases the two fields will decorrelate, each giving little information about the possible values of the other.

### 3.1.3 Correlation functions

Experimentally the correlation functions are computed from photon streams either in software or by means of a digital correlator. The field  $g^{(1)}(\tau)$  and intensity  $g^{(2)}(\tau)$  autocorrelation functions are given below:

$$\begin{aligned} g^{(1)}(\tau) &= \mathbb{E}[E(t)E^*(t + \tau)] \\ g^{(2)}(\tau) &= \mathbb{E}[I(t)I(t + \tau)] \end{aligned} \tag{3.4}$$

In the Fourier domain the ‘autocorrelation’ of a signal can be defined by Eq. 3.5 [28]. This can be related to the power spectral density of the signal via the Wiener-Khinchin theorem.

$$g(\tau) = \int_{-\infty}^{+\infty} f(t)f^*(t + \tau)dt \tag{3.5}$$

Similarly, ‘statistical autocorrelation’ (Eq. 3.6), related to the autocovariance arising in stochastic processes analysis by subtraction of the expectation from the random variable can be defined by incorporating the probability density  $p(f, t, t + \tau)$  [14].

$$g(\tau) = \mathbb{E}[f(t)f(t + \tau)] = \int_{-\infty}^{+\infty} f(t)f(t + \tau)p(f, t, t + \tau)dt \tag{3.6}$$

## Photon correlation spectroscopy

---

These two definitions coincide only when the random process is weakly stationary. In this case the Fourier definition of autocorrelation is equal to the statistical autocorrelation of the random process. The Lorentzian spectral lineshape, produced by Doppler broadening can then be related to the exponentially decaying field correlation via the Fourier transform.

$$K(\omega) \propto \frac{\Gamma}{\Gamma^2 + \omega^2} \quad \xleftrightarrow{FT} \quad k(t) = e^{-\Gamma|t|} \quad \Re(\Gamma) > 0 \quad (3.7)$$

Extending to a continuous particle size distribution the lineshape becomes a weighted integral of Lorentzians, while the field correlation function is given by a Fredholm integral equation of the first kind.

$$\begin{aligned} g^{(1)}(\tau) &= \int_0^\infty G(\Gamma) e^{-\Gamma|\tau|} d\Gamma \\ \text{with } g^{(1)}(0) &= \int_0^\infty G(\Gamma) d\Gamma = 1 \\ \text{where } \Gamma &= q^2 D_t \end{aligned} \quad (3.8)$$

The non-negative particle size distribution is represented by  $G(\Gamma)$  (Fig. 3.2a). At zero lag time, the normalization of the correlation further imposes that the PSD integral be equal to unity. The decay rate  $\Gamma$  scales quadratically with scattering vector  $q$  (Eq. 2.1) and linearly with  $D_t$  (Eq. 2.18).

### Moment theorems and Siegert relation

The field correlation function is the second moment of the electric field. Similarly, the intensity correlation function is the second moment of the intensity, but also the fourth moment of the field. Using  $I(t) = E(t)E^*(t)$  we can write  $g^{(2)}(\tau)$  as:

$$g^{(2)}(\tau) = \mathbb{E}[E(t)E^*(t)E(t+\tau)E^*(t+\tau)] \quad (3.9)$$

In order to relate this to the field correlation function, the second moment we recall that the E-field is a Gaussian process - specified fully by it's first and second moments and that each of the E-field components is a zero-mean GP. Using these assumptions for a general  $v(t) = u^2(t)$  relation one moment theorem is given by [14]:

$$\mathbb{E}[v(t)v(t+\tau)] = \mathbb{E}[u^2(t)u^2(t+\tau)] = \mathbb{E}[u^2(t)]^2 + 2\mathbb{E}[u(t)u(t+\tau)]^2$$

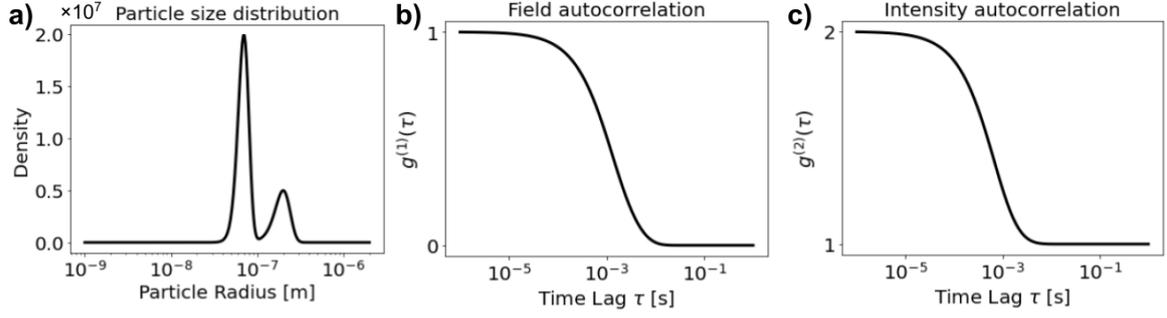


Fig. 3.2 a) Simulated bimodal particle size distribution  $G(\Gamma)$ . b) Field correlation function  $g^{(1)}(\tau)$  computed via Fredholm integral. c) Intensity autocorrelation function  $g^{(2)}(\tau)$  computed via the idealized Siegert relation.

In the case of interest the electric field consists of two components, the real and imaginary parts, with each component forming a GP, and together form a ‘circular complex’ GP. For the case of the circular complex GP the relation between autocorrelation of electric field and intensity, is given by [14]:

$$\mathbb{E}[|E(t)|^2 |E(t+\tau)|^2] = \mathbb{E}[|E(t)|^2] \mathbb{E}[|E(t+\tau)|^2] + |\mathbb{E}[E(t)E(t+\tau)]|^2$$

Applying the definitions of intensity and field correlation functions above and under the assumption of  $\mathbb{E}[|E(t)|^2] = \mathbb{E}[|E(t+\tau)|^2] = 1$  we recover the more conventionally stated ‘Siegert relation’ (Fig. 3.2c) common to PCS literature [29]:

$$g^{(2)}(\tau) = 1 + |g^{(1)}(\tau)|^2 \quad (3.10)$$

Alternative functional forms of this relation exist for diffusion near boundaries, motile object and other experimental configurations [20]. Here we will however limit ourselves to the above form, but with a parameterized baseline  $\alpha$  and coherence factor  $\beta$  [20] (Eq. 3.11). It is worth noting that above results generalize also to higher order  $g^{(n)}$  correlation functions [30] that are utilized in situations when the field does not satisfy the Gaussian assumption.

$$g^{(2)}(\tau) = \alpha + \beta |g^{(1)}(\tau)|^2 \quad (3.11)$$

The modified form increases parametric freedom when fitting experimental data, allowing the model account for experimental non-idealities such as baseline offset  $\alpha$  caused by the presence of static scatterers [31]. Crucially, the ‘coherence’ factor,  $\beta$  quantifies the quality

## Photon correlation spectroscopy

---

of the experimental setup and the achievable signal-to-noise ratio, acting as a measure of speckle contrast. Ricka [32] considered how the  $\beta$  varies with the number of modes imaged onto the detector for different experimental configurations. The three main configurations - two-pinhole, multi-mode fibre supporting  $N$  modes and single ( $N=1$ ) mode fibre cases are tabulated in Table 3.1. For the two pinholes measurement configuration [33] the ratio  $A/A_{coh}$  is the relative size of the imaged area  $A$  to the coherence area  $A_{coh}$ , average area of a single speckle. A similar interpretation of speckle can be applied in the case of the multi-mode fibre. Consequently, most modern instruments utilize a single mode fibre for detection, maximizing  $\beta$ , as this largely determines the initial autocorrelation amplitude relative to the noise baseline. The resulting low intensity requires the use of single photon detectors (discussed in Chapter 5). Variability of scattering signal intensities, due to particle scattering cross-sections or the particle concentration variation requires an large dynamic range of input intensities, up to 120 mW [34]. In this work however we aim for the more conservative 10 mW incident power, similar to the ARN-2 [35] and Zetasizer [36] systems, using CW Helium-Neon lasers.

Configuration	Coherence factor ( $\beta$ )
Single-mode fibre	1
Multi-mode fibre	$1/N$
Two pinhole configuration	$\frac{1}{A/A_{coh}+1}$

Table 3.1 Coherence factor  $\beta$  for different PCS detection configurations. The coherence factor can be considered as being approximately inversely proportional to the number of independent speckles averaged by the detectors.

As stated previously, the challenge of PCS stems from the need to utilize the above formulation in the reverse direction, to go from the measurement and  $I(t)$  and calculation of  $g^{(2)}(\tau)$  through to the recovery of the particle size distribution. This ‘inverse’ problem is covered next.

## 3.2 Inverse problems

An inverse problem is one which requires estimation of a set of parameters from a set of indirect measurements. An inverse problem falls into the class of ‘ill-posed’ problems which in turns are defined as problems that are not ‘well-posed’. The definition of a ‘well-posed’ problem was given in the early 20<sup>th</sup> century by Jacques Hadamard, for the case of linear problems

**Definition 3.2.1** (Well-posed problem). A problem is well-posed if it satisfies three conditions

- **Existence** - a solution to the problem exists
- **Uniqueness** - a solution to the problem must be unique
- **Stability** - the solutions depends continuously on the data

An ‘ill-posed’ problem, on the flip side, is one which does not satisfy **at least one** of these conditions. While Hadamard formulated the criteria for well-posed problems with linear forward operators, since then, many problems with nonlinear forward operators have been found. This is also the case for PCS where the forward model is composed from two constituents, a (linear) Fredholm equation (Eq. 3.8) and the nonlinear Siegert relation (Eq. 3.11). Direct inversion of PCS from  $g^{(2)}(\tau)$  is hence a nonlinear problem. However, a widespread approach to PCS inversion has been to ‘linearize’ the model through an initial estimation of  $g^{(1)}(\tau)$  followed by inversion of the Fredholm equation.

### 3.2.1 Linear inverse problems

#### Fredholm integral equation of the first kind

The Fredholm integral equation that relates the field correlation function to the particle size distribution is a frequent occurrence in inverse problem theory with the general form of Eq. 3.12.

$$g(s) = \int_b^a K(s,t)f(t)dt \quad (3.12)$$

The kernel  $K(s,t)$  represents a smoothing operation on the function of interest  $f(t)$ , with standard form integration bounds  $[a,b] = [0,1]$ . A commonly occurring special case is the convolution integral, occurring for kernels with form  $K(s,t) = K(s-t)$ , in optics a well-known example is the point-spread function. While traditional requirement of square integrability (defined on the infinite interval  $[-\infty,\infty]$ ) of the kernel  $K(s,t)$  does not hold for the PCS kernel (a decaying exponential), a variant of square integrability defined on an interval  $(a,b)$  does,  $K \in L_2(a,b)$  if:

$$\|K\|^2 = \int_b^a |K(t)|^2 dt < \infty$$

Qualitatively, the computation of the forward model, the Fredholm integral, acts as a smoothing operator. The inverse problem is then the task of ‘unsmoothing’ the data, amplifying the noise terms to reconstruct the high frequency features blurred by the integral. This suggests that our ability to reconstruct the high frequency features of the target function  $f(t)$  is determined by the smoothness properties of the kernel as well as the noise level in the measurement that will act to corrupt  $g(s)$ . One intuition that can be gained from this qualitative explanation is that future efforts in improving PCS, aimed at attaining higher resolutions should try to modify the kernel to reduce smoothness. Alternatively, as discussed later, experiments should attempt to minimize the noise in the measurement, a simple statement in theory, but more difficult in practice, due to the extreme sensitivity of reconstruction on the noise level.

### Discretization

The first component of solving the Fredholm integral numerically is its representation in a discrete form that can be analyzed by a computer. To do this we can evaluate the integrand over a finite set of points in order to discretize the integral. This can be achieved through ‘quadrature methods’ or projection of the integrand onto a suitable basis set. Here we consider only quadrature methods. The choice of quadrature method determines two factors, the quality of approximation and its rate of convergence to the exact value. The latter has an effect on the number of points required to reach a given precision when evaluating the integral. In the context of inverse problems a suitable choice of quadrature or basis set is vital as it reduces dimensionality of the inverse problem making solutions more tractable. Conversely, a poorly chosen quadrature method may either lead to large errors in the approximation of the integral or require a large number of ‘abscissa’ increasing problem dimensionality. The simplest quadrature method is the ‘midpoint’ rule, derived from Riemann summation, where uniformly spaced points over a given domain are used.

$$g(s_i) = \int_a^b K(s_i, t) f(t) \approx \sum_j w_{ij} K(s_i, t_j) f(t_j) = \sum_j A_{ij} f_j \quad (3.13)$$

Discretization of the Fredholm integral by a generic quadrature method can be expressed as Eq. 3.13. Here  $w_{ij}$  represent the discretization weights,  $t_j$  are termed ‘abscissa’ while  $s_i$  are the ‘collocation points’, positions where we set value of the output function equal to the approximation [37] [38]. For PCS there are two possible choices for discretization domain, the autocorrelation decay time constants or particle radii, the choice of discretization affects the effective spacing of abscissa in the particle radius domain and thus care must be

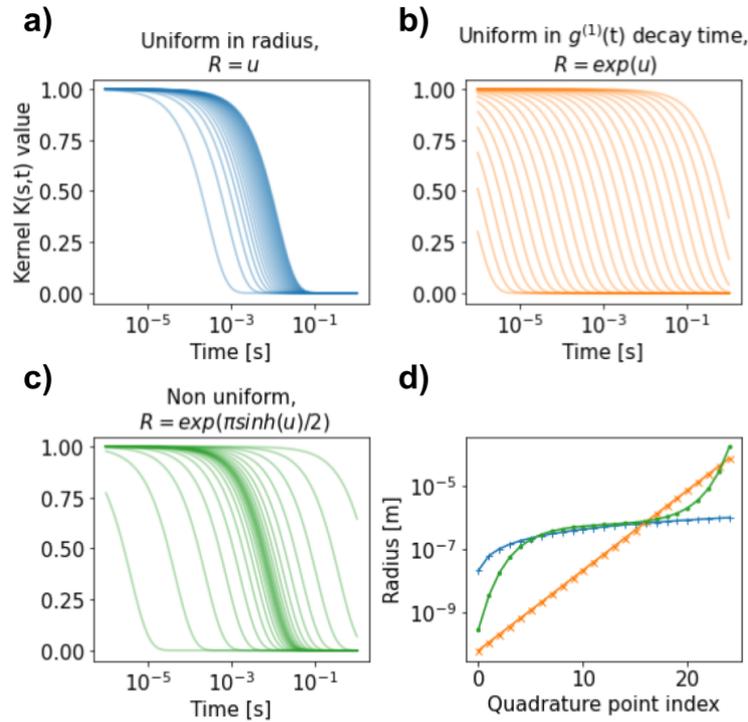


Fig. 3.3 Comparison of different quadrature methods for discretizing the Fredholm equation for the PCS kernel. a) Correlation function components generated by uniformly spaced abscissa (in particle radius,  $R$ ) exhibit nonlinear spacing in autocorrelation decay times. b) Correlation function components generated by exponentially spaced abscissa are distributed uniformly in autocorrelation decay time but lack resolution in the reconstruction domain. c) Nonuniform abscissa spacing spans a large range of decay times and has non-uniform density of point spacing, allowing high resolution reconstructions. d) Abscissa (particle radius  $R$ ) location as a function of quadrature point index, using 25 quadrature points for the discretization.

taken here in relation to forms of the priors imposed to stabilize the problem. Three possible discretizations for PCS and their impact of the resulting ACF components are depicted in Fig.3.3 (discussed in 3.3.3) showing that suitable choice is crucial for decomposition of the measured signal and reconstruction of the underlying size distribution. Once discretized the integral equation becomes a matrix equation, with recovery of a solution possible with a suitably designed inversion procedure.

### Condition number

Stability of the inversion of the linear forward operator (the matrix  $A$  in discrete form) can be characterized in terms of the ‘condition number’ of the  $A$  matrix (in the  $L_2$ -norm), the ratio of largest to smallest singular values

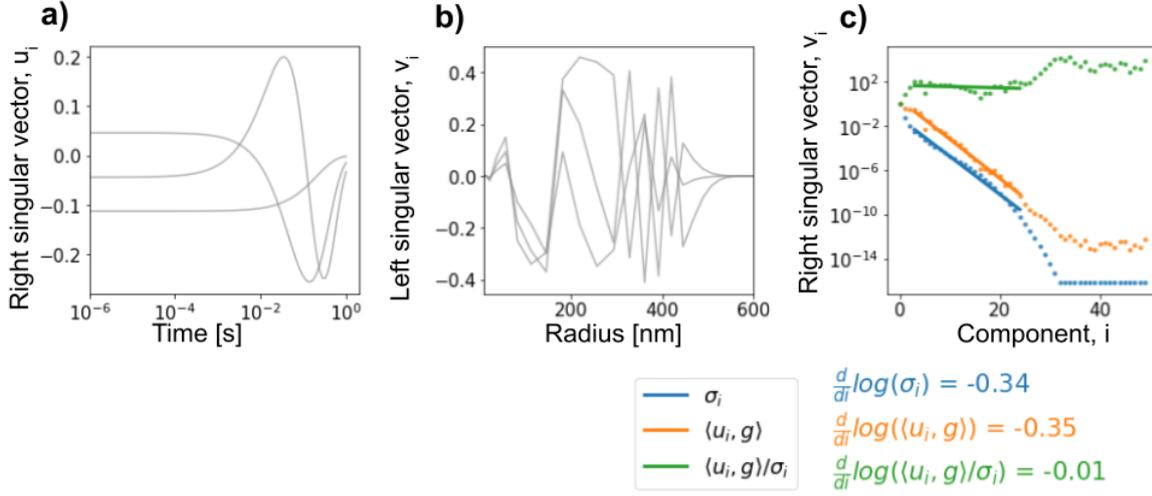


Fig. 3.4 a) Singular value decomposition  $u_i$  components representing the terms that will be used to decompose the measured autocorrelation b) Singular value decomposition  $v_i$  components representing the terms used to decompose the PSD. Note that due to lack of non-negativity constraints, the components can take on negative values. c) Picard plot showing decay of projections of  $\langle u_i, g \rangle$  components relative to  $\sigma_i$  singular values. It can be seen that the Picard condition is barely satisfied with the gradients differing by  $-0.01$  attributable to the smoothness of the kernel.

$$\kappa(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)} \quad (3.14)$$

For a linear problem with additive noise  $e$ ,  $Ax = b + e$  the condition number places an upper bound on the effect of the noise on the resulting reconstruction error [39]

$$\frac{\|x_{\text{exact}} - x\|_2}{\|x_{\text{exact}}\|_2} \leq \kappa(A) \frac{\|e\|_2}{\|b\|_2} \quad (3.15)$$

Thus inversion of matrices characterized by a large condition number is unstable in the presence of even small amounts of additive noise.

### Singular value expansion

A shared basis for inversion techniques is a representation of the integral equation in terms of a singular value expansion (Eq. 3.16). The kernel  $K$  is represented using right ( $u_i$ ) and left ( $v_i$ ) singular functions, with singular values  $\sigma_i$ , arranged in descending order  $\sigma_i \geq \sigma_{i+1}$  scaling the outputs  $u_i$ . Using these basis sets the original equation Eq. 3.12 can be expanded in terms of the left and right basis sets as in Eq. 3.16, [39].

$$\begin{aligned}
 K(s, t) &= \sum_{i=1}^{\infty} \sigma_i u_i(s) v_i(t) \\
 \int_b^a K(s, t) v_i(t) dt &= \sigma_i u_i(s)
 \end{aligned}
 \tag{3.16}$$

From this it can be seen that a smaller singular value  $\sigma_i$  (with increasing index  $i$ ) corresponds to a larger attenuation of  $v_i$  contributions to the final solution. The rate of decay of singular values is related to the aforementioned smoothness of the kernel. For the PCS kernel Fig.3.4b,a show the left and right singular vectors while Fig.3.4c shows the decay rates for both singular values  $\sigma_i$  as well as projections of the output function onto the right singular vectors  $u_i$ .

### TSVD and discrete Picard condition

Using the singular value expansion and selecting the  $k$  largest singular values  $\sigma_i$ , with corresponding functions  $u_i$  and  $v_i$  one can construct a solution to the inverse problem (Eq. 3.17), known as the Truncated Singular Value Decomposition (TSVD). Truncating at the  $k^{th}$  singular values acts to eliminate effects of all smaller singular functions, whose coefficients are assumed to be dominated by noise.

$$x_k = \sum_{i=1}^k \frac{u_i^T b}{\sigma_i} v_i
 \tag{3.17}$$

In order for the solution  $x_k$  to remain bounded as the number of coefficients  $k$  increases, the singular values must remain non-zero while the Fourier coefficients  $|u_i^T b|$  must decay to zero, on average, faster than the singular values  $\sigma_i$ , constituting the discrete Picard condition [40], illustrated in the case of PCS in Fig.3.4c noting that the smoothness of the exponential decay kernel results in a slow decay of  $\frac{\langle u_i, g \rangle}{\sigma_i}$  with increasing component index  $i$ .

The use of TSVD can also be used to explain the reduction of resolution, of the reconstructed solution, with increasing measurement noise as higher index  $i$  singular value functions  $u_i, v_i$  correspond to higher frequencies. The increasing noise, results in fewer components  $k$  being used in the reconstruction, subsequently lowering the resolution of the solution.

### 3.2.2 Tikhonov regularization

Instead of directly computing singular values and vectors as in TSVD, Tikhonov regularization (Eq. 3.18) augments the ill-conditioned problem with a penalty term. This stabilizes

## Photon correlation spectroscopy

---

recovered solution making the problem well-posed while also allowing the recovery of solutions ‘near’ the desired original while simultaneously being less sensitive to noise.

$$x = \operatorname{argmin}_x \left\{ \|Ax - b\|_2^2 + \lambda^2 \|L^{1/2}x\|_2^2 \right\} \quad (3.18)$$

The first term  $\|Ax - b\|_2^2$  represents the model and data misfit, while  $\|L^{1/2}x\|_2^2$  is the penalty term representing our beliefs about desired solution properties, often a ‘smoothness’ penalty, encoded in the precision matrix  $L$ . The regularizer  $\lambda$  tunes the relative contribution of two terms in the (scalarized bi-objective) optimization problem [41] and defines a Pareto-optimal front of solutions, in the case when the solution to the problem is unique and optimal. The latter of these can be guaranteed in the framework of convex optimization that provides guarantees on optimality of the solution, should a solution exist.

Similar to TSVD, the choice of  $\lambda^2$  is determined by the noise level in the measured data  $b$ , aiming to construct solutions from projection coefficients of basis vectors believed to not be contaminated by noise. For a given  $\lambda$  the solutions obtained can be directly related to the basis decomposition (for ‘standard-form’  $L = I$  of Tikhonov) discussed previously, where filter factors  $\phi_i^{[\lambda]}$  determine how each vector  $v_i$  is attenuated in the final reconstruction.

$$x_\lambda = \sum_{i=1}^n \phi_i^{[\lambda]} \frac{u_i^T b}{\sigma_i} v_i \quad (3.19)$$
$$\phi_i^{[\lambda]} = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}$$

While the choice can be done manually, automatic choice of regularizer is desirable. Multiple methods such as cross-validation, discrepancy principles (eg. Morozov) [42] and L-curve [43] are available for this. The latter seeing widespread use and a go-to tool for analysis of inverse problems, as it does not require inputs of noise level estimates (discrepancy) or incurring significant computational costs (cross-validation), but being limited in large part of the bi-objective case where the L-curve can be visualized on a 2D plot.

### Bayesian inference

Some of the aforementioned issues related to regularizer choice, such as model interpretability and noise level estimation may be addressed by reframing Tikhonov regularization into a statistical inference setting, formulating the inversion as a Bayesian inference problem.

$$\underbrace{\mathbb{P}(\theta|D)}_{\text{Posterior}} = \frac{\underbrace{\mathbb{P}(D|\theta)}_{\text{Likelihood}} \underbrace{\mathbb{P}(\theta)}_{\text{Prior}}}{\underbrace{\mathbb{P}(D)}_{\text{Evidence}}} \quad (3.20)$$

In this interpretation, Bayes theorem (Eq. 3.20) can be used to derive a posterior probability distribution  $P(\theta|D)$  for model parameters  $\theta$  given the measured data  $D$  through specifying a likelihood  $P(D|\theta)$ , encoding a noise model. A priori beliefs about distributions of model parameters are encoded in the prior  $P(\theta)$  while the normalizing constant, given by the evidence  $P(D)$ , is often neglected through replacing the equality of Bayes' rule with a proportionality ( $\propto$ ). Applying logarithms to Bayes' theorem allows the connection to Tikhonov regularization to be made (Eq. 3.21), where  $\delta$  and  $\gamma$  represent noise levels, related to regularizer  $\lambda$  through  $\lambda^2 = \delta/\gamma$ .

$$\log(\mathbb{P}(\theta|D)) = \underbrace{\frac{\gamma}{2} \|Ax - b\|_2^2}_{\log(\mathbb{P}(D|\theta))} + \underbrace{\frac{\delta}{2} \|L^{1/2}x\|_2^2}_{\log(\mathbb{P}(\theta))} - \overbrace{\log(\mathbb{P}(D))}^{=const} \quad (3.21)$$

### 3.2.3 Posterior sampling

In order to sample from complex posterior distribution  $q(\theta) = \mathbb{P}(\theta|D)$  a often used approach is Markov Chain Monte Carlo (MCMC). Two common examples are Metropolis-Hastings (MH) and Gibbs sampling.

Construction of the Markov chain by MH, at a point  $\theta_i$  in the parameter space is performed by sampling a new point  $\theta_p$ , conditional on the current point using a proposal distribution  $g(\theta_p|\theta_i)$ . The new proposal is accepted ( $\theta_{i+1} = \theta_p$ ) with probability  $\alpha$  (Eq. 3.22), and rejected ( $\theta_{i+1} = \theta_i$ ) with probability  $1 - \alpha$ . In the asymptotic limit of infinite samples the resulting sequence forms a stationary distribution equal to our desired posterior.

$$\alpha = \min\left(1, \frac{q(\theta_p)g(\theta_i|\theta_p)}{q(\theta_i)g(\theta_p|\theta_i)}\right) \quad (3.22)$$

While Tikhonov regularization computes a maximum-a-posteriori (MAP) of the functional, a point estimate (single instantiation of all variables), MCMC allows construction of both the mean estimator as well as computation of uncertainty bounds (typically 95% confidence intervals), providing solution uncertainty quantification (UQ).

With increasing dimensionality, simple MH proposals simultaneously updating all variables suffer poor scaling as samples produced exhibit high autocorrelation (slowing convergence) or have high rejection rates [44]. Sequential variable updating in Gibbs sampling may be beneficial, but require knowledge of conditional distributions of each variable. Techniques combining the two methods, such as ‘Metropolis-within-Gibbs’, where sequential updates are combined with a MH step can be applied to improve convergence, however convergence may still be slowed for groups of correlated variables requiring adaptive proposals that can transform samples to generate correlated variable block updates.

### 3.2.4 Optimization-based sampling

A limitation of above MCMC methods is that they often fail to leverage technological developments in the closely related fields of optimization, applied to solving conventional regularization problems. This is problematic for MCMC performance given recent high paced developments in gradient based solvers, where variants of stochastic gradient descent have been applied at scale and augmented with various hardware accelerators (ASICs or GPUs).

An additional consideration, arising mainly in optimization problems, is the handling of constraints and variable bounds. In the case of PCS such bounds arise naturally from the physical interpretation of a size distribution (non-negative) of particles, more broadly arising in imaging due to optical intensity non-negativity. Additionally the incorporation of equality constraints  $g^{(1)}(0) = 1$  (Eq. 3.8), vital for solution uniqueness, is an aspect that should be carefully considered.

#### Linear Non-negative Hierarchical Gibbs (LHGS)

In order to generate samples without direct use of a proposal function to perturb the current parameter vector (as in the case of simple MH or Gibbs sampling) one can instead perturb the measured data vector and use this to generate the samples of the hidden state through solving the corresponding optimization problem. The solution of this problem is then a sample from the posterior distribution. Here we describe the methods presented in [45, 46] which I implemented and applied to the problem of inverting PCS data, for the linear case. For the linear inverse problem with a non-negativity constraint on the parameter vector, with forward matrix  $A \in \mathbb{R}^{m \times n}$ , prior precision (inverse of covariance) matrix  $L \in \mathbb{R}^{n \times n}$  and measured data vector  $b \in \mathbb{R}^m$  and corresponding likelihood and prior given by

$$\begin{aligned}\mathbb{P}(b|x, \gamma) &\propto \gamma^{n/2} \exp\left(-\frac{\gamma}{2} \|Ax - b\|_2^2\right) \\ \mathbb{P}(x|\delta) &\propto \delta^{n/2} \exp\left(-\frac{\delta}{2} x^T Lx\right)\end{aligned}\tag{3.23}$$

Placing conjugate Gamma priors on  $\gamma, \delta$ , with locations  $\alpha_\gamma, \alpha_\delta$  and scales  $\beta_\gamma, \beta_\delta$ .

$$\begin{aligned}\mathbb{P}(\gamma) &\propto \gamma^{\alpha_\gamma - 1} \exp(-\beta_\gamma \gamma) \\ \mathbb{P}(\delta) &\propto \delta^{\alpha_\delta - 1} \exp(-\beta_\delta \delta)\end{aligned}$$

It is shown in [45] that samples from the posterior for  $x_k$  can be generated through solving the stochastic optimization problem in Eq. 3.24.

$$x_k = \arg \min_{x \geq 0} \left\{ \frac{1}{2} x^T (\gamma_{k-1} A^T A + \delta L) x - x^T (\delta_{k-1} A^T b + w) \right\}\tag{3.24}$$

Here  $w \sim N(0, \lambda A^T A + \delta L)$  is the sampled random noise perturbing the measured data. Full posterior inference is then performed through embedding the optimization problem in a hierarchical Gibbs sampler followed by iteratively sampling<sup>1</sup>  $P(x_k | \gamma_{k-1}, \delta_{k-1})$  followed by  $P(\gamma_k | x_k)$  and  $P(\delta_k | x_k)$  using the Gamma hyper-priors above:

$$\begin{aligned}\gamma_{k+1} &\sim \Gamma(n/2 + \alpha_\gamma, \frac{1}{2} \|Ax_k - b\|_2^2 + \beta_\gamma) \\ \delta_{k+1} &\sim \Gamma(n_p/2 + \alpha_\delta, \frac{1}{2} x_k^T Lx_k + \beta_\delta)\end{aligned}$$

Here in sampling  $\delta_{k+1}$  the location of the Gamma prior is  $n_p/2 + \alpha_\delta$ , with  $n_p$  being the number of non-zero elements of the of the parameter vector  $x_k$ ,  $n_p = n - |\{i | x_i = 0\}|$  (refer to [45] for in-depth description). From the above one can notice that the optimization problem (Eq. 3.24) is convex, allowing the use of fast iterative solvers and providing optimality guarantees of the solution.

<sup>1</sup>with slight abuse of conditional probability notation

### Randomize-then-Optimize (RTO)

In order to perform inference in the general nonlinear case, a modification of the above hierarchical Gibbs sampler is required. This is motivated by the need to perform general nonlinear inversion in the case of PCS where the linearity arising from the Siegert relation (Eq. 3.11) requires either linearization (first estimating  $g^{(1)}(\tau)$ ) or solution of the nonlinear problem. Below I describe the underlying algorithms implemented in the library based on [47–49].

The first step in generalization is in replacing the linear model  $f(x) = Ax$  in the likelihood of Eq. 3.23 with a general form  $f(x)$  giving a new likelihood:

$$\mathbb{P}(b|x, \gamma) \propto \gamma^{n/2} \exp\left(-\frac{\gamma}{2} \|f(x) - b\|_2^2\right) \quad (3.25)$$

A key feature of this form is the ability to incorporate the prior into an equation of the same form through augmenting the output vector  $b$  and nonlinear model to include the prior:

$$\tilde{f}(x) = \begin{bmatrix} f(x) \\ L^{1/2}x \end{bmatrix} \quad \text{and} \quad \tilde{b} = \begin{bmatrix} b \\ L^{1/2}x_0 \end{bmatrix} \quad (3.26)$$

The RTO method relates the distribution of parameters  $x$  obtained by solving Eq. 3.27 that is shown to be distributed according to  $\mathbb{P}_{x_{MAP}}(x) \propto c(x)\mathbb{P}(x|b)$ , with  $\mathbb{P}(x|b)$  representing the posterior distribution of parameters  $x$  given data  $b$ , with nonlinear forward model  $f(x)$ .

$$x = \operatorname{argmin}_{\theta} \|Q^T(f(\theta) - (b + \varepsilon))\|_2^2, \quad \varepsilon \sim N(0, I) \quad (3.27)$$

The approach proceeds by finding the maximum-a-posteriori (MAP) estimate  $x_{MAP}$  through solving  $x_{MAP} = \operatorname{argmin}_{\theta} \|f(\theta) - y\|^2$ . Evaluation the Jacobian  $J(x_{MAP})$  followed by a QR decomposition of  $J(x_{MAP}) = QR$  yields the matrix  $Q$  that scales the additive noise  $\varepsilon$  in the stochastic optimization problem.

The scaling factor  $c(x)$  can then be shown to be given by [47, 49]

$$c(x) = |Q^T J(x)| \exp\left(\frac{1}{2} \|Q^T(f(x) - b)\|_2^2\right) \quad (3.28)$$

The scale factor  $c(x)$  can then be viewed as the proposal weight term of the Metropolis-Hastings accept-reject step (Eq. 3.22) that can be used to generate samples from the nonlinear

inverse problem posterior distribution, for a given set of noise parameters. This sampler can then be embedded inside the hierarchical Gibbs sampling scheme similar to LHGS above to allow solution of the nonlinear inverse problem with automatically inferred noise estimates and UQ on each parameter, implemented in this work.

## 3.3 Experimental results

We now turn to experimental considerations when constructing the PCS system including requirements for the light source, spectral filtering techniques to generate tuneable incident wavelength. Alongside this we will present computational results of the developed inference engine for  $g^{(1)}$ , applied to synthetic data where we will show how the use of an optimizer allows generalization of the PCS inversion routine to both multi-angle and sequences of measurements inside a single inversion procedure. In addition to this we demonstrate the ability to apply the extension RTO for inverting  $g^{(2)}$ , applying it to experimental data measured by the constructed system. This is followed by a discussion of additional measurement modalities, manifesting as different correlation functions that may be used in future with the inference engine. We leave discussion of the various photon count data acquisition systems used to conduct the measurements to a subsequent chapter.

### 3.3.1 System overview

A diagram of the constructed system is presented in Fig.3.5. The incident light from a laser passes a spectral filtering module, followed by a variable attenuator controlling incident power and linear polarizer, producing a ‘vertical’ ( $\phi_i = 0^\circ$ ) polarization, plane perpendicular to the experiment. The attenuator is introduced for controlling both incident power and the detected count rate, which is influenced by sample dilution and determines the duration of a single measurement. The incident lens acts to focus the Gaussian beam inside the cylindrical (or square, single angle measurements at  $\theta = 90^\circ$ ) sample cell, defining the scattering volume.

The scattered light is relayed by a collection lens through a polarizer (polarizer angle  $\phi_c$ ) to a single mode fibre. The entire collection path assembly is mounted on a rotation stage allowing variation of collection angle  $\theta$ . Coupling into the single mode fibre is achieved through a end-fused graded refractive index (GRIN) lens. The single mode fibre selects a single mode (Table.3.1) maximizing the resulting  $\beta$  coherence factor (Eq. 3.11) while the polarizer allows one to perform both conventional polarized ( $VV, \phi_i=0, \phi_c = 0$ ) and depolarized ( $VH, \phi_i=0, \phi_c = 90^\circ$ ) measurements. The latter of these measurements typically

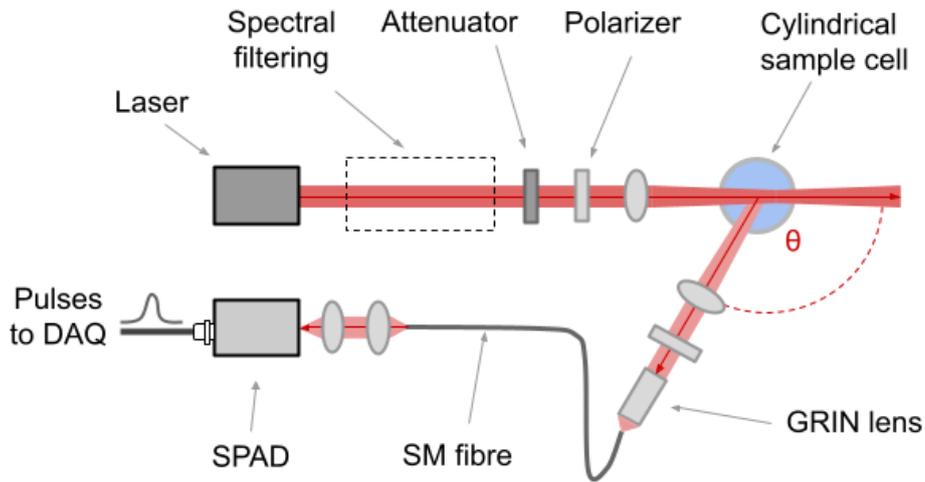


Fig. 3.5 Diagram of the constructed multiangle PCS setup. Additional wavelengths can be added through using a fibre based beam combiner. In the system an incident laser beam is collimated, filtered spectrally and polarized before being focused into a sample cell containing the particles. The scattering signal, collected at a scattering angle  $\theta$  is relayed to a GRIN lens that focuses the scattered light into a single mode fibre transmits the signal to a single photon avalanche diode (SPAD) producing output pulses detected by acquisition electronics.

suffers from significantly reduced count rates comparatively to the  $VV$  polarized experiment, requiring either longer measurement times or increased incident laser intensity.

The light propagated through the single mode fibre is subsequently focused onto a single photon avalanche diode (SPAD) producing pulses recorded by a single photon counting system ('nanosecond' variants, described in Chapter 5). From considerations of detector non-linearity (Eq. 5.1), in order upper bound the count rate error due to detector dead time 1%, the incident intensity is adjusted to keep the detected counts  $\leq 100$ kcps.

### 3.3.2 Spectral filtering

In order to motivate the requirement for spectral filtering we refer back to the theory of light scattering (Section 3.1) noting that constructive or destructive superposition relies on a source with sufficient spatio-temporal coherence properties, producing a detectable contrast between 'constructive' maxima and 'destructive' minima intensity fringes. One key goal of this project was the evaluation of using a supercontinuum white light laser for the purpose of performing PCS measurements at multiple wavelengths either simultaneously or in quick succession. Thus in order to achieve this we investigated techniques for spectral filtering of the supercontinuum spectrum.

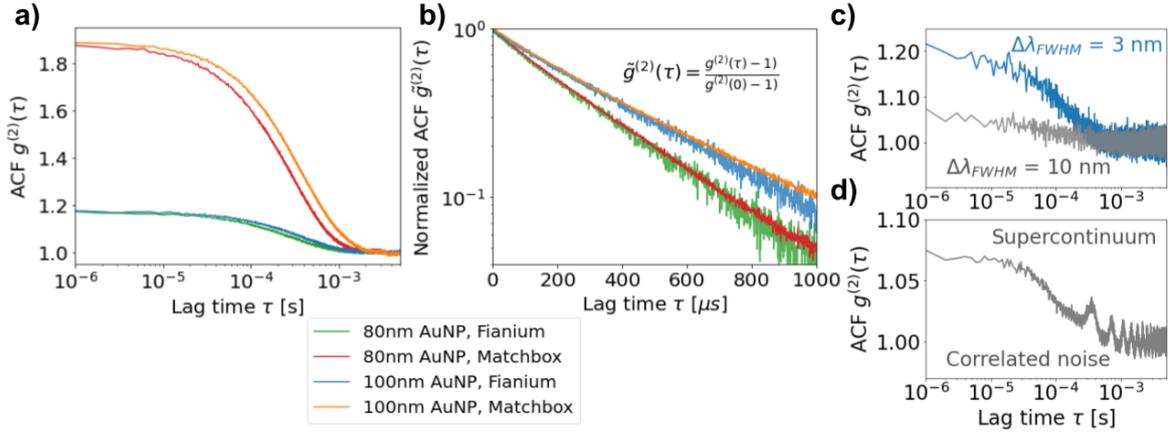


Fig. 3.6 Autocorrelation function measurement for 80 nm and 100 nm diameter AuNPs using a CW,  $\lambda = 633$  nm (Matchbox, Integrated Optics) and a pulsed supercontinuum (SC-450, Fianium) filtered using a  $\Delta\lambda_{FWHM} = 3$  nm interference bandpass filter centred on 633 nm. a) Autocorrelation function measured using the CW (integration time = 1 s) and pulsed supercontinuum (integration time = 20 s) b) Normalized autocorrelation function measurements demonstrating increased noise level of supercontinuum relative to CW laser c) Autocorrelation functions for supercontinuum filtered using  $\Delta\lambda_{FWHM} = 3$  nm and  $\Delta\lambda_{FWHM} = 10$  nm bandpass interference filters d) Supercontinuum correlated noise floor arising from laser intensity fluctuations.

The first benchmark for spectral filtering is a standard interference bandpass filter, with bandwidths of 10-20 nm due to their widespread availability. Bandpass filters of this type can be utilized to create narrow-range tuneable filtering through rotating angle of the filter relative to the incident direction of light. Effects of bandwidth on the quality of the correlation function are shown in Fig.3.6a, where a comparison between a 3 nm and a 10 nm filter (centred on 633 nm) on the correlation function can be seen, with the  $\beta$  factor reducing from  $\sim 0.2$  to  $\sim 0.08$ . The performance of a 3 nm filtered optical signal from the supercontinuum can be contrasted, for measurements of 80 and 100 nm AuNPs with the same samples measured using a CW diode laser (Fig.3.6c,d), exhibiting  $\beta \sim 0.9$  ( $\times 4.5$  gain). Reduction of noise in both measurements is achieved through averaging correlations recorded at 0.1 second integration time per correlation. In the figure the filtered supercontinuum signal required  $\times 200$  averaging while the CW diode laser requires  $\times 10$  averaging but still demonstrates superior performance in terms of noise reduction as seen in Fig.3.6d where  $\beta$ -normalized correlation signals are overlaid. The effect of increased noise manifests even in simple monodisperse fits using the cumulants method [50] with CW diode radius estimates of 87.4 nm and 110 nm for the 80 nm and 100 nm particles respectively, while size estimates from filtered supercontinuum correlations produced 88.8 nm and 115 nm respectively. An

## Photon correlation spectroscopy

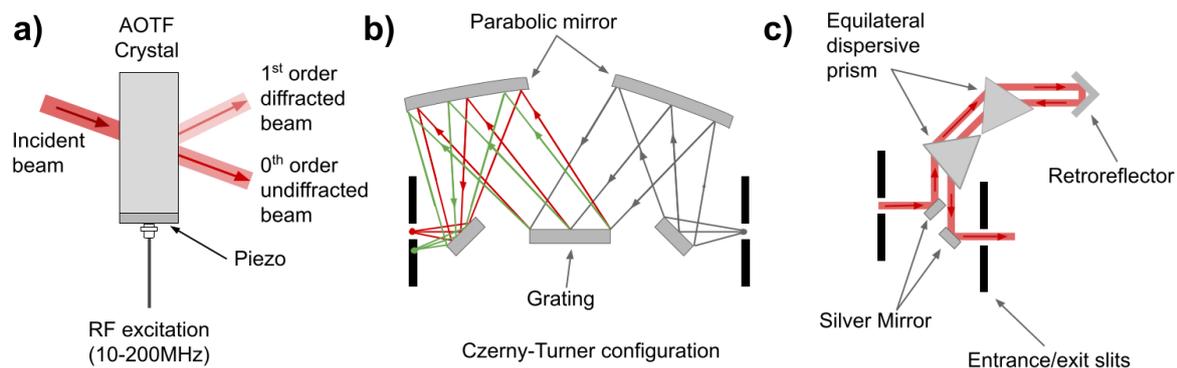


Fig. 3.7 a) Acousto-optic tunable filter. b) Single grating Czerny-Turner monochromator. c) Four pass dispersive prism configuration.

additional consideration in utilizing supercontinuum sources is laser intensity instability, resulting in an increased correlated noise floor, as shown in Fig.3.6b).

While a tuneable system using interference filters could be built, due to providing sufficient optical intensities of 1-2mW suitable for some types of samples, the prohibitive linear scaling of cost for each additional wavelength motivated investigation of alternative sources.

The first of the alternative tuneable systems are devices based on acousto-optic (AOTF) effect (Fig.3.7a) where a volume Bragg grating is created by an acoustic wave propagating through a birefringent crystal ( $TeO_2$ ). The period and amplitude of the acoustic wave are controlled through a piezo actuator driven by a RF signal (20-200MHz, 1-2W) attached at one end of the crystal. While high speed control is possible, chief limitations to this approach are the low spectral power density of the supercontinuum and its unpolarized output, which when combined with reduced optical performance of the AOTF compared to interference filters resulting in sub-mW optical powers onto sample. Consequently the combination of supercontinuum with AOTF lacks the dynamic range required for dilute samples (up to 10mW). A secondary limitation is the broader bandwidth (Fig.3.8a) of the signal compared to the bandpass, resulting in a low coherence factor,  $\beta \leq 0.2$ .

Achieving bandwidths similar/exceeding those of interference filters can be done through grating or prism dispersive elements. The former, arranged in Czerny-Turner (Fig.3.7b) configuration, incorporated into a Bentham DTMc300 (dual turret Czerny-Turner found was capable of achieving  $\sim 1$  nm bandwidths but again, due to sub-mW output power to sample was not suitable for PCS measurements. The final 'home-built' system based on a four-pass dispersive prism configuration (Fig.3.7c) did achieve 1 mW output power at a 3 nm passband (Fig.3.8b) but due to its mechanical complexity and similar  $\beta$  value to the interference filter,

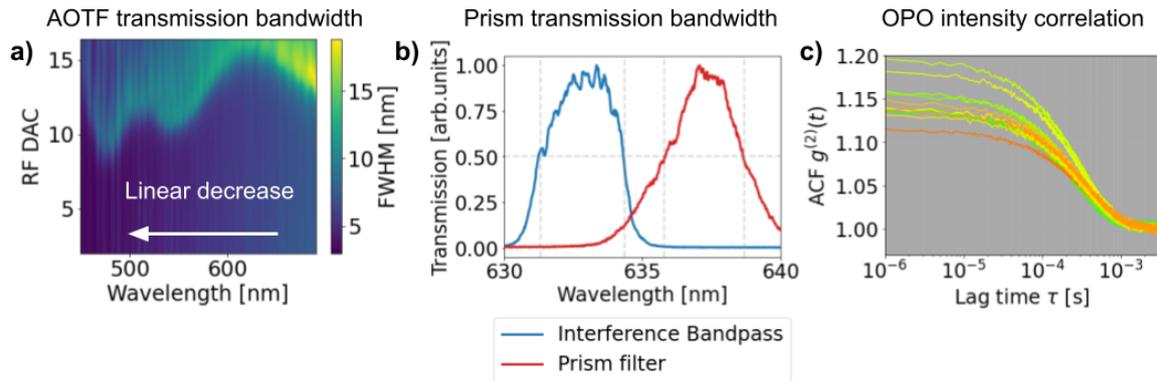


Fig. 3.8 a) AOTF performance as function of RF drive power showing a narrow bandwidth FWHM at low drive powers and nonlinear increase in FWHM at drive DAC values above 10 b) Spectrum comparison between lab-build prism filter and 3 nm interference bandpass filter. c) Intensity autocorrelation recorded using OPO as light source, using wavelength in the 510 nm to 620 nm range. Autocorrelations are coloured by wavelength.

was rejected as a viable option for generating tuneable, narrowband, high power ( $\sim 10$  mW) light.

The issue of spectral power density can be resolved through replacing the supercontinuum laser with an Optical Parametric Oscillator (OPO) system (Inspire, Spectra-Physics). The OPO output exhibits performance comparable to the interference filters, with  $\beta$  values in the range of 0.1 – 0.2 (Fig.3.8c) but allows for higher dynamic range of optical powers, with an average output power of  $\sim 200$  mW at 532 nm. A difficulty with applying such systems is the short optical pulse duration and high repetition rate. This feature of pulsed OPOs that makes characterization of their temporal coherence properties challenging but one can understand the limitation of applying pulsed systems by considering the spectral density, its relation to temporal coherence (that is  $g^1(\tau)$ )

As will be discussed in the subsequent section the quality of measured correlations obtained using this technique was not deemed to significantly improve the accuracy of the obtained results when compared to a conventional CW diode laser, with even simple CW laser diodes found to exhibit coherence factors in the 0.5 – 0.6 range.

The above investigation of light source on properties of  $\beta$  indicate that suitable choice of light source is critical for PCS experiments. Unfortunately common tools available in labs for filtering may not be sufficient for construction of high performance wavelength tuneable PCS systems. Based on comparisons of different light source types it is suggested that a multi-wavelength system with suitable power and coherence factor,  $\beta$ , requirements ( $\geq 0.8$ ) may be constructed from a set of CW diode laser sources with output powers in the 30-100

mW range, merged and power controlled by an AOTF, similar to many configurations used in confocal microscopy. In this configuration one would be able to get the benefits of both temporal coherence and power of the CW lasers rapid switching between the set of available wavelengths using the AOTF. Broad tunability of wavelength may be achieved through control of each of the laser diodes case temperatures which is known to tune center wavelength by  $0.25 \text{ nm}/^\circ\text{C}$ , enabling each diode to be tuned through  $12.5 \text{ nm}$  using a  $50^\circ\text{C}$  temperature range. An alternative to the diode laser array are CW OPO systems such as the Hubner Photonics C-Wave specified to cover the entire visible range with output powers in the 200 mW range and 1 MHz linewidths. This reasoning can be understood by interpreting the field correlation function  $g^{(1)}(\tau)$  as the degree of temporal coherence, with its decay characterized by a coherence time  $\tau_c$ . Assuming a Gaussian optical pulse duration  $\Delta\tau_{FWHM} = 100 \text{ ps}$  generated by the Inspire OPO one can estimate the linewidth  $\Delta f_{FWHM} = 2.86 \times 10^{11} \text{ Hz}$ , allowing the calculation of a coherence time  $\tau_c$  and coherence length (in air)  $l_c = c\tau_c$ , for the Gaussian spectral density using [51]:

$$\Delta f_{FWHM} = \frac{2\sqrt{2\ln(2)/\pi}}{\tau_c} \approx \frac{0.66}{\tau_c} \quad (3.29)$$

For the Inspire OPO this yields  $\tau_c = 2.0 \text{ ps}$ ,  $l_c = 6.9 \times 10^2 \mu\text{m}$ . Comparatively, for a CW system with a linewidth of 1 MHz,  $\tau_c = 0.66 \mu\text{s}$  and  $l_c = 200 \text{ m}$ , five orders of magnitude greater. Thus CW systems, either laser diodes or CW OPOs provide significant improvements over pulsed systems.

### 3.3.3 Inference engine

In order to integrate measurements varying multiple parameters into a single reconstruction of the particle size distribution, as well as recover multi-modal size distributions and estimate polydispersity of the PSD, fitting based on cumulant analysis [50] is replaced by the inverse problem formulation. The main technique for solving this has been the use of regularization methods, formulated as optimization problems. In PCS literature this is referred to as either CONTIN [52] honoring the name of the first software the package implementing a quadratic program solver for weighted least squares regression for PCS, or Non-negative Least Squares (NNLS), a name more accurately describing the form of the optimization objective solved during inference. These methods form the basis of for estimation of PCS size distributions, implemented by PCS instrument manufacturers like Malvern Panalytical (NNLS) [53] and LS Instruments (CONTIN) [54], with suitable generalization to handling multiangle measurements recording autocorrelations at different values scattering vectors

$q$  (Eq. 2.1). While newer techniques involving solution of full nonlinear PCS have been advertised (CORENN, LS Instruments, [55]) at time writing publications describing their features and methodology are lacking, thus quadratic program solvers remain the workhorse of PCS measurements.

Quadratic programs are subsets of convex [41] programs solvable using convex optimization, having desirable convergence properties (polynomial,  $O(n^3)$ ) with optimality guarantees if a feasible solution, satisfying a given set of constraints. However it is limited to the class of convex objective functions that we construct using using Disciplined Convex Programming (DCP) [56] and are suitable only for linearized (ie.  $g^{(1)}(\tau)$ ) inversion in PCS. With the full PCS model being non-convex its inversion has previously been considered in the MCMC framework methods like Metropolis-Hastings [57] or Metropolis-within-Gibbs [58] sampling. While asymptotically convergent to the posterior distribution, such methods scale poorly with increasing dimension and have trouble when sampling highly correlated sets of variables. Consequently such methods require careful tuning, a large number of samples and are not scalable to high dimensional inverse problems (‘curse of dimensionality’).

In this work we propose that high dimensional PCS problem may be solved more efficiently through combining MCMC with optimization techniques. The former lets us stay within the Bayesian framework, allowing for explicit estimation of measured data noise level and posterior probability of the particle size distribution (PSD). The posterior provides uncertainty quantification (UQ) of the recovered PSD and enables an intuitive connection between effect of noise magnitude and quality of PSD reconstruction, not previously considered in PCS. The issue of correlated noise floor in the measurement, while not considered here can be communicated through plotting residual error between experimental data and model fit. In the case when noise floor can be measured independently the measured data can be ‘pre-whitened’ [48] as part of the inversion procedure. The incorporation of optimizers in the sampling allows for increased efficiency in sample generation (less samples, lower sample autocorrelation) and opens the potential to scale the inverse problem to simultaneously invert a sequence of PCS measurements, instead of the more common approach to invert each measurement separately. Dynamic inverse problems like this allow for coupling information across successive PSD estimates and regularize the solutions in time. The temporal regularization may in future be combined with multi-angle measurements to improve the resolution through performing adaptive experiments that utilize the current estimate PSD and its UQ bounds to determine the next experimental parameters such as angle, wavelength, temperature that can maximize the amount of information obtained from the next measurement. In this context staying within the Bayesian framework has the advantage that such data driven experiment can be

formulated from the perspective of minimizing the UQ bounds. Doing this rapidly during an experiment also benefits from the use of optimizer based sampling that enables one to exploit software and hardware frameworks such as GPU based training, in order to accelerate the optimization.

### Linear inverse problem

Part of formulating the linear inverse problem consists in choice of suitable discretization, which in PCS is nontrivial due to the interplay between discretization and representation of the autocorrelations and particle sizes, related through the exponentially decaying kernel (Eq. 3.8), with the particle radius  $R$  and decay time constant  $\Gamma$  being inversely related  $\Gamma \propto 1/R$ .

Using the midpoint rule to discretize uniformly in radius is desirable as this simplifies the form of the  $2^{nd}$  derivative matrix smoothness prior to the standard  $(1, -2, 1)$  form. This quadrature is used in the nonlinear inverse  $g^{(2)}$  inversion due to its simplicity and ease of debugging. For the linear inverse problem where the convex optimizer guarantees solution optimality, we are therefore less cautious to choose the prior without fear of convergence issues, thus we investigated other quadrature methods motivated by the limited ‘dynamic range’ of the correlation functions produced by uniform abscissa spacing in radius  $R$  (Fig.3.3a,d). While various other quadrature techniques exist we were interested in retaining clear interpretability of the quadrature scheme, and considered generalizations of the midpoint quadrature. One such technique, initially proposed in [59], uses transformation of the midpoint quadrature applied over a uniform coordinate  $u$ , to generate different non-uniform quadrature schemes through suitably selecting a transformation function. The first such considered transformation is  $R = \exp(u)$  (Fig.3.3b) producing uniformly spaced correlation functions in measurement domain (time lag  $\tau$ ) and being exponentially distributed in particle radius  $R$ . While having high ‘dynamic range’ in representing vastly different autocorrelation decay rates, this scheme is limited in its resolution in  $R$  due to the exponential spacing of the abscissa. While not explicitly described in commercial PCS instrument manuals, it is likely that a similar quadrature scheme is used there, which we infer from the logarithmic PSD scales that PCS results are plotted on. An alternative transformation that can balance the need for high resolution and high dynamic range is  $R = \exp(\pi \sinh(u))$  [59]. This exhibits a nonlinear spacing of points in  $R$  (Fig.3.3d) along with a nonlinear spacing of autocorrelation components (Fig.3.3c) and represents the intermediate form between  $R = u$  and  $R = \exp(u)$ . We parameterize this form to become  $R = a \cdot \exp(b \cdot \sinh(u))$ , allowing the adaptation of the quadrature to the measured autocorrelation while simultaneously preserving the high dynamic range. It is worth noting

the need for a large dynamic range stems not only from the desire to cover a large range of particle sizes, but to enable the models to incorporate effects of detector afterpulsing (Fig.5.7) which manifest as an increase of the autocorrelation at short lag times ( $< 10 \mu\text{s}$ ). In addition to qualitative arguments it can also be shown that for PCS Fredholm integrals, the above double exponential form achieves faster rates of convergence as the number of quadrature points is increased, compared to either uniform or exponentially spaced abscissa. This improved convergence aids in minimizing the inverse problem dimensionality allowing for more efficient sampling. One challenge in using non-uniformly spaced discretization is also the need for generating adaptive priors that suitably incorporate smoothness constraints on the non-uniform grid where simple forms such as second-order difference matrices are no longer applicable due to non-equidistant abscissa spacing. While generalizations of gradient operators to non-uniform grids are possible [60], alternative options such as GP kernels [27] can be used - investigations of these for PCS inversion are left to future work.

Once discretized the Fredholm integral is converted to a matrix equation (Eq.3.13), with matrix  $A$  representing the forward model. For PCS this can be analyzed through SVD by considering the discrete Picard plot along with the left and right singular vectors (Fig.3.4). Here we verify that the Picard condition is satisfied, with  $\frac{\langle u_i, g \rangle}{\sigma_i} = -0.01$ . This small rate of decay is indicative that the PCS kernel is highly smooth.

The software library developed in this work implements a variant of LHGS described in the previous section along with RTO, implemented as a proposal distribution embedded inside the Metropolis-Hastings step inside the hierarchical Gibbs sampler. As LHGS incorporates a convex problem, we interface with different convex solvers including CVXOPT, ECOS and SCS through building optimization functionals using CVXPY [61]. As RTO in general is non-convex this is not utilized for the nonlinear case, where we instead rely on variants of stochastic gradient descent based gradient solvers. By making use of higher order functions in python we are able to allow great flexibility in supporting different choices such as quadrature type, solver type and options as well as different forward models, making the library applicable to different inverse problems. The interface of the library is separated into two categories ‘LinearModel’ (LHGS) and ‘NonlinearModel’ (RTO), each supporting a *sklearn*-like ‘fit-predict’ [62] interface, simplifying the execution of the inversion algorithms for a end-user.

Before considering more complex cases we demonstrate equivalence between solutions obtained via L-curve method and LHGS for the linear PCS problem in Fig.3.9a. To do this we simulate a correlation function computed from a bimodal ( $140 \pm 20 \text{ nm}$  and  $240 \pm 20 \text{ nm}$ ) particle size distribution. The forward field correlation function is corrupted by additive

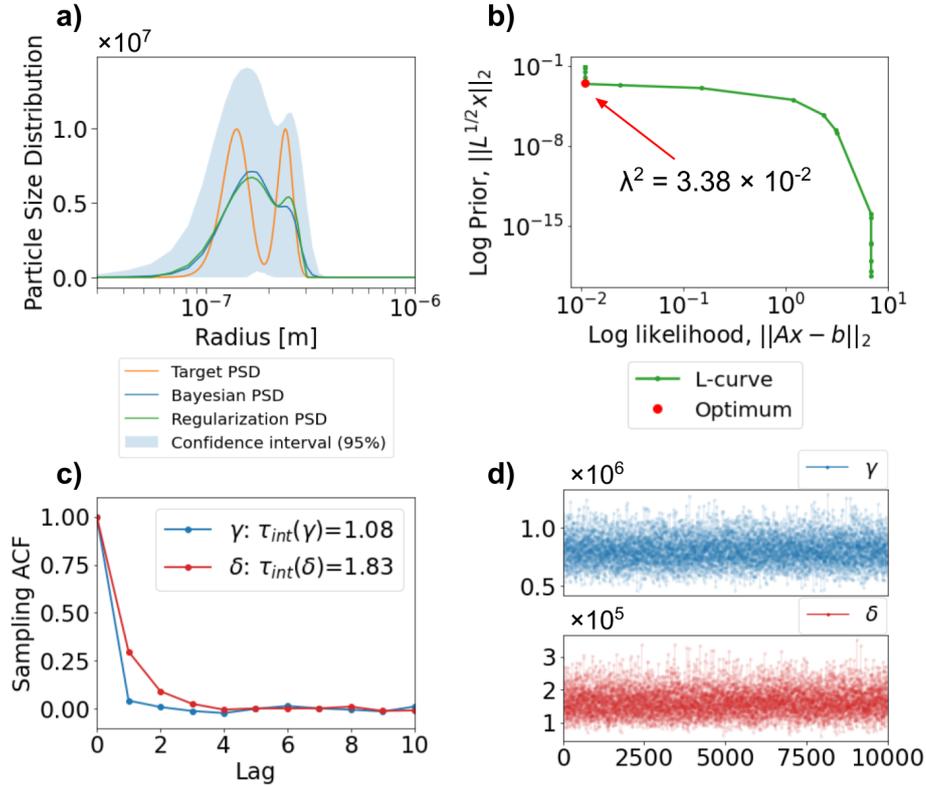


Fig. 3.9 a) Inversion of synthetic bi-modal size distribution demonstrating that LHGS (‘Bayesian PSD’) recovers the same solution as the the Tikhonov regularization using the L-curve criterion for regularizer selection. b) L-curve for the bi-modal PSD inversion showing in red the point of maximum curvature selected as the optimal regularizer c) Auto-correlation of MCMC samples generated during inference exhibit low values of integrated autocorrelation  $\tau_{int}$  resulting in efficient sampling. d) MCMC sample traces for the  $\gamma$  (likelihood) and  $\delta$  (prior) noise precision estimates, showing fast convergence to the stationary distribution.

Gaussian noise  $\varepsilon \sim N(0, 1 \times 10^{-3})$  and subsequently inverted, using either L-curve (Fig.3.9b) or LHGS. In generating the synthetic data we avoid committing an ‘inverse crime’ of using the same discretization for both data generation and inversion, using  $10^4$  uniformly spaced quadrature points for when computing the forward model, while using only 50 in the inverse case, generated using the  $R = a \exp(b \sinh(u))$  quadrature scheme. The consistency between the L-curve technique and LHGS validate the applicability of the technique to PCS data, along with the implementation. The hyperpriors for LHGS are for this case set to  $\alpha_\gamma = \alpha_\delta = 1$  and  $\beta_\gamma = 10^{-8}$ ,  $\beta_\delta = 10^{-4}$ , while the L-curve recovered regularizer value  $\lambda^2 = 3.38 \times 10^{-2}$  is extracted as the point of maximum curvature. While both methods produce a overly smooth solutions relative to the target function, the 95% confidence bound produced by

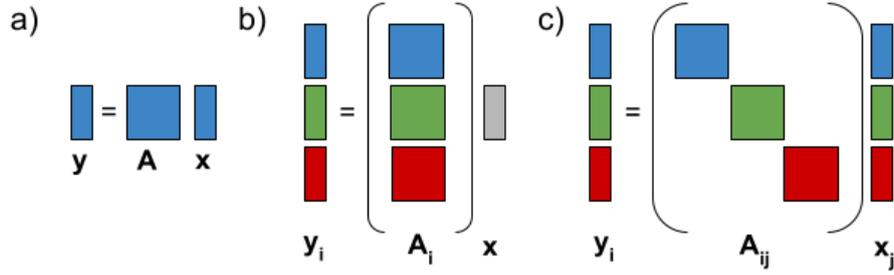


Fig. 3.10 a) Linear inverse problem with single forward model and measurement, a conventional PCS measurement. b) Inverse problem consisting of multiple forward models and measurements, the form of a multi-angle PCS measurement. c) Multiple forward models and reconstructed measurements, a dynamic inverse problem that can include multi-angle measurements but also accounts for temporal variation of the PSD.

LHGS includes the target PSD. We are also able to estimate noise level  $\varepsilon \approx 1.1 \times 10^{-3}$  using samples of  $\gamma = 1/\varepsilon^2$  (Fig.3.9d), similar to the original  $\varepsilon$  value used to corrupt the synthetic  $g^{(1)}(\tau)$  autocorrelation data.

The LHGS framework also supports extension to multiple forward models, as is the case of multi-angle light scattering measurements (Fig.3.10b). The use of L-curve technique in this case becomes more challenging due to the increasing number of hyperparameters  $\gamma_i$ , resulting in increased difficulty in visualization of the L-curve, typically applied in only 2D. The use of Bayesian formalism on the other hand allows easy extension, by considering each of the measurements as independent. The corresponding likelihood (Eq. 3.21) is the product of the individual measurement likelihoods. Thus the optimized functional is generalized to the sum of individual likelihood measurements, scaled by associated noise terms  $\gamma_i = 1/\varepsilon_i^2$  terms:

$$\sum_{i=1}^p \frac{\gamma_i}{2} \|\mathbf{A}_i \mathbf{x} - \mathbf{b}_i\|_2^2 + \frac{\delta}{2} \|\mathbf{L}^{1/2} \mathbf{x}\|_2^2 \quad (3.30)$$

While natural in the case of LHGS, such an extension for regularization is more difficult due to increased number of regularizers (Eq. 3.30, Fig.3.10,b). This makes it harder to apply the L-curve method (often limited to 2D). An example of such a reconstruction using the developed library is shown in Fig.3.11, where the technique is applied to data generated using the same bimodal distribution as in the previous example. Each of the five correlation functions are generated by using the ideal field correlation  $g^{(1)}$  to which a zero mean noise vector drawn from  $\varepsilon \sim N(0, 10^{-3})$  five times and subsequently added to the uncorrupted correlation function. This results in five measurements (Fig.3.11a) that are subsequently

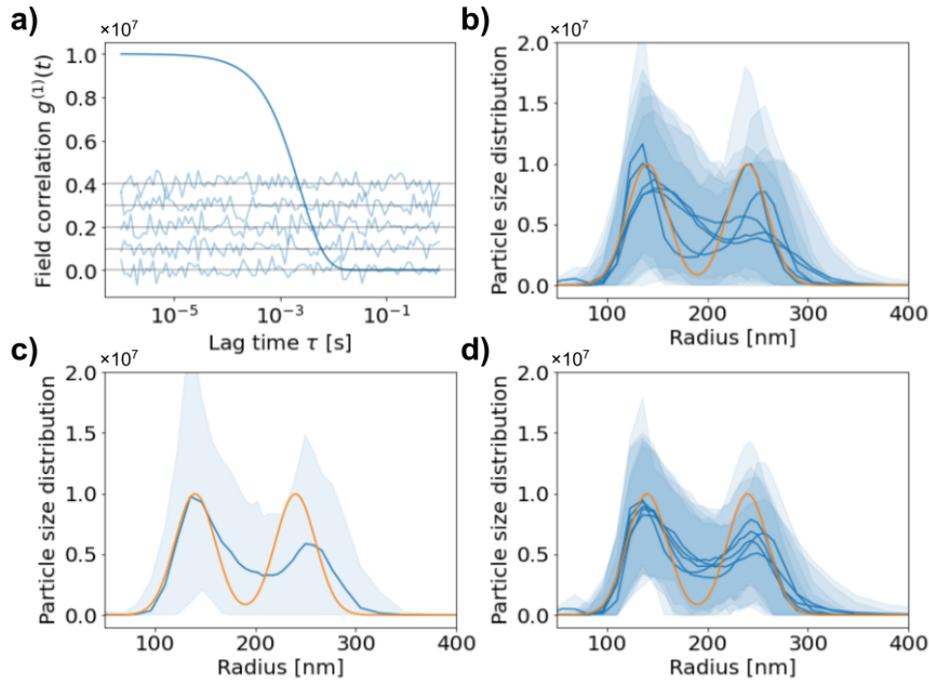


Fig. 3.11 a) Field correlation function with five random noise vectors, each vector added to the correlation function is considered a separate measurement b) Solutions to each of the measurements independently using LHGS c) Single solution vector obtained by integrating all five measurements into a single inversion d) ‘Temporally’ coupled solutions obtained augmenting the prior matrix to couple each of the radius ‘R’ variables across measurements

inverted. Inverting each measurement separately, shown in Fig.3.11b , demonstrates that even at the same noise level recovered results may/may-not resolve the bimodal size distribution. Based on this, future work using LHGS can help determine which of the sampled lag times play a role in this (sensitivity analysis), allowing for optimal lag time selection, as current schemes for lag time selection are limited to using exponentially spaced lag times. Combining the five measurements in a single inverse problem yields the PSD estimate in Fig.3.11c. While resolving the individual peaks the solution does not demonstrate reduced 95% bounds, it is included to show the feasibility of multi-angle/multi-forward matrix inversion using the developed library.

Leveraging the convex optimizers scalability to solving high dimensional problems is the next extension (Fig.3.10c). To do so, each of the individual measurements and forward models reconstruct separate PSDs (similar to Fig.3.11b), with an additional ‘temporal’ coupling introduced between the individual PSD vectors. This has the effect of requiring individual priors to be specified, one for each PSD (Eq. 3.31) along with the temporal coupling prior  $L_T$ . Experimentally this can be understood as simultaneously inverting a sequence of

PCS measurements under the assumption of a smoothly varying/time-evolving PSD. In the case when measurements are equally spaced in time and with a uniform abscissa spacing in  $\mathbf{R}$  the resulting combined prior forms a structure similar to Gaussian Markov Random Field (GMRF) with the prior precision having non-zero entries only for neighbours of a particular random variable. The dimensionality of such a problem increases as  $O(n_q \cdot n_p)$  where  $n_q$  represents the number of quadrature points while  $n_p$  is the number of PSDs being reconstructed. The resulting reconstruction shown in Fig.3.11d exhibits reduced 95% uncertainty along with improvements in the individual PSD estimates, with all having pronounced peaks. This improvement in stability of reconstruction comes at the cost of being the most computationally intensive variant, however is it perhaps the technique with widest appeal as many commercial PCS systems perform measurements at a single angle and wavelength only, thus they are unable to modify the forward matrix  $A$  during an experimental run, if ever.

$$\sum_{i=1}^p \left( \frac{\gamma_i}{2} \|\mathbf{A}_i \mathbf{x}_i - \mathbf{b}_i\|_2^2 + \frac{\delta_i}{2} \|\mathbf{L}_i^{1/2} \mathbf{x}_i\|_2^2 \right) + \frac{\delta_{p+1}}{2} \|\mathbf{L}_T^{1/2} \mathbf{x}_t\|_2^2 \quad (3.31)$$

Before considering the nonlinear inversion, it is useful to note that the multi-angle (Eq. 3.30) and temporal (Eq. 3.31) inversions are not mutually exclusive. In the case when multiple measurements at different scattering angles can be made simultaneously, each individual multi-angle inversion may be embedded inside the temporal structure to allow for monitoring of the timeseries of measurements. Two challenges of scaling are encountered here, first of which is the scaling of the forward model size, while the second is that of the estimated parameter space. While challenging to solve with the current implementation, it is anticipated that with proper code optimization this would be feasible with the proposed technique.

#### Nonlinear inverse problems

The above linearized inverse problem was solved with a measurement of  $g^{(1)}(\tau)$ . In practice estimation of  $g^{(1)}(\tau)$  from experimental data (via Eq. 3.11) is problematic as propagating residual errors and residual correlations from the preliminary fitting step would influence recovered PSDs. An example of the challenges associated with such preliminary processing is the requirement of taking the square root of  $g^{(1)}(\tau) = \sqrt{(g^{(2)}(\tau) - \alpha)/\beta}$ , undefined for negative values, leading to data point censoring in the model.

With the forward model being nonlinear the convex optimization tools are no longer directly applicable. We now require general purpose gradient based solvers<sup>2</sup>. To prevent non-uniqueness of the solution, the formulation now requires introducing the explicit equality constraint  $g^{(1)}(0) = 1$ , in addition to the previous non-negativity constraint imposed on the PSD random variables.

Selection of a suitable solver for such nonlinear least squares problems is again crucial for achieving good performance. The freely available solvers in Scipy [63] as well as variants of SGD in the Jax [64] package are utilized. Due to differing support for equality and inequality constraints in their APIs, some of the solvers require explicit formulation of the constraints, while others do not. To handle constraints in the case when a given solver does not explicitly support this we can rewrite the constrained (inequality and/or equality) optimization as an unconstrained problem with additional ‘penalty’ terms. A theoretical penalty for an inequality constraint is an indicator function  $I_\infty(u)$ , zero on the domain of interest and infinite outside. While correct in theory, this introduces discontinuous gradients at the boundary. Approximating such a penalty function, having finite gradients, is the logarithmic barrier function (Eq. 3.3.3) constraining the variable  $u \leq 0$ . In Eq. 3.3.3  $t$  determines quality of approximation. Unlike  $I_\infty$ , such barrier functions do not introduce gradient discontinuities and allow analytic gradients to be calculated. Iteratively solving a sequence of such unconstrained problems with increasing values of  $t$  allows approximation of the constrained problem solution. This form of handling constraints is incorporated inside the RTO functionals, used for solving the inverse problem, for purposes of non-negativity constraints imposed on the PSD values.

$$I_{\text{barrier}}(u) = -\frac{1}{t} \log(-u) \quad (3.32)$$

Equality constraints can also be handled similarly by introducing penalties to the unconstrained objective through, for example, Lagrange multipliers. In PCS, prior work [57, 58] has instead incorporated  $g^{(1)}(0) = 1$  through direct rescaling of the PSD to satisfy the equality constraint. This form of normalization is problematic as such an objective is non-unique and therefore cannot be solved using an optimizer. A second consideration is that such ‘special’ treatment of  $g^{(1)}(0)$  differs to that of other model outputs and does not allow for more equality constraints on other values of  $g^{(1)}(\tau)$ , or other variables, to be introduced. We opt instead for a probabilistic equality constraint, achieved through appending the tuple

---

<sup>2</sup>An interesting problem here is determining whether it is possible to incorporate a convex sub-program to accelerate solution of the nonlinear problem

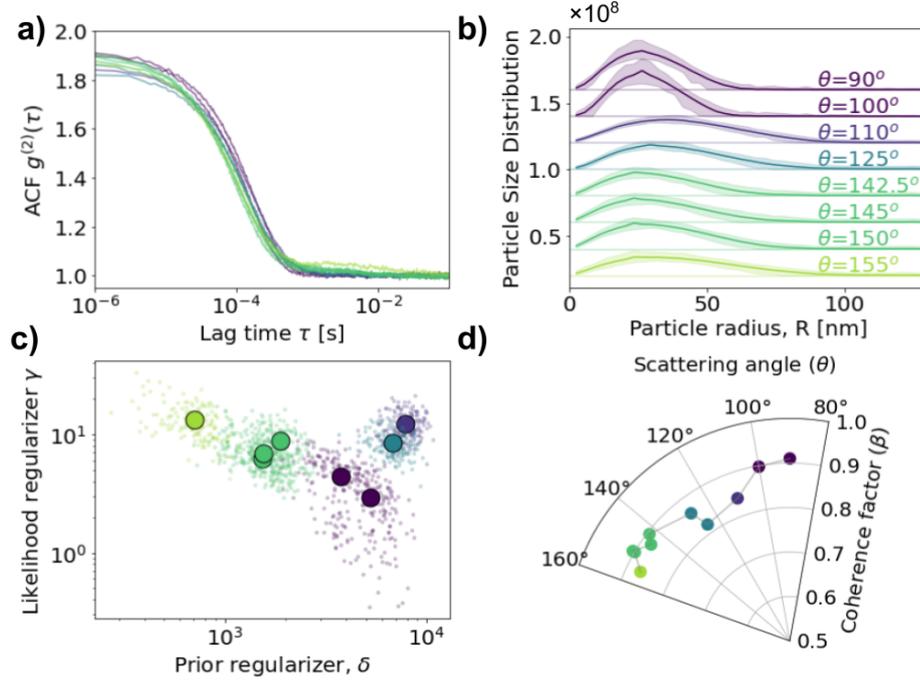


Fig. 3.12 a) Intensity autocorrelation functions recorded for scattering from AuNPs with nominal diameter 40 nm using the constructed multi-angle PCS system. b) Reconstructed particle size distributions for each scattering angle. c) Distribution regularizers generated by the hierarchical Gibbs sampler during process of reconstructing the PSD. d) Coherence factor variation with scattering angle.

$(g_{model}^{(1)}(0), 1)$  to the model output and data vector tuple  $(g_{model}^{(2)}(\tau), g_{data}^{(2)}(\tau))$ . This acts to incorporate the constraint probabilistically into the likelihood, treating all model outputs in a uniform way.

In order to sample from the posterior of the nonlinear inverse problem I implement the RTO algorithm but modify the proposal function to incorporate the effect of the non-negativity constraint, both in the functional being optimized and inside the Gibbs sampling steps, for the hyper parameters  $\gamma$  and  $\delta$ . To simplify and accelerate the implementation, gradient information of different functionals utilized inside RTO is obtained using automatic differentiation provided by Jax. This removes the need to perform multiple gradient estimations inside the sampling routines. Acceleration of computation is also achieved through just-in-time (jit) compilation of python functionals speeding up execution, increasing sampling rate. Further acceleration may be achieved through execution on GPUs, left for future work.

### 3.3.4 Nonlinear inversion and extensions

To verify the ability of the nonlinear model to invert dynamic light scattering data in a consistent way for different forward models (angles) and test the constructed multi-angle PCS system I conducted multi-angle measurements of intensity autocorrelations for nominally spherical 40 nm diameter gold nanoparticles (AuNPs) shown in Fig.3.12, using a vertically polarized 633 nm diode laser. While broad consistency between results in angles ranges  $110 - 155^\circ$  can be observed (with  $90^\circ$  and  $100^\circ$  exhibiting narrower size distributions) one must note that heating induced convection effects make PCS measurements of metallic nanoparticle systems troublesome<sup>3</sup> to analyse due to overestimation of hydrodynamic radius. In order to properly evaluate reliability of the inversion and validate the PCS measurement system would require non-absorbing low polydispersity reference nanoparticle standards, which could be mixed in different ratios create calibration particle size distributions with known particle number ratios. Accounting for scattering intensity anisotropy inside the the inversion procedure would require modification of the forward models to incorporate the Mie angular scattering models, that would enable the recovery of number-weighted, instead of intensity-weighted particle size distributions discussed here. This increased complexity would also require further tuning of the optimizers to ensure good performance on the inversion procedure. In the nonlinear inversion case, the reduced performance of gradient-based optimizers, relative to the convex optimizers is evidenced by an increase in the integrated autocorrelation time, for the noise precisions  $(\gamma, \delta)$  increasing from  $\gamma = 1.08$  and  $\delta = 1.83$  (for LHGS, Fig. 3.9) to  $\gamma = 2.58$  and  $\delta = 3.0$  using RTO. This remains a topic for future work.

A further extension of the technique to utilize polarization control can also be envisioned for extraction of shape properties nanostructures, generalizing from the spherical particle model. An example of such a depolarized PCS [35, 65] measurement with the system built is shown in Fig.3.13b, where gold nanorods (AuNRs) are measured. Robust implementation of this would require an increased degree of automation for controlling polarizers followed by development of quantitative models relating experimental configuration and sample type with signal to noise. An example of  $\beta$  variation for AuNPs with changing analyzer angle is presented in Fig.3.13a. Analysis of autocorrelations for AuNRs require modification of the field correlation function to have the form [35]:

$$g^{(1)}(\tau) = A \exp^{-q^2 D_t \tau} + B \exp^{-(q^2 D_t + 6D_r) \tau} \quad (3.33)$$

---

<sup>3</sup>and generally discouraged

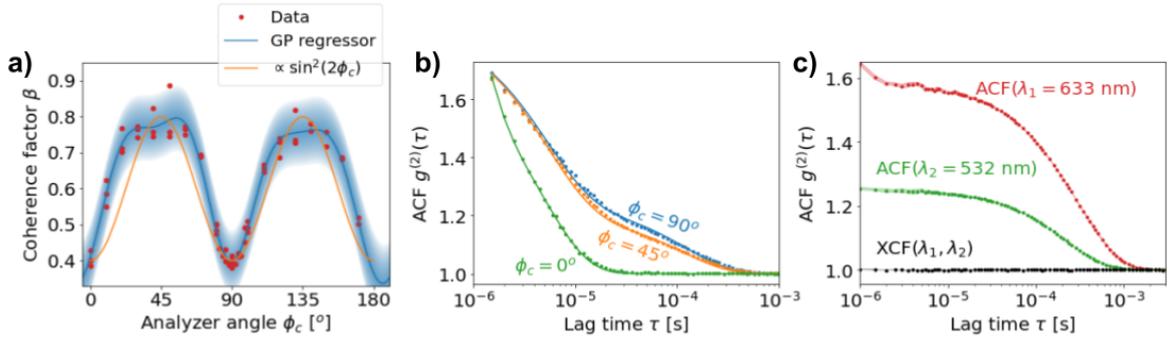


Fig. 3.13 a) Variation of coherence factor  $\beta$ , red points represent measurement, blue represents estimated measurement uncertainty while orange is an analytic function  $\propto \sin^2(2\phi_c)$  b) An example of a depolarized PCS measurement of asymmetric gold nanorods with nominal length,  $L = 40 \text{ nm}$  and diameter  $d = 15 \text{ nm}$ . The form of the autocorrelation function is modified on rotation of the polarization analyzer on the collection path allowing extraction of translational  $D_t$  and rotational  $D_r$  diffusion coefficients by fitting a modified form of  $g^{(2)}$ . c) Single angle dual colour cross correlation between two wavelengths measurement exhibiting no cross-correlation between scattering signals.

Here  $D_r$  is the rotational diffusion coefficient, while  $A$  and  $B$  are functions of experimental configuration and analyzer angle. The modified form allows estimation of an effective length  $L$  and diameter  $d$ , but requires introduction of additional ‘end-effect’ correction factors. From the AuNRs, nominal length  $L = 40 \text{ nm}$ ,  $d = 15 \text{ nm}$  we estimate  $D_t = 1.20 \times 10^{-11}$ ,  $D_r = 2.37 \times 10^4$  giving  $L = 67.5 \text{ nm}$  and  $d = 12.5 \text{ nm}$ . The disparity likely manifests from the combination of ‘end-effects’ along with using an incident wavelength of  $633 \text{ nm}$ , near to absorption resonance of the AuNRs at  $660 \text{ nm}$ .

Observing correlations between two separate wavelengths, at the same scattering angle, the initial premise of this project however is unlikely to produce results when utilizing two distinct lasers (Fig.3.13c). Theoretically this may be investigated from considering the requirements for observing intensity cross-correlation. From the theory described at the start of the chapter the intensity correlation functions are derived from self-similarity of the electric field GP. One would first have to show that such a correlation exists between scattered fields of different wavelengths (corresponding to different scattering wavevectors  $q$ , and thus different decay constants  $q^2 D_t$ ). Given the lack of correlation between two components of the same electric field this seems unlikely. The well defined phase relation between the two scattered fields would need to persist through the central limit theorem applied when calculating random phasor sums. Further, to verify this experimentally laser sources with a

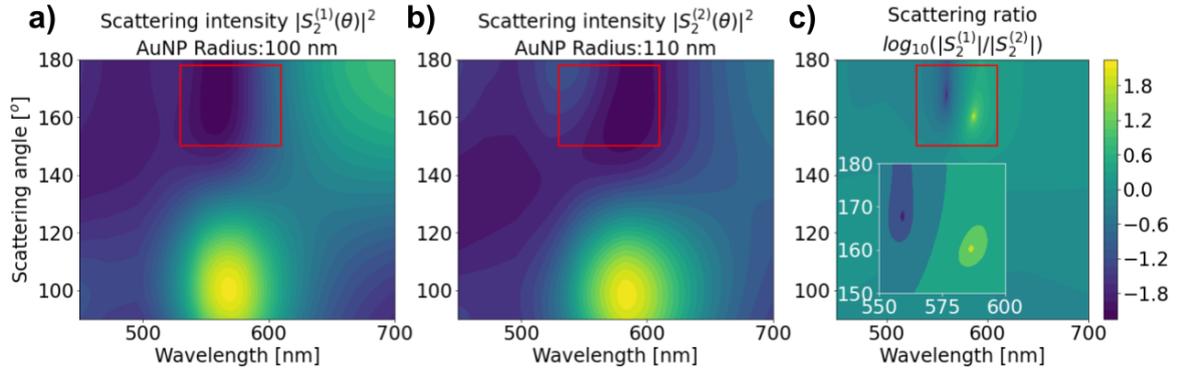


Fig. 3.14 a) Scattering intensity  $|S_2^{(1)}|^2$  of a 100 nm radius spherical AuNP in water as a function of angle and wavelength. b) Scattering intensity  $|S_2^{(2)}|^2$  of a 110 nm radius spherical AuNP in water c) Logarithm of scattering intensity ratio  $|S_2^{(1)}|^2/|S_2^{(2)}|^2$  exhibiting a relative scattering resonance at  $\lambda = 585$  nm,  $\theta = 160^\circ$  with a symmetric anti-resonance at  $\lambda = 560$  nm,  $\theta = 170^\circ$ .

fixed phase offset, coherence, must be generated, ruling out all of the methods investigated here, both the picosecond pulsed supercontinuum and the diode lasers.

Caution must also be exercised when applying this technique to non-spherical aggregates which exhibit complex angular scattering patterns (discussed in the next chapter). This can be understood by considering that an aggregate is formed of multiple Mie scatterers, themselves producing a complex angular scattering pattern (Eq. 2.10). To accurately characterize such a sample these models must be incorporated into the PCS inverse problem. The increased model complexity, combined with the ill-conditioned problem and heterogeneous sample is expected to pose significant theoretical challenges in both acquiring stationary timeseries data and interpreting the autocorrelations in a way that improves the resolution of the current technique. It is likely that such methods will remain limited to coarse estimation of average particle size, and be unable to provide high resolution information on different particle sub-populations. A potential avenue to explore however is the use of a CW tuneable multi-wavelength, multi-angle system for improving resolution of spherical particle characterization that exploits the scattering intensity ratios of two similarly sized AuNPs, for example the particles of radius 100 nm and 110 nm. With full tuneability of both angle and wavelength it may be possible to selectively enhance scattering signal from one of these sub-populations relative to another, illustrated in Fig. 3.14.

## 3.4 Summary

In this chapter I have covered the development of a PCS light scattering system and computational tools for inverting intensity autocorrelations measured by the instrument. The contribution of lies primarily on the computational side, where the proposed method allows quantification of uncertainty of the recovered size distribution, providing automatic estimate of measurement noise and generalizes the conventional approach to sequences of measurements spread in time. The proposed approach is also shown capable of inverting the full nonlinear inverse problem in PCS, removing the need to perform preprocessing of the autocorrelation data. This chapter also detailed the construction of a PCS system and demonstrated a set of preliminary measurements performed using it. For future attempts to construct similar systems it is suggested careful attention is paid to selection of illumination source, where designers should favour continuous wave over pulsed systems. In addition, given the ill-posed nature of the PCS inverse problem, care must be taken in minimizing noise sources as these reduce resolution size distribution estimates. In the case of temporally evolving samples the issue of integration time is also crucial as there one has to balance the implicit assumption of stationarity against the noise level requirement.

It is anticipated that gains in the resolution of PCS may be made, but would stem primarily from advancement of inference engines that integrate acquired data and enable automated experiments that are maximally informative. The development of such engines for inverse problems will likely be addressed in the near future. Such techniques are however unlikely to be a panacea in our ability to study ‘complex’ systems, as they still require a well formulated model.

An attractive approach to resolving sub-populations of within an ensemble is to eliminate ensemble averaging at the instrument level, requiring techniques to measure properties of single particles. While previously such techniques were not feasible due to limited computing capabilities, the increases in computing power in recent decades and its projected future growth makes the development of such techniques viable, in the case when analysis can be automated to performed at scale, to achieve sufficient statistical accuracy. An approach based on this idea is discussed in the next chapter.

# Chapter 4

## Single particle tracking

### 4.1 Introduction

A major challenge of PCS is the requirement to resolve different sub-populations in an ensemble measurement, where the smoothness of the Fredholm kernel combined with the scaling of scattering intensity ( $I \propto V^2$ ) and measurement noise act to limit resolution of the reconstructed particle size distribution. Instead of monitoring the superposition of fields scattered by the ensemble one can attempt to measure individual particular trajectories and use the Brownian path to estimate the hydrodynamic radius. This principle forms the basis of ‘nanoparticle tracking’ analysis (NTA) [18, 19]. Through tracking individual particles, such techniques are capable of rejecting outliers and estimate the true number weighted size distribution, as opposed to the intensity weighted PSD reported by PCS. A conventional NTA system utilizes segmentation of localization of particle centres for reconstructing Brownian trajectories of primarily spherical systems, however for tracking complex structures such as AuNP aggregates additional information may be extracted from the intensity fluctuations of structures caused by the reorientation of the AuNP aggregates as they undergo rotational motion. As such signals are typically transient and stochastic in nature, capturing them requires increasing the time resolution of NTA systems above the conventional 30 fps. In this chapter we investigated the connection between the scattering resonances and diffusive motion, both determined by the structure of the nano-object, and how this may give rise to transient signatures in the scattering intensity timeseries recorded as the particle undergoes Brownian motion. We focus our investigation on plasmonic metal nanoparticles and their aggregates, an example SEM image of which is depicted in Fig.4.1, consisting of 40-100 nm diameter AuNPs and displays resonances across the visible range related to dipole-dipole

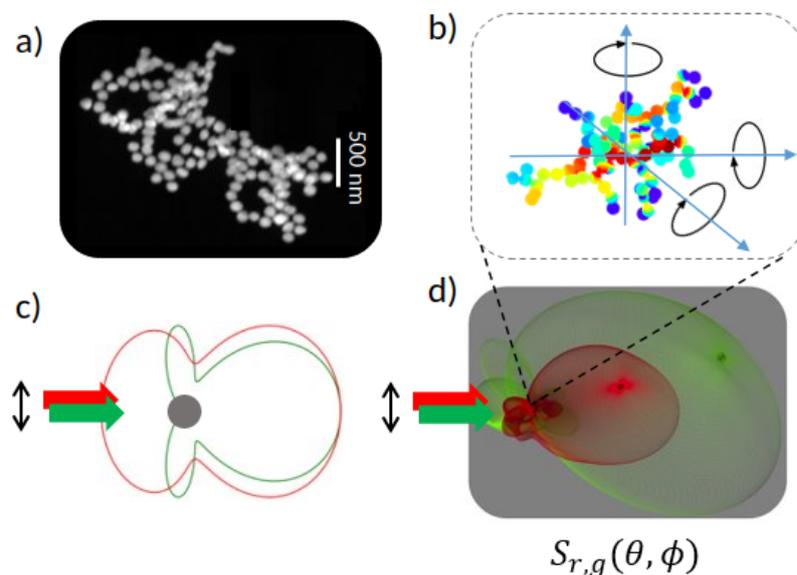


Fig. 4.1 (a) SEM of typical Au NP aggregate dried down onto Si substrate, taken from [66]. (b) Optical modes (red around 800 nm, blue around 550 nm) from simulations of Au NP aggregate with 0.9 nm gaps. As the Au NP aggregate undergoes diffusive motion it will rotate in the laboratory reference frame, giving rise to scattering intensity fluctuations. Adapted from [66] (c,d) Schematic scattering from (c) single colloidal sphere, and (d) colloidal complex nanostructure, under red and green illumination with different scattering patterns that change with orientation under rotational and translational diffusion in solution. Black arrow indicates polarization

interactions along the linear chains (Fig.4.1b) [66] with the resonance being determined by the chain length. With simultaneous measurement of scattering intensities at two wavelengths, from the AuNP aggregate, we show that it is possible to track rotational diffusive motion of the structure owing to the relation between scattering profiles of two wavelengths incident on the particle (Fig.4.1c,d).

## 4.2 Nanoaggregate scattering and diffusion

Owing to the increased complexity of the AuNP aggregate structure, modelling diffusive motion and scattering pattern also increases in complexity. This theoretical background is however a strict requirement for anyone who would attempt to model the motion of such structures or attempt to construct meaningful features which could be reliably extracted from the stochastic timeseries generated by experimental measurements. In this section we thus introduce the basic theory for both generalized diffusion and multi-particle Mie scattering required to perform calculations of such forward models, noting that the real challenge is in

mapping the experimental results back to physically meaningful quantities, which we do not attempt to address.

### 4.2.1 Generalized diffusion

In the introductory theoretical background we considered diffusion of a spherical particle under stochastic forcing produced by collisions with solvent molecules. When generalizing to non-spherical structures the relation between the diffusion constant and hydrodynamic friction (Eq. 2.18) is extended with both diffusion constant and friction coefficient being promoted to  $6 \times 6$  tensor variables. The diffusion tensor takes the form of Eq. 4.1 [67] with  $D_T, D_R, D_C \in \mathbb{R}^{3 \times 3}$  forming sub-tensors representing translational ( $D_T$ ), rotational ( $D_R$ ) and rotational-translational coupling ( $D_C$ ) terms.

$$D = \begin{pmatrix} D_T & D_C^T \\ D_C & D_R \end{pmatrix} \quad (4.1)$$

The connection to the spherical particle can be made by summing over suitable elements of the diffusion tensor  $D$ . For example the translation diffusion coefficient ( $D_t \in R$ ) is the trace of  $D_T$ ,  $D_t = \text{Tr}(D_T)/3$ . Similar to the one dimensional case the Brownian motion of the structure is again governed by equations relating second moments of the, now generalized, coordinates  $q$  to the components of the diffusion tensor [67] via Eq. 4.2

$$\langle q_i, q_j \rangle = 2D_{ij}\Delta t \quad (4.2)$$

In this form the generalized coordinate vector  $q = (\Delta x_1(t), \Delta x_2(t), \Delta x_3(t), \phi_1(t), \phi_2(t), \phi_3(t))$  incorporated translation increments  $\Delta x_i$  and angular rotation increments  $\phi_i$ . The elements  $\Sigma_{ij} = \langle q_i, q_j \rangle$  can be seen as elements of a covariance matrix defining a multivariate normal, draws from which can be used to generate Brownian trajectories. Unlike the case of pure translational motion where translations along each coordinate axes are commutative, the same cannot be said for rotational motion, meaning that order of application of rotation matrices affect the final structure orientation, making accurate simulation of 3D diffusion challenging.

In order to simulate motion of an arbitrary AuNP aggregate we implement the Brownian dynamics algorithm presented in [67] for the case of free diffusion only. Using a time step  $\Delta t = 1 \mu s$  in our simulation, the algorithm generates rotation and translation coordinate updates according to Eq. 4.2 in the particle coordinate system, with translation steps  $\Delta r(t_0)$ .

## 4.2 Nanoaggregate scattering and diffusion

---

Iterative updates for the centre of mass position,  $r_L(t)$  in laboratory frame, and the rotation matrix, mapping from particle to laboratory reference frames, starting at an initial time  $t_0$  up to a time  $t$  is given by Eq. 4.3. Here  $M_{t_0 \rightarrow L}$  represents the rotation matrix of an arbitrary coordinate in the particle reference frame to the laboratory reference frame at an initial time  $t_0$ . Rotation around each of the axes is represented by  $R_i$ ,  $i \in \{1, 2, 3\}$ .

$$\begin{aligned} r_L(t) &= r_L(t_0) + M_{(t_0 \rightarrow L)} \Delta r(t_0) \\ M_{(t \rightarrow L)} &= M_{(t_0 \rightarrow L)} \cdot R_3 \cdot R_2 \cdot R_1 \end{aligned} \quad (4.3)$$

For the case of rotating AuNP aggregates we neglect any translational effects (both diffusional and fluidic) and focus only on the rotational part, applying the sequence rotation matrices  $M_{(t \rightarrow L)}$  to the coordinates of the individual AuNPs, modelled as spheres to generated different AuNP aggregate orientations.

### Hydrodynamic bead model

To compute the coordinate increments of an arbitrary shapes particle one is required to estimate the diffusion tensor  $D_{ij}$  required in Eq. 4.2. As the AuNP aggregates consist of many connected spheres, with a fixed spacing, it is natural to model such systems using the rigid bead model such as the one implemented in Hydro++ [68] package. An example of a simulated aggregate is shown in Fig.4.2a, where four 80 nm spheres separated by 0.9 nm gaps are arranged in an L-shape. The sphere size is set to nominal size of AuNPs used experimentally, while the 0.9 nm inter-particle spacing is set according to the spacing defined by the cucurbit[n]uriluril (CB[n]) molecules used to trigger aggregation (we use CB[5]). For this structure the translational ( $D_T$ ), rotational ( $D_R$ ) and rotational-translational coupling ( $D_C$ ) sub-tensors calculated by Hydro++ are given below (unit:  $cm^2/s$ ) with ‘.’ representing zeros.

$$D_T = \begin{pmatrix} 3.05 \times 10^{-8} & -1.06 \times 10^{-9} & . \\ -1.06 \times 10^{-9} & 2.79 \times 10^{-8} & . \\ . & . & 2.60 \times 10^{-8} \end{pmatrix}$$

$$D_R = \begin{pmatrix} 3.63 \times 10^{-5} & -6.22 \times 10^1 & . \\ -6.22 \times 10^1 & 2.02 \times 10^2 & . \\ . & . & 1.64 \times 10^2 \end{pmatrix}$$

$$D_C = \begin{pmatrix} . & . & 3.69 \times 10^{-5} \\ . & . & -3.52 \times 10^{-5} \\ 3.69 \times 10^{-5} & -3.52 \times 10^{-5} & . \end{pmatrix}$$

### 4.3 Multiparticle Mie Scattering

While individual particles can be considered point scatterers for the purposes of nanoparticle tracking, the aggregates exhibit a complex structure consisting of multiple scattering centres giving a measured scattering intensity distribution that depends both on particle orientation relative to the incident wavevector and collection cone of the objective.

#### Multiparticle formulation

To simulate the detected intensity fluctuations of light scattered by the aggregate the first step is to determine the angular scattering distribution of the structure as a function of incident wavevector, light polarization and its orientation, computed at each time step in the diffusion simulation.

To compute the angular scattering distribution we employ computational techniques generalizing on analytical results of Mie theory (2.1.2) known as Generalized Multiparticle Mie Theory (GMMT) [69] [70], which treats both the scattering from individual particles as well as cross-coupling between the scatterers. From earlier work on characterizing such aggregates it has been shown that chain modes of plasmonic aggregates produce characteristic signatures with the resonance frequency of the chain mode being determined by the length of the linear chain [66]. This resonant scattering effect was expected to give a characteristic signature in the scattering intensity timeseries recorded as the AuNP aggregate undergoes diffusive motion with a timescale determined by the diffusive dynamics - spanning microseconds to milliseconds for the aggregate sizes considered here. In order to perform such simulations we utilize the MiePy [70], using which we are also able to calculate the optical cross sections of the aggregate (Fig.4.2b).

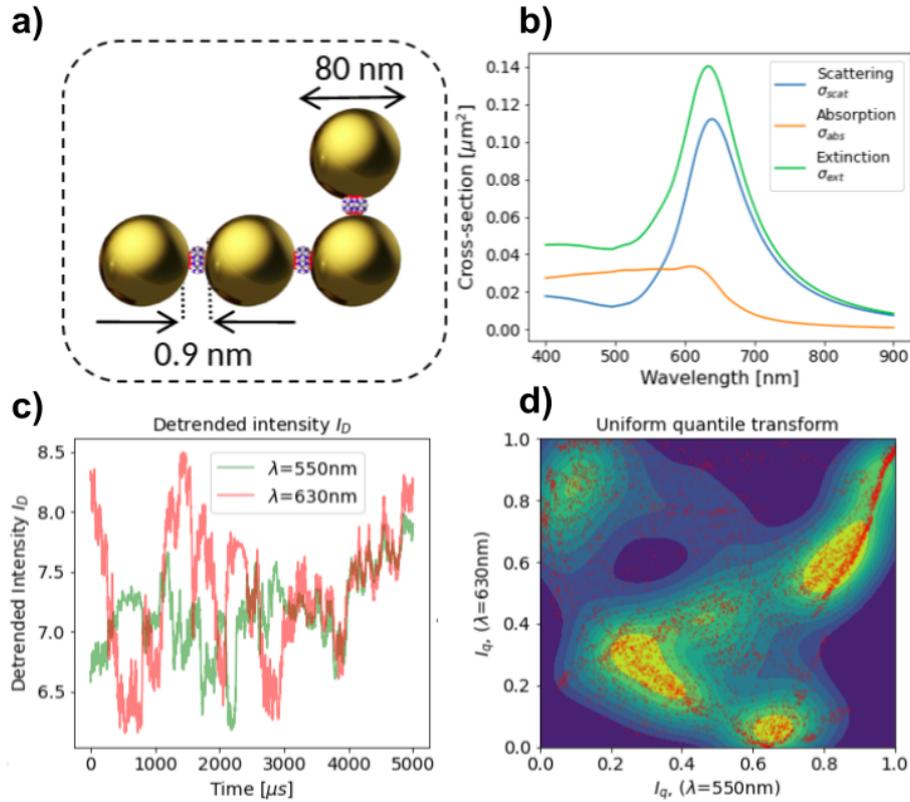


Fig. 4.2 a) An example of a ‘L’ shaped AuNP aggregate, having four equally sized spherical particles (diameter=80 nm) with spacings of 0.9 nm b) Optical cross sections of nanoaggregate for light with an incident polarization parallel to the x-axis, the long three nanosphere chain c) Simulated intensity timeseries for the aggregate undergoing three dimensional diffusion d) Quantile-transformed joint distribution of scattering intensity of the aggregate

To simulate intensity timeseries we can expect to see experimentally the intensity distribution must be integrated over a scattering cone defined by microscope numerical aperture (NA, set to 0.8 in simulation) that models the finite collection cone of the microscope. The intensity distribution at a particular scattering angle given by  $I(\phi, \theta)$  in spherical polar coordinates can then be integrated over the suitable sphere surface using Monte-Carlo (Eq. 4.4) sampling on the unit sphere, where we evaluate the intensity at a given coordinate  $(\theta, \phi)$  followed by rescaling by the volume element ( $dV$ ) of the space, corresponding to the area element of the unit sphere  $dV = dA = r^2 \sin(\theta) d\theta d\phi$ , with  $r = 1$ . For simplicity the integral is evaluated by sampling  $\theta$  and  $\phi$  from uniform distributions between fixed bounds, corresponding to a ‘square’ integration domain. The use of the area element  $dA$  scales each points contribution

## Single particle tracking

---

in order to account for the non-uniformity of points near the poles of the sphere.

$$\int f dV \approx V \langle f \rangle$$
$$\text{for which } \langle f \rangle = \frac{1}{N} \sum_{i=0}^{N-1} f(r_i) \quad (4.4)$$

By performing the above integral at multiple wavelengths one can simulate a vector timeseries of intensity signals for a particle undergoing diffusive motion. For purposes of simulation we neglect any incident intensity variation that may arise from drifts in alignment or laser intensity noise. One aspect to note from considering the intensity integral over the unit sphere is that, similar to the case of the Fredholm kernel, integrating over a large area acts as a smoothing function which reduces the amplitude of possible intensity fluctuations caused by reorientation of the AuNP aggregate. Thus choice of objective NA represents design variable controlling the trade-off between fluctuations in the timeseries, caused by particle reorientation, against intensity of the recorded signal, required for high speed acquisition. A intensity timeseries produced by our simulation is shown in Fig.4.2c, showing strong anti-correlation between the 550 nm and 630 nm signals from 0-3 ms, followed by strong correlation in the 3-5 ms interval.

## 4.4 Experimental setup

To perform tracking experiments at microsecond timescales I constructed a dual-colour scattering system depicted in Fig.4.3a. Incident lasers L1 (Integrated Optics, Matchbox 633 nm, continuous wave) and L2 (Spectra Physics Inspire OPO, centre wavelength 550 nm) are combined via a dichroic beam splitter and coupled into a hollow-core anti-resonant optical fibre (HC-ARF) by an under-filled 0.1 NA (4 $\times$ ) objective. The light coupling and fluid flow into the fibre is facilitated by a custom flow cell [71] containing a sapphire glass optical access window. To fill the fibre, the flow cell inlet is attached to a syringe with an analyte solution containing nanoparticles or aggregates. In order to monitor the filling of the fibre we image the end facet of the HC-ARF, fixed in a V-groove mount, using a 20 $\times$  objective, with the coverslip used to both protect the objective from the analyte and eliminate distortion due to droplet formation on the end-facet.

To measure scattering from individual nanoparticles or aggregates we place a 60 $\times$  objective perpendicular to the long axis of the fibre. Water immersion is used to reduce refractive index mismatch between the fibre surface and the surrounding medium - reducing the effects

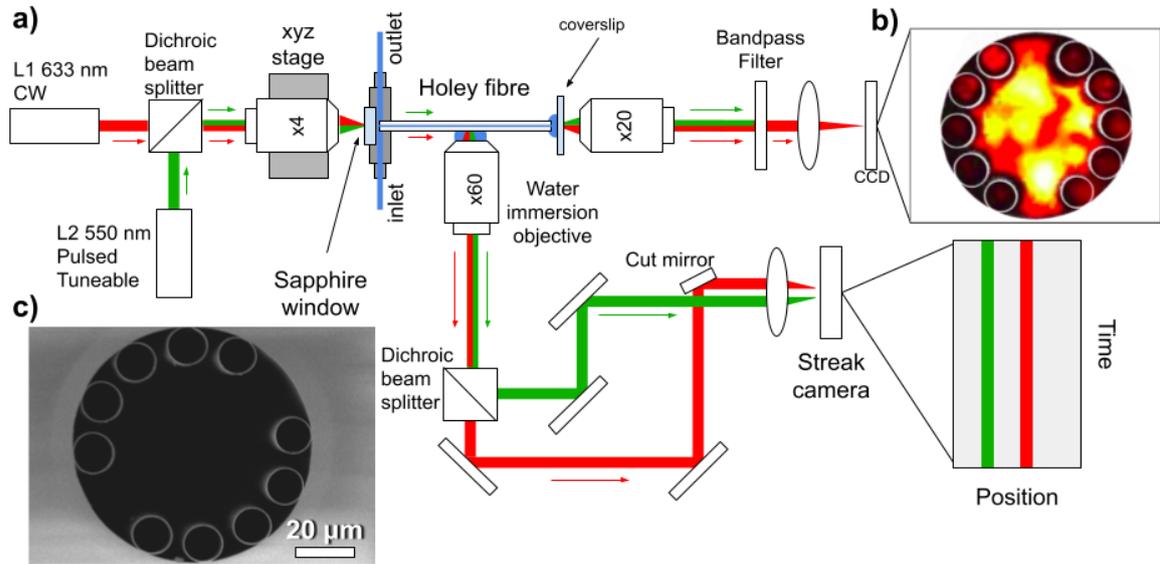


Fig. 4.3 a) Diagram of experimental setup constructed for two-colour high speed light scattering. b) Optical mode,  $\lambda = 633\text{nm}$ , propagating through HC-ARF fibre imaged at fibre end facet. c) Scanning electron microscopy image of HC-ARF structure.

of a curved upper surface of the optical fibre on the imaging. Optical distortion due to the internal structure of the fibre is minimized by rotating the fibre such that the intentional gaps in the ring of small capillaries shown in Fig.4.3b are aligned to the imaging direction of the 60 $\times$  objective so that imaging occurs through these gaps. The collected light is relayed to a dichroic beam-splitter separating the 633 and 550 nm wavelengths which are subsequently focused onto the entrance slit of the Streak camera (Hamamatsu, C5680). The recombining mirrors are used to spatially offset the two wavelengths - creating two images of each tracked particle on Streak camera, allowing simultaneous recording of both wavelengths

### Streak camera

To easily access a wide range of timescales for measuring intensity fluctuations, from millisecond, to nanosecond, using a single instrument we employed a Streak camera. Broadly, Streak camera designs utilize either rotating mirrors or optoelectronic elements to achieve the high time resolutions. The optoelectronic variant uses a microchannel plate (MCP) photomultiplier to convert incident photons to electron pulses. The electron pulses then impact a phosphor screen that is imaged by a CCD to generate a frame. To achieve the high time resolution, the electron pulse generated by the MCP is subject to a time varying acceleration voltage perpendicular to the direction of propagation. The swept voltage produces a time-dependent deflection of the electrons relative to a global trigger signal

## Single particle tracking

---

initiating the sweep. The pulse arrival time (relative to a trigger signal) is thus converted to a spatial coordinate, trading one of the two spatial coordinates for a temporal dimension. Thus 2D images generated using the Streak camera in fact record only a single spatial coordinate, with the other coordinate designated for monitoring time variation, allowing for high temporal resolution.

### Hollow core anti-resonant fibre (HC-ARF)

A hollow-core optical fibre is used in this work to act as a sample holder for the solution containing AuNP aggregates. Such systems have previously been used for optical trapping experiments where position of particles was tracked by video imaging of the side scattering [72], similar to the current configuration. The HC-ARF used in our work (Fig.4.3c) consists of an interrupted circular arrangement of 10 silica capillaries within a larger silica capillary fibre with an outer diameter of  $176\ \mu\text{m}$ , resulting in a core diameter of  $55\ \mu\text{m}$ . The average wall thickness of capillaries is  $315\ \text{nm}$ , resulting a first anti-resonant reflection condition [73] at  $747\ \text{nm}$ . The first resonant loss wavelength occurs at  $372\ \text{nm}$ , well below the  $550\ \text{nm}$  and  $633\ \text{nm}$  wavelengths used in our experiment. Strong wave-guiding is thus expected across the visible wavelength range, confirmed experimentally with a 50% coupling efficiency in our experiment, where an average incident power of  $40\text{mW}$  of resulted in  $20\text{mW}$  output at the end facet. Use of a HC-ARF reduces the volume of analyte required to  $0.24\ \mu\text{L}/\text{cm}$  and enables control of flow through the measurement region, allowing for flow-through measurements and pressure-controller based feedback. Potential future opportunities for control of the propagating mode would enable spatial intensity control [74], allowing for comparison between diffusion in the centre and near the walls of the channel in a single configuration. We note that the excited mode is sensitive to the launch conditions, and the fibre thus needs to remain stable during any run. A constant light-filled transverse mode distribution is beneficial as scattered intensity variation is then smoothly varying within the central core of the fibre.

### Timeseries analysis

After recording of a sequence of frames by the Streak camera we extract the scattering intensity at each wavelengths by first identifying the positions of particle. This is performed by a developed segmenting algorithm identifying the two intensity peaks using a variant of region growing on the one-dimensional projection along the Streak image spatial coordinate (Fig.4.5d). Once segmented the raw scattering intensity timeseries are generated by summing the columns within the segmented region.

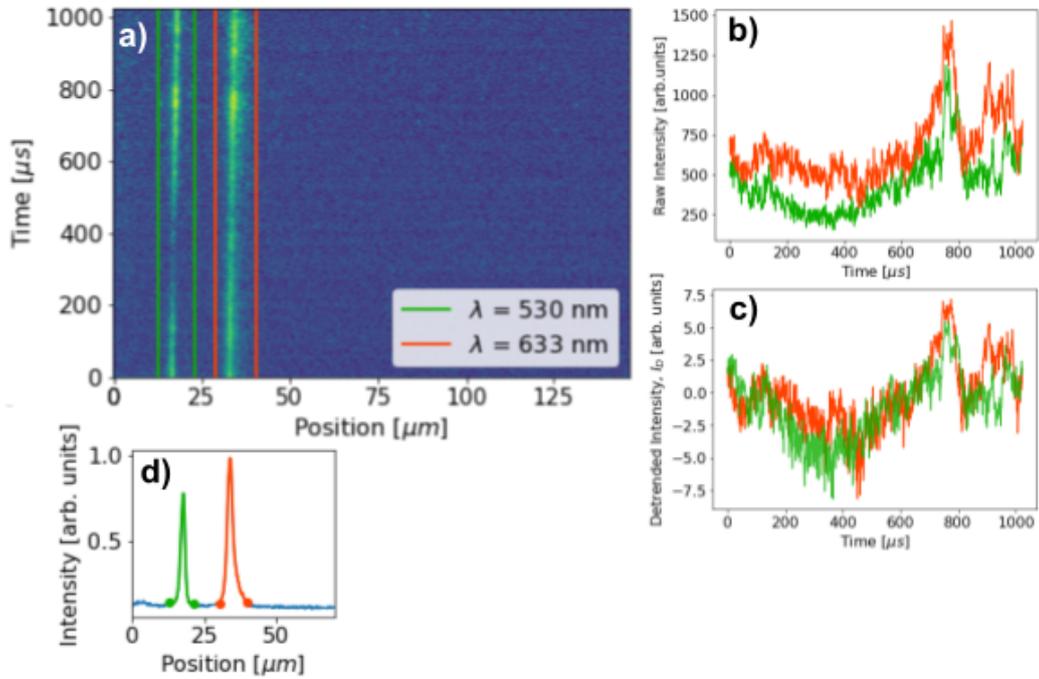


Fig. 4.4 a) Segmentation traces (red and green lines) overlaid on a image captured by the Streak camera, pixels are preprocessed using a  $\log(1+x)$  transformation before plotting. b) Raw intensity timeseries constructed from each segmented trace. c) Stationary timeseries constructed by de-trending log-transform of raw intensity timeseries. d) Row summed intensity for performing region-growing based segmentation of the two signals.

With the recorded intensities exhibiting an exponential-like trend in time arising as an artefact of the detection, the timeseries serial correlation and do not satisfy stationarity requirements discussed in the previous chapter. Such trends manifest as artificial dependence between the two intensities and must be removed by preprocessing the timeseries to be stationary prior to attempting to infer dependence structure. Many methods exist for preprocessing timeseries signal with selection determined by features exhibited by the timeseries, two common methods are  $n^{th}$  order differencing, analogous to gradient estimation, and de-trending. With the series also exhibiting non-Gaussian noise, where signal standard deviation is correlated with the mean, we also stabilize the noise variance in the the timeseries by applying the logarithm transform prior to de-trending. The latter is achieved through z-normalization of the 1<sup>st</sup> order differenced timeseries. In normalizing the series in this way we attempt to minimize the correlations induced artificially through processes involving model-based de-trending and stabilization. It is worth noting that the approach described above is by no means optimal, but instead proposed as a good starting point for processing the observed

## Single particle tracking

data. The resulting transformed intensity timeseries we term the de-trended intensity  $I_D$  (Fig.4.6c).

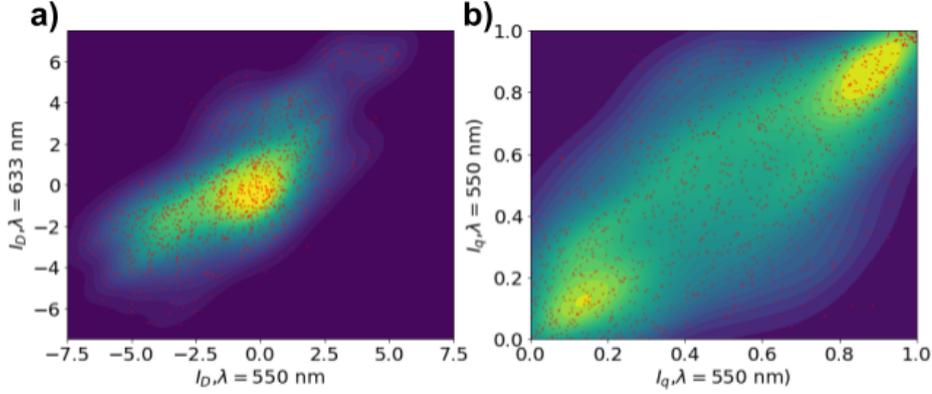


Fig. 4.5 a) Joint distribution of intensities  $I_D$  for 550 nm and 630 nm b) Quantile transformed density revealing dependence structure (uniform univariate marginals).

Once stabilized we can plot the joint distribution of the intensity timeseries (Fig.4.5a), that encodes the individual marginal distribution of each intensity variable along with dependence between variables. To quantitatively differentiate between the effect of the marginals and isolate the variable interdependence we leverage the copula formalism, formulated in Sklars theorem (Eq. 4.5) [75] where  $F(x, y)$  is the joint cumulative distribution,  $F_X, F_Y$  are marginal distributions, with  $C(., .)$  being the copula.

$$F(x, y) = C(F_X(x), F_Y(y)) \quad (4.5)$$

An alternative form of this theorem, connecting the joint probability density  $f(x, y)$  with univariate marginal densities  $f_X(x), f_Y(y)$ , through a copula density  $c(., .)$  is

$$f(x, y) = c(F_X(x), F_Y(y))f_X(x)f_Y(y) \quad (4.6)$$

The ability to decompose of joint density into univariate marginals and a copula allows us to decouple the distributions of each scattering intensity from the dependence structure and analyse each separately. Without this conventional analysis such as correlation or covariance estimates may lead to incorrect results, as they are often based on the assumption of univariate marginals being Gaussian distributed.

In this work we use copula density estimation algorithms implemented in the *kdecopula* [76] package. A similar visual effect to performing copula density estimation is to apply a

quantile transform of each coordinate, converting  $I_D$  (de-trended intensities) to rescale each axes into the range  $[0,1]$ , which we term ‘intensity quantiles’  $I_q$ , followed by conventional kernel density estimation [62] (Fig.4.5b). Use of copula density estimation would in future allow classification of dependence types by identification of the family the copula density belongs to. The same processing is also applicable to simulated data, an example of which is shown in Fig.4.2d.

## 4.5 Particle tracking

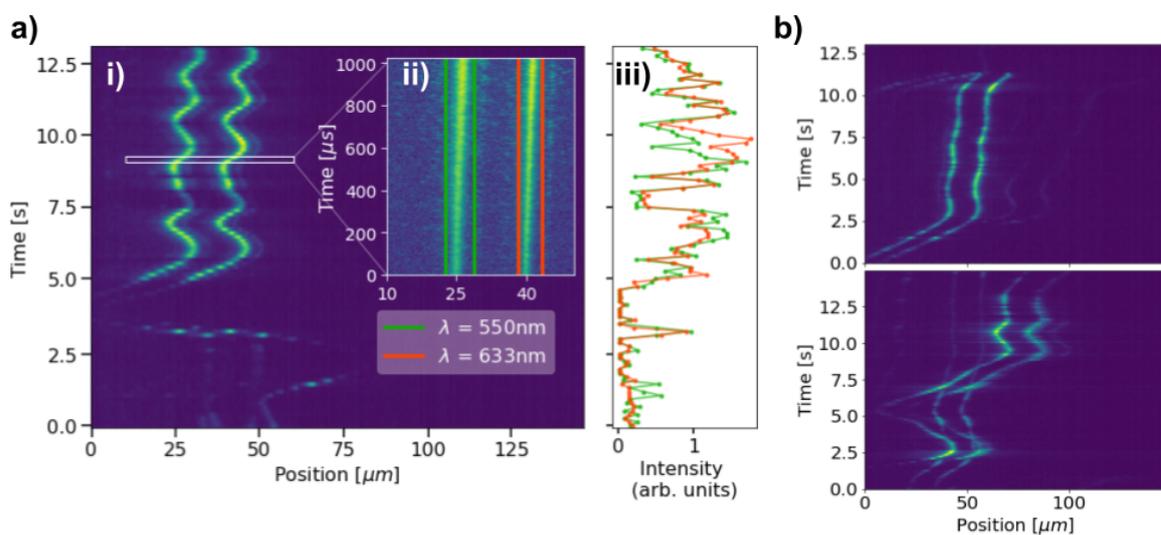


Fig. 4.6 a) Single particle track constructed by stacking multiple images showing motion control of a single nanorod in fibre i) Long time trace consisting of multiple frames ii) Single frame, highlighting the segmented regions of the 532 nm (left) and 633 nm (right) scattering intensities iii) Total intensity of each scattering wavelength per frame. b) Example long time intensity traces recorded in fibre showing ability to hold particles in sample volume for up to 10 seconds.

Initial testing is performed with dilute solutions of AuNRs with nominal length 40 nm and diameter 12 nm. The Streak sweep time set to 1 ms, for 1024 pixels along the time axis, producing a temporal resolution of  $1\mu s$ . The camera's maximum frame rate of 8 fps gives a 125ms inter-frame interval and results in the capturing 0.8% of the timeseries (per unit time). This temporally sparse sequence due to large interframe-spacing precludes application of common imputation methods and requires the particle to be held in a imaged region for extended periods to enable sufficient samples of the intensity process to occur. This confinement is achieved for durations exceeding 10 s of seconds (Fig.4.6) through manual feedback with a syringe connected to the fluid inlet port, Fig.4.3a. We anticipate that this

## Single particle tracking

simple scheme be extended significantly through integration of video based feedback with automatic pressure controller.

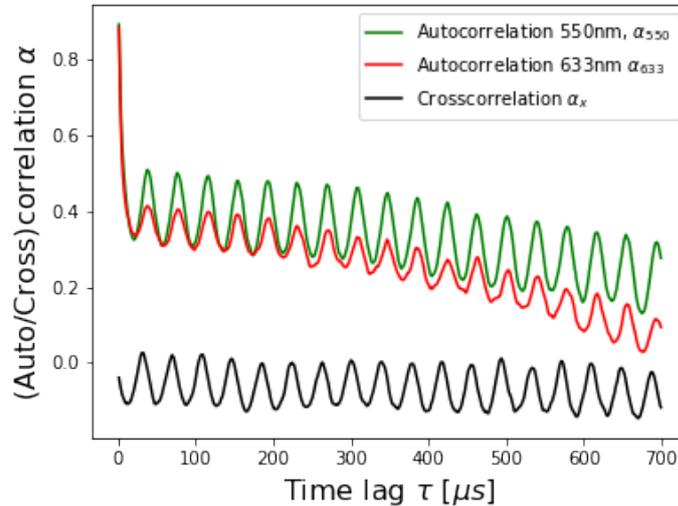


Fig. 4.7 Average of 93 autocorrelation and cross-correlation computed from dual colour scattered intensity timeseries recorded from AuNP aggregates. Autocorrelations exhibit serial decay in correlation, corrupted by a periodic signal. Cross-correlation does not exhibit a serial correlation structure but exhibits the same frequency periodicity as the autocorrelations.

Aside from copula density estimations, in order to assess the self-similarity of the scattered intensities we calculate the auto- and cross-covariances (Fig.4.7) of the intensity traces for each frame. The autocovariance of the 633 nm and 550 nm scattered intensities,  $\alpha_{633}$  and  $\alpha_{550}$  exhibit similar behaviour with an initial fast decay of self-similarity, which we attribute to the finite extent of the particle image on the Streak CCD. This is followed by a slower decay of self similarity from 0.4 for both wavelengths, modulated by a 26 kHz sinusoidal signal with an oscillation period of  $38.5\mu s$ . This noise term may arise from readout noise in the intensified CCD or through a non-Gaussian point-spread function convolved with the temporal intensity fluctuations. The latter may be introduced by distortions caused by the HC-ARF used as the sample holder. For the cross-covariance  $\alpha_x$  we find a similar pattern of a 26kHz sinusoidal signal, which however exhibits a phase lag  $\phi = 69^\circ$  relative to both sinusoidal oscillations of  $\alpha_{550}$  and  $\alpha_{633}$ .

Richer dynamics are detectable when analyte contains AuNP aggregates, as demonstrated in Fig.4.8 from four frames from a single AuNP aggregate, where examples of correlated, anti-correlated and bi-modal distributed intensity traces are shown. Such complex dynamics with time varying dependence structure are unlikely useful in estimating physical properties of the AuNP aggregates without strong assumptions on the shape of the particles or introduction of

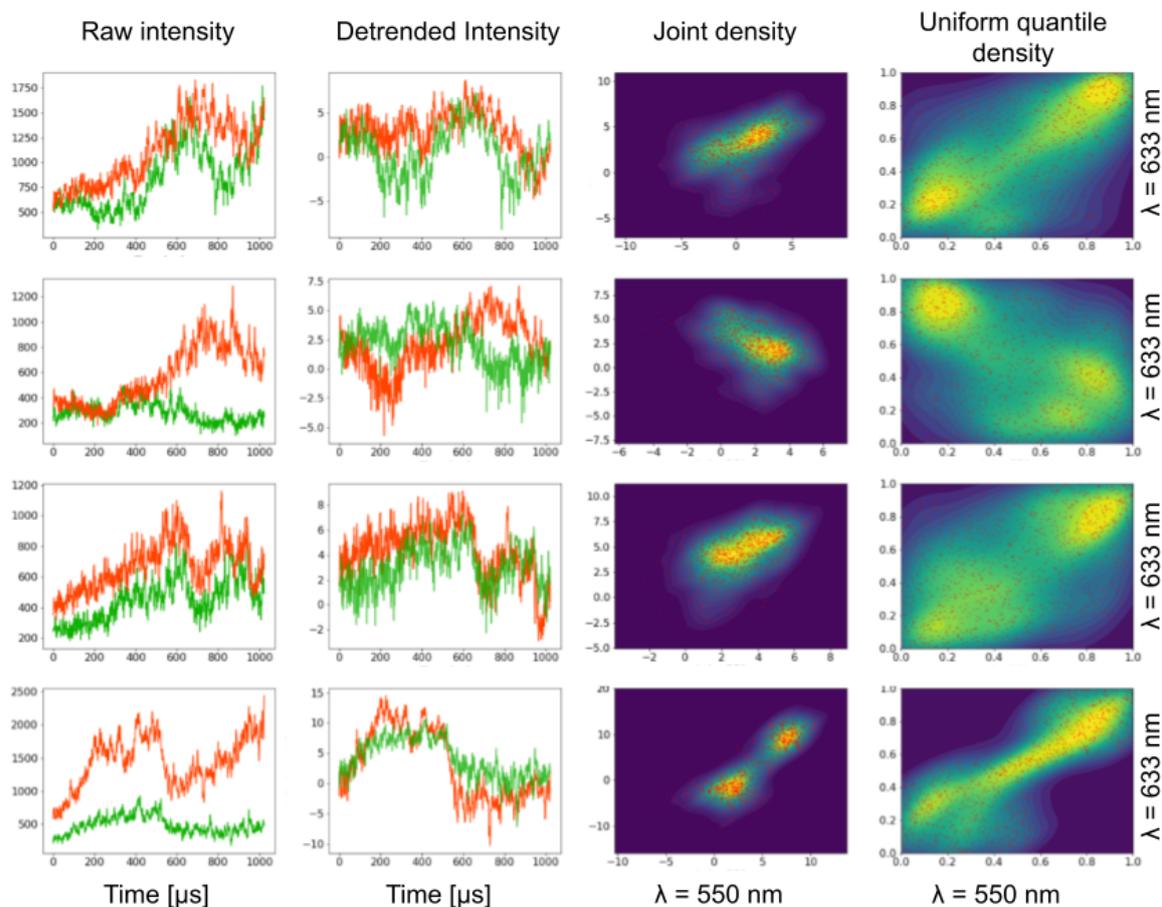


Fig. 4.8 Dual colour AuNP aggregate tracks exhibiting correlated and anti-correlated properties

additional experimental control variables capable of inducing deterministic motion of the AuNP aggregates. Given such as modification it may be possible in future to reconstruct properties of a single AuNP aggregate, however achieving this would initially require careful formulation of a forward model that would map time-invariant AuNP aggregate parameter (latent variables) to a set of features that can be computed from the intensity timeseries. Given the stochastic nature of the diffusive motion and complex angular scattering patterns the challenge of any such approach would then be the relation between these latent variables and suitable features measurable from the extracted from the timeseries. In addition to a suitable forward model, a central limitation of the above approach is the Streak camera, with a 125 ms inter-frame time. A more appropriate choice of equipment, commercially available, are high speed imaging cameras, capable of reaching frame rates of  $10^5$  fps. Using such systems would allow for continuous recording of intensities covering 100% of the timeseries in comparison to the 0.8% afforded by the Streak. When such systems are not available,

## Single particle tracking

---

techniques to slow down the diffusive dynamics of the AuNP aggregates by increasing solution viscosity may also be applied. Variation in spatial intensity distribution can be controlled through replacing the HC-ARF with a conventional cuvette and using fibre beam combiner to generate a collinear multi-wavelength beam without needing to complex adaptive optics to control fibre mode profile. If flow control is required, a square or circular glass capillary embedded in suitable refractive index matching fluid may be utilized.

## 4.6 Summary

In this chapter I have presented an single particle tracking technique applied to studying the dynamics of complex AuNP aggregates. The constructed system was capable of simultaneous monitoring of two colour scattering intensities at time resolutions down to the microsecond timescales in a hollow core optical fibre geometry. The low sample volume and simple means of controlling flow enabled monitoring of individual particles up to tens of seconds. The proposed analysis technique based on copula methods allows extraction of dependence structure between two variables.

While single particle tracking has been successfully applied to determine particle size distributions extracting information about complex geometries using scattering relying on only basic intuitions about properties of the resulting timeseries will most likely not result in success. To relate the features of the timeseries to a physical quantity a greater degree of mathematical modelling would be required. Based on the complexity of the forward model, as described above, attempts to measure more complex dynamics should approach the issue first from the theoretical perspective to determine a-priori what features can be reliably extracted from an experiment. Without this an elaborate system may be constructed but measurements may remain uninformative.

In part the limitation of the above technique is the low readout rate afforded by the Streak camera, combined with the need to track intensities both in time and space. This is primarily a challenge for electronics readout instrumentation. For this reason, in the next chapter we turn to considering an electronic readout system, for recording a single photon stream at both high temporal resolution and high data rates.

# Chapter 5

## Time correlated single photon counting

### 5.1 Introduction

In the previous two chapters we considered high temporal resolution measurement techniques recording dynamics of scattered fields. To a large extent the ability to record such signals comes from the readout devices utilized, such as the SPADs or the Streak cameras. In performing such experiments one quickly realises that the ability to perform a certain measurement types is limited by the detection system. Thus, part of the capabilities in performing novel measurements is the need for highly capable instrumentation. While the prevalence of high speed detection systems is increasing they are still primarily limited to few research laboratories, in turn limiting applicability of the new techniques to laboratories possessing the instruments.

Photon counting systems are a prime example of this, enabling multiple different measurement modalities, from PCS and Fluorescence Correlation Spectroscopy through optical ranging and LIDAR, time-of-flight spectroscopy [77], fluorescence lifetime imaging (FLIM) [78] to super-resolution microscopy [79] and quantum optics. While systems capable of measuring time intervals with the required sub-nanosecond resolution have been implemented on custom application specific integrated circuits (ASICs) there have typically had a high entry barriers both in terms of price, design cost and skill sets. This has limited the prevalence of techniques requiring such instruments relatively few groups and organization. With increasing adoption of low-level reprogrammable integrated circuits (ICs), known as Field-Programmable Gate Arrays (FPGA) this is rapidly changing due to falling entry barriers.

## Time correlated single photon counting

---

ers required for developing capable photon counting systems. The increased accessibility comes from the reduced cost and complexity of development alongside FPGA performance improvements and integration of the programmable logic elements into Systems-on-Chip (SoC) which incorporate processor cores and peripherals hardware such as random-access memory (RAM). Simplification in the required toolchains and their increased availability have also lowered barriers to designing the systems themselves. Based on this reasoning it is anticipated that the next 5-10 years will see broader adoption and development of FPGA-based systems in small laboratories.

In this chapter we describe the photon counting systems utilized in PCS measurements, achieving nanosecond-scale time resolutions. In the remainder of this section we will introduce key concepts used in development of sub-nanosecond photon counting systems. The rest of the chapter describes the development of a such photon counting system, incorporating a **time-to-digital converter (TDC)**, capable of resolving time of arrival within a few tens of picoseconds, having a continuous single photon readout rates of 3 million hits per second, from up to five separate channels. A developed pulse conditioner for converting between NIM and TTL pulse standards, required to maintain the high temporal resolution generated SPADs [80], is also presented and evaluated.

### 5.1.1 Detector technologies

Early designs for high temporal resolution measurements were based on photomultiplier tubes (PMT) [81] - a type of vacuum tube. Here, a photon striking a photocathode results in the emission of a photoelectron via the photoelectric effect. A focusing electrode imparts kinetic energy on the photoelectron by accelerating it through an applied field to an electron multiplier plate (dynode). The acquired kinetic energy causes emission of secondary electrons, through ‘impact ionization’, who in turn undergo a series of pulse amplification steps at subsequent dynodes. The event is finally detected when the electrons strike the anode of the PMT. The analog current pulse is then processed by a ‘front-end’ electronic system to extract pulse properties and relay them to the data acquisition system. Examples of such properties include pulse amplitude, duration, rise time. While the above describes a generic method used in particle detectors, for low-light level photon detection the quantity is the arrival time of the leading edge of the pulse, considering each photon to be ‘identical’ for purposes of detection.

To achieve this the front-end for photon detection converts the analog pulses to a digital signal with a pulse shaper. In the case of a PMT we must remove effects such as pulse-height and width variation inherent in the amplification process, commonly done using a ‘constant

fraction discriminator' (CFD) triggering when the analog signal reaches a fixed fraction of its maximum amplitude. Using a CFD on the output of a PMT can be used to record single photons, given that the rate of photon arrival is sufficiently low for the individual pulses to not overlap temporally. At this point it is worth noting that the key distinction between a conventional analog PMT and a 'Photon counting' PMT is in effect based not on the mechanism of detection but on the device specific parameters, such as dark current and presence of such pulse-shaping electronics. The dark current is an inherent feature of the detectors defined as the leakage current measured from a PMT when no light is incident on input port of the device, arising from thermal excitation of charge carriers (electrons). This dark current translates into a 'dark count' on the output of the device setting the baseline signal that is detectable from a PMT or other detector. While conventional PMTs, have and still are, widely used detector technology, new variants such as Micro-Channel-Plate (MCP-PMT) improve the time resolution of this technology down to the picosecond range.

Challenges associated with construction of PMTs have however results in development of solid-state semiconductor devices. Semiconductor based detection utilize the same principle of impact ionization as a PMT which now occurs inside a p-n junction of photodiodes (PD). A silicon photodiode is formed using a p-n junction where a photon absorbed inside the depletion region under reverse bias will create an electron-hole pair that is accelerated by the applied potential. For low-level light detection the reverse-bias is increased so that the each charge carrier in the electron-hole pair created after the absorption event acquires kinetic energy and on collision with the lattice inside the junction produce secondary charge carriers giving rise to an avalanche, resulting in the naming of such systems 'Avalanche photodiode' (APD). The APDs are operated at or near breakdown voltage of the material and can be used as replacements for analog PMTs. Further increasing the reverse bias to above the breakdown voltage of the material increases the kinetic energy and the amount of impact ionization with each photon producing the current pulse. This increases sensitivity and allows for single photon detection, giving rise to the 'Single-Photon Avalanche Diode' (SPAD). One key distinction between APDs and SPADs is that SPADs are reverse biased, well above the breakdown voltage and require 'quenching' by temporarily lowering the reverse bias below breakdown in order to allow device recovery, performed either by an active or passive 'quench circuit' located in close proximity to the junction. Through increasing the number of SPADs on a single silicon chip and connecting the outputs of individual SPADs in parallel one arrives at Silicon-Photomultipliers (SiPMs), discussion of which we omit here.

## Time correlated single photon counting

---

### Single Photon Avalanche Diodes

Before describing the detection systems utilized in this project it is important to understand properties of the SPAD detectors available for experiments in order to determine suitable parameters for the data acquisition system.

The first key feature of SPADs is the quenching circuits that, on detection of an avalanche current, lower the reverse bias to below threshold in order to quench the avalanche and bring the SPAD to equilibrium. In the case of a SPAD device (PDM, Micro Photon Devices) the SPAD has an integrated active quench circuit which deactivates the bias voltage for a fixed time interval after the detection. This fixed time interval, of the order of 80 ns determines the saturability of the SPAD and the maximum count rate, given by the inverse of the dead time.

In addition to limiting the maximum count rate through saturation of the detector, the dead time introduces a nonlinearity of scaling in the count rate. This arises from the stochastic nature of photon arrival times and the ‘censoring’ of photons arriving within the deadtime from a previously detected photon. It can be shown that this produces a relation between the effective count and real count rate given by [82]:

$$R_{actual} = \frac{R_{measured}}{1 - R_{measured} \cdot \tau_{deadtime}} \quad (5.1)$$

This effect can be both beneficial, increasing the dynamic range of the SPAD [83, 84], as well as detrimental in the case when the true photon variation with time is required. Effects of deadtime for paralyzable and non-paralyzable [85] detector systems used to compute correlation functions was investigated by Schatzel [86] where increases in the deadtime parameter and count rate introduce increases autocorrelation decay rate and lower the coherence factor  $\beta$  - leading to increased estimates of polydispersity for PCS measurements. To determine a suitable operational count rate, forming part of the requirements for the further developed TDC system, consider the requirement of a 1% error in count rate according to the above equation. Using this as a criterion, for an approximate 80 nanosecond deadtime, can estimate the required count rate to be 100 kcps per SPAD. This forms one of the requirements for the photon counting system, in that the readout rate must be able to sustain continuous streaming of 100 kcps per channel.

A second parameter of SPAD detection is the effect of charge carriers trapped in the lattice during an avalanche [87] that can retrigger a second avalanche, some time after the first. This effect is termed ‘after-pulsing’ [88] and manifests as an increased probability of a

second pulse during a short time interval (Fig.5.7) after the primary detection event giving rise to an increased probability in the short-time tail of the inter-arrival time distribution, discussed in theoretical background (Eq. 2.16). The increased detection probability results in an increase in the number of counts detected, for the MPD PDM [80] modules specified as a 1-3% increase of absolute count rate [80] relative to the true count rate. The skewing of the inter-arrival time distribution also manifests as an increase in the autocorrelation function at short timescales ( $\leq 1 \times 10^{-5}$  seconds), problematic for PCS and Fluorescence correlation spectroscopy (FCS) measurements. In practice this can be removed experimentally through cross-correlating two SPAD photon streams, based on the assumption that after-pulsing for the two detectors is independent. The effect may also be treated computationally through algorithmic filtering [89] techniques on the count timeseries. The conceptually simpler technique of cross-correlating to remove afterpulsing from measured correlations motivates the need for a large number of channels in the developed system. The design of a five channel systems enables simultaneous recording of up to two channels in cross-correlation mode, with one remaining channel recording trigger events, for example from a high speed trigger diode anticipated to be useful in FLIM where the fluorescence decay is referenced to the incident laser pulse.

The final metric that is of interest to the experimenter is the time resolution of the detectors. For SPADs the conventional method of measuring the temporal resolution is through the measurement of photon arrival times from a pulsed laser and characterizing the distribution of arrival times relative to the reference clock generated by another fast photodiode with high time resolution. The measured distribution of arrival times is then a convolution of responses of both the detectors and the timestamping electronics as well as the optical pulse width, negligible in the case of femto-second pulsed lasers with pulse-widths down to the 100 femtosecond range. In the case of a perfectly fixed time interval between successive pulses the resulting time distribution forms a  $\delta$ -function. In practice, variation in inter-pulse arrival times, the ‘jitter’ in a system produces a finite width distribution. If considered in the frequency domain, the analogue of ‘jitter’ is ‘phase noise’ of a periodic signal.

**Definition 5.1.1** (Jitter). Deviation from true periodicity, typically considered for a reference clock signal. The concept of jitter in timing systems is an equivalent of phase noise in the frequency domain.

**Definition 5.1.2** (Phase noise). Deviation from true periodicity considered in the frequency domain as variation of phase of sinusoidal wave. Frequency analog of jitter.

Inside the detector, key design decisions that affect temporal resolution include both sensor design and signal conditioning circuits, such as the pulse discriminator, an analogue of

## Time correlated single photon counting

---

the CFD for PMTs. The output pulse standard driver circuit also determines the temporal resolution. For the case of the MPD PDM there are two options available. The first, a positive polarity, 3.3V active high, Transistor-Transistor Logic (LVTTTL) signal with time resolution of 250 ps and pulse duration of 20 nanoseconds [80]. The second, high resolution output, is a negative polarity Nuclear Instrumentation (NIM) output with time resolution down to 30 ps [80].

In part the properties of the output pulses are determined by the logic technology on which the pulse standards are implemented. Newer logic families such as emitter-coupled (ECL) are capable of faster switching and thus shorter rise times. Rise time and jitter in the leading edge of the pulse are related by [90] Eq 5.2, where  $dv/dt$  is rate of change of voltage signal (related to rise time),  $\sigma_v$  is the RMS value of noise voltage, and  $\sigma_t$  is the standard deviation of the timing error from a single point measurement of a step transient (we assume leading edge).

$$\sigma_t = \sigma_v (dv/dt)^{-1} \quad (5.2)$$

Finally, for low count rate detection, the dark-counts [91] represent the background events due to thermal carrier excitation [87] inside the junction of the SPAD. This scales with detector active area and junction temperature, requiring TEC cooling of the sensor to achieve the specified value of 15 counts per second.

### 5.1.2 Time interval measurement

We now provide a brief overview of time-interval measurement techniques in order to demonstrate that this in itself is a diverse and varied field.

#### Counter

Perhaps the most simple method for measuring time intervals is through counting the number of cycles between the ‘start’ and ‘stop’ signals defining the interval using a high frequency reference oscillator. In this case the counter returns the total expired time, with an uncertainty of the true pulse width given by 1 cycle of the oscillator. For a reference clock with frequency of 100MHz this gives a TDC resolution of 10 nanoseconds. Increasing the clock frequency above 1GHz allows sub-ns resolutions, such as the TDC demonstrated by Kirichenko et al [92] using a 33GHz superconductive TDC. While technologically possible, implementation of such systems in practice is complex. An alternative is the use of multiple counters with

phase-offset clocks that allow time-interleaved measurements [93] and thus enable resolution beyond that of each single clock. A key consideration is then the need to maintain high stability in relative phases of the individual clocks.

### **Time-to-Amplitude converter**

Picosecond time interval measurements can also be achieved using a time-to-amplitude converter (TAC) [94, 95]. Such systems measure intervals on the nanosecond level by monitoring the (dis)charging of a capacitor initiated on incidence of a start pulse. For constant current charging, the time interval between is proportional to the voltage across the capacitor. Readout occurs on the incident of a stop photon using a high speed analog-to-digital converter (ADC).

Such systems have the limitation that the capacitor charging time constant sets the limit of maximum time interval measurable. Secondary limitations are requirements to reset the capacitor and sample the voltage by the ADC. These place limits the deadtime of such systems to the tens of nanosecond range.

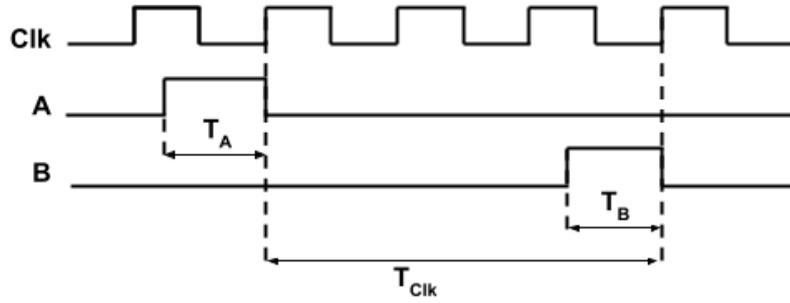
### **Nutt method**

In this work we use the Nutt method for time interval recording, an extension of the counting technique. Instead of starting and stopping the counter at the arrival of the start and stop signals, we instead determine their arrival times relative to a free running counter (/reference oscillator). Subtracting the difference between the two times recovers the duration of the time interval. Initially proposed by R.Nutt [96] in 1968 this method has been applied to many TDC designs [97] [98]. In order to achieve the fine time resolution required the method measures a relative ‘fine’ time difference between the start/stop pulse edges and rising edge of a reference oscillator ( $T_A, T_B$  of Fig.5.1). The reference oscillator with frequency  $f_{Clk}$  is then used to count the number of elapsed cycles, giving a time interval of  $T_{Clk}$ .

This technique is advantageous as the duration of interval measurable is limited only by the maximum number bits used to represent the reference oscillator counter  $N$ . In practice increasing the duration of the measured interval incurs a reduced resolution due to reference oscillator instability, arising from various sources, such as voltage supply noise or thermal drifts.

In the case of the TDC, system clock is used as a reference oscillator. The ‘fine’ time intervals  $T_A, T_B$  therefore represent values obtained by performing ‘sub-clock-cycle’ interpolation. Similarly to the counter values  $N_A, N_B$  the times  $T_A, T_B$  are also encoded as binary strings,

## Time correlated single photon counting



$$T_{interval} = T_{Clk} - T_B + T_A \quad (5.3)$$

$$T_{Clk} = (N_B - N_A) / f_{Clk} \quad (5.4)$$

Fig. 5.1 Time interval measurement using Nutt method

with the number of bits used determined by the properties of the method used to perform the interpolation. The ‘coarse’ values are obtained from a counter at the system clock frequency  $f_{Clk}$ .

### Tapped Delay line

In order to measure the fine time interval, either between a start/stop or using the Nutt method for a digital system one needs to perform interpolation inside the clock cycle. To achieve this differing techniques, implemented both in ASICs and on FPGAs have been proposed, with much effort devoted to utilizing FPGA based delay lines.

A delay line consists of a linear chain of delay elements, with delays  $\tau_i$ . The output of each delay stage is split with one end connected to the next delay element while another is connected to a readout ‘tap’ where its arrival is recorded. As the pulse propagates through the delay elements it triggers tap outputs. Synchronous readout of the taps allows for the measurement of how far along the delay line the pulse has propagated.

Figure 5.2 shows a generic tapped delay line (TDL) with readout taps. Here the start and stop pulses propagate through delay elements with delays  $\tau_i$  and  $\delta_i$  (Fig. 5.2) with the stop pulse triggering state transition transferring D input to the Q output. In the case when the stop signal is the system clock and  $\forall i. \delta_i = 0$  we obtain a ‘synchronous-stop’ delay line implemented in the TDC. This however isn’t the only architecture - alternatives such as a ‘Vernier’ delay line where  $\delta_i \leq \tau_i$  use the smaller stop delays to reduce the interval between the start and stop at the readout flip flop as the pulses propagate.

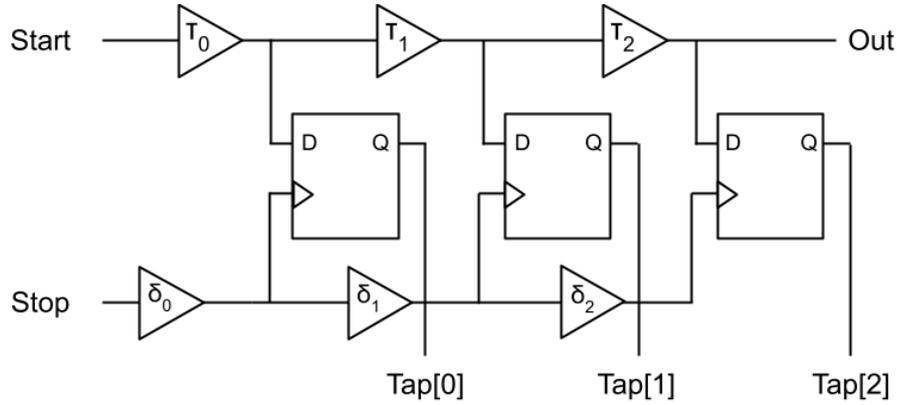


Fig. 5.2 Tapped delay line

In the general case, the total propagation delay to the  $n^{th}$  stage for the start pulse is given by

$$\tau_n = \sum_{i=0}^n \tau_i \quad (5.5)$$

When using the ‘synchronous-stop’ delay line, with readout by a clock with frequency  $f_{Clk}$  and assuming the same delay for each of the elements  $\forall i. \tau_i = \tau$  it is possible to lower bound the number of delay elements  $L$  required for the delay line to generate delays of  $T = 1/f_{Clk}$ :

$$L = \frac{T}{\tau} = \frac{1}{f_{Clk}\tau} \quad (5.6)$$

The above ideal cases are complicated by real situations where manufacturing variability may give rise to different propagation delays  $\tau_i, \delta_i$ . In FPGAs, effects like clock skew may also mean that the readout is not synchronous and metastability may result in incorrect readout. Additionally, implementation details such as ‘look-ahead’ logic may mean that the propagation delays may also, in effect, be negative. For these reasons delay line implementations have been largely restricted to ASICs where the designer has greatest control over components types, their layout and connections.

To evaluate the performance of a TDC one must first calibrate the timescale. When using the Nutt method the coarse time is set by the system clock. The fine time however must be calibrated using a ‘statistical code density test’ (Fig.5.3a) where pulses, uncorrelated with the system clock, are fired into the delay line and the resulting timestamps are recorded.

## Time correlated single photon counting

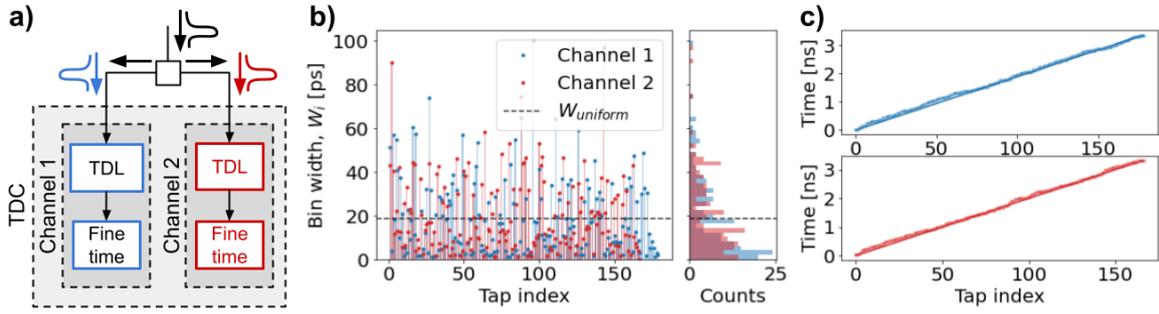


Fig. 5.3 Two channel code density test a) An input pulse is split and received by two TDLs inside the five channel TDC. Each TDL produces a fine timestamp, corresponding to the number of taps the pulse has propagated before the next rising clock edge. b) Temporal bins widths of each ‘tap’ calculated as fraction of total events with the fine timestamp equal to the tap index. c) Calculated calibration table, mapping between a tap index and a time of arrival, with sub-clock cycle resolution.

Each timestamp encodes the maximum tap index to which the pulse propagated, a form of ‘thermo-binary’ coding often seen in ADCs. The temporal ‘width’ of each delay element can be calculated by counting the number of pulses that reach the given tap, under the assumption that pulses are uncorrelated with system clock.

$$W_i = \frac{N_i}{N} T \quad (5.7)$$

Here  $N_i$  is the number of pulses recorded at the  $i^{th}$  tap,  $N$  is total number of pulses and  $W_i$  is the bin width. In the code density test shown in Fig.5.3a we use  $9.3 \times 10^6$  hits for both channel 1 and 2, using a 800 kHz hit rate per channel. Combining with the previous bin widths one can then compute the fine time calibration table for each bin as (Fig.5.3b) [97] by:

$$T_i = \sum_{j=0}^{i-1} W_j + \frac{W_i}{2} \quad (5.8)$$

In order to assess the linearity of the delay line, we can compute the relative deviation of each bins width from the ideal uniform delay termed ‘differential nonlinearity [99]:

$$DNL_i = \frac{W_i - W_{uniform}}{W_{uniform}} \quad (5.9)$$

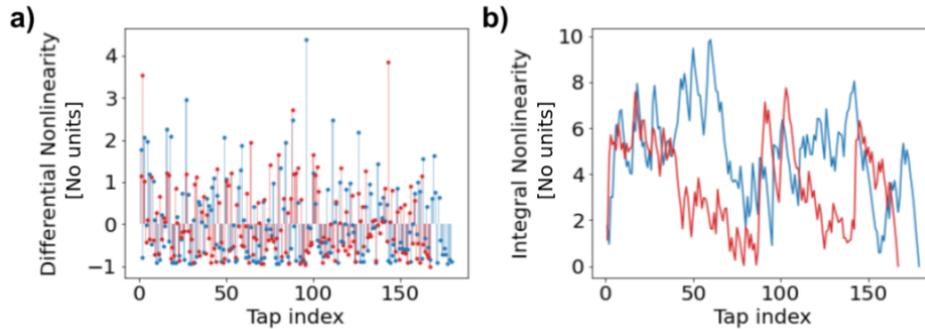


Fig. 5.4 Differential and integral nonlinearities computed from code density test data for two channels of the TDC.

The uniform delay  $W_{uniform}$  is the average expected delay per tap. For the data from code test of Fig.5.3 the resulting  $DNL$  is depicted in Fig.5.4a. Summation of the per tap  $DNL$  is termed ‘integral nonlinearity’ [99] shown in Fig.5.4b.

$$INL_i = \sum_{n=0}^i DNL_n \quad (5.10)$$

### Wave union

In absence of skews, the resolution or the TDC impulse response, follows a Gaussian distribution, characterized by the standard deviation. A simple scheme to improve the resolution, beyond that of a single TDC, is to average multiple,  $n$ , measurements of time interval and exploit  $1/\sqrt{n}$  scaling of the standard deviation - allowing the measurement of a single time interval to a precision better than the resolution of a single TDC element. On FPGAs, limitations of the TDL resolution also arise from non-uniformity of propagation delay induced by placement considerations, look-ahead logic, clock boundaries and skew, producing ‘ultra-wide’ bins which act to limit the resolution of the TDCs.

Instead of measuring a pulse multiple times, or with multiple delay lines, Wu et al [100] proposed the launching of multi-edge patterns into the delay lines. These patterns, consisting of a train of 1’s and 0’s allow for multiple measurements to be made from a single pulse input into a delay line, the ‘Wave-Union’ method. The improved accuracy comes at the cost of increased complexity of the decoder stages replacing the thermo-binary decoder used to decode the single edge pattern. This introduces more complex logic and increases per-TDL resource utilization, but has been applied to large-scale TDC systems [101].

### Ring oscillators

One limitation of delay lines is the scaling of the length of the delay line with the maximum interval required to be measured. If the delay line is folded back in on itself, the result is a ring-oscillator [102, 103] with the inverse of the total propagation delay giving the oscillation frequency. Using two such ring oscillators with differing frequencies is another technique for fine time interval measurement. Here, the start enables the slow oscillator  $f_{slow}$  while the stop pulse enables a fast oscillator  $f_{fast}$ . For a single cycle of each oscillator the temporal separation between the start and stop pulses decreases by  $\Delta\tau = 1/f_{slow} - 1/f_{fast}$ . The initial time interval is then proportional to the total number of oscillator cycles required before the stop pulse overtakes the start pulse, at which point the oscillators are stopped.

## 5.2 Field-Programmable Gate Arrays

Field-programmable gate arrays are reconfigurable integrated circuits whose logic level functionality can be configured post-manufacture through the upload of a configuration ‘bitstream’. Unlike traditional computer processors where the architecture of the processor is fixed at a logic level and where programmability is achieved through a program executed using a fixed instruction set. The programmable logic (PL) of FPGAs has simplified the cost and speed of prototype development in comparison to development of application-specific integrated circuits and is already broadly adopted in a range of industries. In the last decade the entry barrier to design of FPGA systems, and their availability is rapidly increasing to the extent that a development of prototypes for applications like photon counting becomes possible.

At a high level the FPGA architecture can be described as consisting of a set of configurable logic blocks (CLBs) connected together by a programmable interconnect (switch matrix). Each CLB contains elements such as look up tables (LUTs), flip-flops and multiplexers which can be configured to encode simple logic functions in each CLB. Linking multiple CLBs is then used to construct processing pipelines using the configurable switch matrix. Different architectures of FPGAs utilize differing CLB architectures. The Zynq-7000 CLB components are shown the Table 5.1 where the CLB is further subdivided into two slice elements, individually connected to the switching matrix (Fig.5.5). Crucially, each of the CLB slices contains a ‘CARRY4’ element, connected vertically to other carry elements, conventionally used to allow efficient implementation of high speed arithmetic operations, which we exploit in the construction of the delay line.

Slices	LUTs	Flip-flops	Carry chains
2	8	16	2

Table 5.1 Zynq-7000 CLB logic

Aside from these basic elements both CLBs and more broadly the PL include additional specialized components such as distributed RAM, shift registers and DSP elements. These can be exploited by high level synthesis toolchains to reduce CLB resource utilization in the PL and increase overall performance of the design, through accelerating the more specialized functionality using these elements. A thorough overview of the CLB architecture of the Zynq PL is given in Xilinx UG474 [104].

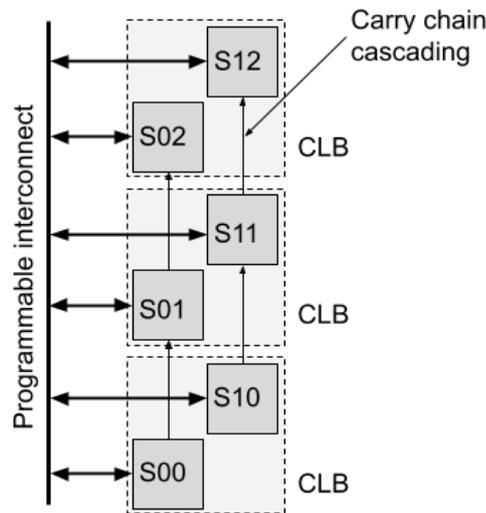


Fig. 5.5 Slice carry cascading within single CLB column

### 5.2.1 Zynq System-on-Chip

Early FPGA technology was based on a standalone FPGA ICs on which developers often implemented ‘softcore’ processors mimicking conventional CPUs. This came at the cost of reduced performance and increase complexity of the FPGA design and had a high resource utilization. The integration of both an FPGA-based ‘Programmable logic’ (PL) component with a more conventional ‘Processor system’ (PS) on a single chip resulted in a ‘System-on-Chip’ (SoC) design. The PL enables the low-level logic prototyping while the PS runs a more conventional system managing access to resources like main memory (RAM), a CPU register and common peripheral hardware as well as interfaces to the PL. This greatly reduces

## Time correlated single photon counting

---

the learning curve and complexity of interfacing to the PL inside the SoC, with one needing only to implement the PS-PL communication. Communication with external devices, such as other PCs can then be handled by the more conventional networking protocols included in operating system distributions running on the PS. This makes the SoCs a good choice for prototyping instrumentation like photon counting acquisition systems as it strikes a balance between performance, design complexity, flexibility, cost and learning curve.

In order to build a prototype, a SoC is integrated into a development board containing components like RAM memory, hardware peripherals, external port and interfaces. To this extent, the choice of development board is dictated not only by the SoC chip itself, but also by the properties board. Key among these are timing properties and PCB trace routing to the PL input ports.

In this project we selected the Xilinx Zynq-family of SoCs. These include a dual-core ARM Cortex A9 processor (PS) integrated with a Xilinx Artix-7 family of FPGAs (PL). The SoCs are integrated into the Arty Z7-20 and MicroZed 7020 evaluation boards, with the former being used for initial prototype proof of concept while the latter was used for a optimized version due presence of FPGA Mezzanine ports (FMC) and increased RAM memory capacity relative to the Arty. The Table 5.2 notes some key characteristics of the XC7020 SoC in comparison to both the XC7010 from the same family as well as the Altera Cyclone-2 board used previously [105] for TDC applications.

Name	LUTs	PLLs/MCMMs	PS	Technology
Cyclone-2(EP2C5)	4608	2	-	90 nm
Artix-7(XC7010)	17600	4	Cortex A9	28 nm
Artix-7(XC7020)	53200	8	Cortex A9	28 nm

Table 5.2 Parameter table for different PLs architectures and variants.

### 5.2.2 Time in digital systems

In order to measure time using a SoC one must understand the basics of timing in digital systems. Here, the passage of time is measured by a clock giving a fixed frequency of oscillation. Synchronous circuits such as flip-flops utilize a clock to determine when to transition between states. Asynchronous designs such as combinatorial logic on the other hand do not rely on a clock directly, but are typically connected at inputs and outputs to synchronous components.

### Synchronous circuits

A synchronous design ensures that state transitions in the circuit occurs based on a common trigger condition, typically the edge of a clock. Using this one can consider all state changes to occur ‘simultaneously’, simplifying analysis of the circuits and masking effects due to components variability. Incorrect synchronous design can lead to circuits producing undefined ‘metastable’ states.

**Definition 5.2.1** (Metastability). is a condition occurring when inputs to a flip flop change at similar times to the clock making the circuit enter an unstable equilibrium state between 0 and 1, representing thresholds in a given voltage range.

Such metastability may result in unpredictable behaviour of digital systems. In a synchronous design one aims to avoid this through ensuring that the input to an element such as a flip-flop is stable on the input for a given time interval before and after the sampling clock edge. These are known as the ‘setup’ and ‘hold’ times respectively.

**Definition 5.2.2** (Setup time). The minimum time required, before sampling (active clock edge) that an input must be stable for sampling to be reliable.

**Definition 5.2.3** (Hold time). The minimum time required after the sampling event (active clock edge) that input must be stable for sampling to be reliable.

Physical components like flip-flops also have a finite duration which is required for the signal to propagate through the component, from the input to the output, termed ‘propagation delay’.

**Definition 5.2.4** (Propagation delay). The time between the sampling (clock active edge) event and the output of the flip flop changing to equal the input

Combined, these form a set of timing constraints that a synchronous digital design must satisfy. These constraints become harder to satisfy with an increasing clock frequency as the available time for signal propagation, setup and hold falls.

### Clocks

At high frequencies non-ideal effects and finite speed of light cause delay and phase offsets between the clock edges arriving at different components resulting in the breakdown of the assumption of simultaneous switching.

**Definition 5.2.5** (Clock skew). an effect occurring in digital systems where the arrival of the leading edge of a clock into two or more components has a temporal offset.

## Time correlated single photon counting

---

This is critical to avoid for TDC applications as the skew between clock inputs on the readout flip flops can result in errors in the fine resolution bits of the system. Counteracting skew can be achieved by using multiple clock domains on a single system. For such systems, a global master oscillator drives multiple phase locked loops, each of which is configured to have a different phase offset to counteract the skew due to propagation delays. Skews can be further minimized through specialized clock distribution paths on the ICs. Issues with crossing between different clocking regions can be problematic for high time resolution measurements on TDCs and can manifest as ‘ultra-wide’ bins of the carry chains, discussed later. A second parameter of interest is the long-term stability of a system clock. Effects such as temperature variation and power supply noise can produce small shifts in oscillator frequency, resulting in drift of the time recorded by the system.

**Definition 5.2.6** (Clock drift). a feature of any pair of timing devices meaning that over time the clocks will desynchronize. This feature manifests due to heating effects such as CPU load.

The third feature of clocks in digital systems is the need to move between clock domains running at different frequency - Clock domain crossing (CDC). When the clocks are asynchronous a signal crossing over from one domain to another without adequate CDC techniques may give rise to metastability in the destination domain. We consider this in a later part of this chapter.

### 5.2.3 Advanced eXtensible Interface (AXI)

Similar to the programmable interconnect at the CLB level of the PL, a modular design of PL subsystems is beneficial. In order for individual components to communicate they must implement a common series of commands and signals, an ‘interface’. For the case of the Zynq SoC, which incorporates a dual core ARM Cortex A9 CPU the communication uses the ARM AMBA (Advanced Microcontroller Bus Architecture) standard. Part of this architecture is the Advanced eXtensible Interface (AXI4) which the individual modules, known as ‘Intellectual Property’ (IP) cores, have to implement in order to communicate over the bus. Different modes of communication are enabled through different variants of AXI: AXI4, AXI4-Lite and AXI4-Stream. One or more of these must be used inside each module implemented inside the PL. The IP cores can then be connected together and interfaced to the PS using AXI interconnects and other generic IP cores such as FIFO queues or DMA engines. Two types of communication are possible, a ‘master-slave’ transaction based communication and a continuous ‘streaming’ between cores. The former is often used to enable interfacing of the IP cores with user-space applications, by using ‘memory-mapping’ which allows

user-space programs to read or write register values of the IP inside the PL through linking the IP core registers to regions of the main memory address space.

### **Memory mapped AXI**

For purposes of communication the IP cores implementing AXI are separated into a ‘master’ initiating the data transfer, and ‘slave’, responding to the data transfer in both the read and write cases. For both transaction types, the cores perform a two-way handshake with the master first setting to logic high a ‘valid’ signal indicating that it has placed a valid signal onto the transfer lines. The slave sets its ‘ready’ signal high when it is ready to receive information from the master. The transfer begins when both ‘valid’ and ‘ready’ are asserted and lasts for a set number of data transfers ‘bursts’. The AXI4 protocol has a limit on the size of the burst transaction to 256 data transfers, with the AXI-Lite variant allowing for only one data transfer, reducing throughput but increasing simplicity and lowering PL resource utilization. When memory mapped the IP core register values mapped to a region of main memory. Access to the registers of the IP cores to user-level code is then provided by the kernel drivers that interface the user-space applications with the main memory, through the operating system.

### **AXI Stream**

While useful for controlling IP cores the memory mapped devices communication is often too slow for continuous acquisition. A more suitable variant of AXI communication is the AXI Stream. Here the protocol allows for unidirectional communication between a master and slave(s) with an unlimited burst size avoiding the overhead of requiring to frequently re-initiate the handshake. With the potential for an unlimited burst size comes the caveat that the AXI stream, unlike AXI and AXI-Lite is not memory mapped, it does not directly write to a fixed location in memory. Instead the location of the next packet is controlled through configuring the ‘Direct memory Access’ (DMA) engine, connected on the Zynq through the high bandwidth (HP0,HP1) ports to the main memory. During data acquisition, end of a given packet is indicated by the master generating a ‘TLAST’ flag, in turn triggering an interrupt, handled by a PS-level interrupt service routine (ISR) that reconfigures the DMA engine. The reconfiguration takes the form of setting the next memory locations to which data is transferred from the PL. Bypassing the PS during data transfer, apart from the ISR, allows this technique to significantly exceed the memory mapped method in terms of maximum data transfer rate.

### 5.2.4 HDL/Verilog

Two possible routes for designing IP cores are available, a ‘Hardware description language’ (HDL) based design or a ‘High level synthesis’ (HLS) design. The latter uses a high level language like C or MATLAB to specify the functional behaviour of the module while the former is a special purpose language like VHDL or System Verilog. We opt for the HDL path based on the basic subset of System Verilog, Verilog. Unlike conventional high level programming languages where execution of the code is sequential, Verilog is event driven. In Verilog models of the system exist at the level of individual registers acting as variables storing state. Data transfer between registers is controlled by clock signals, while asynchronous components are modelled by combinational logic. This abstraction level is known as the ‘register-transfer’ level (RTL).

#### Design flow

The photon counting system is complex. To deal with the complexity we work on multiple abstraction levels using a range of different tools included as part of the development toolchain.

Logic level module design including the TDC itself is performed using Verilog. Each of the modules is represented as a set of source files. These are packaged into IP cores using the Vivado IP packer. Module level testing is carried out using testbench files defined in a similar way to the source verilog. Simulation of the module is also performed in the testbench allowing visualization of the signal traces. Packaged IP cores after testing are stored in an IP repository, for later integration into the block design.

Next, the custom cores are imported into a separate IP-centric design flow using the Vivado IP integrator. Here, the custom cores are combined inside a block design with Xilinx cores such as AXI controllers, DMA engines and Interrupt controllers. At this stage a model of the processing system (PS) in the Zynq SoC is introduced into the design. Various parameter customizations and interfacing of the PS core to the rest of the PL design is performed at this stage. The parametrization of the PS core allows the specification of key quantities such as number and frequency of system clocks, enabled interrupt and reset lines and number of AXI masters originating in the PS that are connected to the PL AXI interconnect. It is at this stage that we enable the high speed HP ports for later configuration by the DMA. Other AXI cores are similarly customized at this stage, for example when using AXI Stream FIFOs one must specify FIFO bus width as well as depth, the latter in part determining the maximum burst rate that the readout system can sustain, as interrupts must be serviced by the PS before the FIFOs fill up. In addition to this, the block design also includes the specification of the

memory mapping, with each AXI/AXI-Lite core connected to the PS being mapped into the main memory address space. Once complete the block design is validated by the Vivado IP integrated and ‘compiled’ and output products generated.

The process of converting the block design into a bitstream consists of the generation of a system-level ‘Register-Transfer-Level’ (RTL) representation of the system through combining the HDL code of each of the IP cores using the specification in the block design. The RTL representation of the system undergoes ‘synthesis’ that converts the HDL design to a ‘gate-level’ netlist, performing logic simplification and optimization as well as introducing gate-level timing information and incorporating constraints into the design. This netlist is then further processed in an ‘implementation’ step where further, technology specific mappings and optimizations are performed along with a place-and-route, a step that assigning positions to elements onto the target platform. The design is again augmented with more in-depth timing information more specific to the device technology.

### 5.2.5 Linux

The generated bitstream is integrated into a Xilinx Linux variant ‘Petalinux’ which is built using a command line script utilizing tools provided by Xilinx (thanks to Mohammadsadegh Sadri for providing tutorials on how to build the Petalinux distribution and how to integrate the IP cores into the build, as well as other examples of Zynq development!). At this stage of the build kernel level drivers and application code is incorporated into the system. In addition to this the kernel build parameters are customized to enable functionality, such as increasing the amount of useable memory available to processes, required for the ring-buffer storing the photons streaming into main memory by the DMA engine.

Another aspect of the Linux build is the specification of mappings between the kernel driver and the peripheral devices, in our case the AXI components which have been mapped to certain memory positions inside the memory. This mapping is determined by a ‘device tree’ a configuration file encoding devices as nodes of a directed acyclic graph. On boot the device tree is unfolded and each peripheral devices driver is determined. The device tree mapping is in part generated by the Vivado pipeline, but at this point needs to be modified to allow integration of modified kernel drivers. A particular example of this the DMA driver and associated DMA IP core. In order to allow control of the DMA driver used we manually wrap the DMA engine at the IP core level, in order to allow control of tags used in mapping between IP cores and drivers in the device tree. Without this the DMA IP core would be mapped to generic Xilinx DMA drivers.

## **Time correlated single photon counting**

---

In our case the DMA kernel driver is a simplified version of a DMA driver with fixed block size designed for the basic implementation of photon stream readout. Similarly to the ADC readout time-tagging scheme described below DMA driver operates in a round-robin/circular buffer scheme. In this scheme each interrupt generated by the DMA engine inside the PL is handled by an ISR defined in the driver. The ISR increments the pointer to the next buffer to be filled by the incident photon timestamps. One caveat of this process is that while the ISR is being handled the AXI FIFOs are not being read out by the DMA engine, as it waits for a new memory block to be allocated into which to write. While not a significant problem in the current implementation this means that short AXI FIFO depths can lead to the FIFOs becoming full in the time interval that the PS takes to handle the ISR. We anticipate that at high readout rates this may indeed occur, but suggest that such a limitation can be addressed by exploiting the dual-core ARM processor in the PS and modifying the core-affinity settings of the ISR, allocating it to its own core. Simultaneously to the ISR, an application process is reading from the device and incrementing its own buffer pointer. The application code reads from the buffers and sends the read photon stream without additional processing to a UDP socket for readout to another computer, pausing when its buffer pointer is equal to the current ISR pointer.

## **Networking**

The motivation for incurring the extra complexity of building a Linux distribution onto the Zynq SoC where a ‘bare-bone’ implementation of a PS, without an operating system, may have performed better is the ease of integration with higher level networking protocols. The requirement to implement network protocols at the PL level has been noted by Traxler et al [101], during their design of the Trb-3 on Lattice FPGAs, to be prohibitively complex for the case of the TCP protocol. On the Trb system the design was subsequently limited to UDP. By utilizing the TDC on a Zynq SoC running a Linux distribution we bypass the need to implement networking protocols, instead relying on the libraries supplied with the operating system. This allowing us to potentially use either TCP or UDP. While this design does not utilize TCP, opting for UDP, in principle this can be implemented in a future iteration.

While there have been qualitative discussions that the use of a bare-bone UDP implementation may be faster in practice the incurred complexity and the availability of optimizations on the kernel or application level to achieve faster readout with UDP up to 250 Mbit/second without additional PL components was demonstrated in this project. This corresponds to an approximate continuous readout rate of 3 Mhits/second using direct ethernet link connecting the Zynq board to a acquisition computer. In future work we hope to increase

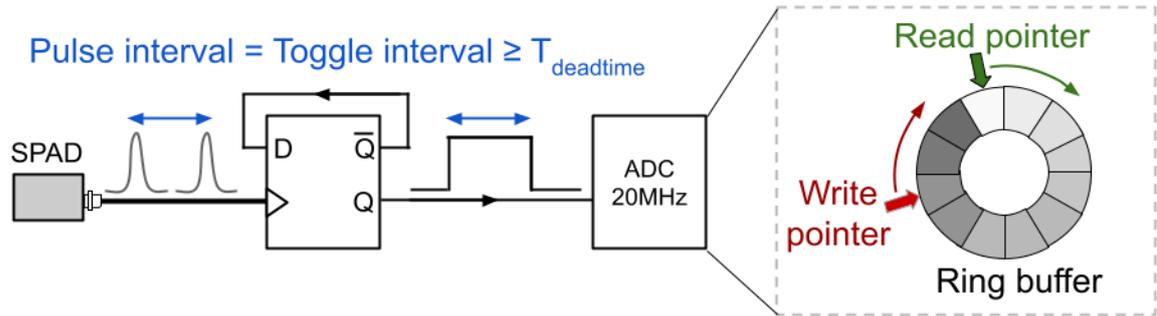


Fig. 5.6 ADC based readout system exploiting SPAD deadtime. The flip-flop output transition occurs on receiving a detector pulse. Given a pulse duration of 20 ns produced by the SPAD, with a deadtime  $> 70$  ns, a 20 MHz sample rate ADC can acquire the continuous photon stream without missing a photon as the interval between sampling events is lower than the maximum toggling rate of the flip flop.

the maximum hit rate to be above 10 Mhit/second, corresponding to the saturation limit of SPAD detectors.

This increase in performance of data readout can be achieved on the same Microzed platform through on-board compression of photon timestamps combined with improved performance optimization such as processor affinity for the UDP sender application and the utilization of a PL UDP core (outside the PS) that will increase the 'Maximum Transmission Unit' (MTU) of the UDP protocol beyond the current value, reducing per-packet overhead and increasing the transmission rate up to the specified maximum of 1Gb/s [106] by using 'jumbo' frames [107].

## 5.3 Nanosecond TDC

Before discussing the Zynq SoC based photon counting system, we will consider features and limitations of other systems, utilized primarily in the PCS experiments of Chapter 3.

### 5.3.1 Analog-to-digital converter

With the prevalence of analog to digital (ADC) converters in most experimental laboratories a widely available solution to recording the photon count rate is to use properties of the output digital pulses to digitize them using an ADC. This type of system was the first system constructed for PCS measurements in this thesis and used a 20 Msample/second ADC, Adlink 9812, PCI base acquisition system.

## Time correlated single photon counting

---

For a sample rate of 20 Msps the interval between individual samples is given by 50 nanoseconds, more than double the 20 ns pulse duration of the output LVTTTL pulse. Through exploiting the knowledge of the SPAD deadtime being greater than 70 nanoseconds we can achieve low resolution high rate detection of the photon stream by inserting a d-type flip-flop as an intermediate between the SPAD output and the ADC. The flip-flop toggles once for every detected photon, acting as a ‘frequency divider’ where the input clock is the rate of photon arrival. The lower bound on the toggling interval is set by the dead time, thus it becomes feasible to record photon arrival times from the SPADs with the temporal resolution of 50ns and a maximum count rate of 10 Mcps. To sustain high speed readout (20Msps x 12 bits) the ADC is interfaced to a computer via PCI interface storing memory in a ring-buffer. The topology allows continuous streaming of data out from the device either into RAM or onto disk. While allowing for, in principle, highest count rates of all systems investigated, the ADC technique suffers from linear scaling of data volume with acquisition duration, as well as linear scaling with number of channels. This makes multichannel detection difficult due to high (100s of Mb per second) data rates involved. For ‘low’ count rate experiments with rates < 100 kcps the majority of the data volume contains little information, and requiring a large amount of post-processing to extract photon data.

### 5.3.2 High speed counter

The only available open source alternative to photon counts data is the FPGA based timetagging system [105] developed initially by S. Polyakov, A. Migdall and S.W. Nam and further extended by B. Gamari [108]. The principle of operation is to assign a timestamp to each arriving pulse using a Altera, Cyclone 2, FPGA. The system operates by using a Crystal oscillator with a 32MHz crystal oscillator that is up-multiplied by (4×) using a phase-locked loop (PLL) to produce a system clock of 128 MHz frequency, giving a period of 7.8 ns.

This enables a time resolution of 7.8 ns and allows continuous monitoring of arriving photon data. With a single photon producing a timestamp of 48 bits the approach also scales linearly in data rate with the number of detected photons. With four channels, this system exceeds the capabilities of the ADC-based approach for many photon counting applications.

However, being based on an FDTI USB controller implementing USB protocol, such a system also has limitations from the perspective of read out rate. The theoretical USB2.0 maximum theoretical signalling rate is 480 Mbit/s (10Mcps). In practice this rate for the system is limited to below approximately 1 Mcps, above which we begin to lose photons during transmission from FPGA fabric to the USB host controller. A ‘loss’ occurs when the USB protocol is not able to keep up with the incoming rate and is forced to overwrite a photon

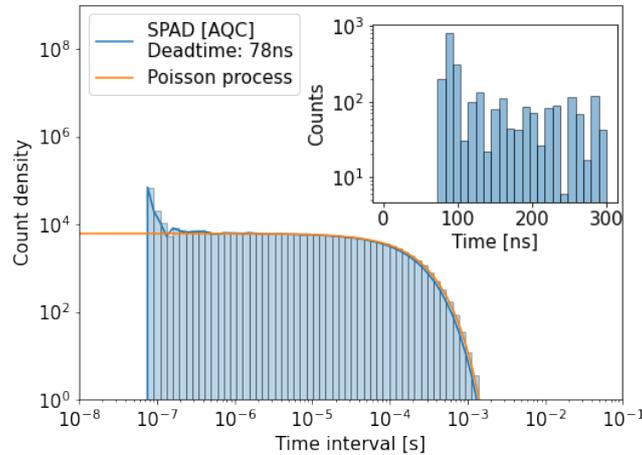


Fig. 5.7 An example of a SPAD dead-time measurement using Cyclone 2 FPGA system. The threshold for truncation of the inter-arrival distribution is the measured dead-time  $\sim 80$  ns. The increase in arrival probability at short times is the effect of afterpulsing. It should be noted that the active quench circuits are "non-saturable" detectors whose dead-time is not extended by incidence of another photon during a dead period and hence they are not sensitive to "pile-up" effects [86]. As reference we plot the ideal Poisson process inter-arrival distribution, for a 4 kcps count rate.

in the output queue of the device. In order to notify the user of this the system sets a 'reset' flag inside the newly recorded photon. As a single photon can be rewritten multiple times while in the queue, this system only provides a lower bound on the number of missed photons, but is a good indicator of problems in practice. Implementation of such a 'lost-photons' mechanism is a candidate for future implementation in the Zynq TDC system.

The ability to measure arrival time with 7.8 ns resolution enabled us to characterize the SPAD deadtime accurately, as well as measure the effect of afterpulsing, both recording the distribution of photon interarrival times from a incoherent bulb (Fig.5.7). In the distribution the dead-time manifests as a truncation of the distribution at time delays below the dead-time, while the after-pulsing acts to increase the short-time arrival probability.

## 5.4 Picosecond TDC

In this section we describe key components in the developed sub-nanosecond TDC using the Zynq SoC, primarily at the RTL level, as these are most specific to the photon counting architecture. Steps in integrating these Verilog elements into higher level modules at the block, device and instrument level are omitted. To understand these the reader is referred

## Time correlated single photon counting

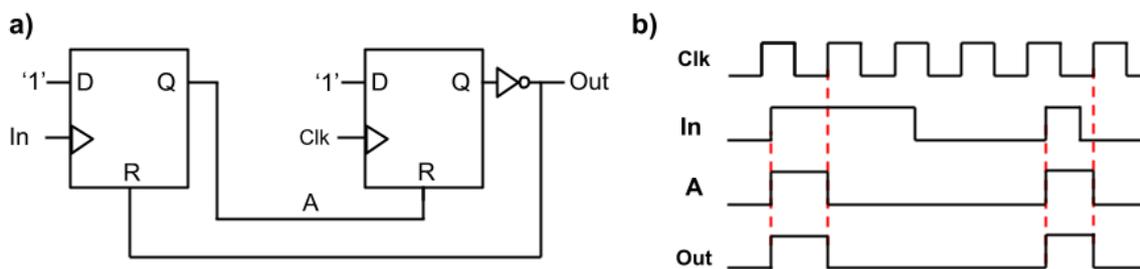


Fig. 5.8 Pulse conditioner

to many resources available from Xilinx and the broader electronic systems and software development and communities.

### 5.4.1 Signal conditioning

The first key element of a photon counting system, prior to timestamping pulses inside the FPGA fabric with the TDL is the pulse ‘conditioner’. In the case of pulses incident into the TDL the conditioner (Fig.5.8a) is used to bound the pulse duration, truncating the conditioned pulse at the next rising edge (Fig.5.8b). The conditioner preserves the asynchronous rising edge key to the detection, but acts to prevent retriggering of the TDL at the subsequent clock cycle. Motivation for the conditioner stems from the potentially variable duration of pulses produced by different detectors or pulse standards. An example is the NIM standard, where the duration of the pulse is not standardized, with the standard applying only to the rising edge.

The demonstrated design can also be applied to stretching the pulse, in the case when pulse duration is shorter than a single clock period, such as the case of a MCP PMT exhibiting  $< 1$  ns pulse durations [81]. Pulse stretching to more than one clock cycle can be achieved through introducing additional flip-flops at the output, using the final output as the reset signal.

### 5.4.2 Carry chain

The second component, forming the building block of the delay line are ‘CARRY4’ elements, included in each the two CLB slices (Fig.5.5). These elements are introduced into CLBs due to the need for efficient arithmetic operations, lying at the heart of all modern computing systems. Perhaps most basic among these is addition of two binary integers. In order to construct an adder for a N-bit binary word one begins with considering the addition of two bits (A and B), forming what is known as a ‘full-adder’, shown in Table 5.3.

Input			Output	
$C_{in}$	A	B	$C_{out}$	S
0	0	0	0	0
0	0	1	0	1
0	1	1	1	0
1	0	0	0	1
1	0	1	1	0
1	1	0	1	0
1	1	1	1	1

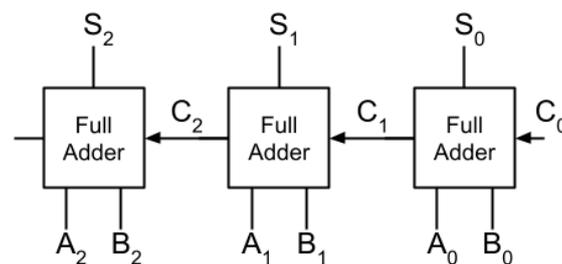
Table 5.3 Full-adder for adding bits  $A$  and  $B$ 

Fig. 5.9 Ripple carry adder

Here the  $C_{in}$  represents the input carry bit,  $C_{out}$  represents the output carry bit and  $S$  represents the sum produced by adding  $A$  and  $B$  given the input carry. Chaining  $N$  full-adder circuits allows the design of an  $N$ -bit binary word adder (Fig.5.9) . In the full-adder table, particular attention should be paid to the last row of the table where an incoming carry bit triggers the addition of  $A$  and  $B$  while also producing a output carry bit that can propagate to the next adder in the chain, ‘rippling’ through the chain. Encoding the input pulse rising edge using the bits carried through the full-adder allows the implementation of an  $N$ -stage delay line using the same architecture as an  $N$ -bit adder. Alternative types of delay lines, such as Wave-Union pattern launchers, similarly be considered as adders computing a summation of two binary strings. Care must however be taken as the logical picture presented above likely does not represent the physical layout, routing and implementation at device level. At this level the propagation delay through elements is detrimental for speed of arithmetic circuits, and thus techniques are used to propagate information faster through the carry chain. Examples of this include ‘lookahead’ logic where a ripple-carry is accelerated by precomputing carry bits. While beneficial for arithmetic this may have a negative impact on tapped-delay line designs, such as creating time bins with negative widths.

## Time correlated single photon counting

As mentioned previously, in the Xilinx Virtex-7 FPGA family, dedicated fast carry lookahead logic is incorporated on the CLB level and exposed to the designer in the form of a ‘CARRY4’ primitive. Inside each CLB, each of the two slices contains this carry element. These carry elements can be cascaded with carry elements in column-adjacent CLBs through linking their ‘carry-in’ (CIN) and ‘carry-out’ (COUT) ports. The caveat to cascading carries on the Zynq is that this cascading is possible only column-wise, and not inside individual CLBs, meaning that a column of CLBs in principle can implement two carry chains, but is limited in extent to the number of number of CLB rows in the SoC layout.

### 5.4.3 Taps and synchronizers

A signal from the pulse propagating through the carry chain produces a series of 1’s on the output of the ‘taps’. The taps change asynchronously relative to the system clock, with the propagation of the input pulse leading edge and can thus lead to metastable states at the outputs of the carry chain. In order to reduce the probability of metastability, due to setup and hold time violations, of the tap output we use a synchronizer [109] depicted in Fig.5.11. This acts to transfer the asynchronous signal into the TDC clock domain.

The probability of occurrence of a metastable state is characterized by a ‘Mean-time-between-failure’ (MTBF) [110–112]. Estimation of MTBF is typically performed during the process of compilation of the SoC design onto the architecture, and with multiple possible bounds that can be computed the calculation of MTBF is nontrivial. A simple equation for the MTBF can however give an indication of key parameters and motivate the need for the synchronizer:

$$MTBF = \frac{e^{(N-1)T_c/\tau}}{T_0 f_c f_d} \quad (5.11)$$

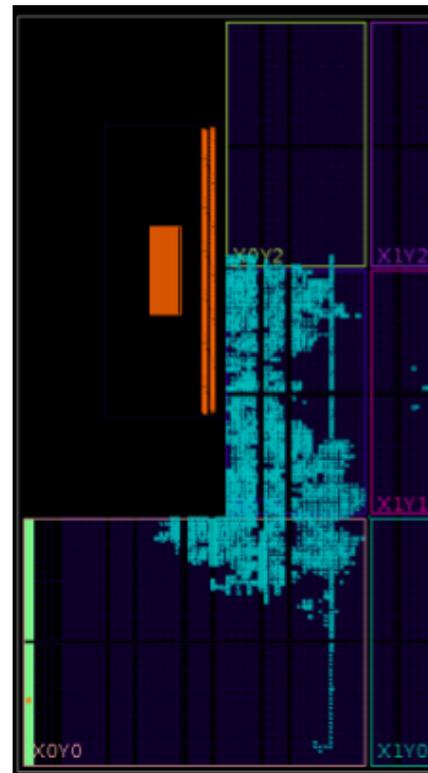


Fig. 5.10 Single tapped delay line TDC in the Zynq SoC PL after place-and-route.

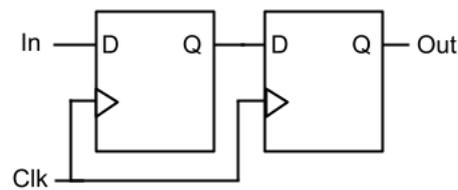


Fig. 5.11 Dual stage synchronizer

Here  $N$  is the number of latches in the chain,  $T_c = 1/f_c$  is the inverse of the receiver circuit clock frequency  $f_c$ ,  $f_d$  is the sender circuit clock frequency (hit rate). The parameters  $\tau$  and  $T_w$  are termed ‘resolution time’ and ‘metastability aperture’.

Using the Eq 5.11 we can estimate the MTBF for dual flip-flop synchronizer on the Zynq 7020 SoC with a 300 MHz clock frequency and a 100 MHz hit rate based on a 76 ps resolution time ( $\tau$ ) and a 0.55 ps aperture ( $T_w$ ) [113] (for the XC7Z020) we can estimate the MTBF to be  $2.2 \times 10^7$  years. This estimate may help explain the ability of some developers of TDCs to utilize only one layer synchronizing flip-flops [97] by considering the first set of latches in the thermo-binary decoder as elements of the synchronizer. This form of simplification, if more broadly applicable, would allow a reduction of resource utilization increasing the number of TDC channels implemented on a single chip. However in order to analyse the MTBF in this case, a rigorous analysis would need to include propagation delays and effects of intermediate combinational logic on metastability of the resulting system.

Practical considerations to increasing MTBF through layout minimizing propagation delay may act to make the implementations more robust. An example, utilized in this design is the ‘ASYNC\_REG = TRUE’ flag that groups together the synchronizer latches during place and route - reducing propagation delay. In addition to this, there is also the potential to utilize Xilinx-defined macros for implementing CDC primitives such as synchronizers, which may more fully exploit the Zynq architecture implementation and support the optimizers performing the place-and-route step of the bitstream compilation. Investigation of these optimizations is left for future work.

#### 5.4.4 Delay line decoding

The output from the delay line is a binary sequence of 1’s and 0’s encoding how far the pulse propagated through the delay line, with the synchronizer eliminating the metastability of the resulting code in the TDCs clock domain. In the case of a single pulse propagating through the delay line, the sequence of 1’s extends from the first tap in the delay line to some  $N^{th}$  tap. This sequence encodes the number  $N$ , and is termed a ‘thermo-binary’ encoding,

## Time correlated single photon counting

---

etymologically stemming from analogy to mercury thermometer readings where a increased level of mercury corresponds to a higher temperature. The thermo-binary decoding is not unique to delay lines, but arises also in ADCs where the encoding is the output from an array of comparators digitizing a analog signal.

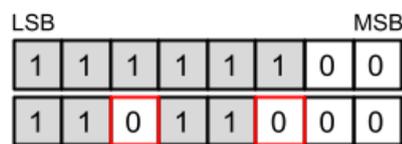


Fig. 5.12 Bubble errors and LUT decoding

For ADCs effects of noise may produce unwanted 0's on the outputs of some comparators that should have, under ideal conditions, produced a 1. This is termed a 'bubble error' (Fig.5.12) and can arise on any of the 1's in the sequence. To design an efficient decoder, converting from a potentially bubble-contaminated binary string, a number of aspects must be taken into account - propagation delay, number of elements, power consumption and error tolerance. Robustness to bubble errors is key in both ADCs and TDCs. In the latter such errors can arise from timing skew, metastability and potentially lookahead-logic effecting the resulting timestamps.

A number of techniques for designing decoders with differing levels of error correction have been analyzed [114]. In the design of the decoder for the TDC we use a ones-counter decoding scheme previously used in [115, 116] which itself has origins in implementation of high speed ADCs [117] [118], where it is known as a 'Wallace tree'. Here, instead of determining the position of the leading 1 in the encoded binary sequence the counter simply returns the total number of 1's present in the sequence. This implementation has two key benefits, global bubble error correction and simplicity of implementation.

The first part of the decoder, performed by LUTs inside CLBs is to convert a 6-input thermo-binary encoded sub-sequence into a count, of 1's. For this, three LUTs (Fig.5.13) are used, with each LUT coding for a single bit in the binary word (thanks to Michel Adamič for providing the LUT decoder values). The summation of the resulting 3-bit binary words can be structured in a tree topology, with each addition being performed between two pairwise binary words of length  $M$  and producing an output binary word of length  $M + 1$ . The tree structure of the adder is advantageous in this case depth scales logarithmically with delay line length and allows hardware pipelining of the timestamps. This latter feature, is advantageous as it has maximum throughput and forms an implicit FIFO, allowing 'pre-triggering' of the fine resolution timestamps by the TDC when an external trigger signal arrives after the hits.

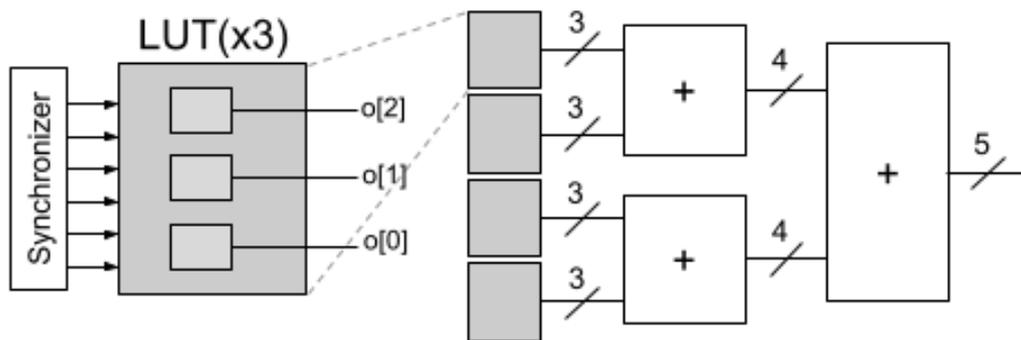


Fig. 5.13 LUT decoder and Wallace tree adder

A caveat of the above decoder description is the constraint that the number of delay elements must be a multiple of 4, the number of taps in a single CARRY4 element, and 6, the number of inputs to a 3-stage LUT decoder. With divide-and-conquer methods being prototypical recursive programs used in functional programming, we use the same recursive approach to define the pipelined adder, a feature supported by the Verilog standard.

### 5.4.5 Trigger and Hit signals

Due to the delay line value being decoded in a pipelined manner using the ones-counter, there is a need to distinguish timestamps representing real hits and empty reads. To do this a 'hit' flag is also introduced and pipelined along with each of the thermometer codes. On the output of the pipeline into the readout logic from the core, this flag is inspected and used to determine if the corresponding fine timestamp is sent further or discarded. With the empty reads discarded, the remaining hits correspond to real hits in the TDC. In many experimental measurements not all hits are of interest. One way of performing this type of filtering is through the use of 'trigger' signal. Given a particular trigger, we may define an time interval around the trigger as the 'trigger window' within which events are recorded for read out for analysis, while events outside the trigger window are discarded. Events recorded before a trigger event are said to lie in a 'pre-trigger' window, while events after a trigger are in a 'post-trigger' window. To implement triggering in the TDC, each core includes a conditioner asynchronous input, identical to the pulse routed to the TDL. The output of the conditioner is used to start a decrementing counter that sets the trigger window based on the condition that the counter value is above zero. In this way, the TDC is capable of performing post-triggering with the limit set by the value the counter is initialized to. Pre-triggering is also possible through exploiting the pipelined decoder allowing for a pre-trigger window set by the number

## Time correlated single photon counting

---

of layers in the adder tree. The aggregated ‘hit’ (*hit*) signal is generated by combining the pipeline ‘hit’ flag (*hit<sub>p</sub>*) and the trigger flag (*trigger*) using the logical and operator,  $hit = hit_p \wedge trigger$ . Continuous readout, without triggering, is achieved through setting the trigger flag high and is the mode used for benchmarking and validation of TDC performance. Future work would develop and test the full triggering functionality and integrate the TDC into an experimental measurement where this triggering can be exploited to gate the detected photons, for applications such as FLIM.

### 5.4.6 Clock domain crossing

The final consideration for developing the TDC is the length of the delay line, where minimizing the number of CARRY4 elements through reducing the TDC clock frequency is key both for improving timing performance, by reducing number of CARRY4 elements, and resource usage optimization, shrinking the size of the thermo-binary decoder pipeline. In order to achieve this, a 300 MHz clock generated by a Mixed-Mode Clock Manager (MCM) is used on the PL. A challenge of utilizing this high clock frequency for the TDC is the need to interface the TDC IP core with the broader set of AXI IP cores, typically operating at only 100 MHz for the Zynq SoC of the variant used. With each of the TDC cores embedded into a AXI core there is thus a requirement to perform clock-domain crossing (CDC) [119] [120], in both direction, from the slow (100MHz) to the fast (300MHz) and vice versa. The former, slow to fast, allows the TDC core to be controlled from the memory mapped AXI interface enabling configuration of core channel number and trigger parameters. The latter, fast to slow, is required to interface the output of the TDC core with the rest of the readout system, and enable transfer of measured timestamps into main memory, via the AXI Stream FIFOs and DMA engine. The first of these can be solved by the same technique as used to convert the asynchronous delay line tap values to synchronous thermo-binary codes - a dual stage synchronizer, a CDC technique in itself. In the transition from the fast to the slow clock domain the simple dual-synchronizer scheme is however no longer possible, as it is unable to guarantee reliable transmission as the signal. This is due to the sender (high frequency) domain being able to change multiple times before a single clock cycle of the receiver domain, giving rise to packet loss.

Multiple techniques have been developed to enabling the fast-to-slow CDC for different types of data. Transfer of a single bit can, for example, be performed using handshake between the sender and receiver domains [120]. We choose to use an asynchronous FIFO to go from the 300 MHz to the 100 MHz domain, generated using a parametrized macro supported by Xilinx, integrated into the TDC core on the Verilog level. Alternatively, the sender can stretch

the data pulse in time so that its duration exceeds the receivers clock period, with correct accounting of setup times.

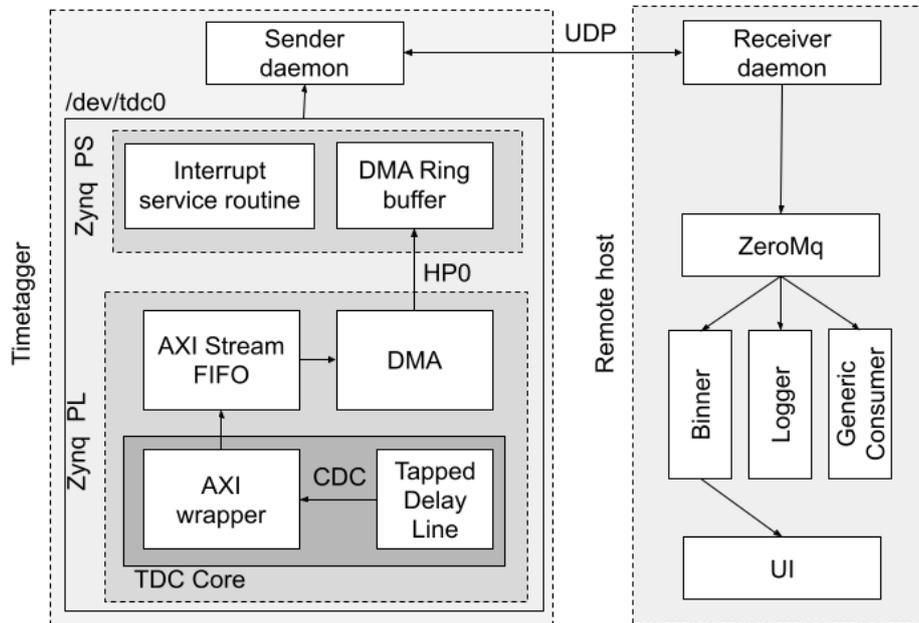


Fig. 5.14 Overview system diagram of constructed picosecond timetagging system. The low level implementation consisting of the tapped delay line feeds into the readout system and is presented to user-space applications as a character device. This is monitored by a sender daemon that reads the data stream and sends it to a receiver host that performs data processing, storage and analysis.

### 5.4.7 System architecture

The majority of the above techniques are incorporated at the Verilog level to generate the AXI interfaced TDC core. In order to perform a successful detection and acquisition of a photon stream, a significant architecture is required, encapsulating the IP core that transfers the acquired timestamps from the PL into the PS, and subsequently onto a machine capable of longer term storage of the data. A coarse schematic for this system is illustrated in Fig.5.14. The initial input pulse is timestamped at in the ‘Tapped Delay Line’ module, incorporating delay line, decoder, coarse counter. The timestamps are subsequently transferred from the fast (300 MHz) clock domain into the slow (100 MHz) domain inside the TDC core. Transfer between the TDC core and the AXI Stream FIFO is controlled through a AXI wrapper interface that sets flags associated with stream flow control, initialization, core enabling and core reset functionality. Subsequently to being ingested by the FIFO the timestamps stream converted to a memory mapped entity, through connecting to the S2MM interface of the

## Time correlated single photon counting

---

DMA IP. To simplify the process of data transfer at this stage, we ensure that a AXI Stream packet size is equal to the DMA transfer size, removing complexities associated with data alignment. Correspondingly each packet triggers an interrupt, which is handled on the kernel driver level by an interrupt service routine, indicating that the current element of the DMA ring buffer is full, requiring updating of the corresponding increments, and reprogramming of the DMA IP, performed on the PS level. Having been built into the Petalinux distribution kernel and mapped during system boot each of the five TDC cores appears in user space as a device `/dev/tdc[i]` where  $i$  indicates the index of the given TDC core. To transfer data from the Microzed board onto a remote host, capable of long-term storage of the data, another PC, a sender daemon continuously reads from the DMA ring buffer being written to by the TDC, writing the data to a UDP socket connected to by the remote host. On the receiver side this process is reverse with a receiver daemon reading the socket and transferring the data to a ZeroMQ pipe. The use of the ZeroMQ pipe allows distribution and duplication of the data stream to separate consumer processes. An example of such is the binner process, that monitors the stream counting the number of incident count in a given  $\sim 10$  ms interval, producing a low data rate output count stream that can be easily visualized by a UI. Another example is the logger process that records the received counts to a file for offline processing. In general the use of ZeroMQ allows an extensible processing pipeline to be developed on the remote host tailored to the application.

### 5.4.8 Performance evaluation

To evaluate the performance of the system I performed a series of time interval measurement tests (Fig.5.15c). Instead of using the previously described technique of simultaneous recording of photons detected by a SPAD from a pulsed laser, generating a trigger signal, the method uses a single function generator creating LVTTTL pulses with a controllable pulse rate. Each pulse is split and directed to two channels of the TDC, recording the arrival time of the pulses. This allows a measurement of both propagation delays and width of the interarrival distribution, characterizing the time resolution of the system. To decorrelate input pulses from the TDC clocks, I use a voltage controlled oscillator (VCO) to determine repetition rate of the pulses. Feeding in a pseudo-random signal into the VCO removes beating effects between the input pulses and TDC clocks, through randomly varying the pulse frequency.

To recover estimates of each channels uncertainty, without assuming equal variance between channels, one consider each channel as a node in a undirected graph (Fig.5.15b), with edges representing possible pair-wise measurement. With 5 channels, this allows up to 10 pairwise

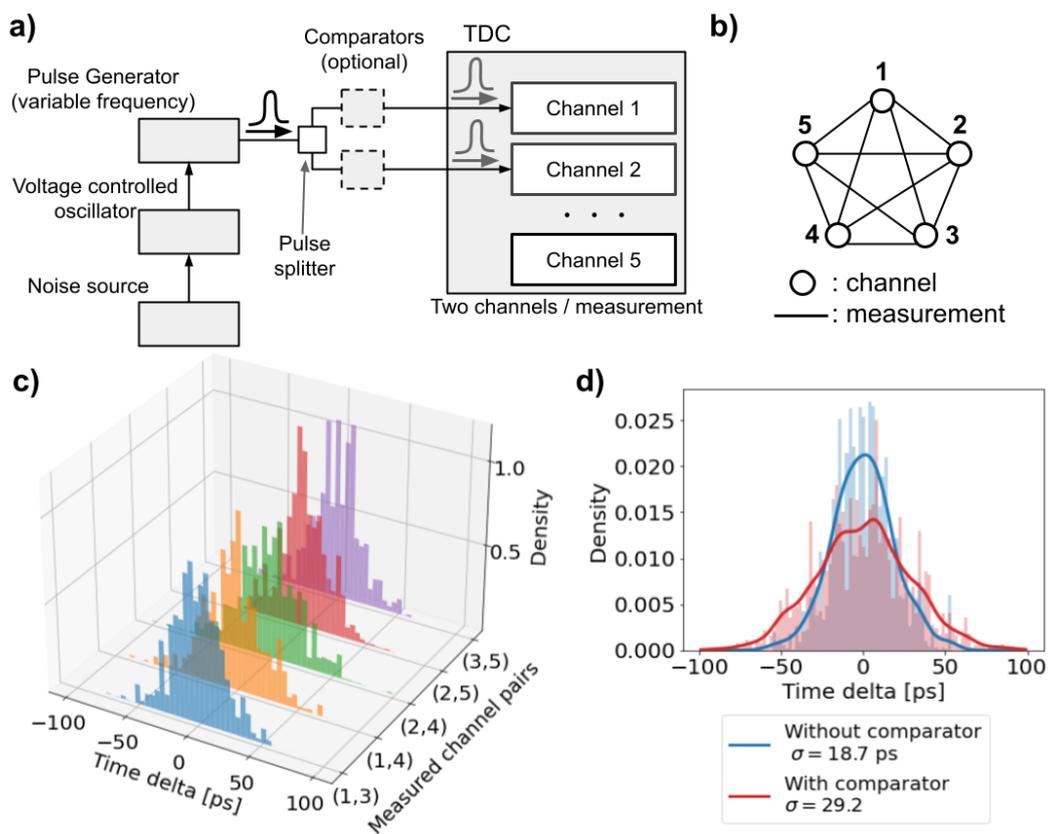


Fig. 5.15 a) Time-interval measurement configuration. A pulse generator injects the initial pulse, which is then split and passes either through comparators or directly into the TDC. The difference between the arrival times allows the calculation of both, length of delay between the two arms and the total jitter in the system. Comparators preprocessing the pulse are optional. b) Fully connected undirected graph representing possible pairwise measurements configurations. For  $n$  nodes, the number of edges is given by  $\frac{n(n-1)}{2}$ . c) Five pairwise inter-hit distribution measurements, including at least one measurement containing each channel. d) Interhit distributions showing effect the increased jitter in a measurement including comparator. Comparator board jitter standard deviation is estimated as 22.5 ps.

## Time correlated single photon counting

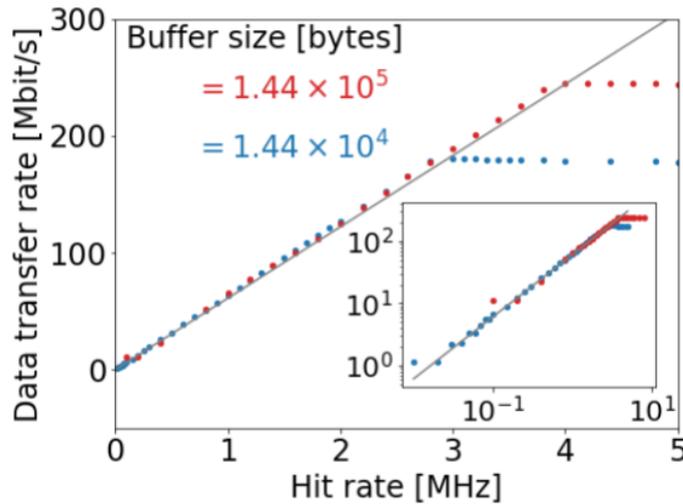


Fig. 5.16 Continuous streaming readout rate as a function of incident hit rate. Plot (red) shows capability to maintain a 3 MHz continuous readout rate without packet loss ( $\times 30$  readout rate for single SPAD). A parameter influencing maximum readout rate is the AXI Stream packet size, set to an integer multiple of the UDP frame size (1440 bytes). Lower AXI stream packet sizes (blue,  $\times 10$  UDP frame size) incur overheads and exhibit reduced performance relative to larger stream packet sizes (red,  $\times 10^2$  UDP frame size). When readout rate is exceeded, packets are lost during acquisition, resulting in asymptoting of the transfer rate.

measurements, allowing the individual channel variances to be computed using additive property of variances  $Var(X + Y) = Var(X) + Var(Y)$  to formulate this as a linear problem. In the figure we use 5 measurements. Unlike PCS this linear problem has a condition number of 3.2, solved by computing matrix inverse to give, Table.5.4. Introducing the comparator signal conditioning boards into each of the paths allows characterization of their contribution to the total system jitter, shown in Fig.5.15d, with estimated jitter contribution of 22.5 ps. Propagation delay through the comparator is estimated as 2.9 ns by considering changes in mean delay when inserting comparator into different arms of the measurement. Combined with the TDC resolution the resulting system standard deviation is upper bounded at 30 ps.

In addition to time resolution, two other key parameters are maximum readout rate and deadtime. For the developed system, shown in Fig.5.16 where the effect of varying the AXI Stream packet size is shown to have an effect on the maximum readout. Without compression or a larger UDP frame size the current system has a maximum readout rate at approximately 250 Mbit/second, corresponding to a total incident hit rate of 3.9 Mhit/second. The deadtime results for the TDC are shown in Fig.5.17, where a low frequency rate of pulse pairs is sent

Channel number	Standard Deviation [ps]
1	12.6
2	14.6
3	15.0
4	19.7
5	12.2

Table 5.4 Individual channel time resolution standard deviations computed from five pairwise interhit measurements between channels  $(i, j) = [(1, 3), (1, 4), (2, 4), (2, 5), (3, 5)]$  with standard deviation values  $\sigma_{(i,j)} = [19.58, 23.43, 24.53, 19.02, 19.35]$  ps.

into the channels of the TDC. The pulse pairs are generated using a 125 MHz function generator, and their interval is tuned until a halving of the count rate is observed. For each channel the deadtime computed by determining the smallest detected time interval, with average deadtime being  $16.5 \pm 0.5$  ns.

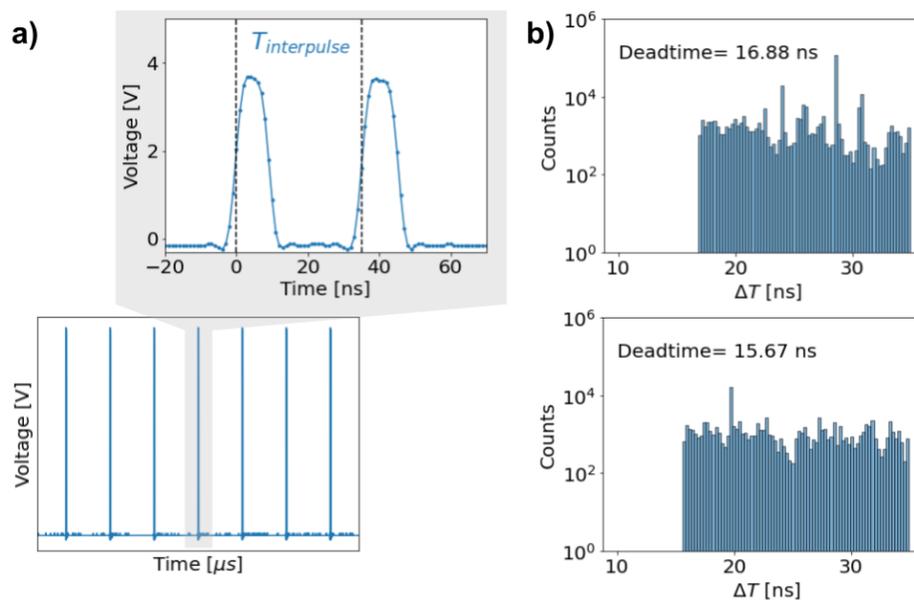


Fig. 5.17 a) Diagram of deadtime measurement principle. A pulse generator creates a kilohertz frequency train of pulse pairs, with variable interpulse separation ( $T_{interpulse}$ ) in the tens of nanosecond range. Tuning the interpulse separation allows determination of TDC deadtime without requiring high data rate. b) Examples of interpulse distribution histogram generated by the TDC showing an upper bound on the deadtime to be of order 16 ns.

## 5.5 Signal conditioning board

In order to achieve the high time resolutions of the SPADs we must utilize the high speed NIM output, having a specified resolution of 30 ps. To do so we must be able to convert the NIM signal, whose properties are specified in Table 5.5 into a positive polarity pulse that can be interfaced to the FMC-DIO front-end accepting pulses with a configurable threshold between  $V_{in} = 1V$  and  $V_{in} = 5V$ .

Output driver current (mA)		Receiver Input Voltage
Logic 1	-14 to -18	-0.6 (max) to -1.8 (min)
Logic 0	-1.0 to +1.0	-0.2 (min) to +1.0 (max)

Table 5.5 Fast, negative polarity , NIM pulse standard

To achieve this, we designed and constructed a high speed comparator module, consisting of two boards, a power and a comparator (Fig.5.18). The former accepts a 5-12V DC voltage supply from mains adapter to two high stability  $\pm 3.3V$  supply voltages via a low drop out (LDO) regulator and inverter. The comparator uses these as references in two high speed LVDS data line receiver, which trigger a positive polarity logic pulse when an input signal goes outside either lower ( $V_{lb} = -0.8V$ ) or upper bounds ( $V_{ub} = 1.8V$ ). The threshold voltages are generated using a combination of resistor dividers and high stability voltage references ( $V_{ref} = 2.048V$ , accuracy= $\pm 0.1\%$ ) that minimize effects from supply power variation and thermal drift. The outputs from each comparator are combined in a high speed OR-gate to generate a trigger pulse controlling a SN74 family line driver that producing the final output.

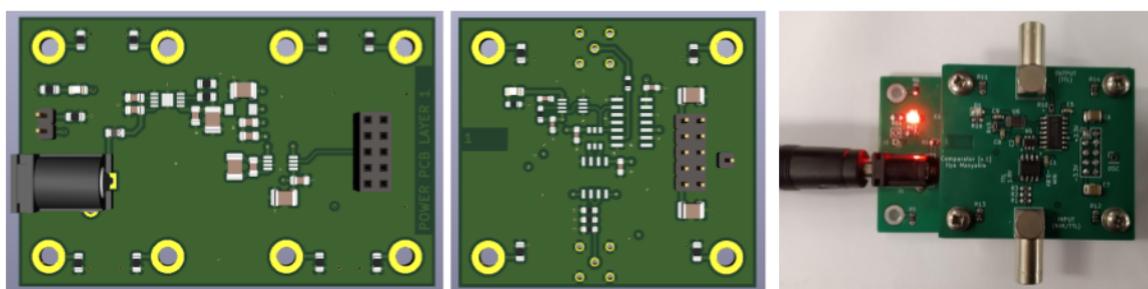


Fig. 5.18 Models of submodule boards and assembled image of full comparator system

Motivation for creating both positive and negative polarity bounds is the less widespread use of NIM in signal generating equipment, with TTL being significantly more widespread. By incorporating the ability to trigger on a positive polarity pulse we can easily test the systems with a function generator without requiring to construct a NIM pulser. The modularity

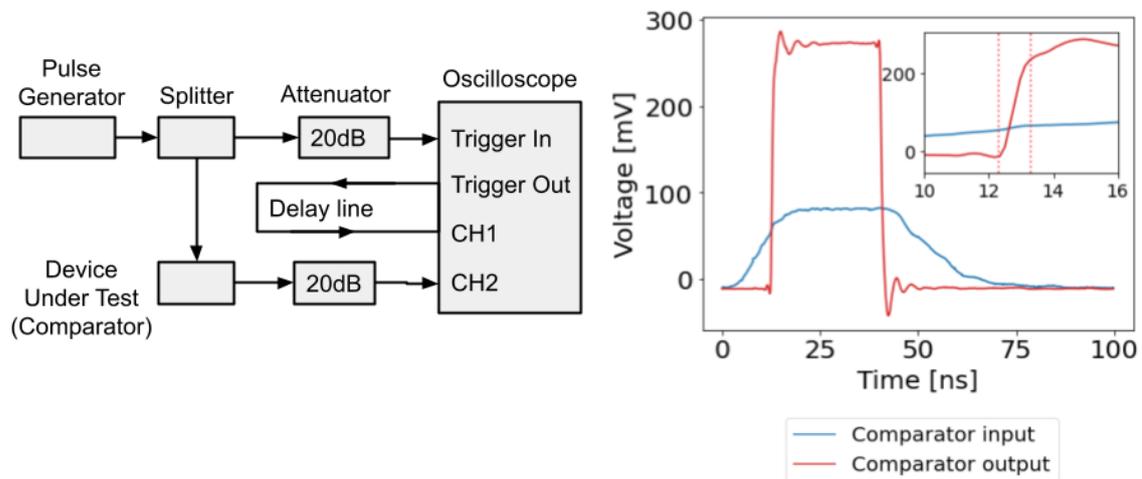


Fig. 5.19 a) Diagram of measurement configuration for comparator evaluation b) Measurement of comparator and input and output pulses, with comparator output exhibiting 1 ns rise time, corresponding to a 350 MHz bandwidth.

enables independent upgrades to each of the board designs without affecting the other and allows the re-use of the power board for different applications.

Key parameters for the testing of this module are the signal rise time and jitter introduced into the measurement. The former, due to the high bandwidth requirements is non-trivial to evaluate using a conventional digital-storage-oscilloscope (DSO) due to typical bandwidth limitations of such systems to 100-200 MHz. Instead, periodic pulse train on the input and output of the module can be measured using a ‘sampling oscilloscope, which exploits the periodicity of the measured signal to measure a high bandwidth signal through successive measurements of the waveform. Using this method the sampling oscilloscope can greatly exceed the performance of a DSO, in our case allowing bandwidths of up to 11 GHz. The testing configuration for the board is shown in Fig.5.19 where the input signal from a pulse generator is split, and used to both trigger the oscilloscope acquisition as well as the comparator. Both signals are attenuated to bring them into the suitable voltage range for the sampling oscilloscope and recorded. The measured pulse on the output of the comparator showing a sub-nanosecond rise time (orange) along with the corresponding input pulse (blue) are shown. Estimation of the bandwidth of the comparator using  $f_{bandwidth} = 0.35/t_{rise}$  [90], for a rise time of 1 ns gives an estimate 350MHz, with the rise time being measured between the the 10%-90% signal amplitude thresholds.

### 5.6 Future work

In the context of this thesis next step in development of the TDC system is the integration of photon counting into the transmission microscope described in Chapter 7, where the TDC would be applied to recording transmission intensity variation of a nanoaperture. This would enable multiple measurement modalities, including fluorescence correlation and lifetime measurements, enabling temperature sensing near the nanoaperture [78]. Increased time resolution would also allow high speed particle dynamics inside the aperture to be studied, extending the work of [12], where intensity correlation decay times applied to classify proteins. Motivated by depolarized PCS, this form of measurement may allow extraction of rotational coefficients of particles in the aperture.

Beyond this work, the reconfigurable nature of the system, consisting of software defined SoC IP blocks and standardized FMC front-end interfaces, enable development of a platform for constructing high performance scientific instrumentation. In this platform, reuse of shared data transfer mechanisms and related software tools, as well as low-level electronic reconfigurability, may allow construction devices like high speed cameras and spectrometers, potentiostats or optoelectronic device controllers whose performance requirements cannot be accommodated by the current electronic modules like Raspberry Pi or Arduino. Developments like this are already being fielded, and thus we are likely to see increased adoption of such software-reconfigurable instrumentation in the near future where the developed TDC may be useful for nanosecond synchronization between multiple devices - a challenge having multiple applications spanning domains from aerospace to finance.

### 5.7 Summary

In this chapter I described a time-to-digital converter photon timetagging system and high speed comparator board developed for performing high temporal resolution photon counting experiments. The described system is capable of simultaneously recording five independent channels continuously, with hit rates up to 3 million photons per second. The time resolution of each channel lies in the tens of picoseconds range. The developed system provides a cost-effective alternative to commercial instrumentation and lowers the entry barriers to constructing high speed photon counting experiments.

# Chapter 6

## Nanoaperture fabrication

### 6.1 Introduction

While broadly successful and commonly employed, scattering techniques such as PCS or NTA have nevertheless left room for the development of complementary techniques for probing particle properties. The length scale between 1 nm and 100 nm is particularly important, encompassing many biologically relevant structures like proteins, viruses, and DNA. Here however optical scattering techniques such as PCS and NTA are limited by the volume ( $\propto V^2$ ) scaling of the scattering cross-section, requiring either high sample purity (PCS) or being limited in resolution and requiring fluorescent labels (NTA). This limitation, combined with the practical applications often involving heterogeneous samples containing multiple sub-populations, motivates the search for techniques capable of resolving such populations and characterizing their properties.

Two key considerations for any technique performing single-particle characterization are the control parameters and extractable signal features. The former defines a design space of possible experiments available to customize the technique for a given application. The latter determines the robustness of the approach and the possible information that can be extracted from a measurement. For both PCS and NTA, the control parameters manifest primarily in properties of the illuminating beam, such as polarization, wavelength, beam shape. At the same time, hydrodynamic radius, computed from the recorded stochastic intensity (PCS) and position (NTA) data, is the main feature extracted from such measurements. Instead of ‘direct’ measurement of particle properties, one can instead consider a sensor-based approach where the response of a sensor device as it interacts with a nanoparticle is measured. By introducing the sensor element, we can expand the design parameter space in an attempt to

## Nanoaperture fabrication

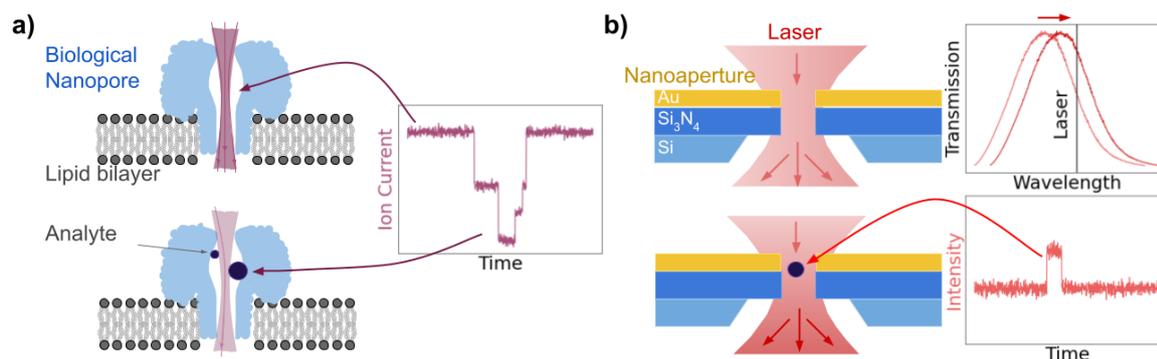


Fig. 6.1 a) Ion current sensing of analyte translocation through biological nanopore. b) Laser transmission modulation from analyte refractive index change in solid state nanopore, a nanoaperture.s

circumvent issues associated with scattering measurements, including volume scaling and diffraction-limited resolution of such techniques.

To motivate the nanoaperture, we first consider a close analogue of such a system, a biological pore. Such ‘nanopore’ systems consist of a protein embedded inside a lipid membrane, separating two fluid chambers. The protein, such as alpha-hemolysin, forms a channel capable of passing a small (pA-nA) ionic current between the two chambers when an external voltage is applied. Molecules such as DNA or RNA flowing through this channel produce ‘current blockades,’ modulations in the electronically detected current. The ability to map between characteristics of current blockades and properties of the molecules translocating through the pores has enabled such systems to be applied to high throughput DNA and RNA sequencing, instrumental to modern biological sciences. The success of nanopore sensing has spurred efforts to extend this technique to include optical probing methods and allow the nanopores to be deployed in a broader range of environmental conditions by using ‘solid-state’ instead of biological pores. Nanoapertures are an instance of such a technique. A particle translocating through the nanoaperture creates a shift in the transmission spectrum of the structure, modulating the intensity of the beam passing through it.

In the remainder of the introduction we will cover nanoapertures and challenges associated with scaling their detection rates to enable measurement of a large number of nanoparticles. This will be followed by an overview of methods employed to fabricate and characterize the nanoapertures, the enclosing fluidic chip and the assembly of the resulting sample.

### 6.1.1 Nanoapertures

The localized nature of the nanoaperture combined with the ability to probe its properties optically enables multiple measurement modalities beyond transmission intensity modulation. One such example is fluorescence detection [121], where the  $100\text{nm}^3$  volume of the nanoaperture reduces the detection volume by  $\times 10^3$ , relative to a conventional focused laser spot, while simultaneously enhancing the fluorescence intensity up to  $\times 20$  [122, 123], enabling the of probing single-molecule interactions. The confinement of electric fields near the aperture also extends the optical trapping range through ‘self-induced back action’ (SIBA) [9, 124], allowing one to trap single proteins and particles [12, 10] down to the 10-20 nm range. The high electric fields created by the nanoaperture also enhance Raman [125] scattering signals, thus facilitating local chemical probing and discrimination of DNA bases [126]. This localized nature of measurement combined with multiple possible signals [127] and features measurable motivates the use of nanoapertures for nanoparticle characterization.

The ability to tailor the optical properties, such as spectral response, polarization selectivity [128, 129], field distribution, by controlling the nanoapertures size, geometry and material is a crucial benefit of solid-state systems over the biological pores used in nanopore sequencing technology. The range of parameters of a nanoaperture that can be modified produces an ample design space we can explore to engineer the nanoapertures for a given nanoparticle characterization task. This design space can be explored using a range of ‘top-down’ and ‘bottom-up’ techniques for fabricating nanoapertures, including focus-ion-beam (FIB) [130] milling, electron-beam lithography (EBL) [131], controlled dielectric breakdown (CDB) [132, 133], and colloidal lithographies [134, 135] [136]. In addition to controlling geometry and material, one can modify the surface properties [137] of these systems to increase the stability of these structures in complex environments.

A key challenge for ion-current measurements in solid-state nanopores is the increased noise level compared to biological nanopores [138, 139]. While similar challenges for the low frequency regime are also present with optical techniques relying on intensity modulation, these have been shown to be more robust to high frequency noise, exhibiting  $f^0$  scaling compared to the  $f^1$  scaling for ion-current measurements [140].

This favourable scaling, in combination with the above mentioned detection modalities makes nanoapertures a system worth considering for nanoparticle characterization if the sensors can be operated in a ‘high throughput’ mode, observing a large number of nanoparticle translocation events.

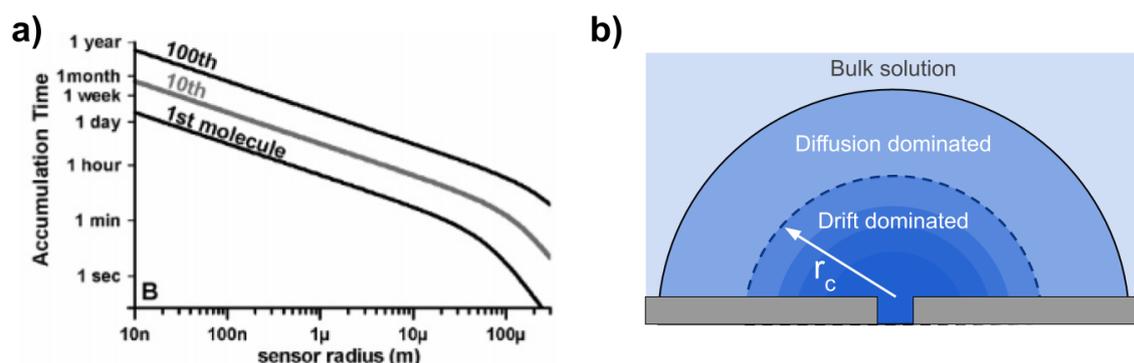


Fig. 6.2 a) Time required for hemispherical sensor to accumulate a fixed number of analyte molecules with a 1 fM concentration solution (taken from [141]). Time required scales linearly with radius. b) Nanopore hemispherical capture/access region where drift dominates diffusion processes.

### 6.1.2 Translocation control

In order to gather significant statistics about the distribution of nanoparticles in the sample one must consider the time required for a single nanoparticle to enter the region of the nanoaperture, resulting in a detection. The effect of nanoaperture size on time required for an analyte to diffuse into the region of the nanoaperture can be related to a similar problem considered in [141]. In this work a hemispherical sensor was shown to require an ‘accumulation time’ growing linearly with decreasing radius  $r_c$  of the sensor. The authors show that in order for a sensor with  $r_c \leq 1 \mu\text{m}$  to accumulate a single analyte particle, of radius  $\sim 2 \text{ nm}$ , present in solution at femtomolar (fM) concentrations, the required accumulation time ranges from 1 hour, for  $r_c \approx 1 \mu\text{m}$  to  $\geq 1 \text{ day}$  for  $r_c \approx 10 \text{ nm}$ .

For DNA translocation of nanopores the hemispherical model was also considered in [142] where the capture rate  $R$  was given as a function of diffusion constant  $D$ , molar concentration  $C$  and nanopore ‘capture radius’ [143], of the hemisphere,  $r$ :

$$R = 2\pi CD r_c$$

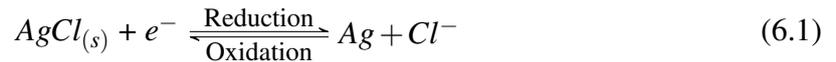
Unlike the case of a fixed dimension hemispherical sensor, in the case of the nanopore the capture radius can be understood as a radial distance around the nanopore where the drift regime dominates the diffusive motion. Multiple definitions of capture radius exist [143], such as:

$$r_c = \frac{\mu_{ep}\Delta V}{D}r_e$$

This form exhibits a linear dependence on the potential applied across the nanopore membrane  $\Delta V$  along with characteristic size of the nanopore  $r_e = r_{pore}/(2l_{pore}/r_{pore} + \pi)$ , where  $r_{pore}$  and  $l_{pore}$  are the radius and length of the nanopore. The scale factor of  $\mu_{ep}/D$  encodes the electrophoretic mobility and diffusion constant of the captured analyte, such as DNA or nanoparticle. The above form shows that, unlike the fixed size hemispherical sensors, nanopore capture radii can be tuned through application of a bias potential across the membrane containing the pore.

With event rates for nanoaperture tweezers on glass reported to lie in the  $10^{-1}$  -  $10^{-4}$  Hz region [144] the need to increase particle translocation event rates through actively driving particles towards the nanoapertures is clear. Achieving this in practice however is no easy feat as it requires detailed understanding, not only of nanoaperture and particle properties, but also the properties of the electrolyte, electrodes and interactions between the different sets of parameters.

In the case of electrolyte key parameters are ion concentrations and mobilities, while for electrodes parameters such as their effective surface area, type and frequency response are crucial. A common electrode of choice is Ag/AgCl (anode and cathode) combined with a chlorine-containing ( $Cl^-$ ) electrolyte solution such as NaCl or KCl capable of driving the reduction-oxidation (redox) reaction in Eq. 6.1.



In nanoaperture literature the range of electrolyte concentrations spans from 10 mM [11] to 150mM [145]. A similar but broader range of concentrations nanopore literature spans from the 1 mM [146] through to the 3 M [147], the latter likely in part due to the reduction electrode interfacial capacitance ( $C_{DL}$ ) lowering the effective RC time constant and allowing the authors of [147] to achieve a 10 MHz measurement bandwidth.

## 6.2 Device Fabrication

Before performing measurement of nanoparticle translocation through a nanoaperture, such a system must be built. In our case this consisted of fabricating the nanoapertures in a commercially available silicon nitride ( $Si_3N_4$ , SiNx) membrane, reducing deposition of suitable

## Nanoaperture fabrication

metallic layers. To house the silicon frame containing the SiN<sub>x</sub> membrane we designed and manufactured a fluidic chip with both electrical access to the reservoirs separated by the membrane, and optical access to the nanoaperture. Following this the transmission properties were measured by a constructed transmission microscope, described in the following chapter.

### 6.2.1 Silicon Nitride Membranes

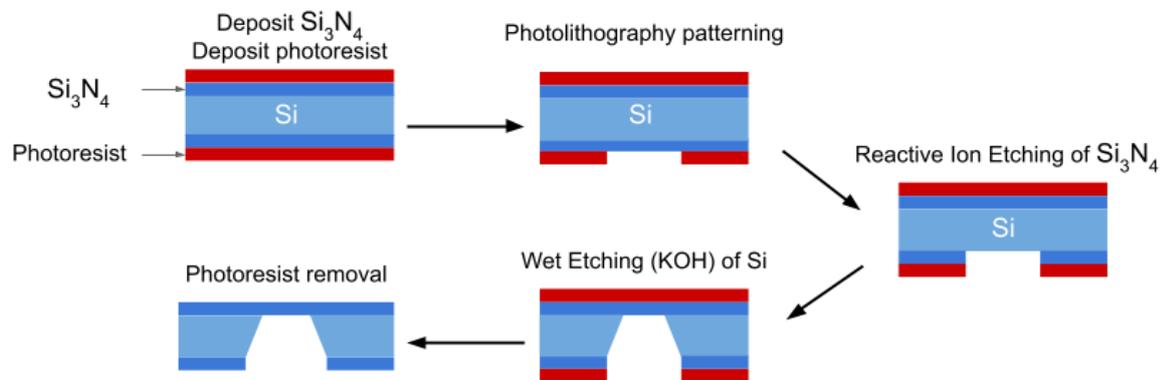


Fig. 6.3 Steps in creation of thin free standing  $Si_3N_4$  membrane [148].

In order to manufacture solid state nanopores one requires a thin membrane which can be micromachined to create the nanopores. A versatile system for this are silicon nitride 'SiN<sub>x</sub>', which can be fabricated with membrane thicknesses from 20 nm to 1 μm. These have been widely applied in fields ranging from tests of quantum mechanics to nanopores, microfluidics and microelectronics.

Procedures for SiN<sub>x</sub> membrane fabrication have been extensively studied in the past 30 years [149, 150]. The first step in the process is growth of the silicon nitride layer using chemical vapour deposition (CVD) techniques (Plasma-Enhanced, PVCVD or Low-Pressure LPCVD [150]) on a silicon wafer. Control of deposition rate and measurement of thickness allows thickness tolerances down to 1%, which optically can be characterized by the transmission spectrum of the resulting membranes [151]. Following deposition photolithography is used to define, with a positive photoresist the etch geometry which, for a given wafer thickness, typically in the 100 μm range, determines the geometry of the resulting window. The removal of the photoresist followed by reactive ion etching (RIE) steps (Fig.6.3) exposes the silicon nitride layer under the photoresist free area. Subsequently, wet etching using potassium hydroxide (KOH) [148] creates free-standing SiN<sub>x</sub> membranes, relying on anisotropic etching of the silicon wafer. The etch rate is significantly faster for the <100> crystal orientation

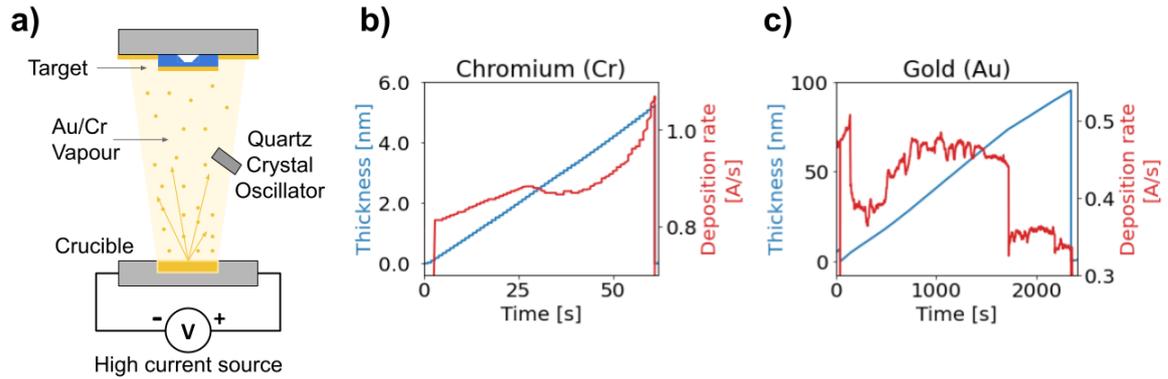


Fig. 6.4 a) Thermal Physical Vapour Deposition (PVD) system. b) Chromium adhesion layer deposition, 5 nm thickness. c) Gold layer deposition, 95 nm thickness.

compared to the  $\langle 111 \rangle$  direction giving the ability to etch through the wafer in a controlled way, without etching the SiNx.

## 6.2.2 Metal layers

Next step in the fabrication is the deposition of metallic layers, into which the nanoapertures are milled, Achieved through thermal evaporation or Physical Vapour Deposition (PVD) of metal layers onto the SiNx surface. For this we place a silicon nitride target into a vacuum chamber with pressures down to  $1 \times 10^{-6}$  mBar. Two crucibles containing Cr and Au are resistively heated by passing a high ampere current. Upon reaching the boiling point of the metal, the vapour produced coats the deposition target (Fig.6.4). Deposition rate of each layer is monitored by measuring the shift in resonance frequency of a quartz crystal microbalance (QCM) [152]. As the deposited layer thickness increases so does the mass of the crystal, in turn reducing the resonance frequency. Combined with knowledge of the materials atomic mass allows for estimation of layer thickness. The rate of deposition (Fig.6.4)  $\sim 1A/s$  (Cr) and  $\sim 0.5A/s$  (Au) determines uniformity, roughness and stress of the resulting surface [153]. The latter of these may result in mechanical deflection of the membrane [154] as a consequence of the difference in films deposited on either side of the membrane. With a suitable device such as a Michelson or Mirau interferometer, this can be measured quantitatively.

In order to prevent delamination and compensate the lattice mismatch between the Au and SiNx surfaces, we apply an intermediate ‘adhesion layer’ of Cr. The adhesion layer, initially introduced in microelectronics for improving uniformity and reducing delamination of Au layer, is deposited between the SiNx and the Au layers. The properties of the adhesion to

## Nanoaperture fabrication

---

SiN<sub>x</sub> substrates has been studied in [155] where mechanical keying, from surface roughness, and oxygen pre-annealing were found to improve Cr-SiN<sub>x</sub> adhesion. The latter forms silicon oxynitride reacting strongly with Cr, increasing bond strength. Such preprocessing is not necessary for all adhesion layers, like Ti, where an interface TiN layer forms. Optical properties of various adhesion layers on gold nanoaperture arrays were studied in [156] where both Cr and Ti adhesion layers were shown to broaden the spectral response in the > 700nm wavelength region relative to both no adhesion layer or a etched layer. For the case of a single pore, the variation in thickness and composition of adhesion layer is found to affect fluorescence intensity, with a TiO<sub>2</sub> achieving the best performance (~ 25×), while 10nm Ti and 5nm Cr pores achieve a ~ 7× intensity enhancement relative to solution [123]. Local temperature surrounding the single nanoaperture is also found to increase with the use of the adhesion layer [157] and thus this is a key design parameter for nanoaperture systems.

As the nanoapertures in this work will be immersed in an electrolyte solution, care must also be taken in considering the interaction of the adhesion layer with the electrochemical environment. Electrochemical studies using QCMs have shown that effects of the Cr/Ti layers are non-trivial in solutions where, upon dissolution of the adhesion metal, the Au surface roughness increases. With the adhesion layer applied, the Au layer is subsequently deposited.

In this work we use a 5 nm Cr adhesion layer on top of which we deposit 95 or 195 nm of Au, noting that the ‘tooling factor’ (ratio between thickness deposited on target and on QCM) of our instrument is assumed to equal 1. We note that considering the above, the properties of the adhesion layer such as material thickness form a set of parameters that may be tuned to optimize the nanoaperture properties. Potential effects arising from interdiffusion between the adhesion layer and gold may also play a role in the nanoaperture optical response [158].

More generally, the number of layers, their composition and thickness forms a set of design space that may be explored to generate the required optical response, such as extraordinary optical transmission [159]. While not depicted in diagrams for simplicity we also apply a 5 nm Cr layer to the bottom of the membranes, to minimize charging effects during micromachining.

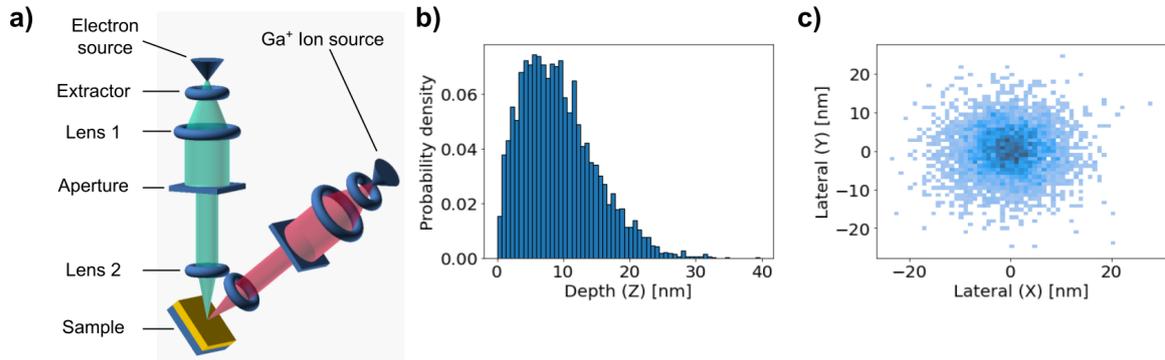


Fig. 6.5 a) Diagram of dual beam FIB/SEM configuration. b) Simulated Gallium ( $Ga^+$ ) ion range c) Simulated straddles for 30kV acceleration voltage into semi-infinite Au.

### 6.2.3 Microfabrication

Following deposition, nanoapertures in the metal coated SiNx membrane are made by milling through the Au/Cr coated SiNx membrane using a Helios Nanolab 650 dual beam system. This system combines an electron and a  $Ga$  ion beam source (Fig.6.5). The principle of ion beam milling is based on ion sputtering, where a  $Ga^+$  ion flux, evaporated from a liquid ion source - a small droplet of gallium, is accelerated towards a target by an electric field. The high energy ions undergo a cascade of collisions with atoms in the target surface, resulting in recoil of atoms from the surface, ejecting some of the surface atoms, resulting in material removal - ‘milling’. Such techniques have been extensively used [160] to fabricate nanostructures at the  $\sim 10$  nm (for  $Ga^+$  FIB) resolution, while newer variants utilizing  $He^+$  ions have manufactured features smaller still.

Key parameters in the process of FIB milling are the ion acceleration voltage ( $V_{Acc}$ ) and beam current ( $I_{beam}$ ) with the former determining the kinetic energy of the  $Ga^+$  ions while the latter measures total ion flux. Higher resolution is attained with increasing acceleration voltages and decreasing beam currents. The ion kinetic energy determines penetration depth (stopping power) of the  $Ga^+$  ions in the substrate. The distribution of penetration depths can be simulated using Monte-Carlo methods using the SRIM package [161]. The depth distribution of  $Ga^+$  ions into a semi-infinite Au layer, for a 30kV acceleration voltage, has mean (range) and standard deviation (straddles) of  $9.3 \pm 5.6$  nm. As this layer is near the surface, and comparable to the skin depth of Au, the ion implantation of  $Ga^+$  during milling will affect optical response of nanoapertures due to proximity of  $Ga^+$  ions to the surface, reducing transmission for  $Ti_2O$  [162],  $Al$  and  $Mg$  [163] structures. Comparatively,  $He^+$  ion with the same acceleration voltage and same geometry also simulated using SRIM have a range and straddles of  $85 \pm 40$  nm - leading to improved plasmonic response [164]. While

## Nanoaperture fabrication

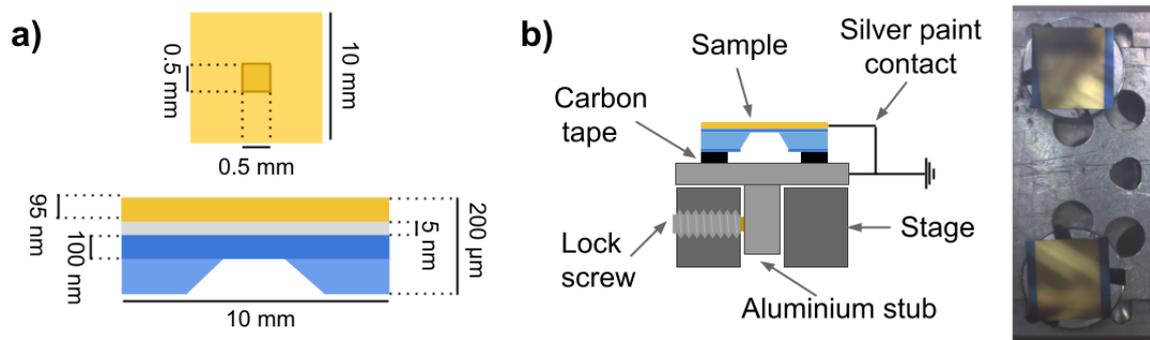


Fig. 6.6 a) Schematic of metal coated SiN<sub>x</sub> membrane b) Mounting of SiN<sub>x</sub> membrane inside Helios system

acceleration voltage is typically maximized, improving resolution, a smaller beam current that balances resolution with milling time must be chosen. For nanoapertures this is typically in the range 5-15 pA [165]. Control of beam current is achieved through the use of differently sized mechanical apertures in the ion beam path, whose relative offsets in the beam path produce different focusing conditions for the beam, requiring realignment for each current setting, thus a single beam current is typically used for milling. The mechanical nature of the apertures means that milling current is tuneable in discrete steps, and is prone to increase over time as the high energy ions widen it. In our work we use a 7.7 pA current, as this is in the specified range above.

Two additional parameters in milling are the beam 'dwell time' ( $\tau$ ) and 'volume per dose' ( $V_D$ ) (sputter rate). The former determines the duration of the beam at a single pixel of the pattern, while the latter is a material dependent quantity (with units  $\mu^3/nC$ ) quantifying the rate of material removal [166] per charge, which determines the total milling time  $t$ . When specifying pattern parameters  $V_D$  is included indirectly through the 'z-size' pattern parameter calculating depth of the pattern.

Mechanical system alignment also plays a critical role in quality of the obtained results. A key location, termed the 'eucentric position' is defined as the intersection of the axis of rotation of the tilt stage, mounting the sample, with the axes of the two beams. When a sample is aligned in the eucentric position, tilting of the sample does not induce defocusing, allowing one to switch between imaging and milling of the sample at normal incidence via tilting sample by  $52^\circ$  without needing to realign each of the beams to focus on the sample. The maximum tilt angle is set by the arrangement of electron and ion beam columns (Fig.6.5). The system alignment proceeds by an initial coarse focusing of the electron beam on an edge of the sample (tilt= $0^\circ$ , perpendicular to the electron beam) at low magnifications, followed by

a 'Link Z to FWD', linking the z coordinate of the stage to the Free Working Distance (FWD) of the electron objective. This is followed by fine focusing of the electron beam at higher magnifications with relinking and correction of beam astigmatism along with zeroing of beam offsets. Next, the sample stage is gradually tilted up to a maximum of  $52^\circ$  (perpendicular to the ion beam), in steps of  $5-10^\circ$ . During this a surface feature is kept centred on the image plane by z-axis translation of the stage to bring the stage to the eucentric point, achieved when tilt of the sample from  $0-52^\circ$  does not create vertical translation of the feature. At this point the linked z-height of the stage should be approximately 4.1 mm. With the sample perpendicular to the ion beam (tilt= $52^\circ$ ) the ion beam condition is adjusted to bring the sample into focus, removing astigmatism and zeroing beam offsets (for a fixed value of beam current). Next, electron beam offset used to finish the co-alignment of the beams to image the same feature on the sample surface (we find that in our system, this is only possible for 2kV electron beam acceleration voltage, while for the 30kV acceleration voltage the maximum beam offset is not sufficient compensate the beam offsets and bring beams into co-alignment). Once co-aligned on the edge, the sample is translated near the membrane where test nanoapertures, not passing through the membrane, are milled and imaged to fine tune the beam conditions. Finally, we move onto the edge of the membrane and mill the through-membrane nanoapertures using a predefined pattern. Given the simplicity of the nanoaperture patterns milling takes less than one minute per aperture, potentially enabling tens to hundreds to be milled without significant time investment - given the process is optimized and the machine is tuned.

As shown by Fürjes [167] in order to achieve precise control over geometry of the milled shape, in our case a circular aperture, charging effects of the dielectric membrane, leading to beam defocusing, during ion bombardment must be minimized. We achieve this by depositing a metallized layer on the side facing the beam and apply silver paint, forming a connection to the grounded stage (Fig.6.5b), with the bottom of the SiNx frame attached to the stage via black carbon tape, preventing charge build up. While this allows for minimization of charge build up the displayed configuration does not allow active measurement of perforation time of the membrane, preventing the development of a quantitative model of beam and SiNx parameters required to optimally tune properties of the resulting nanoapertures. Such a configuration has however been demonstrated in [167] where transmitted ions impact a metal foil producing secondary electrons that are then detected. This sample holder allowed a quantitative model based on ion beam current, layer composition and thickness to be developed, thus future work in fabrication of nanopores in our lab would require construction of such as sample holder opening the possibility to develop a similar model. Further

## Nanoaperture fabrication

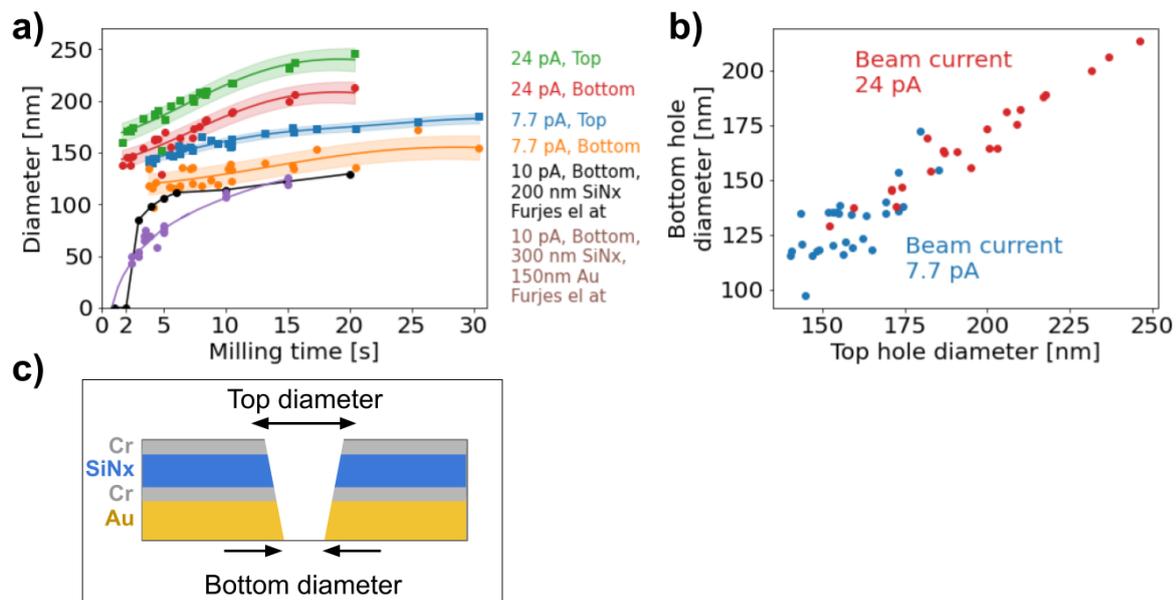


Fig. 6.7 a) Nanoaperture diameter evolution as a function of milling time. The incident ion beam creates a conically shaped aperture with differing top and bottom diameters, determined by membrane thickness and composition. b) Relation between top and bottom diameters for 7.7 pA and 24 pA ion beam currents. c) Schematic of nanoaperture in SiNx membrane showing the tapering of the aperture cross-section left by the ion beam milling.

extensions may then generalize to measure penetration times for milling at non-perpendicular angles of ion incidence.

## Electron microscopy

While the ion beam is used primarily for milling the sample, the electron beam is used for imaging, both the top and back sides of the membrane, to verify the creation of the nanoapertures. Prior to milling, to align the system without knowing exact offsets of the membrane from the frame edges we use a 30 kV electron acceleration voltage. This electron beam is used to determine the location of the membrane,  $0.5 \times 0.5$  mm, within the larger frame,  $10 \times 10$  mm (Fig. 6.6a). The mechanism behind this is similar to the stopping power and penetration depth calculations of  $Ga^+$  ions in Au. At higher acceleration voltages of 30 kV the penetration depth of  $e^-$  in a semi-infinite Au film can be estimated as  $\sim 200$  nm [168] [169] [170]. As the image of the target surface is formed by recording secondary electrons emission in some ‘extraction depth’ the higher acceleration voltages produce a contrast between regions of different thickness, the 200  $\mu$ m Si frame and the 200 – 300 nm metal coated SiNx membrane. With the extraction depth of secondary electrons in the SiNx membrane limited by membrane thickness this region appears darker (Fig. 6.5) than

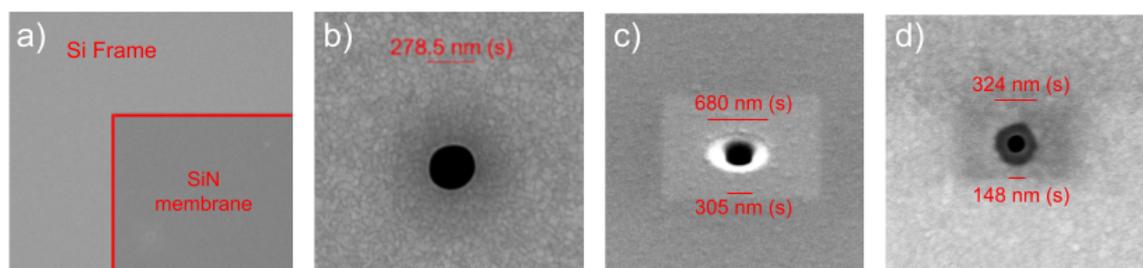


Fig. 6.8 a) SEM image of SiNx membrane with 30 kV acceleration voltage showing contrast between membrane and frame. b) Front Au coated surface of SiNx aperture back-milled through SiNx. c) Back of SiNx membrane milled, coated with a 5 nm Cr layer minimizing charging. d) Front of Au-first milled SiNx membrane exhibiting step-like profile due to higher rate material removal for Au vs SiNx.

the surrounding bright Si frame, allowing easy identification of membrane edges. Once identified, the coordinates of the membrane corners are recorded and stored in the control software to allow easy ‘homing’.

#### 6.2.4 Surface treatment

Contaminants adsorbing to a surface during sample storage (both in ambient atmospheric conditions as well as vacuum [171]) or implanted during nanoaperture fabrication ( $Ga^+$  and SEM (carbon) [172, 173]) may modify surface properties of the gold coated membranes, including reduction of hydrophilicity, as well as optical response of the nanoapertures. Lack of control of surface properties may lead to reduced reproducibility as well as having a fundamental impact on fluid flows near membrane surface. Effects such as hydrophobicity induced by contaminants may increase rates and durations of nanobubble formation at the membrane-liquid interface [174] which may modify both optical measurements and nanoparticle dynamics as well as affect the electrochemical reactions. Control of surface properties is therefore key to achieving reliable operation of the nanoaperture systems.

Understanding behaviour of adsorbed contaminants at surfaces and their removal is a large varied field with a multitude of techniques including chemical treatment, plasma cleaning and annealing and thus development of a robust, quantitative, protocol is challenging. However, some simple steps can be applied during assembly of the membranes into the microfluidics that can qualitatively help to reduce effects of surface contamination and remove impurities added by the fabrication process.

The first set of the techniques is low-temperature plasma cleaning, where the surface is exposed mixtures of ionized gases created via application of a electric field to a neutral

## Nanoaperture fabrication

---

gas [175]. The ions interact with the surface either through chemical reaction or physical ablation. Oxygen plasma, falling into the former category, removes organic compounds and carbon [172], through reactions of oxygen radicals with carbon producing  $H_2O$ ,  $CO$  and  $CO_2$  which desorb from the surface due to their high vapour pressure and are removed by the vacuum pump [173]. Argon ( $Ar$ ) plasma instead removes contaminants through sputtering of the surface.

Aside from impurity removal the process of plasma cleaning can modify the chemical composition of the surface as well as physical properties. In the case of oxygen plasma treatment, a gold oxide ( $Au_2O_3$ ) layer is formed on the surface, while for argon the physical sputtering acts to increase surface roughness at a faster rate [176]. Key parameters in the treatment are ion energy, gas flow rate, exposure time and gas ratios and chamber pressure. The latter parameter, gas ratios, may result in super-additive etch-rate due to synergy between ion etching mechanisms [177].

In our work we base the exposure time on parameters determined in [178] which considered  $O_2$  and  $Ar$  plasmas separately. The nanoapertures are treated by 40 sccm flow rate, 100% power,  $O_2$  plasma (Henniker Plasma, HPT-100) for 1 minute. This protocols aims maximize duration of wettability induced by the treatment and remove carbon implanted during imaging [172] without increasing surface roughness [176] [178]. The use of  $O_2$  over  $Ar$  has significant advantage in the former with the reduction of hydrophilicity as for  $Ar$  treatment the contact angle restores to  $30^\circ$  within 5 minutes, while for  $O_2$  treatment the contact angle remains below  $30^\circ$  for over 100 minutes [178].

In addition to plasma treatment the effect of annealing on Au surfaces for removal of  $Ga^+$  ions implanted during FIB is an additional step that can be performed. Here one can consider the ion range of  $Ga^+$  ions implanted into Au shown in Fig.6.5b. Based on this estimate of depth and results on gallium diffusion presented in [179], is may be possible to develop a protocol for removal of implanted ions through annealing - determining temperature and annealing time required based on diffusion timescales and depth of implantation, balanced against the need to preserve nanoaperture geometry.

## Finite-Difference Time-Domain Modelling

To predict approximate positions of the nanoaperture scattering cross sections, both in transmission and reflection one can perform finite-difference time-domain (FDTD) simulations of the nanoaperture (Fig. 6.9a) using Lumerical FDTD software. The aperture is illuminated with a Total-Field Scattered-Field (TFSF) element having a 400-1000 nm broadband pulse with a 2.6 fs duration. The TSFS source separates the simulation domain into a total field

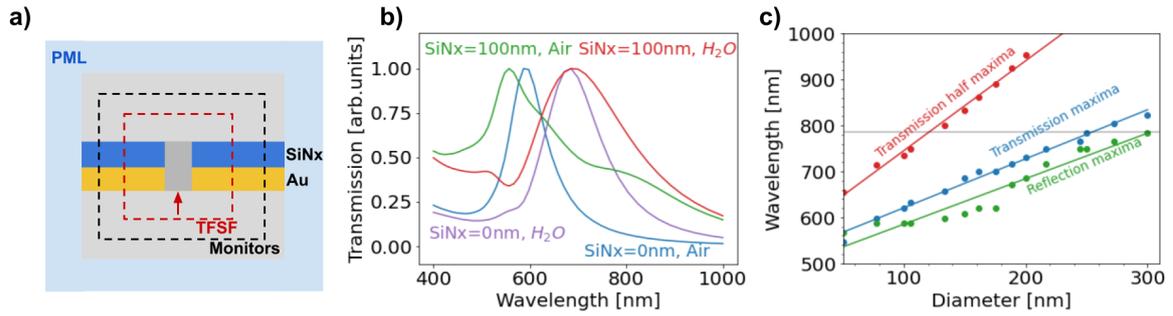


Fig. 6.9 a) Diagram of FDTD configuration used to simulate nanoaperture scattering and transmission. b) Simulated normalized transmission of a 150 nm diameter nanoaperture in a 100 nm Au film for 100 nm SiNx and a free-standing Au membrane. c) FDTD results for evolution of transmission and reflection resonances with increasing aperture diameter in a 100 nm Au film on a 100 nm SiNx membrane, immersed in water ( $n = 1.33$ ).

(inside source) and a scattered field (outside source) regions, allowing recording of scattered fields by placing scattering monitors externally to the source, with automatic power normalization returning total scattering cross section, . By recording the field backscattered from the surface, and transmitted through, we can compute the transmission and reflection spectra expected for the nanoaperture, normalizing to the each spectrum's peak. To remove effects of reflections from boundaries we use Perfectly Matched Layers (PMLs) around the simulation domain and extend the simulation to  $5 \mu m$  around the centred nanoaperture. To retain accuracy of the simulation near the aperture a mesh override region is defined with a uniform element size of  $2.5 \times 2.5 \times 2.5$  nm, qualitatively determined by lowering element size until no changes in cross-sections are observed. Effects of different media can be incorporated by changing the parameters of the background medium of the simulation domain and aperture (separately), shown in Fig. 6.9b where a redshift of the spectrum is observed, both for Au/SiNx and Au-only nanoaperture configurations when medium refractive index changes from  $n = 1$  to  $n = 1.33$ . For a nanoaperture in a medium surrounded by water the simulated evolution of transmission and reflection peaks, along with the transmission half-maxima, are shown in Fig. 6.9c, with the grey line indicating the 785 nm wavelength of the incident laser.

### Darkfield Imaging

After fabrication the nanoapertures are characterized optically to determine individual nanoaperture spectra - allowing easy calculation of sample variability and prediction of optical response when under laser illumination. To measure each apertures spectrum we use darkfield spectroscopy, where a broadband incoherent hollow beam, generated by collimated

## Nanoaperture fabrication

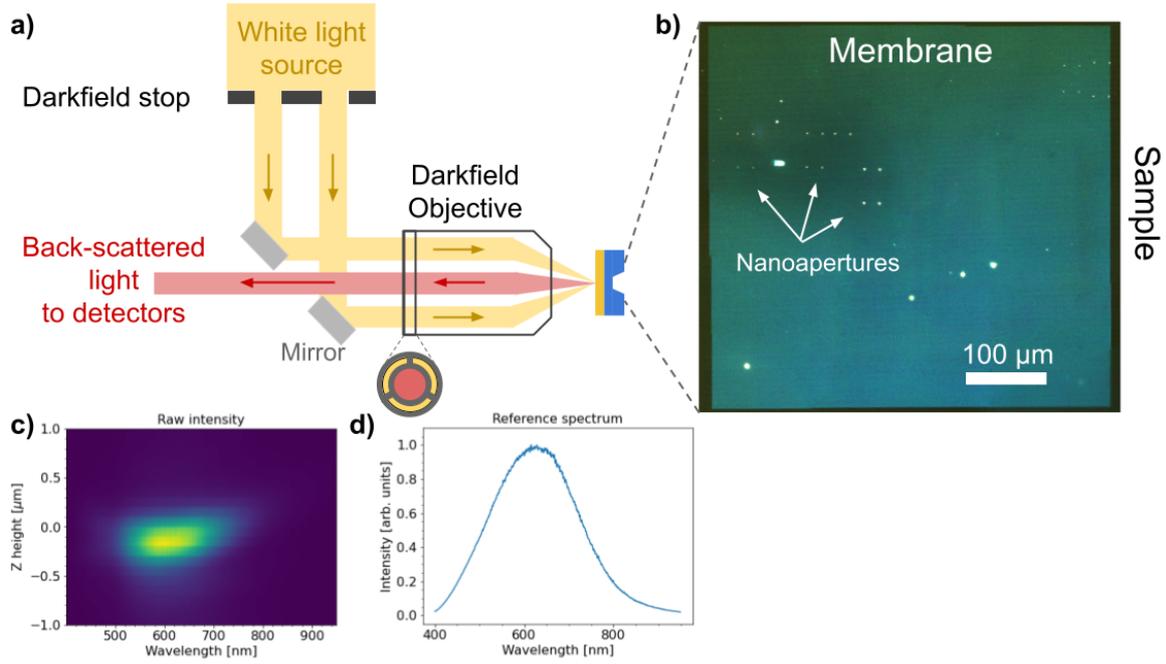


Fig. 6.10 a) Darkfield imaging configuration, illumination through the backside allows imaging of apertures in transmission. b) Image of a SiN<sub>x</sub> membrane in transmission showing nanoapertures. c) Spectral ‘z-stack’ used for correction of chromatic aberration by computing the maximum signal along the ‘z-height’ in array column. d) White light reference spectrum used to normalize measured signal.

output from a filament passing through a darkfield stop, is focused by an objective onto the sample. The backscattered light or transmitted light is collected by the objective and relayed to a spectrometer (Ocean Optics), as illustrated in Fig.6.10.

Filament source spectral variability and system optical transmission are incorporated through a measurement of a ‘reference’,  $I_{ref}$ , white light spectrum (Fig.6.10d) from a white light diffuse scattering surface. Effects such as stray light and spectrometer dark pixel counts are accounted for by measurement of a ‘background’,  $I_{bg}$ , with microscope shutter closed. Finally, chromatic aberration, induced by wavelength dependent dispersion through optical elements, is compensated through measurement of a ‘z-stack’  $I_{meas}$  (Fig.6.10c) of spectra obtained by scanning the objective perpendicular to the sample plane through  $2\mu m$  in steps of 50 nm. For a single  $z$  height of the objective the normalized spectrum computed using Eq. 6.2.

$$I_{norm}(\lambda) = \frac{I_{meas}(\lambda) - I_{bg}(\lambda)}{I_{ref}(\lambda) - I_{bg}(\lambda)} \quad (6.2)$$

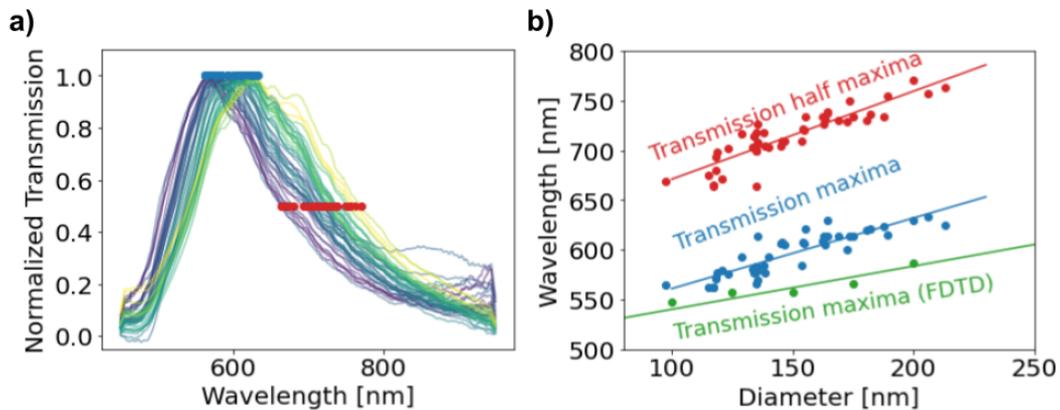


Fig. 6.11 a) Experimentally measured transmission spectra for nanoapertures, coloured by nanoaperture average diameter measured using SEM. b) Variation of resonance peak with increasing nanoaperture diameter.

For a given z-stack the resulting spectrum of a single nanoaperture is obtained by calculating the 99<sup>th</sup> percentile for each wavelength. When performing measurements in transmission, an additional additive component is the transmission spectral of the membrane, removed in a way similar to the electronic background.

Combining the transmission spectra with the SEM imaging allows mapping between optical properties of the nanoapertures and their geometry, extracted from SEM images using ImageJ [180] shown in Fig.6.11. While broadly consistent the positions of the transmission maxima deviate from the prediction produced by the FDTD. This disparity may arise from the lack of adhesion layer in the FDTD model, ellipticity of the nanoaperture or variation of deposited gold layer thickness due to non-unity tooling factor of the PVD system. Resolving this disparity requires increased precision measurements of film deposition parameters and extension of the FDTD modelling to account of ellipticity and cross-section profile of the nanoaperture, as well as accounting for effects of surface contamination. Other potential sources of this disparity are the carbon and gallium deposited in the Au surface during fabrication [172]. With an improved FDTD model and fabrication protocol involving additional characterization steps it is expected that one would be able to more accurately predict properties of nanoaperture systems, quantify effects of fabrication parameters on the optical properties of the nanoapertures and obtain estimates for manufacture tolerances.

## 6.3 Fluidics

In order to perform simultaneous optical and electrical measurements of processes near the nanoapertures, the SiN<sub>x</sub> membrane system must be incorporated into fluidic system. Similar

## Nanoaperture fabrication

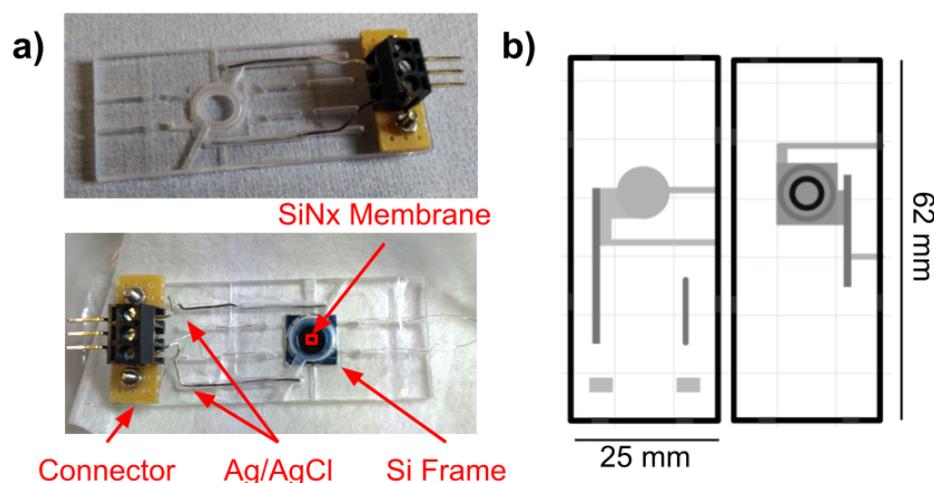


Fig. 6.12 a) Fluidic chip with and without SiNx membrane. b) Planar milling design of fluidic chip, greyscale indicates depth of cut, ranging up to 3 mm.

systems to the one developed in this thesis have been demonstrated before [11] [124] [181] [182] but none are available commercially, thus a fluidic chip (Fig.6.14a) was designed and manufactured as part of the project.

In a preliminary investigation we considered the standard route of using Polydimethylsiloxane (PDMS, Sylgard 184). To do this, a curing agent was mixed with the PDMS in a 10:1 ratio and poured into an aluminium inverse mould master. After a period of degassing and curing (at  $80^\circ$  for 2 hours) the PDMS fluidic chip was removed from the mould. Subsequent assembly included sealing the SiNx membrane to the PDMS, followed by plasma oxygen treatment which activated the surface of the PDMS to enable its bonding to glass coverslips in order to create two sealed chambers. In(/out)let ports, made from polymer coated glass capillaries, and electrodes were integrated into the chip by puncturing the PDMS to gain access to the fluid chambers. Following application of a thin layer of PDMS to seal the in(/out)let connections and electrodes in the chamber the assembled chip was again cured at  $80^\circ$  for up to an hour to prevent leaks.

Creation of PDMS moulds with suitable surface quality and thermal stability requires machining aluminium. This increased development time and costs, making fast design iteration challenging. Combined, with long cycle times for producing a single chip, limited by degassing and curing times (Fig.6.14a), we considered alternatives to PDMS - polymers.

Unlike PDMS, polymers are often directly machined to produce fluidic systems, without requiring a mould. The central benefit of polymers for fluidic chips is manufacturability, with the ability of the chip to be machined directly out of the chosen polymer. This removes

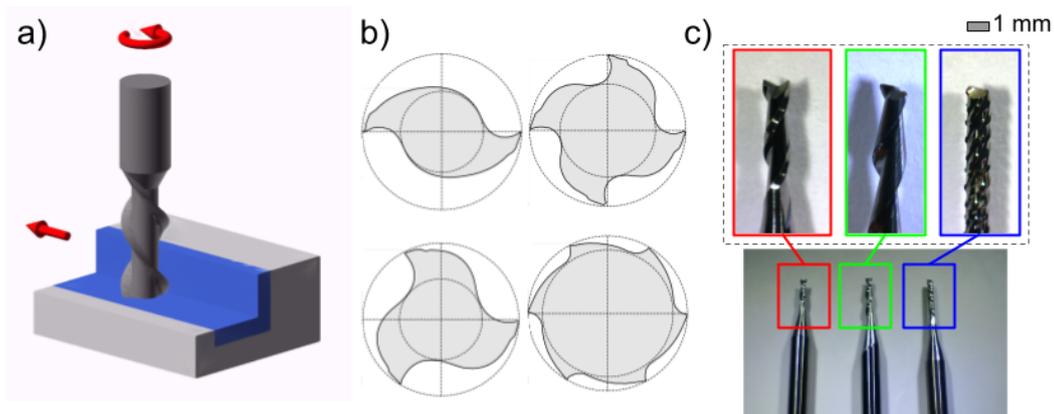


Fig. 6.13 a) Diagram of milling process, regions where material is removed highlighted in blue. b) Schematic diagram of flute end-mills with varying numbers of flutes. c) Image of 1 mm diameter two flute end-mills.

Table 6.1 Properties of polymers utilized in microfluidics [183]

Polymer (Acronym)	$T_m$ ( $^{\circ}C$ )	Water absorption(%)	Solvent resistance	Thermal expansion coefficient ( $10^{-6}/K$ )
Polymethylmethacrylate (PMMA)	250 - 260	0.3 - 0.6	Good	70 - 150
Polycarbonate (PC)	260 - 270	0.12 - 0.34	Good	60 - 70
Polystyrene (PS)	240 - 260	0.02 - 0.15	Poor	10 - 150
Polyetheretherketone (PEEK)	340 - 350	0.1 - 0.5	Excellent	47 - 54

degassing, mixing and curing steps associated with PDMS drastically cutting on cycle times, down to the order of minutes. The mechanical properties of plastic also lower requirements for machining as compared to metal moulds, allowing the use of cheap desktop CNC machines, which have lower rigidity and accuracy but, like 3D printers, are drastically dropping in cost and increasing in capability.

Aside from ease of fabrication, the use of polymer based microfluidics expands the range of available materials allowing customization of material properties to both application (solvent resistance, water absorption) and production process (melting temperature,  $T_m$ , coefficient of thermal expansion, CTE) as well as allowing for long-term storage without material degradation. Some commonly available materials are tabulated in Table 6.3 [183]. The ability to select a polymer also addresses potential issues with water evaporation from PDMS and the materials adsorption of small molecules [184].

## Nanoaperture fabrication

---

While plastics place less stringent requirements on the equipment, a number of key parameters set the quality of the resulting chip. Key among these are spindle rotation speed (rpm), measurable using a tachometer, and feed rate (mm/min) determining the rate of material removal along with forces exerted on the tool during milling. For small diameter tools (Fig.6.13c) large forces exerted at high feed rate causes tool breakage while low feed rates result in insufficient loading, producing chatter and poor surface finish. In this work, a feed rate of 200 mm/min and 9000 rpm was used for cutting. A related parameter in calculation of feeds and speeds is flute count, the number of cutting edges of a mill bit (Fig.6.13b). Increasing flute count decreases material removal by each edge, reducing chip size and increasing local heating. For soft materials a low flute count is beneficial. If incorrectly selected melting of the plastic onto the tool prevents cutting. Depth of cut, along z-axis, and step-over (perpendicular to z-axis) determine depth and thickness of material removed per pass, indicated by blue in Fig.6.13a. These quantities in turn influence total milling time and layers required to reach a given depth. For the 3018 CNC we experimentally determined optimal values to be 0.2 mm depth of cut 30% step-over.

Material choice is also key for manufacturability and can vary within a given polymer type (Table 6.3) based on the method used to manufacture the sheet. An example of such variability is ‘cast’ and ‘extruded’ PMMA sheet, with the former exhibiting improved thermal stability but reduced mechanical sheet tolerance ( $\pm 0.7\text{mm}$  for 3 mm sheet) while the latter was found to be less thermally stable and melt during machining and clog endmills. After testing with various milling parameters and materials the final choice for fluidic chips is 3mm polycarbonate (PC), due to its good temperature stability, eliminating material melting during machining, combined with higher sheet dimensional tolerance ( $\pm 0.1\text{mm}$ ) removing the need for facing operations.

The design of the chip consists of two sides (Fig.6.14b) milled in a sheet in two stages. The CNC is controlled using g-code commands sent to the microcontroller through Grbl control software.

### Sealing and Assembly

Sealing the silicon nitride in the fluidic chip creates two chambers into which the electrolyte containing the nanoparticles is loaded. An preliminary approach was to utilize PDMS in order to create the inter-chamber seal and bond the chip to the coverslips. However, significant amounts of outgassing resulted in dew formation on the membrane surface (Fig.6.14b). A viable alternative is the use of an optically curable adhesive (Norlands NOA 68 [185]).

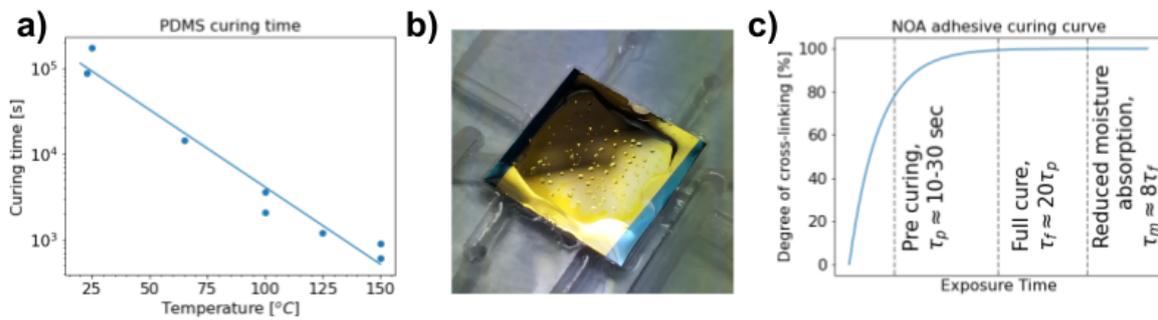


Fig. 6.14 a) PDMS curing time as a function of curing temperature. b) Dew formation of Au surface after curing c) Diagram of NOA 68 curing as a function of UV exposure time.

This single mixture adhesives have an initial rapid pre-curing phase when exposed to a UV (350-380 nm) light source. The pre-curing is followed by a full curing phase, a function of UV dose and curing volume (Fig.6.14c). Unlike PDMS, the ability to rapidly cure the adhesive enables assembly of the chips in separate stages. The assembly proceeds by sealing the silicon frame containing the membrane into the polycarbonate chip followed attaching the coverslips. Finally the electrodes are inserted into the side channels and also sealed into the chip to prevent leaking, leaving only an inflow/outflow port in each chamber. The resulting assembly is then cured fully with 365nm illumination for 30 minutes.

Potential extensions of this approach include spin coating the adhesive in order to surface roughness left by machining. Additional modification of the resulting surface properties can also be achieved through introduction of additives into the formulation, as in [186].

## 6.4 Electrical model

The final element of the nanoaperture system are the electronic parameters of the system, including the solution, electrodes and SiNx membrane. Below we discuss properties of the individual elements and how they map to an effective circuit model, that in future can be used in making quantitative measurements of the nanoaperture system.

### 6.4.1 Electrodes

The electrodes incorporated into the chip, are both Ag/AgCl. To fabricate the electrodes we submerge 0.5 mm diameter silver (99.99%) wire in household bleach consisting of sodium hypochlorite ( $NaOCl$ ), sodium hydroxide ( $NaOH$ ), sodium laureth sulfate (SDS). During the submersion (30 minutes - 1 hour) the silver reacts with the chloride ions producing AgCl, which is white in colour. The observed darkening of the electrodes we attribute to the  $OH$

## Nanoaperture fabrication

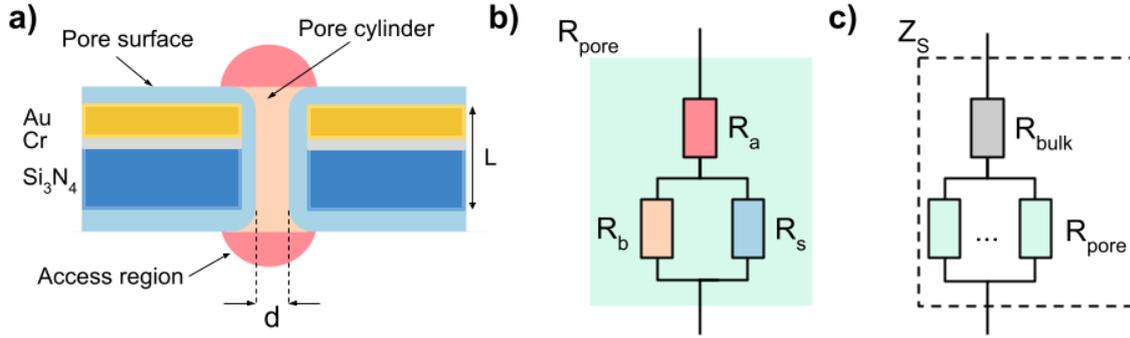


Fig. 6.15 a) Diagram of nanopore cross-section. b) Equivalent circuit of a single nanopore, with  $R_a$  representing the ‘access resistance’,  $R_b$  being the resistance of the cylindrical pore while  $R_s$  incorporates surface effects. c) Combined effective impedance  $Z_s$ , incorporating bulk solution resistance and contribution of nanopores.

groups of  $NaOH$  producing  $Ag_2O$ , removed by cycling the electrode in a  $KCl$  solution. With two electrodes, we can model their contribution to the measurement according to the Randles cell model introduced in Chapter 2 (Fig.2.3b). While quantitative parameter measurement is possible through electrical impedance spectroscopy (EIS), this was not possible with the available equipment and is left for future work.

### 6.4.2 Nanoaperture conductance

To complete the equivalent circuit model, we must incorporate the effect of the nanoapertures in the  $SiN_x$  membrane into the Randles cell model and combine the membrane resistance with that of the bulk. The first element of a nanopore model (Fig.6.15a) is an analogue of a conventional resistor with  $R = \rho A/L$  where  $A$  is the cross-section area and  $L$  is the length. Applying this to a cylindrical pore in a solution with bulk conductivity  $\sigma = \rho^{-1}$  we obtain:

$$R_b = \frac{1}{\sigma} \frac{\pi r^2}{L} = \frac{1}{\sigma} \frac{\pi d^2}{4L}$$

Corrections to this basic model are incorporated through separate consideration of ‘access region’ (similar to the ‘capture region’ discussed previously) and surface charge interaction effects, such as Debye layer. The latter, surface effects, are modelled by introducing a surface resistance term  $R_s$  in parallel with  $R_b$  [187, 188], which has the form:

$$R_s = \frac{L}{\pi \cdot \sigma_{surface} \cdot \mu_K \cdot d}$$

This acts to model the modification of ion distribution at the liquid-pore interface, due to charge rearrangement near the surface where ions shield the pores surface charge. Key in this is the surface charge density  $\sigma_{surface}$ , a function of the surface  $\zeta$ -potential and ion concentration. This can be estimated from the Grahame equation and varies as a function of ion concentration  $c$  [189] [190], we however will assume it to be constant, for simplicity.

Combining the two, in parallel, we obtain an effective resistance of pore as  $R_b || R_s = (R_b^{-1} + R_s^{-1})^{-1}$ . As the nanopore increases in aspect ratio ( $l \ll d$ ) the flow of ions through the pore becomes limited by the diffusion time for ions to move from bulk solution to the entrance of the pore. This is ‘access resistance’ is a contribution added in series to bulk and surface terms as takes the form [187] [189] [191] [190]:

$$R_a = \frac{1}{\sigma d}$$

Integrated into total pore resistance as we obtain  $R_{pore} = R_a + (R_b || R_s)$ . Multiple non-interacting pores are modelled through parallel addition of individual pore contribution. Incorporating bulk solution resistance gives the effective impedance element  $Z_s$ . This allows the modelling of the measured system as a Randles cell (Fig.2.3c).

An indicative parameter range for a single nanopore using the above model can be obtained by considering pore diameters in the  $d = 100\text{-}300$  nm range. Combined with total membrane thickness set by the sum of the SiNx and metal layers,  $L = 200\text{nm}$ , and a constant surface conductivity of  $|\sigma_{surface}| = 15\text{mS}/\text{m}^2$  (from [190]) we obtain a estimate for the conductivity of nanoapertures (Fig.6.16b). the plot demonstrates that KCl concentration range 1-100 mM KCl represents the transition regime between bulk and surface conductivity. It can also be shown that access resistance can be neglected, by noting that  $R_a \ll (R_b || R_s)$ . A simple way of estimating bulk conductivity in the above example is using ion electrophoretic mobility values reported in [189],  $\mu_K = 7.616 \times 10^{-8}\text{m}^2/\text{Vs}$  and  $\mu_{Cl} = 7.909 \times 10^{-8}\text{m}^2/\text{Vs}$ . Using the relation [188]:

$$\sigma = 10^3 \cdot (\mu_K + \mu_{Cl})n_{KCl}(c) \cdot e$$

Here  $n_{KCl}(c) = N_A c$  is ion number density,  $e$  is elementary charge and  $N_A$  is Avogadro’s constant. We estimate solution conductivity shown in Fig.6.16a, where circular markers denote reference standard KCl values from [192] at  $T = 20^\circ\text{C}$ . Computing the Debye length for the same concentration range we observe that the on-set of crossover between the bulk  $R_b$  and surface  $R_s$  dominated regime at  $c = 3\text{mM}$  corresponding to the Debye length of  $6\text{nm}$ .

## Nanoaperture fabrication

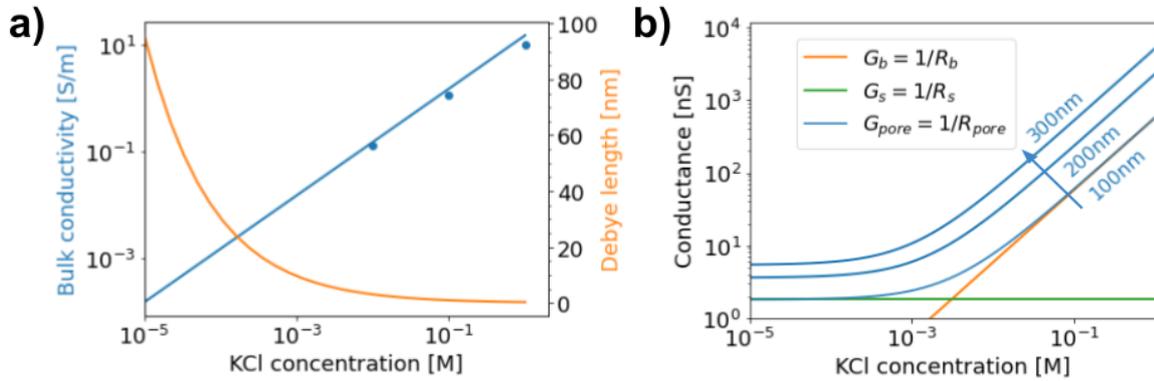


Fig. 6.16 a) Bulk solution conductivity and Debye length with KCl concentration. b) Single nanopore conductance with concentration for three diameters, 100, 200, 300 nm and membrane thickness  $L = 200\text{nm}$  and surface conductivity  $|\sigma_{surface}| = 15\text{mS}/\text{m}^2$ .

With our experiments we operate at the 10 mM, 50 mM and 100 mM, thus being in the bulk conductivity regime, with Debye layer being  $\leq 3\text{nm}$ . The solution is prepared by adding the KCl salt into deionized water (pH=7). Nanoparticle stability to aggregation is a function of particle size, surface properties, pH, and electrolyte concentration [193].

Based on the focusing of the ion beam and the deposited metallic layers, a natural extension of this cylinder model of the nanopore would be introduction of asymmetry parameters modelling tapering of cross-section and a change in surface charge properties, cause by the change in material layers between  $\text{Si}_3\text{N}_4$ ,  $\text{Cr}$  and  $\text{Au}$ . This asymmetry will result in an asymmetric charge distribution inside the channel, termed ‘ion concentration polarization’ (ICP) [194]. We leave development of this model to future work, noting that applying a potential to the membrane by contacting the metal surface is possible in the developed fluidic chip. Instead of this we will work within the membrane equivalent circuit model  $R_{pore} = R_a + (R_b || R_s)$ .

## 6.5 Summary

In this chapter we have described key steps for fabricating nanopores on  $\text{SiN}_x$  membranes, including effects of adhesion layers, surface activation, micromachining and optically characterized nanopore. The fabrication of a fluidic chip for allowing both optical and electrical access was also detailed. Finally we have introduced a unified equivalent circuit model for the nanoporous membrane capturing both surface and volume effects present in nanopores and investigated key associated parameters and their ranges showing that surface and bulk resistance terms dominate over access region effects.

# Chapter 7

## Nanoaperture scattering

Subsequent to fabrication and assembly of the nanoaperture containing SiNx membranes into the fluidic chip we investigated properties of nanoaperture transmission in solutions containing nanoparticles in the size range 20 to 110nm in diameter. To perform such transmission measurements we constructed a transmission microscope capable of high speed recording of intensity modulation, along with implementing the required control and data acquisition software for this purpose, which we describe in this chapter. Along with details of the microscope we present preliminary measurements obtained with this system, including the expected intensity modulation in transmission on translocation as well as demonstrating a number of other, related effects that indicate the complexity of dynamics occurring near the nanoapertures. Finally, we suggest some future directions for work in this area.

### 7.1 Microscope

In order to perform measurements of nanoaperture transmission a microscope depicted in Fig.7.1c was constructed. An initial version relied on an upright construction similar to previous designs developed in our group. This design, while suitable for low speed measurements such as darkfield spectroscopy or Raman scattering, where large >100 ms integration times are used, is susceptible to vibrations and transients present in the environment. In part caused by a combination of poor vibration isolation of the optical table from the floor and high centre of gravity of the construction. The improved design depicted in Fig.7.1c uses a layout in the plane of the optical table removing the vertical frame mounting the components. This reduces the centre of gravity of the microscope, decouples mechanical motion of components, allows shorter beam paths between elements and easier to extend to accommodate new beam

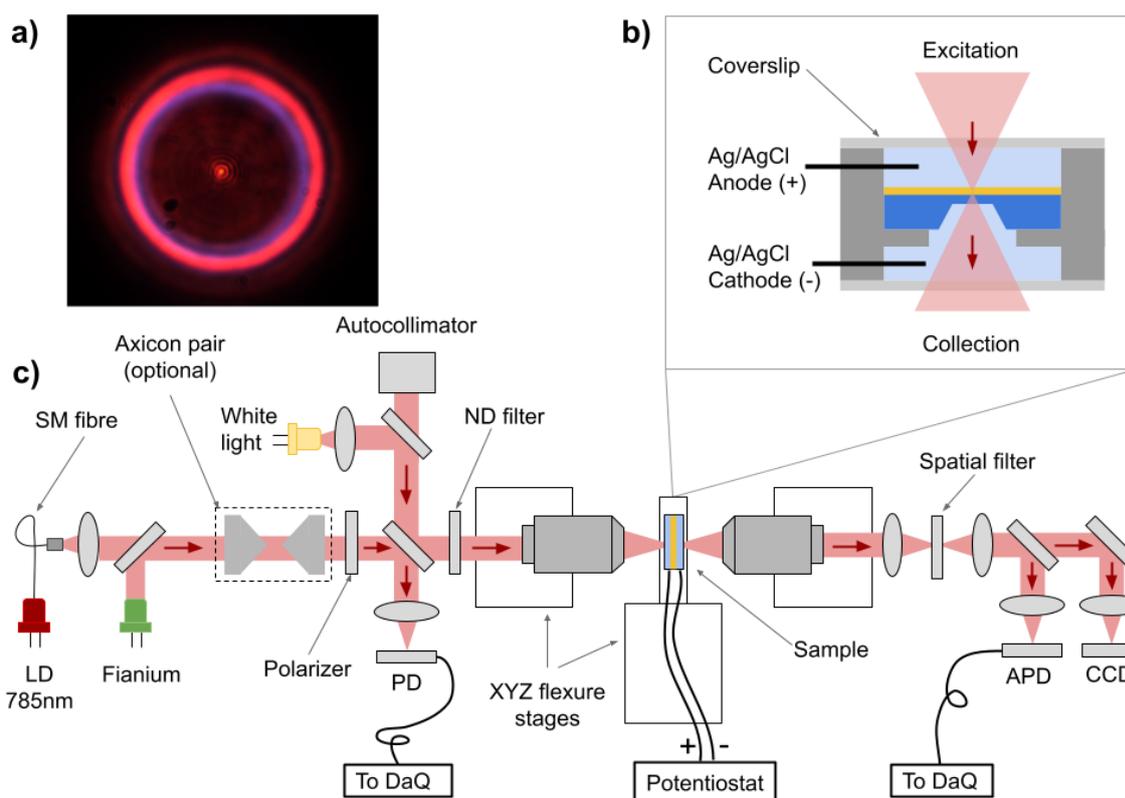


Fig. 7.1 a) Image of hollow beam generated by axicon pair. b) Configuration of SiNx membrane inside optical microscope. c) Diagram of constructed optical microscope for measurement of nanoaperture transmission intensity modulation.

paths. Broadly, the microscope can be divided into an excitation path, a collection path and the sample mount.

The excitation path incorporates two laser sources and an incoherent fibre coupled lamp, used for coarse alignment of the SiNx membrane. The lasers, a 785 nm home-made CW diode laser and a pulsed supercontinuum (Fianium), are individually collimated and combined in the main beam path, selecting the polarization with a linear polarizer. Stability of incident optical power is monitored by a photodiode (PD), while power control is achieved using a continuously variable neutral density filter, inserted before the excitation objective. Light entering the 100 $\times$  excitation objective is focused through the sample coverslip, onto the surface of the Au coated SiNx membrane containing the fabricated nanoapertures.

The signal transmitted through the nanoaperture is collected by a 100 $\times$  ultra-long working distance objective, required due to the <3 mm thickness of the fluidic chip. The collected beam then propagates through a 1:1 lens relay system, where a pinhole (50-150  $\mu$ m diameter), attached using a magnetic mount to a XY translation, acts as a spatial filter, allowing the

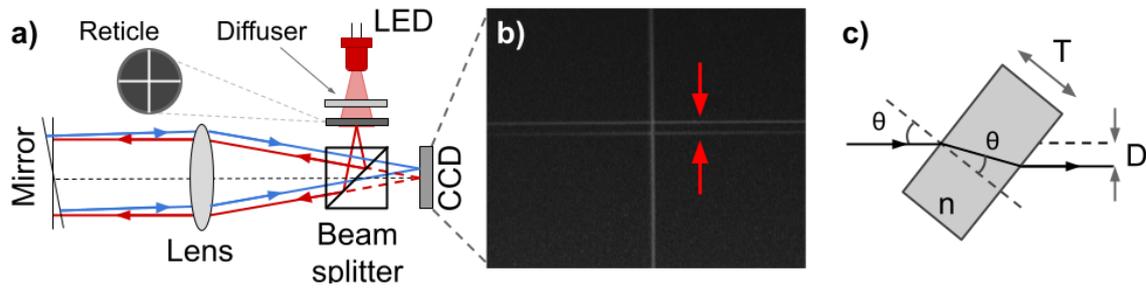


Fig. 7.2 a) Autocollimator schematic. b) Autocollimator image showing effect of vertical misalignment of objectives producing vertical displacement of formed image. c) Beam displacement as a result of passed through a tilted glass slide.

selection of a single nanoaperture at a time for imaging. The spatially filtered signal is then transmitted and focused on an avalanche photodiode (APD) for high speed recording and CCD camera, used for alignment and slow monitoring.

With the microscope lying in the plane of the optical table and not being based on a commercial system, we use two XYZ flexure stages to allow collinear alignment of the excitation and collection objectives along the same optical axis. To achieve this reliably, we constructed autocollimator [195] allowing measurement of small offsets in alignment. To perform the alignment the first step is to place a planar mirror at entrance of illumination objective and using steering mirrors to bring the autocollimator beam reflected from the mirror to form an image of the reticle at the centre of the autocollimator CCD. This ensures that the autocollimator beam propagation is perpendicular to the plane of the first objective. Next, placing the same mirror at the exit of the second, infinity corrected, collection objective results in a back reflection that propagates through the two-objective system twice, again forming an image on the autocollimator CCD. Based on the symmetry of the reflected image as well as its focusing one can determine the lateral offset of the second objective relative to the first, and align the collection objective using the flexure, completing the alignment. Once aligned, effects such as tilt of the objectives under load (Fig.7.2b), vibration and drift can in principle be measured quantitatively, we however leave this for future work. It should be noted that prior to performing the above alignment one should ensure the collinearity of travel of the XYZ flexures, achievable using a rail mounted dial indicator.

To understand the basis of the alignment technique, consider the autocollimator optical layout (Fig.7.2a) consisting of a cross-hair reticle illuminated by a red LED passed through a glass diffuser. Here, the image of the reticle is projected to infinity by the collimating lens and reflects from a mirror surface to be aligned. The reflected image is refocused by the lens onto a CCD, positioned in the same focal plane as the reticle. Tilt of the mirror reflecting the

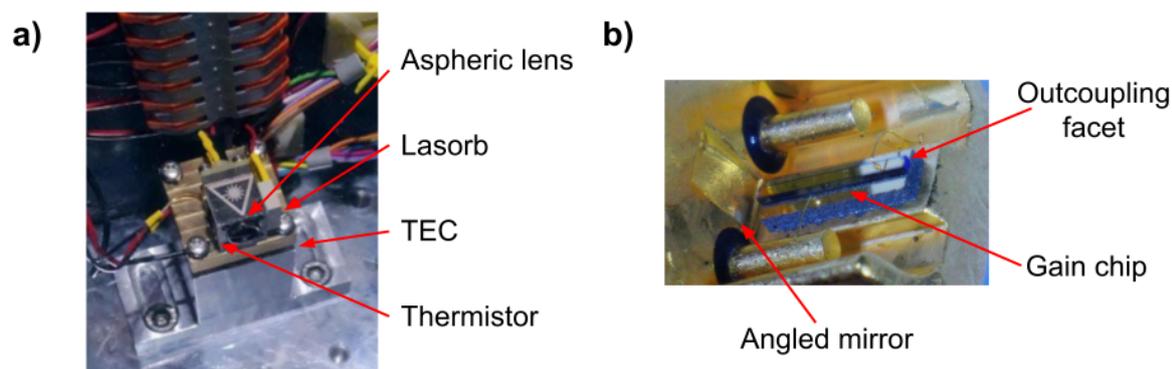


Fig. 7.3 a) Diode laser assembly mounted inside hermetically sealed enclosure. b) Fabry-Perot laser diode chip.

light produces a lateral offset of the resultant image on the CCD. Aside from measurement of angular tilts the autocollimator can be applied to many metrological measurements. Future work on developing the autocollimator would aim to quantitatively characterize its angular resolution properties and define proper operating procedures for its use.

The final component, the sample, is mounted on a three axis flexure stage with open loop piezos enabling automated positioning with a  $20\mu\text{m}$  range (per axis) and  $20\text{nm}$  open loop resolution. Electrical interfacing of the fluidic system with the data acquisition is achieved through connecting the Ag/AgCl electrodes to the working and counter connections of a potentiostat (Rodeostat, version 1.0). The orientation of the anode and cathode relative to the incident beam is shown in Fig.7.1b. A complication arising due to the finite thickness of the fluidic chip ( $\sim 3\text{mm}$ ) the focal shift and lateral offset induced by inserting the sample into the beam path. The former is simply explained by considering the optical path length, modified by the refractive index change. The latter, can be understood by considering a beam displacement plate (Fig.7.2c) shifting the beam laterally. The result of this is a need to perform a second alignment with a dummy chip prior to inserting a real sample.

A feature of the constructed system that we leave for future work, is the axicon lens pair. Introducing this on the incident beam path allows for hollow beams to be generated, similar to the darkfield illumination described in the previous chapter. An image of the produced beam, with a central Airy disk is shown in Fig.7.1a. Preliminary experiments have indicated that this can increase the contrast of the nanoaperture reducing the background contribution of the incident beam.

### 7.1.1 Diode laser

During the initial construction and operation of the microscope the supercontinuum laser was used, due to being the only available system capable of producing  $\lambda \geq 750nm$  wavelengths, red-detuned relative to the nanoaperture transmission resonance. To select a suitable region of the supercontinuum spectrum the laser was spectrally filtered using interference bandpass filter. Similarly to the case of PCS the use of interference bandpass filters introduces a trade-off between bandwidth of the experiment and available power onto the sample. For a given incident laser a baseline transmission intensity can be computed through the overlap integral between the transmission and laser spectra,  $I_{ref}$ . Shifts of the transmission spectrum and the resulting new transmission intensity can be computed in the same way using a new shifted spectrum, or alternatively, shifting the laser wavelength in the opposite direction by the same amount. Relative intensity change is then given by  $\frac{I_{new} - I_{ref}}{I_{ref}}$ . By considering the ‘effective’ laser wavelength we can generalize the above reasoning to describe the effect of finite bandwidth as producing a convolution between the transmission spectrum and the laser spectrum.

While possible to measure transmission modulation using the supercontinuum this introduces a number of complications into the apparatus, potentially affecting the physical device under test, the corresponding analysis, and design of the data acquisition system. The first of these is the pulsed nature of the supercontinuum laser producing picosecond pulses. This pulsed optical input may induce fast transients near the nanoaperture that may affect particle motion through the nanoapertures. With a pulse rate of 60 MHz acquisition of the baseline intensity is also complicated by the need to consider the finite bandwidth of the monitor PD and integrate additional low-pass filters prior to recording by the DAQ. A second related issue is the complexity of the supercontinuum itself, which exhibits high frequency (10-30 kHz) DAC setting dependent noise, something we are unable to remove due to the turn-key nature of the system. To alleviate these issues and remove unnecessary parameters, such as pulsed excitation from the measurement, we constructed a CW diode laser using an inexpensive Fabry-Perot (FP) diode (Fig.7.3). This significantly boosted the range of incident powers we could generate, up to 200 mW. The increased power budget allowed the introduction of the monitor PD, generation of an improved spatial mode profile using a single mode fibre as well as simplifying the acquisition electronics through removing errors associated with measurement of pulse power with limited bandwidth ( $< 10MHz$ ) electronics.

Key requirements for construction of a stable diode laser using a FP laser diode (Panasonic LNC728PS01WW, TO56,  $I_{op} = 180 - 265mA$ ,  $V_{op} = 2.0 - 3.0V$ ) are low input drive current noise along and high temperature stability. The former is achieved using a low noise constant

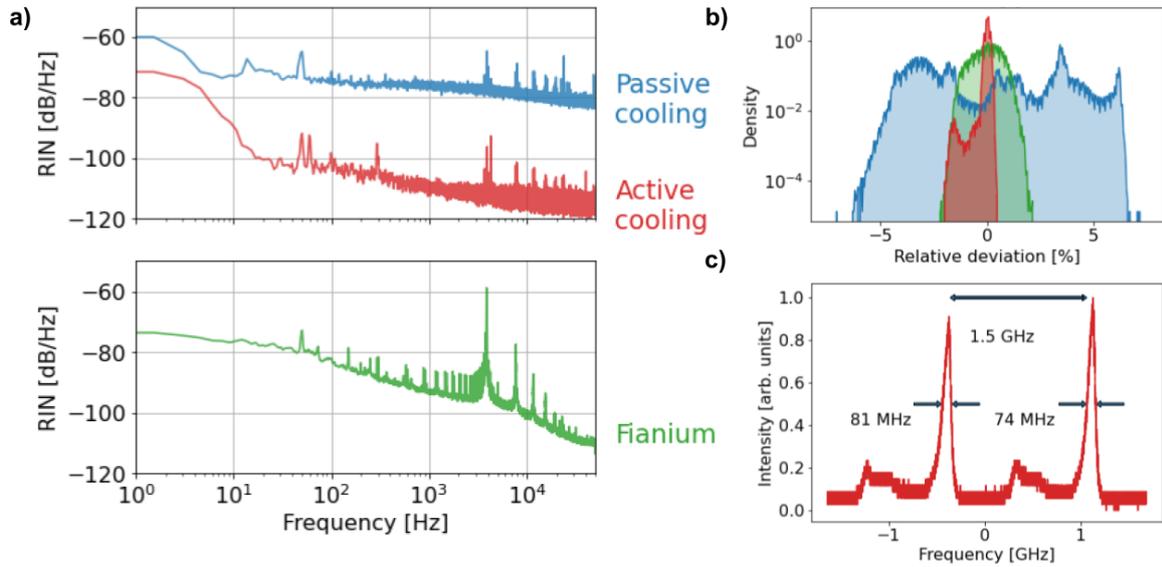


Fig. 7.4 a) Relative intensity noise (RIN) measurement for CW laser diode (blue, red) compared to a 100kHz low-pass filtered supercontinuum exhibiting significant (green) demonstrating the improved performance of the actively cooled laser diode over the supercontinuum. b) Relative deviation histogram of passive and actively cooled laser diode compared to the supercontinuum, bi-modality of active cooled distribution is an indicator of mode hopping. c) Cooled laser diode bandwidth measured by a scanning Fabry-Perot etalon.

current laser diode driver, delivering up to 200 mA currents with a 3  $\mu$ A RMS current fluctuation. High temperature stability (5 mK RMS) is achieved through mounting the laser diode in a copper diode mount placed on top of a Peltier element controlled using a proportional-integral-derivative (PID) loop. The TEC is mounted between the brass diode mount and aluminium block attached to an optical breadboard. Temperature measurement is performed using a  $\sim 10k\Omega$  thermistor ( $\pm 1^\circ C$  accuracy) attached to the side of the laser diode mount using thermal tape, the value of the thermistor is read by the TEC controller, converted to a temperature and used as the feedback signal to determine the TEC drive current required to achieve a given temperature set-point. With an ambient lab temperature of  $19.2 \pm 0.5^\circ C$  the use of a single stage Peltier element (Thorlabs TECF2S) with a 1A maximum TEC driver allows a  $\Delta T \approx 45^\circ C$  for a heat load of 1W ( $V_{op} \times I_{op} = 3V \times 265mA \approx 0.8W$ ). In practice this is limited by air convection, minimized by mounting the laser assembly inside a hermetically sealed enclosure. Optical access out of the enclosure is a 1 inch diameter AR coated window, tilted to prevent back reflection into the FP diode. The enclosure isolates the laser diode from the environment temperature and humidity fluctuations and allows us to achieve the 5 mK temperature stability. Hermetic sealing allows for prospect of wavelength tuning of the laser, with  $\frac{d\lambda}{dT} = 0.25 \text{ nm}/^\circ C$ , in principle enabling a 10 nm tuning range using

a  $\Delta T = 40^\circ C$ . During normal operation however we keep temperature within a few degrees of ambient to minimize possibility of condensation, from air inside the enclosure, forming on the laser diode or collimating lens, in future a desiccant can be used to remove this moisture. As laser diodes are highly sensitive to electrostatic discharge (ESD) and current transients, such as those caused by switching on/off the power supplies, we use a Lasorb ESD protection system connected in parallel across laser diode pins.

The results of active and passive temperature control on laser diode performance are illustrated in Fig.7.4a where passive cooling results in an approximate 5% intensity fluctuation, while active feedback on temperature reduces maximum fluctuation intensity fluctuations to below 2% in worst case and below 1% on average (Fig.7.4b). The asymmetry of the active temperature controlled fluctuations is suspected to arise from mode hops which can be observed by monitoring the intensity as a function of time. Minimization of mode hopping can be achieved through optimizing the drive current and temperature settings or through introducing additional feedback into the laser cavity by means of a grating or using a DFB laser diode, left for future work. Even without DFB the central peak of the laser linewidth, as measured by a 1.5 GHz scanning Fabry-Perot interferometer is found to be below 100 MHz (Fig.7.4c) something that is not possible to achieve by spectral filtering a supercontinuum source. Owing to the ‘high’ power output of the laser diode ( $> 100mW$ ) the elliptical mode profile can be filtered by propagating the beam through a single mode fibre, with a coupling efficiency of 30% (30 mW output) can be reliably achieved. The supercontinuum, comparatively, exhibits lower output power 3mW/nm and is unpolarized, resulting in a 50% loss on passing the polarizer. The higher noise level of this system is clearly illustrated in Fig.7.4a where it can be seen to exhibit a sinusoidal frequency in the  $kHz$  region, dependent on the driving DAC value of the laser. A final advantage of the diode system is the ability to exchange diodes to change wavelength of the laser, to optimize performance of the transmission measurements for a given nanoaperture geometry while retaining high output powers. Low system costs also enables easy duplication in the case when multiple beams are required.

### 7.1.2 Data acquisition and automation

To control the various elements of the microscope an automation python codebase was developed, with each instrument mapped to character devices in */dev*, similar to the TDC. A per instrument daemon process controls communication with the character device providing an intermediate abstraction layer supporting interprocess communication (IPC). This acts to decouple the instrument control from data logging, post-processing and user interfaces. The

## Nanoaperture scattering

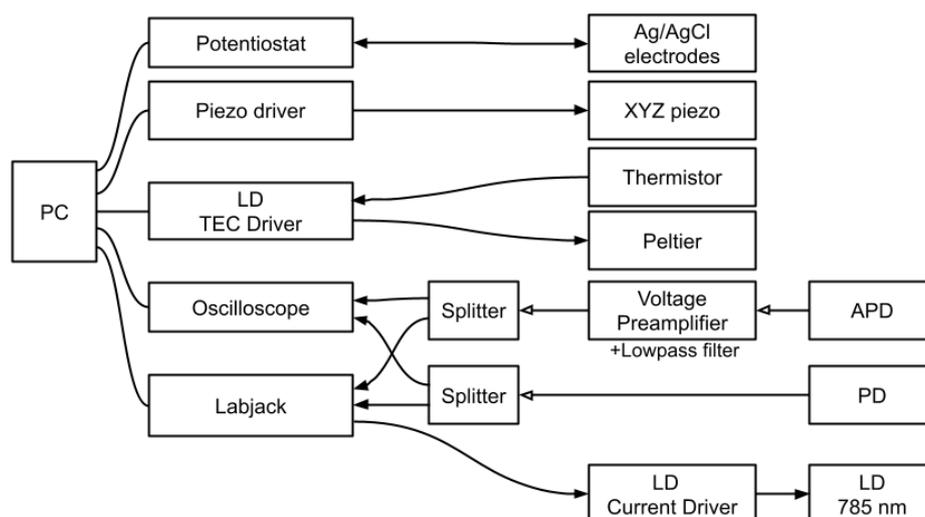


Fig. 7.5 Measurement system control and data acquisition diagram showing key elements used to record signals, control nanoaperture position, applied potential across electrodes and laser source.

decoupling enables each instrument to be separately restarted during a run without affecting others instruments or the UI. Streaming data visualization during an experimental run is implemented using a python Bokeh application running in a browser removing reliance on UI libraries like Qt.

An overview of the instruments controlled from the host machine is presented in Fig.7.5. Key components are the photodiode (PD) and avalanche photodiode (APD) detectors measuring incident and transmitted power fluctuations. Both are connected to the oscilloscope and high speed visualization of the signals. Electronic power splitters are used to simultaneously route each of the signals to a Labjack DAQ, capable of continuous ('stream') recording of up to 100 ksp/s (combined) from all channels, 100/N ksp/s for N channels, using a  $\pm 10V$  amplifier range of the DAQ. The voltage preamplifier (Stanford Research Systems, SR560) is introduced to act as a buffer, filter and amplifier between the APD and DAQ. The primary use of the preamplifier is provide a high termination impedance to the APD signal, required to exceed  $10k\Omega$  to prevent signal ringing. The ability to low-pass filter the APD (bandwidth of 1 MHz) to the range of  $< 50kHz$  removes noise contributions frequencies above the Nyquist sampling rate, with the roll-off in the low-pass set of 12 db/decade. Quantization due to finite ADC resolution (16 bits, resolution index=1, gain = 1, inter-channel time delay= $13\mu s$ ) can be reduced through using the SR560 to amplify the recorded signal.

A key aspect of the automation is the need for fine optimization of nanoaperture position in the focal point of the incident beam. This guarantees this maximizes transmitted signal along

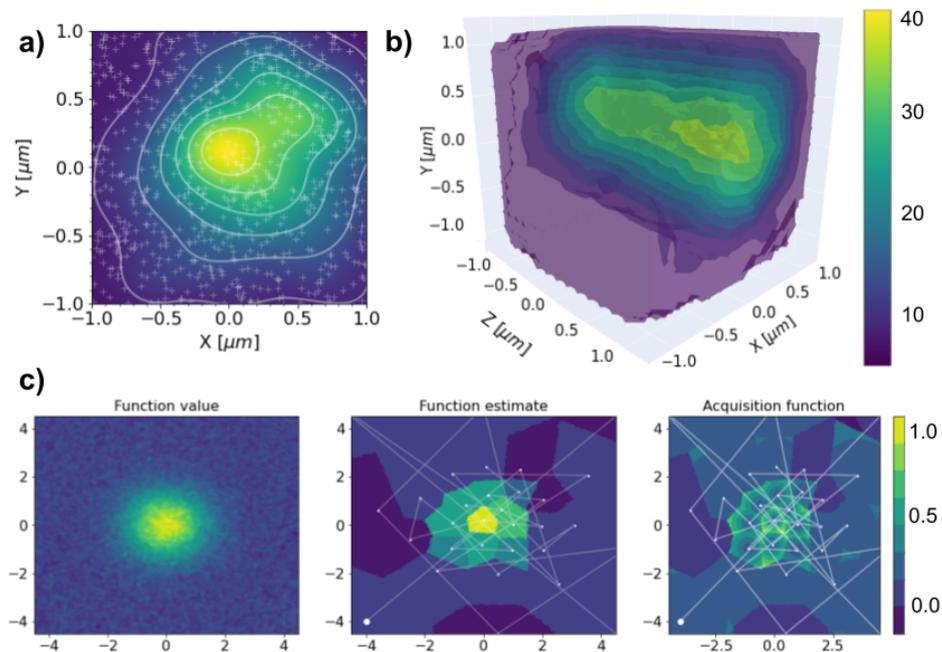


Fig. 7.6 a) Two dimension scan of a nanoaperture intensity distribution, white crosses indicate sample points. b) 3D isocontour rendering of nanoaperture intensity distribution. c) Model based alignment of nanoaperture optimizing a synthetic Gaussian objective in presence of measurement noise, showing the current function estimate and the acquisition function, maximized to select the next sample point.

with preventing artificial intensity increases due to system drifting through the focus. The current version implements a sequential line scan method, maximizing mean signal, queried from the oscilloscope. To explore the local region efficiently, a random scan mode is also implemented to map the transmitted intensity distribution and characterize the transmission as a function of nanoaperture position (Fig.7.6a,b). Future implementation of a robust fully automatic realignment procedure will enabled accurate drift quantification over extended time periods, and enable rapid automated transitions between different nanoapertures within a  $20 \times 20 \mu m$  area, limited by piezo range. An example of preliminary implementation based on active learning is illustrated in Fig.7.6c [196] where a piecewise constant function estimates transmission intensity. The selection of the next position is performed by numerically maximizing an acquisition function. This procedure can be adapted to both automate alignment as well as explore the intensity distribution as a function of experiment parameters such as polarization, opening the door to data driven classification of the nanoaperture and optimization of experiment to maximize signal.

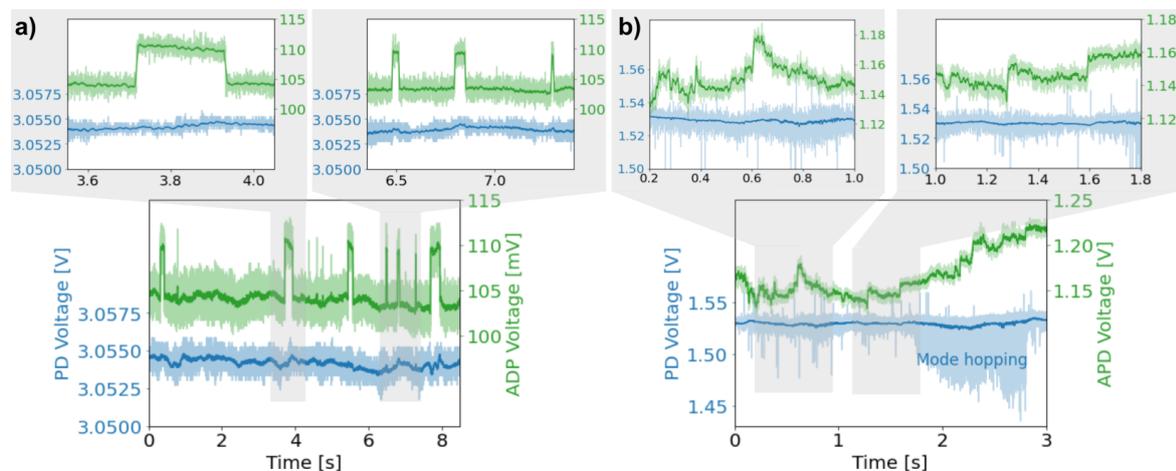


Fig. 7.7 Nanoaperture transmission (green) and monitor photodiode (blue) signals showing increase in transmission signal of the nanoaperture relative to the baseline intensity. In both traces the transmission signal exhibits fast step changes not seen in monitor photodiode signal. a) Nanoaperture transmission intensity changes by  $\sim 6\%$  relative to stable baseline, suggesting translocation of nanoparticles. b) Jumps in transmission ( $\sim 2\%$ ) with changing baseline, suggesting more complex interaction of nanoparticles with the aperture as intensity does not return to stable baseline.

## 7.2 Experimental results

Loading the fluidic chip with a solution (10-150 mM KCl concentration) containing polystyrene nanoparticles (Fluosphere F8801/F8783, Carboxylate coated, diluted number concentrations of  $10^{10}$  -  $10^{12}$  particles/mL) we detect intensity modulations in transmission Fig.7.7 (APD, green) on the APD, relative to the monitor photodiode intensity (PD,blue). The amplitude of modulations observed is in the 1-6% range. In Fig.7.7a baseline intensity drift is negligible, with discrete events clearly observable, suggesting particle translocation through the nanoaperture. On the other hand, Fig.7.7b, shows a slow increase in baseline due to individual step-like transitions, arising perhaps from adsorption of nanoparticles to the surface. The modulation can be understood by considering the effect of a particle entering the nanoaperture as modifying the effective internal refractive index (Fig. 7.8a), making the aperture ‘optically larger’. For a 150 nm diameter aperture the shift in normalized forward scattering cross section, normalized to geometric aperture area, is shown in Fig. 7.8b where increasing the effective refractive index results in a redshift and increase of transmission as the refractive index is increased from  $n = 1.33$  (water) to  $n = 1.59$  (polystyrene). The intensity change for illumination with  $\lambda = 785$  nm, relative to the transmission at  $n = 1.33$  is shown in Fig. 7.8c where  $\Delta n = +0.019$  results in a 5% power transmission increase. While this model is coarse grained, neglecting geometry, size and position of the nanoparticle, it provides a

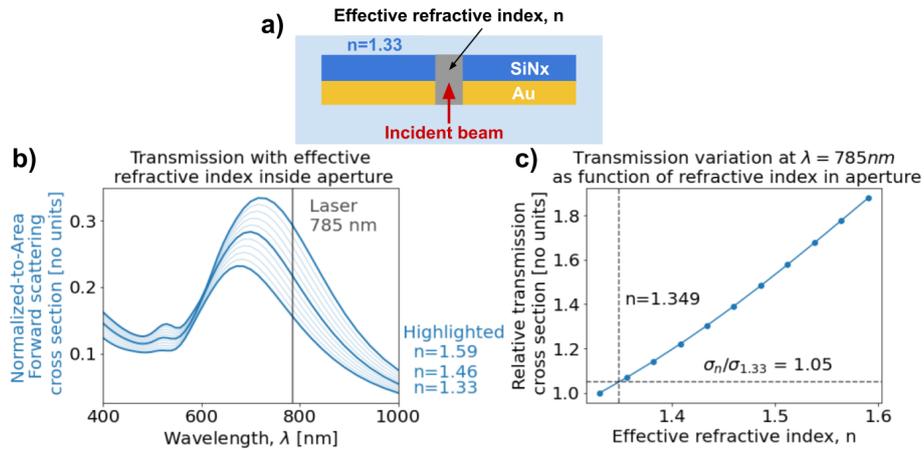


Fig. 7.8 a) An effective refractive index inside an nanoaperture surrounded by water models effects of nanoparticle translocation. b) Increasing the effective refractive index inside the cylinder defining a 150 nm diameter nanoaperture relative to water ( $n = 1.33$ ) up to the polystyrene ( $n = 1.59$ ) in a increase and redshift of transmission of the nanoaperture. c) Referencing the increase in transmission at 785 nm to the baseline transmission at  $n = 1.33$  shows that to achieve a 5% transmission increase requires a  $\Delta n = +0.019$  inside the aperture.

starting point for further investigations. Further work can quantify effects of varying particle size and position through extending the domain with a modified refractive index to include the region surrounding the nanoaperture. Iteratively subdividing this extended domain into smaller elements and perturbing their refractive indices individually would enable sensitivity analysis to identify regions contributing most to spectral changes. Cylindrical symmetry used to both speed up FDTD simulation, and reduce the number subdivision cells that need to be considered.

A key future development of the system relates to the ability to controllably initiate translocation events by inducing flows through the applied field. The increased event rate would enable quantification of timeseries features such as pulse height variation, inter-event time allowing the application of this system to characterization of nanoparticle suspensions. While robust implementation of this remains for future work, preliminary results on such control have been observed (Fig.7.9) in the system. Here the response of the system to an applied potential is observed, both in the mean signal intensity (Fig.7.9a, blue) as well as the increase in the intensity noise (red). The increase in intensity indicates an effective average refractive index increase while the increased standard deviation is an indication of refractive index modulation, potentially caused by motion of nanoparticles near/in the nanoaperture. This can be verified by computing a autocorrelation of the timeseries in regions where the noise is low (RMS Intensity  $\approx 1$  mV) or high (RMS Intensity  $\approx 2$  mV), Fig.7.9c. In the former

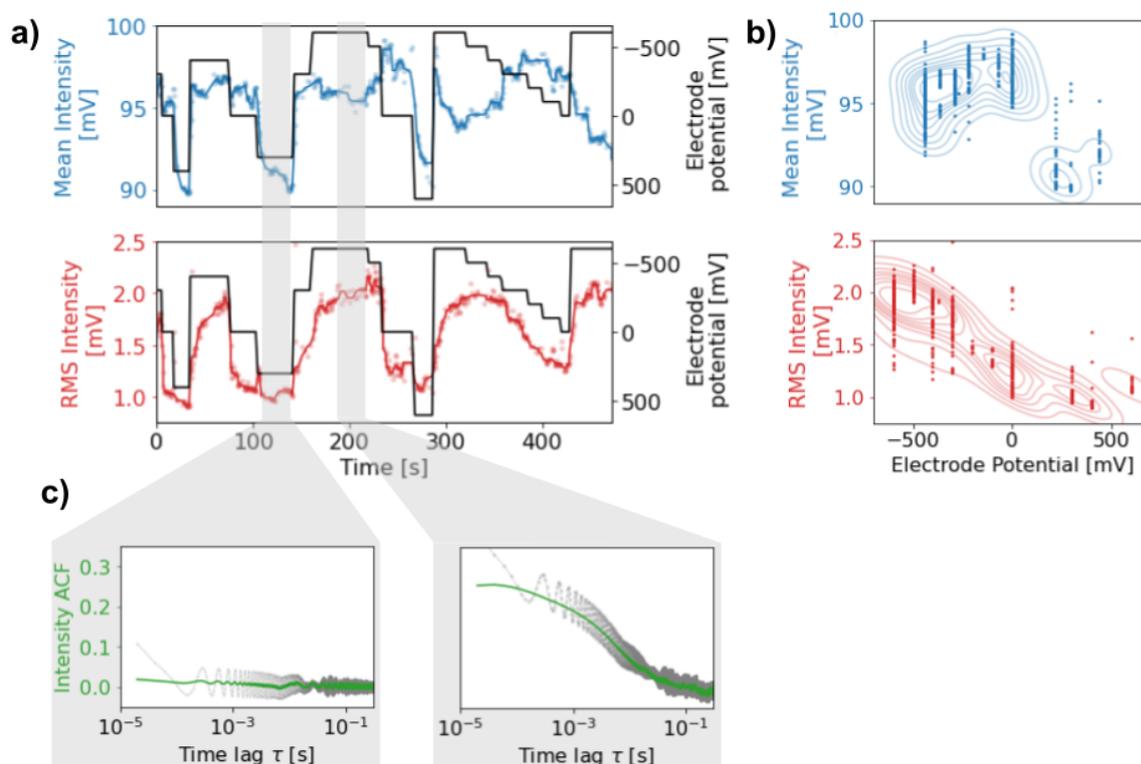


Fig. 7.9 a) Variation of intensity signal mean and standard deviation with application of electrical bias in time. b) Anti-correlation exhibited by intensity standard deviation and applied bias. c) Autocorrelation of highlighted timeseries segments showing that regions of high standard deviation also exhibit different temporal dynamics to regions of low standard deviation.

case one observes that the noise is temporally uncorrelated, while the latter case exhibits a correlation decay time  $\tau \sim 1 \times 10^{-2} s$  and thus exhibits different dynamics. With these two regimes differing only by the sign of the applied potential, the systems response to applied electrode potential is asymmetric. This is evidenced by the strong ( $\rho = -0.79$ ) anti-correlation between intensity noise and applied potential (Fig.7.9b). This asymmetry may be a sign of an ion concentration polarization (ICP) discussed in the previous chapter, arising from asymmetry of the nanoaperture cross sections or difference between surface charge density of the metal and SiNx layers. The latter hypothesis may be tested by application of a uniform coating around the nanopores by atomic layer deposition (ALD).

When monitoring the transmission intensity on the second timescale, fast switches of the transmission intensity in response to a change in applied potential have also been observed (Fig.7.10c). It is hypothesised that these rapid changes, triggered by a step change in applied potential correspond to ‘undocking’ of larger nanoparticle aggregates, sterically

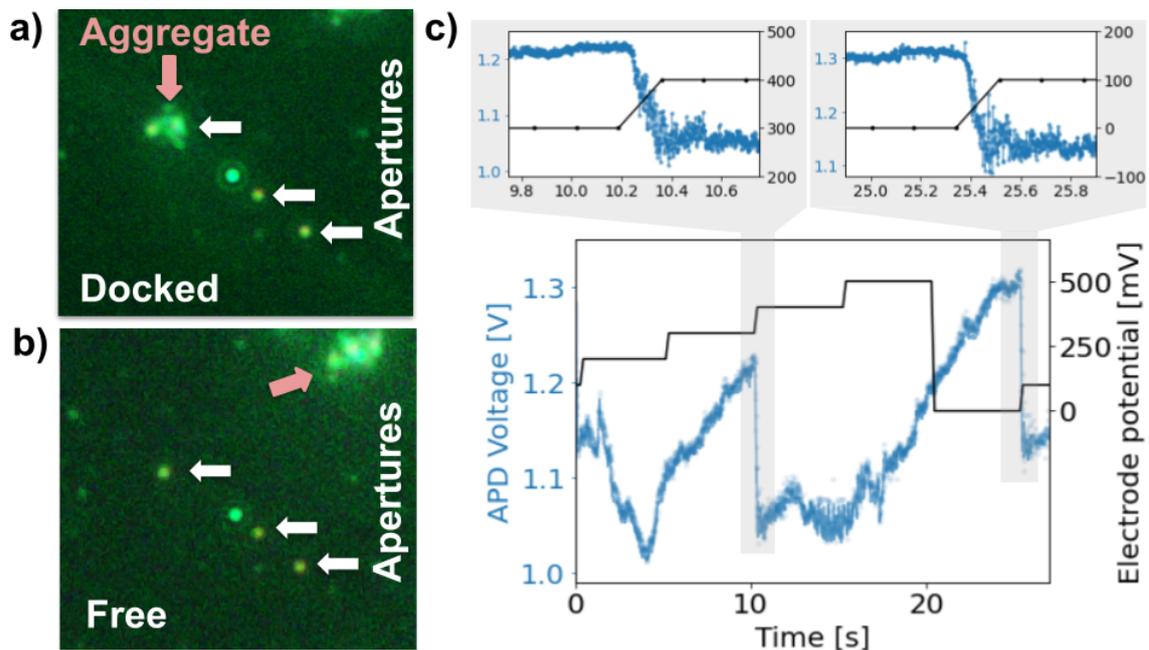


Fig. 7.10 a) Image of initially docked nanoparticle aggregate. b) Image of nanoaperture after release of the aggregate. c) Overlaid optical intensity and applied bias traces for two undocking events showing rapid intensity change upon undocking.

prevented from translocating through the nanoapertures. A set of nanoapertures when the aggregate is docked is shown in Fig.7.10a. Following application of a potential difference to the electrodes this structure is ejected from the aperture, in Fig.7.10b. A separate event, for the same nanoaperture, is shown in Fig.7.10c where the change in potential results in a rapid ( $\sim 0.1$  s) decrease of transmission intensity by  $\sim 20\%$  at  $t = 10$  s, followed by a slow increase in transmission intensity as the aggregate re-docks, and is ejected again at  $t = 25$  s. Validation of the above claim requires simultaneous video imaging and transmission recording to validate the above claim, allow ruling out of other sources of modulation such as bubble formation. Alternatively, measurement of ion current blockade created by the docking event can be used, this however requires a higher resolution than available with the current potentiostat subsystem. If verified, this mechanism could enable applications discussed in [197], where instead of polystyrene aggregates, DNA origami ‘cork’ is docked to the aperture.

Aside from effects associated with nanoparticles, illumination of the nanoapertures with the supercontinuum (bandpass filtered to  $750 \pm 10$  nm) exhibits an ‘optical charging’ effect, where upon opening of a shutter, an initial intensity increase is followed by a slower charging (Fig.7.11a). A discharging effect can also be observed after closing the shutter, exhibiting an exponentially decaying intensity, determined by the time interval between shutter closing and

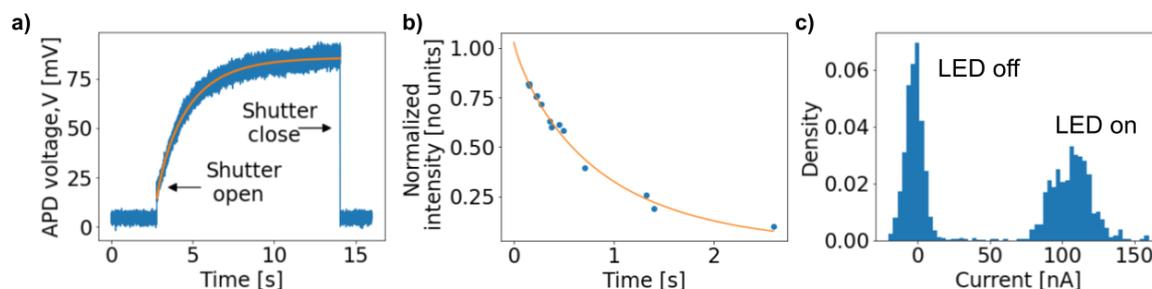


Fig. 7.11 a) Gradual increase in optical transmission, ‘optical charging’, after opening of laser shutter. b) Baseline intensity closing laser shutter. c) Electrical current induced by illuminating whole SiNx membrane with white light LED source.

opening (Fig.7.11b). Using a charging equation similar to the tradition capacitor charge, with a baseline,  $V = V_b + V_0(1 - \exp(-t^\delta/\tau))$  we can recover estimates for charging timescales of  $\tau = 0.98 \pm 0.31s^{-1}$ ,  $\delta = 0.84 \pm 0.08$ . A similar discharging model  $V/V_0 = \exp(-t^\delta/\tau)$  can be used to estimate the discharging properties  $\delta = 0.86$ ,  $\tau = 0.87s^{-1}$ . Further work to investigate this effect should consider how application of the light modifies the distribution of ions both near the surfaces (Au, SiNx) of the aperture, as well as ‘bulk’ solution inside the aperture. In the case of SiNx (with bandgap in the range of 1.85 - 5.15 eV, [198]), the charge carriers excited by absorption of the incident photons may create a net negative surface charge, balanced by an corresponding increase in potassium ion ( $K^+$ ) density inside the electrical double layer [199] leading to variation of refractive index. Alternative effects including local heating may also be partially responsible, as can the previously mentioned charge concentration polarization. In addition to observing an increase in transmitted signal upon illumination, a modulation of ionic current at zero voltage bias across the electrodes is detectable when illuminating the membrane using a white LED (Fig.7.11c) with current magnitude being in the 100 nA range when a  $10 \times 10$  mm area is illuminated.

A final consideration is the advantageous ability to free nanoapertures that become clogged during an experimental run. This can be detected electrically by estimating the ohmic resistance of the nanoapertures through measuring current at different applied voltages (Fig.7.12b) during an experimental run. Once detected, previous literature has suggested that application of a high amplitude periodic voltage to the nanoaperture system [124] to unblock the nanoapertures. We have experimented with this technique for clearing the nanoapertures but have found that for voltages  $> 5V$  produce permanent increases in nanoaperture transmission intensity, illustrated in Fig.7.12a (signal frequency 10Hz, amplitude 5V,  $KCl$  concentration 50 mM) not appearing to produce the desired unclogging effect. While higher frequency signals were considered these did not appear to produce any

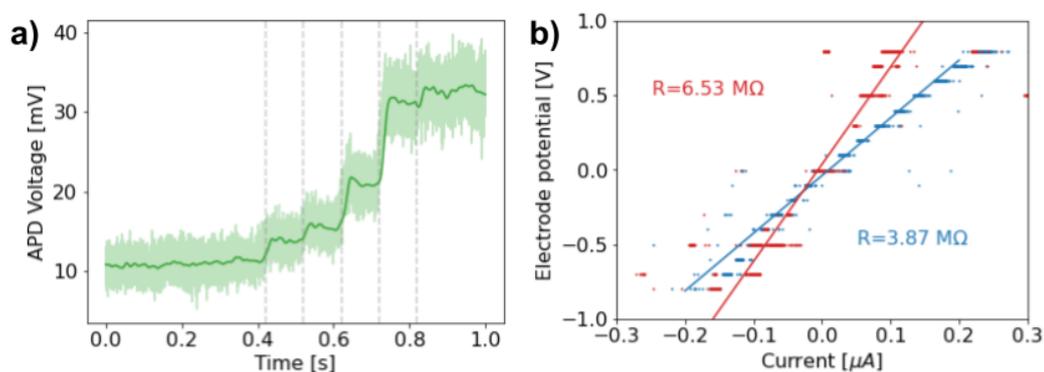


Fig. 7.12 a) Permanent pore enlargement after application of high ( $>5\text{V}$ ) voltage to electrodes, steps indicate the period of driving signal. b) Clogging of membrane during an experimental run detected by increased effective ohmic resistance.

impact on transmission. A potential cause of the permanent transmission increase is etching of the Cr adhesion layer discussed in the previous chapter, motivating the investigation of alternative materials such as Ti.

## 7.3 Summary

In this chapter we detailed the construction of an optical microscope for performing optical transmission measurements of nanoapertures and demonstrated preliminary results on detecting nanoparticle translocation and controlling of flow through the nanoaperture. We anticipate an increased degree of automation, control of experimental parameters and resolution improvements of the electrical components, integrated with a unified data acquisition system, would enable rapid characterization of nanoaperture properties, both optically and electrically. This in turn would enable rapid scanning of the nanoaperture system design parameter space in order to optimize the nanoaperture platform for high throughput single nanoparticle characterization.

The first extension to the optical detection would be integration of SPAD based photon correlation using the TDC developed in Chapter 5. Combined with pulsed excitation this enables measurement modalities like FLIM and FCS and detection of fluorescent tags. Combined with trapping this may reveal high speed dynamics of the nanoparticles near the apertures and allow the study of pairwise hydrodynamic interactions [144]. On the electrical side, deployment of the constructed ‘Dstat’ and its further integration within the DAQ would increase current resolution and high speed readout of ion current measurements, enabling electrical detection of translocation events.

# Chapter 8

## Conclusion and outlook

In this thesis, we have considered optical techniques and instrumentation for optical characterizing nanoscale systems, with a particular focus on techniques for measuring small particles, progressing from ensemble to single-particle analysis techniques.

The first of the approaches, Photon Correlation Spectroscopy, recorded autocorrelation decay times of intensity fluctuations caused by the diffusion of particles, through which the particle size distribution was estimated. While offering robust and reliable results, this traditional approach suffers from low resolution due to both a large condition number of the associated inverse problem and the quadratic scaling of scattering intensity with particle volume - resulting in the measurement being biased to larger particles. Chapter 3 of this thesis investigated the feasibility of constructing a multiangle, tunable wavelength PCS system and the development of associated computational approaches using the acquired data to increase the accuracy and resolution of the estimation of particle size distributions. The experimental component of this work underscores the critical need for tuneable high stability, narrow bandwidth laser systems with a long coherence length required to maximize the signal-to-noise ratio of each measurement. This consideration is shown to be critical in its impact on the resolution of the recovered size distribution and intimately connected to the contribution of this chapter - the development of an inference procedure for estimating particle size distributions by solving the PCS inverse problem, both its linear and nonlinear variants. The proposed method integrates Markov-Chain Monte-Carlo and optimization techniques, increasing the efficiency of posterior exploration through adaptive proposal generation, performs an automatic selection of regularization level, and provides uncertainty quantification of the recovered estimates. Where previous works report only the estimated size distributions, the current approach enables the calculation of error bounds on the recovered estimate and makes explicit the

---

connection between increasing noise and reducing measurement resolution. The increased efficiency, enabled by the utilization of optimization techniques, allows scaling of the problem to higher dimensions, thereby allowing the generalization of the inverse problem to a dynamic setting, where sequences of PCS measurements can be jointly reconstructed to estimate temporally evolving size distributions. This extends the previous approach of multiangle PCS, where multiple measurements are used to estimate a single particle size distribution. The proposed use of temporal priors, coupling information between separate measurements, results in a reduction of estimate uncertainty, as demonstrated for the linear inverse problem formulation.

Future extensions of this work on improving PCS should incorporate both the development of robust and scalable inference engines and low noise experimental instruments that minimize sources of measurement noise, the central factor affecting the resolution. To aid in the future development of such PCS systems, Chapter 5 of this work describes the design of a Time-to-Digital Converter based photon counting system for recording detected photon time series from which intensity correlation functions are computed. The availability of such TDC systems should drastically simplify the development of PCS instrumentation as it forms the core acquisition system required for PCS.

It is anticipated that while complex multimodal size distributions are likely to remain challenging, owing to properties of the corresponding inverse problems, progress can be made through algorithmic improvements relying on data-driven optimization of experimental parameters, maximizing information gained from each recording. Another potentially fruitful extension is the generalization of the approach to depolarized PCS, where extending the nonlinear inverse problem to handle the modified functional forms of correlation functions, challenging for conventional linear inverse problem frameworks, may allow more quantitative analysis of non-spherical particles. In addition to this, the integration of PCS with size separation techniques capable of pre-processing different particle sizes from a heterogeneous solution is another approach that may yield an improvement compared to the PCS configurations without such pre-separation.

Directly tracking individual particles followed by reconstructing their Brownian trajectory, termed Nanoparticle Tracking Analysis, bypasses the aforementioned limitations of PCS, such as nonlinear intensity scaling and ensemble averaging, allowing characterization resolution multimodal size distributions. While NTA uses measured intensity for particle localization, an extension of the technique to tracking time variation of intensity signals may, in the future, allow inferences about the structure of the individual particles by combining the estimated trajectory with the recorded intensity variation. One origin for intensity variation

## Conclusion and outlook

---

in scattering from complex structures is particle reorientation relative to the incident illumination. In Chapter 4 of this thesis, NTA was applied to tracking noble metal nanoparticle aggregates, whose complex fractal structure is qualitatively connected to their optical response through optical resonances of the linear chains forming the aggregates. To study such structures as they diffuse, a dual-wavelength system for particle tracking was constructed to measure the predicted transient signatures caused by particle reorientation. The intensity time series, acquired at microsecond time resolution combined with fluidic control of particle position, enabled long-term tracking of nanoparticles in the focal region. While detectable, quantitative interpretation of the transient signals and their relation to the physical properties of the nanoaggregate remains a significant challenge, partly due to the stochastic nature of these signals. Other factors responsible for the interpretation challenges are that these signals combine multiple effects - from rotational diffusion, multi-particle Mie scattering, and spatial variation of propagating mode supported by the hollow-core fibre. These considerations make even the formulation of a suitable forward model challenging. Future work may progress experimentally by simplifying the detection geometry and increasing the quality of acquired data by replacing the Streak camera used in this work with a modern high-speed imaging system, eliminating the discontinuities in the recorded time series. Alternatives such as slowing diffusion dynamics through increasing medium viscosity or inducing deterministic motion through external forcing to replace stochastic diffusion with deterministic motion are plausible directions.

Both PCS and NTA, as optical techniques, rely on the ability to rapidly acquire, process, and store high rate data streams produced by detectors like single-photon avalanche diodes or camera sensors. Historically such acquisition systems have relied on application-specific integrated circuits integrated into proprietary designs optimized for a given domain. This results in high investment requirements and often lacks customizability, constraining the set of possible measurements that can be performed with a given instrument. In turn, this restricts the range of possible experiments a scientist can perform to the set of acquisition systems they have at their disposal. One solution to address this issue, based on increased performance and lowering entry barriers to Field-Programmable Gate-Array technology, is the development of software reconfigurable modular instrumentation.

In this work, this advancement is leveraged to develop a photon counting system based on a Time-to-Digital Converter implemented inside a Xilinx Zynq SoC, described in Chapter 5, where the tapped delay line TDC is constructed from carry-logic elements. The resulting device achieves a time resolution in the tens of picosecond range, simultaneous acquisition on five channels, and continuous readout rates of up to 250 Mb/s, corresponding to a photon

---

incidence rate of 3 Mhits/second. With a measured deadtime in the 15 ns range, a potential maximum photon count rate of 66 Mhits/s per channel may in the future be sustained by the device, rivaling many commercial instruments. Such a capability greatly simplifies the construction of a range of optical instrumental, including the aforementioned PCS, owing to its ability to record individual photon arrival times, unlike older TCSPC systems that perform on-board data aggregation, allowing arbitrary offline software analysis on the raw data. Some of the applications enabled by the developed system are fluorescence correlation and lifetime imaging, super-resolution optical fluctuation imaging, quantum anti-bunching measurements, and LIDAR. While not exceeding the performance of high-end commercial instrumentation, an advantage of the developed system is its implementation on generic development boards and the ability to modify core aspects of the design without any hardware change. The modular signal conditioning front-end also enables the repurposing of the current TDC for other applications requiring high time resolution, including digital delay generation and sub-nanosecond multi-instrument synchronization, similar to the White Rabbit project [200], or for the development multi-GHz oscilloscopes.

Future work may attempt to re-use the data transfer submodules in combination with custom front-ends to develop other optical acquisition devices such as high-speed cameras, a suggested future development of NTA mentioned above. In this thesis, the developed TDC enables optical probing of local temperature changes near the nanoapertures using FLIM. At the same time, improved sub-nanosecond time resolution can permit optical probing of nanoparticle diffusion dynamics in the nanoaperture, investigated in the final chapters. Combined with optical trapping, such capabilities would enable studies of multi-particle hydrodynamic interactions and probe of single-particle rotational diffusion - particularly relevant for protein characterization [201] exhibiting rotational diffusion times in the nanosecond range. Future work may also consider developing alternative front-ends for ion current sensing, exploiting the high readout rate and high time resolution to precisely synchronize optical and ion current readout from the nanoapertures. Similarly, developing a front-end for linear imaging sensors may allow the design of high-speed spectrometers for label-free chemical characterization of the analyte particles as they diffuse near plasmonically active edges of the nanoapertures.

Detection of analytes using nanoaperture sensors in the last two chapters of this thesis considered the characterization of individual small particles as they translocate through the nanoapertures sensors giving rise to transmission intensity modulation originating from refractive index contrast between the surrounding medium and the particle. A vital advantage of this sensor over scattering-based techniques lies in its ability to detect and manipulate

## Conclusion and outlook

---

smaller particles with length scales down to the tens of nanometers without requiring high input power densities, spatial tracking, or ensemble averaging. The label-free single-particle measurements afforded by this technique make nanoaperture sensors suitable for the characterization of heterogeneous samples, while the ability to detect smaller particles compared to NTA makes the technique suitable for biological applications where particle sizes in the nanometer to tens of nanometer range are common. Having sub-wavelength size and thus not requiring spatial tracking, nanoaperture sensors enable higher detection bandwidths. At the same time, the ability to control device geometry allows optimization of sensors for a required measurement modality. The requirement for sensor fabrication also induces pitfalls for the approach, owing to the need to fabricate the sensor, control device variability, and disentangle this from the variability in analyte properties. Comparatively, scattering-based techniques such as PCS do not require frequent calibration and utilize only basic consumables making their deployment simple for routine characterization tasks, where ensemble properties are of greater interest.

The design and fabrication of nanoaperture sensors, considered in Chapter 6 of this thesis, developed the pipeline for fabrication and characterization of nanoaperture sensors, detailing steps involved in both fabrication of the nanoapertures and their subsequent integration into a designed flow cell. The proposed flow-through geometry, with two fluidic chambers separated by the membrane containing the sensors, may increase the rate of nanoparticle translocating by allowing electrical control of fluid flows through the sensor. To understand the optical response of the sensor, both in the liquid medium and to an analyte particle, the chapter presents FDTD simulation results. It compares them with spectral measurements of the sensors. The design and manufacture of the flow cell, including both optical and electrical interfacing, are also presented. To aid in future development of this geometry and the control of flows through the sensors, the chapter also details an electrochemical equivalent circuit model of the nanoaperture.

As noted in the chapter, a key challenge to further development of the technology is the reduction in nanoaperture variability, requiring the development of robust fabrication procedures combining quantitative models for in-process parameter control to reduce variability between apertures both within a single and between separate fabrication runs. Subsequent work should focus on the precise characterization of nanoaperture properties post-fabrication and, in combination with computational models, attempt to decouple the variability of nanoaperture sensors from the variability of analyte particles being detected. A promising direction for improvement in nanoaperture fabrication using ion beam milling is the replacement of Gallium ions with Helium will increase resolution and reduce surface refractive index modi-

---

fication caused by ion implantation. The effects of surface modifications on the fabricated nanoapertures' optical response and sensor performance and the analyte-surface interactions are also considered a useful direction. Finally, the microfluidic side and fabrication of flow cells followed by their integration with electronic and optical systems are another challenge identified in this work, where the development of automated design tools, broadened the range of available materials, and standardized interfaces would help enable rapid prototyping of microfluidic systems without limitations imposed by conventional PDMS-based techniques.

To apply the developed nanoaperture sensor to studying analyte particles, the sensor is interrogated using an optical transmission microscope designed and constructed in Chapter 7 - detailing the optical, electrical, and software configuration for signal acquisition and instrument control. Also presented are recorded analyte translocation events demonstrating the possibility of detecting analyte particles in the proposed configuration. Presented measurements of asymmetric correlation response of transmitted intensity magnitude and noise properties to an applied electrical potential may indicate effects of surface charge variation between the gold and silicon nitride sensor layers or be related to the nanoapertures conical cross-section left by the ion beam milling - each giving rise to asymmetric flows through the nanoaperture in response to the applied bias. In addition to this observed asymmetry, switches of applied bias are shown to be capable of ejecting larger analytes from the nanoaperture producing fast optical transient modulations. In the future, such a capability to selectively dock and undock particles from the sensor may be helpful when combined with nanomanipulation methods capable of positioning analytes on the sensor. Finally, the chapter presents measurements of 'optical charging' effects where, upon illumination and the corresponding increase of transmitted signals, the optical transmission exhibits a 'charge-up' similar to that of a capacitor, increasing beyond the initial value until a saturation level is reached. When illumination is blocked, a corresponding 'discharging' is also observed, whereupon subsequent re-illumination, the initial transmission is a function of the interval between illuminations. Further efforts to determine the origin of this effect may consider the effects of illumination on the spatial distribution of ions surrounding the nanoaperture sensors.

Building on these feasibility demonstrations should attempt to incorporate high resolution and bandwidth ion current measurements to correlate the optical and electrical signals. These are anticipated to enable robust event detection and allow identification of different modes of interactions of analytes with the nanoapertures, similar to different event classifications in nanopore sensing. As discussed above, future developments can integrate TDC-based

## Conclusion and outlook

---

photon-counting to increase the range of measurement types that the given microscope can perform, extending the current system to detect fluorescently labeled analytes and allowing characterization of nanoaperture temperature. Integrating optical trapping into the microscope would increase the interrogation time of individual analyte particles and enable probing analyte rotational diffusion times - potentially allowing the study of protein conformational changes at a single-molecule level. Augmentation with spectroscopic measurements is also possible and is already under active investigation [202]. Such chemically-sensitive methods have the potential to open the door for studies of analyte surface chemistry and, in the future, may form a key technology for next-generation sequencing.

The techniques examined in this work can be considered part of a continued evolution of optical instrumentation for probing nanoscale systems. A common feature among techniques considered, and more broadly - optical methods are that they are enabled by the continued performance improvements in electronics and are thus, with electronics, likely to continue their trajectory of increasing acquisition bandwidths, performance, and integration with other technologies. These advances, combined with continued increases in available computational resources allowing the application of more powerful statistical tools, is expected to open new possibilities for optical techniques to uncover previously hidden aspects of the world at the nanoscale.

While each of the techniques discussed in this work is aimed at particle characterization, combined, they exist on a trade-off curve where differing measurement requirements bias the choice of preferred instrument. Thus, all of the techniques are likely to see continued development and application to probing nanoscale systems. In the case of PCS, future work is likely to focus on improving the robustness and quality of size distribution estimates. Nanoparticle tracking is likely to see increased use of high-speed imaging and differing illumination schemes to improve time resolution and lower detection limits while simultaneously enriching the extracted Brownian trajectories with other feature dimensions, including multispectral techniques. Further development of nanoaperture sensing, spurred by recent commercial success and limitations of nanopore sensing, in turn, has the potential of becoming a key technology for biomolecule sensing and is expected to benefit from improvements in fabrication techniques as well as continued miniaturization and co-integration of optics, electronics, and microfluidics into lab-on-chip systems. From the viewpoint of both physics and engineering, most, if not all, of these developments can be viewed as a continued evolution rather than a revolution. Thus a new system designer is well advised to consider that which came before carefully. Everything new is but a well forgotten old.

## References

- [1] H. C. Hulst and H. C. van de Hulst, *Light scattering by small particles*. Courier Corporation, 1981.
- [2] V. Octeau, L. Cognet, L. Duchesne, D. Lasne, N. Schaeffer, D. G. Fernig, and B. Lounis, “Photothermal Absorption Correlation Spectroscopy,” *ACS Nano*, vol. 3, pp. 345–350, Feb. 2009. Publisher: American Chemical Society.
- [3] V. Tuchin, “Tissue optics and photonics: Light-tissue interaction II,” *J. Biomed. Photonics Eng*, vol. 2, no. 3, p. 030201, 2016.
- [4] H. A. Bethe, “Theory of Diffraction by Small Holes,” *Physical Review*, vol. 66, pp. 163–182, Oct. 1944. Publisher: American Physical Society.
- [5] F. J. Garcia-Vidal, L. Martin-Moreno, T. W. Ebbesen, and L. Kuipers, “Light passing through subwavelength apertures,” *Reviews of Modern Physics*, vol. 82, no. 1, p. 729, 2010. Publisher: APS.
- [6] A. Roberts, “Electromagnetic theory of diffraction by a circular aperture in a thick, perfectly conducting screen,” *JOSA A*, vol. 4, no. 10, pp. 1970–1983, 1987. ISBN: 1520-8532 Publisher: Optical Society of America.
- [7] T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, and P. A. Wolff, “Extraordinary optical transmission through sub-wavelength hole arrays,” *nature*, vol. 391, no. 6668, pp. 667–669, 1998. ISBN: 1476-4687 Publisher: Nature Publishing Group.
- [8] A. Y. Nikitin, D. Zueco, F. J. García-Vidal, and L. Martín-Moreno, “Electromagnetic wave transmission through a small hole in a perfect electric conductor of finite thickness,” *Physical Review B*, vol. 78, no. 16, p. 165429, 2008. Publisher: APS.
- [9] M. L. Juan, R. Gordon, Y. Pang, F. Eftekhari, and R. Quidant, “Self-induced back-action optical trapping of dielectric nanoparticles,” *Nature Physics*, vol. 5, pp. 915–919, Dec. 2009.
- [10] S. Schmid, P. Stömmmer, H. Dietz, and C. Dekker, “Nanopore electro-osmotic trap for the label-free study of single proteins and their conformations,” *bioRxiv*, p. 2021.03.09.434634, Jan. 2021.
- [11] X. Shi, R. Gao, Y.-L. Ying, W. Si, Y.-F. Chen, and Y.-T. Long, “A Scattering Nanopore for Single Nanoentity Sensing,” *ACS Sensors*, vol. 1, pp. 1086–1090, Sept. 2016. Publisher: American Chemical Society.

## References

---

- [12] N. Hacoheh, C. J. X. Ip, and R. Gordon, "Analysis of Egg White Protein Composition with Double Nanohole Optical Tweezers," *ACS Omega*, vol. 3, pp. 5266–5272, May 2018. Publisher: American Chemical Society.
- [13] M. P. Jonsson and C. Dekker, "Plasmonic Nanopore for Electrical Profiling of Optical Intensity Landscapes," *Nano Letters*, vol. 13, pp. 1029–1033, Mar. 2013. Publisher: American Chemical Society.
- [14] J. W. Goodman, *Statistical optics*. John Wiley & Sons, 2015.
- [15] T. A. Laurence, A. N. Kapanidis, X. Kong, D. S. Chemla, and S. Weiss, "Photon Arrival-Time Interval Distribution (PAID): A Novel Tool for Analyzing Molecular Interactions," *The Journal of Physical Chemistry B*, vol. 108, pp. 3051–3067, Mar. 2004. Publisher: American Chemical Society.
- [16] R. G. Gallager, *Discrete stochastic processes*, vol. 321. Springer Science & Business Media, 2012.
- [17] M. J. Rhodes, *Introduction to particle technology*. John Wiley & Sons, 2008.
- [18] A. Malloy and B. Carr, "NanoParticle tracking analysis–The halo™ system," *Particle & Particle Systems Characterization*, vol. 23, no. 2, pp. 197–204, 2006. ISBN: 0934-0866 Publisher: Wiley Online Library.
- [19] V. Filipe, A. Hawe, and W. Jiskoot, "Critical evaluation of Nanoparticle Tracking Analysis (NTA) by NanoSight for the measurement of nanoparticles and protein aggregates," *Pharmaceutical Research*, vol. 27, pp. 796–810, May 2010.
- [20] R. Pecora, *Dynamic light scattering: applications of photon correlation spectroscopy*. Springer Science & Business Media, 2013.
- [21] A. Delgado, F. González-Caballero, R. Hunter, L. Koopal, and J. Lyklema, "Measurement and interpretation of electrokinetic phenomena," *Elkin 06, International Electrokinetics Conference, June 25-29, Nancy, France*, vol. 309, pp. 194–224, May 2007.
- [22] M. H. Bles, "Foundations of colloid science - R.J. Hunter, second ed., clarendon press, oxford, 2001. 806 pp., ISBN 0-19-850502-7; @\$ 60.00," *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, vol. 210, p. 125, 2002.
- [23] N. Sekar and R. P. Ramasamy, "Electrochemical impedance spectroscopy for microbial fuel cell characterization," *J. Microb. Biochem. Technol. S*, vol. 6, no. 2, 2013.
- [24] K. C. E. Tjon and J. Yuan, "Impedance characterization of silver/silver chloride micro-electrodes for bio-sensing applications," *Electrochimica Acta*, vol. 320, p. 134638, Oct. 2019.
- [25] P. R. F. Rocha, P. Schlett, U. Kintzel, V. Mailänder, L. K. J. Vandamme, G. Zeck, H. L. Gomes, F. Biscarini, and D. M. de Leeuw, "Electrochemical noise and impedance of Au electrode/electrolyte interfaces enabling extracellular detection of glioma cell populations," *Scientific Reports*, vol. 6, p. 34843, Oct. 2016.

- 
- [26] T. Gwanoya, “Quantitative Risk Management: Concepts, Techniques, Tools. By Alexander J. McNeil, Rüdiger Frey & Paul Embrechts (Princeton University Press, 2005),” *Annals of Actuarial Science*, vol. 2, no. 1, pp. 187–189, 2007. ISBN: 1748-5002 Publisher: Cambridge University Press.
- [27] C. E. Rasmussen, “Gaussian processes in machine learning,” in *Summer school on machine learning*, pp. 63–71, Springer, 2003.
- [28] K. F. Riley, M. P. Hobson, and S. J. Bence, *Mathematical methods for physics and engineering*. American Association of Physics Teachers, 1999.
- [29] A. J. F. Siegert, *On the fluctuations in signals returned by many independently moving scatterers*. Radiation Laboratory, Massachusetts Institute of Technology, 1943.
- [30] P.-A. Lemieux and D. J. Durian, “Investigating non-Gaussian scattering processes by using nth-order intensity correlation functions,” *Journal of the Optical Society of America A*, vol. 16, pp. 1651–1664, July 1999. Publisher: OSA.
- [31] H. Ruf, “Treatment of Contributions of Dust to Dynamic Light Scattering Data,” *Langmuir*, vol. 18, pp. 3804–3814, May 2002. Publisher: American Chemical Society.
- [32] J. Rička, “Dynamic light scattering with single-mode and multimode receivers,” *Applied optics*, vol. 32, no. 15, pp. 2860–2875, 1993. ISBN: 2155-3165 Publisher: Optical Society of America.
- [33] R. Pecora, “Dynamic light scattering measurement of nanometer particles in liquids,” *Journal of nanoparticle research*, vol. 2, no. 2, pp. 123–131, 2000. ISBN: 1572-896X Publisher: Springer.
- [34] I. D. Block and F. Scheffold, “Modulated 3D cross-correlation light scattering: Improving turbid sample characterization,” *Review of Scientific Instruments*, vol. 81, p. 123107, Dec. 2010. Publisher: American Institute of Physics.
- [35] A. D. Levin, E. A. Shmytkova, and B. N. Khlebtsov, “Multipolarization Dynamic Light Scattering of Nonspherical Nanoparticles in Solution,” *The Journal of Physical Chemistry C*, vol. 121, pp. 3070–3077, Feb. 2017. Publisher: American Chemical Society.
- [36] A. R. Karow, J. Götzl, and P. Garidel, “Resolving power of dynamic light scattering for protein and polystyrene nanoparticles,” *Pharmaceutical development and technology*, vol. 20, no. 1, pp. 84–89, 2015. ISBN: 1083-7450 Publisher: Taylor & Francis.
- [37] A. Manzhirov and A. Polyanin, *Methods of solutions of integral equations: handbook*. Factorial, 1999.
- [38] J. G. McWhirter and E. R. Pike, “On the numerical inversion of the Laplace transform and similar Fredholm integral equations of the first kind,” *Journal of Physics A: Mathematical and General*, vol. 11, pp. 1729–1745, Sept. 1978. Publisher: IOP Publishing.
- [39] P. C. Hansen, *Discrete inverse problems: insight and algorithms*. SIAM, 2010.

## References

---

- [40] P. C. Hansen, “The discrete Picard condition for discrete ill-posed problems,” *BIT Numerical Mathematics*, vol. 30, no. 4, pp. 658–672, 1990. ISBN: 1572-9125 Publisher: Springer.
- [41] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [42] S. W. Anzengruber and R. Ramlau, “Morozov’s discrepancy principle for Tikhonov-type functionals with nonlinear operators,” *Inverse Problems*, vol. 26, no. 2, p. 025001, 2009. ISBN: 0266-5611 Publisher: IOP Publishing.
- [43] P. C. Hansen, “Analysis of discrete ill-posed problems by means of the L-curve,” *SIAM review*, vol. 34, no. 4, pp. 561–580, 1992. ISBN: 0036-1445 Publisher: SIAM.
- [44] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, *Handbook of markov chain monte carlo*. CRC press, 2011.
- [45] J. M. Bardsley and C. Fox, “An MCMC method for uncertainty quantification in nonnegativity constrained inverse problems,” *Inverse Problems in Science and Engineering*, vol. 20, no. 4, pp. 477–498, 2012. ISBN: 1741-5985 Publisher: Taylor & Francis.
- [46] J. M. Bardsley and P. C. Hansen, “MCMC algorithms for computational UQ of nonnegativity constrained linear inverse problems,” *SIAM Journal on Scientific Computing*, vol. 42, no. 2, pp. A1269–A1288, 2020. ISBN: 1064-8275 Publisher: SIAM.
- [47] J. Bardsley and T. Cui, “Optimization-based MCMC methods for nonlinear hierarchical statistical inverse problems,” *arXiv preprint arXiv:2002.06358*, 2020.
- [48] J. M. Bardsley, A. Solonen, H. Haario, and M. Laine, “Randomize-then-optimize: A method for sampling from posterior distributions in nonlinear inverse problems,” *SIAM Journal on Scientific Computing*, vol. 36, no. 4, pp. A1895–A1910, 2014. ISBN: 1064-8275 Publisher: SIAM.
- [49] J. M. Bardsley, T. Cui, Y. M. Marzouk, and Z. Wang, “Scalable optimization-based sampling on function space,” *SIAM Journal on Scientific Computing*, vol. 42, no. 2, pp. A1317–A1347, 2020. ISBN: 1064-8275 Publisher: SIAM.
- [50] A. G. Mailer, P. S. Clegg, and P. N. Pusey, “Particle sizing by dynamic light scattering: non-linear cumulant analysis,” *Journal of Physics: Condensed Matter*, vol. 27, no. 14, p. 145102, 2015. ISBN: 0953-8984 Publisher: IOP Publishing.
- [51] B. E. Saleh and M. C. Teich, *Fundamentals of photonics*. John Wiley & sons, 2019.
- [52] S. W. Provencher, “CONTIN: a general purpose constrained regularization program for inverting noisy linear algebraic and integral equations,” *Computer Physics Communications*, vol. 27, no. 3, pp. 229–242, 1982.
- [53] J. Austin, C. Minelli, D. Hamilton, M. Wywijas, and H. J. Jones, “Nanoparticle number concentration measurements by multi-angle dynamic light scattering,” *Journal of Nanoparticle Research*, vol. 22, no. 5, pp. 1–15, 2020. Publisher: Springer.

- 
- [54] A. e. Scotti, W. Liu, J. Hyatt, E. Herman, H. Choi, J. Kim, L. Lyon, U. Gasser, and A. Fernandez-Nieves, “The CONTIN algorithm and its application to determine the size distribution of microgel suspensions,” *The Journal of chemical physics*, vol. 142, no. 23, p. 234905, 2015. Publisher: AIP Publishing LLC.
- [55] S. Schintke and E. Frau, “Modulated 3D cross-correlation dynamic light scattering applications for optical biosensing and time-dependent monitoring of nanoparticle-biofluid interactions,” *Applied Sciences*, vol. 10, no. 24, p. 8969, 2020. Publisher: Multidisciplinary Digital Publishing Institute.
- [56] M. Grant, S. Boyd, and Y. Ye, “Disciplined Convex Programming,” in *Global Optimization: From Theory to Implementation* (L. Liberti and N. Maculan, eds.), pp. 155–210, Boston, MA: Springer US, 2006.
- [57] L. A. Clementi, J. R. Vega, L. M. Gugliotta, and H. R. Orlande, “A Bayesian inversion method for estimating the particle size distribution of latexes from multiangle dynamic light scattering measurements,” *Chemometrics and Intelligent Laboratory Systems*, vol. 107, no. 1, pp. 165–173, 2011. ISBN: 0169-7439 Publisher: Elsevier.
- [58] M. Naiim, A. Boualem, C. Ferre, M. Jabloun, A. Jalocho, and P. Ravier, “Multiangle dynamic light scattering for the improvement of multimodal particle size distribution measurements,” *Soft matter*, vol. 11, no. 1, pp. 28–32, 2015. Publisher: Royal Society of Chemistry.
- [59] H. Takahasi and M. Mori, “Double exponential formulas for numerical integration,” *Publications of the Research Institute for Mathematical Sciences*, vol. 9, no. 3, pp. 721–741, 1974. ISBN: 0034-5318 Publisher: Research Institute for Mathematical Sciences.
- [60] L. Jianchun, G. A. Pope, and K. Sepehrnoori, “A high-resolution finite-difference scheme for nonuniform grids,” *Applied Mathematical Modelling*, vol. 19, pp. 162–172, Mar. 1995.
- [61] S. Diamond and S. Boyd, “CVXPY: A Python-embedded modeling language for convex optimization,” *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2909–2913, 2016. ISBN: 1532-4435 Publisher: JMLR. org.
- [62] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg, “Scikit-learn: Machine learning in Python,” *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011. ISBN: 1532-4435 Publisher: JMLR. org.
- [63] P. Virtanen, R. Gommers, and SciPy 1.0 Contributors, “SciPy 1.0: fundamental algorithms for scientific computing in Python,” *Nature Methods*, vol. 17, pp. 261–272, Mar. 2020.
- [64] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, “JAX: composable transformations of Python+NumPy programs,” 2018.
- [65] Alexander D. Levin and Ekaterina A. Shmytkova, “Nonspherical nanoparticles characterization by partially depolarized dynamic light scattering,” vol. 9526, June 2015.

## References

---

- [66] R. Esteban, R. W. Taylor, J. J. Baumberg, and J. Aizpurua, “How Chain Plasmons Govern the Optical Response in Strongly Interacting Self-Assembled Metallic Clusters of Nanoparticles,” *Langmuir*, vol. 28, pp. 8881–8890, June 2012. Publisher: American Chemical Society.
- [67] M. X. Fernandes and J. G. de la Torre, “Brownian Dynamics Simulation of Rigid Particles of Arbitrary Shape in External Fields,” *Biophysical Journal*, vol. 83, pp. 3039–3048, Dec. 2002.
- [68] P. Zipper and H. Durchschlag, “Hydrodynamic multibead modeling: problems, pitfalls and solutions. 3. comparison of new approaches for improved predictions of translational properties,” *European Biophysics Journal*, vol. 42, no. 7, pp. 559–573, 2013. ISBN: 0175-7571 Publisher: Springer.
- [69] Y.-I. Xu and B. Gustafson, “A generalized multiparticle Mie-solution: further experimental verification,” *Light Scattering by Non-Spherical Particles*, vol. 70, pp. 395–419, Aug. 2001.
- [70] G. Pellegrini, G. Mattei, V. Bello, and P. Mazzoldi, “Interacting metal nanoparticles: Optical properties from nanoparticle dimers to core-satellite systems,” *EMRS 2006 Symposium A: Current Trends in Nanoscience - from Materials to Applications*, vol. 27, pp. 1347–1350, Sept. 2007.
- [71] A. M. Cubillas, S. Unterkofler, T. G. Euser, B. J. Etzold, A. C. Jones, P. J. Sadler, P. Wasserscheid, and P. S. J. Russell, “Photonic crystal fibres for chemical sensing and photochemistry,” *Chemical Society Reviews*, vol. 42, no. 22, pp. 8629–8648, 2013. Publisher: Royal Society of Chemistry.
- [72] T. G. Euser, M. K. Garbos, J. S. Y. Chen, and P. S. Russell, “Precise balancing of viscous and radiation forces on a particle in liquid-filled photonic bandgap fiber,” *Optics Letters*, vol. 34, pp. 3674–3676, Dec. 2009. Publisher: OSA.
- [73] N. M. Litchinitser, A. K. Abeeluck, C. Headley, and B. J. Eggleton, “Antiresonant reflecting photonic crystal optical waveguides,” *Optics Letters*, vol. 27, pp. 1592–1594, Sept. 2002. Publisher: OSA.
- [74] A. Ruskuc, P. Koehler, M. A. Weber, A. Andres-Arroyo, M. H. Frosz, P. S. Russell, and T. G. Euser, “Excitation of higher-order modes in optofluidic photonic crystal fiber,” *Optics Express*, vol. 26, pp. 30245–30254, Nov. 2018. Publisher: OSA.
- [75] U. Cherubini, E. Luciano, and W. Vecchiato, *Copula methods in finance*. John Wiley & Sons, 2004.
- [76] T. Nagler, “kdecopula: An R Package for the Kernel Estimation of Bivariate Copula Densities,” *arXiv:1603.04229 [stat]*, May 2017. arXiv: 1603.04229.
- [77] M. Loumagne, P. Vasanthakumar, A. Richard, and A. Débarre, “Time-of-Flight Photon Spectroscopy: A Simple Scheme To Monitor Simultaneously Spectral and Temporal Fluctuations of Emission on Single Nanoparticles,” *ACS Nano*, vol. 6, pp. 10512–10523, Dec. 2012. Publisher: American Chemical Society.

- [78] Q. Jiang, B. Rogez, J.-B. Claude, G. Baffou, and J. Wenger, “Temperature Measurement in Plasmonic Nanoapertures Used for Optical Trapping,” *ACS Photonics*, vol. 6, pp. 1763–1773, July 2019. Publisher: American Chemical Society.
- [79] T. Dertinger, R. Colyer, G. Iyer, S. Weiss, and J. Enderlein, “Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI),” *Proceedings of the National Academy of Sciences*, vol. 106, p. 22287, Dec. 2009.
- [80] M. P. D. MPD, “PDM Series SPAD,” Nov. 2019.
- [81] T. Hakamata, H. Kume, K. Okano, K. Tomiyama, A. Kamiya, Y. Yoshizawa, H. Matsui, I. Otsu, T. Taguchi, and Y. Kawai, “Photomultiplier tubes: basics and applications,” *Hamamatsu Photonics KK Electron Tube Division, Hamamatsu City*, 2006.
- [82] D. Stoppa, D. Mosconi, L. Pancheri, and L. Gonzo, “Single-Photon Avalanche Diode CMOS Sensor for Time-Resolved Fluorescence Measurements,” *IEEE Sensors Journal*, vol. 9, pp. 1084–1090, Sept. 2009.
- [83] E. Vilella, J. García, O. Alonso, and A. Diéguez, “Dynamic Range Extension of a SPAD Imager Using Non-Uniformity Correction Techniques,” *IEEE Sensors Journal*, vol. 16, pp. 2988–2992, May 2016. Conference Name: IEEE Sensors Journal.
- [84] I. M. Antolovic, C. Bruschini, and E. Charbon, “Dynamic range extension for photon counting arrays,” *Optics Express*, vol. 26, pp. 22234–22248, Aug. 2018. Publisher: OSA.
- [85] F. Ceccarelli, G. Acconcia, A. Gulinatti, M. Ghioni, and I. Rech, “Fully Integrated Active Quenching Circuit Driving Custom-Technology SPADs With 6.2-ns Dead Time,” *IEEE Photonics Technology Letters*, vol. 31, pp. 102–105, Jan. 2019. Conference Name: IEEE Photonics Technology Letters.
- [86] K. Schätzel, R. Kalström, B. Stampa, and J. Ahrens, “Correction of detection-system dead-time effects on photon-correlation functions,” *Journal of the Optical Society of America B*, vol. 6, pp. 937–947, May 1989. Publisher: OSA.
- [87] S. Kasap, *Optoelectronics and photonics : principles and practices*. Upper Saddle River, NJ: Prentice Hall, 2001.
- [88] M. Moreno-García, L. Pancheri, M. Perenzoni, R. del Río, Ó. Guerra Vinuesa, and Á. Rodríguez-Vázquez, “Characterization-Based Modeling of Retriggering and Afterpulsing for Passively Quenched CMOS SPADs,” *IEEE Sensors Journal*, vol. 19, pp. 5700–5709, July 2019.
- [89] M. Zhao, L. Jin, B. Chen, Y. Ding, H. Ma, and D. Chen, “Afterpulsing and its correction in fluorescence correlation spectroscopy experiments,” *Applied Optics*, vol. 42, pp. 4031–4036, July 2003. Publisher: OSA.
- [90] T. H. Wilmshurst, *Signal recovery from noise in electronic instrumentation*. CRC Press, 1990.

## References

---

- [91] M. Liu, X. Bai, C. Hu, X. Guo, J. C. Campbell, Z. Pan, and M. M. Tashima, “Low Dark Count Rate and High Single-Photon Detection Efficiency Avalanche Photodiode in Geiger-Mode Operation,” *IEEE Photonics Technology Letters*, vol. 19, pp. 378–380, Mar. 2007.
- [92] A. F. Kirichenko, S. Sarwana, O. A. Mukhanov, I. V. Vernik, Y. Zhang, J. Kang, and J. M. Vogt, “RSFQ time digitizing system,” *IEEE Transactions on Applied Superconductivity*, vol. 11, pp. 978–981, Mar. 2001.
- [93] T. Olsson and P. Nilsson, “A digitally controlled PLL for SoC applications,” *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 751–760, May 2004.
- [94] M. Kanoun, Y. Berube-Lauziere, and R. Fontaine, “High precision time-to-amplitude converter for diffuse optical tomography applications,” in *2008 3rd International Conference on Design and Technology of Integrated Systems in Nanoscale Era*, pp. 1–4, Mar. 2008. Journal Abbreviation: 2008 3rd International Conference on Design and Technology of Integrated Systems in Nanoscale Era.
- [95] M. L. Simpson, C. L. Britton, A. L. Wintenberg, and G. R. Young, “An integrated CMOS time interval measurement system with subnanosecond resolution for the WA-98 calorimeter,” *IEEE Journal of Solid-State Circuits*, vol. 32, pp. 198–205, Feb. 1997.
- [96] R. Nutt, “Digital Time Intervalometer,” *Review of Scientific Instruments*, vol. 39, pp. 1342–1345, Sept. 1968. Publisher: American Institute of Physics.
- [97] E. Bayer and M. Traxler, “A High-Resolution ( $< 10$  ps RMS) 48-Channel Time-to-Digital Converter (TDC) Implemented in a Field Programmable Gate Array (FPGA),” *IEEE Transactions on Nuclear Science*, vol. 58, pp. 1547–1552, Aug. 2011.
- [98] S. Bourdeauducq, “A 26 ps RMS time-to-digital converter core for Spartan-6 FPGAs,” *arXiv preprint arXiv:1303.6840*, 2013.
- [99] H. Chen and D. D. Li, “Multichannel, Low Nonlinearity Time-to-Digital Converters Based on 20 and 28 nm FPGAs,” *IEEE Transactions on Industrial Electronics*, vol. 66, pp. 3265–3274, Apr. 2019.
- [100] J. Wu and Z. Shi, “The 10-ps wave union TDC: Improving FPGA TDC resolution beyond its cell delay,” in *2008 IEEE Nuclear Science Symposium Conference Record*, pp. 3440–3446, Oct. 2008. Journal Abbreviation: 2008 IEEE Nuclear Science Symposium Conference Record.
- [101] A. Neiser, J. Adamczewski-Musch, M. Hoek, W. Koenig, G. Korcyl, S. Linev, L. Maier, J. Michel, M. Palka, M. Penschuck, M. Traxler, C. Uğur, and A. Zink, “TRB3: a 264 channel high precision TDC platform and its applications,” *Journal of Instrumentation*, vol. 8, pp. C12043–C12043, Dec. 2013. Publisher: IOP Publishing.
- [102] S. S. Junnarkar, P. O’Connor, and R. Fontaine, “FPGA based self calibrating 40 picosecond resolution, wide range Time to Digital Converter,” in *2008 IEEE Nuclear Science Symposium Conference Record*, pp. 3434–3439, Oct. 2008. Journal Abbreviation: 2008 IEEE Nuclear Science Symposium Conference Record.

- 
- [103] Z. Cheng, M. J. Deen, and H. Peng, "A Low-Power Gateable Vernier Ring Oscillator Time-to-Digital Converter for Biomedical Imaging Applications," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 10, pp. 445–454, Apr. 2016.
- [104] "7 Series FPGAs Configurable Logic Block User Guide (UG474)," p. 74, 2016.
- [105] S. Polyakov and J. Peters, *Simple and Inexpensive FPGA-based Fast Multichannel Acquisition Board*. 2015.
- [106] M. Christensen and T. Richter, "Achieving reliable UDP transmission at 10 Gb/s using BSD socket for data acquisition systems," *Journal of Instrumentation*, vol. 15, pp. T09005–T09005, Sept. 2020. Publisher: IOP Publishing.
- [107] M. Bencivenni, D. Bortolotti, A. Carbone, A. Cavalli, A. Chierici, S. Dal Pra, D. De Girolamo, L. dell'Agnello, M. Donatelli, A. Fella, D. Galli, A. Ghiselli, D. Gregori, A. Italiano, R. Kumar, U. Marconi, B. Martelli, M. Mazzucato, M. Onofri, G. Peco, S. Perazzini, A. Prosperini, P. P. Ricci, E. Ronchieri, F. Rosso, D. Salomoni, V. Sapunenko, V. M. Vagnoni, R. Veraldi, M. C. Vistoli, and S. Zani, "Performance of 10 Gigabit Ethernet Using Commodity Hardware," *IEEE Transactions on Nuclear Science*, vol. 57, pp. 630–641, Apr. 2010.
- [108] B. D. Gamari, D. Zhang, R. E. Buckman, P. Milas, J. S. Denker, H. Chen, H. Li, and L. S. Goldner, "Inexpensive electronics and software for photon statistics and correlation spectroscopy," *American journal of physics*, vol. 82, no. 7, pp. 712–722, 2014. ISBN: 0002-9505 Publisher: American Association of Physics Teachers.
- [109] S. Beer, R. Ginosar, M. Priel, R. Dobkin, and A. Kolodny, "The Devolution of Synchronizers," in *2010 IEEE Symposium on Asynchronous Circuits and Systems*, pp. 94–103, May 2010. Journal Abbreviation: 2010 IEEE Symposium on Asynchronous Circuits and Systems.
- [110] S. Beer, J. Cox, T. Chaney, and D. M. Zar, "MTBF Bounds for Multistage Synchronizers," in *2013 IEEE 19th International Symposium on Asynchronous Circuits and Systems*, pp. 158–165, May 2013. Journal Abbreviation: 2013 IEEE 19th International Symposium on Asynchronous Circuits and Systems.
- [111] J. Zhou, D. J. Kinniment, C. E. Dike, G. Russell, and A. V. Yakovlev, "On-Chip Measurement of Deep Metastability in Synchronizers," *IEEE Journal of Solid-State Circuits*, vol. 43, pp. 550–557, Feb. 2008.
- [112] R. Najvirt, T. Polzer, and A. Steininger, "Measuring Metastability with Free-Running Clocks," in *2017 23rd IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC)*, pp. 18–24, May 2017. Journal Abbreviation: 2017 23rd IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC).
- [113] T. Polzer and A. Steininger, "An Approach for Efficient Metastability Characterization of FPGAs through the Designer," in *2013 IEEE 19th International Symposium on Asynchronous Circuits and Systems*, pp. 174–182, May 2013. Journal Abbreviation: 2013 IEEE 19th International Symposium on Asynchronous Circuits and Systems.

## References

---

- [114] Z. Jaworski, “Verilog HDL model based thermometer-to-binary encoder with bubble error correction,” in *2016 MIXDES - 23rd International Conference Mixed Design of Integrated Circuits and Systems*, pp. 249–254, June 2016. Journal Abbreviation: 2016 MIXDES - 23rd International Conference Mixed Design of Integrated Circuits and Systems.
- [115] M. Adamič and A. Trost, “A Fast High-Resolution Time-to-Digital Converter Implemented in a Zynq 7010 SoC,” in *2019 Austrochip Workshop on Microelectronics (Austrochip)*, pp. 29–34, IEEE, 2019.
- [116] Y. Wang, J. Kuang, C. Liu, and Q. Cao, “A 3.9-ps RMS Precision Time-to-Digital Converter Using Ones-Counter Encoding Scheme in a Kintex-7 FPGA,” *IEEE Transactions on Nuclear Science*, vol. 64, pp. 2713–2718, Oct. 2017.
- [117] E. Sail and M. Vesterbacka, “A multiplexer based decoder for flash analog-to-digital converters,” in *2004 IEEE Region 10 Conference TENCON 2004.*, vol. D, pp. 250–253 Vol. 4, Nov. 2004. Journal Abbreviation: 2004 IEEE Region 10 Conference TENCON 2004.
- [118] F. Kaess, R. Kanan, B. Hochet, and M. Declercq, “New encoding scheme for high-speed flash ADC’s,” in *1997 IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 1, pp. 5–8 vol.1, June 1997. Journal Abbreviation: 1997 IEEE International Symposium on Circuits and Systems (ISCAS).
- [119] C. E. Cummings, “Clock domain crossing (cdc) design & verification techniques using systemverilog,” *SNUG-2008, Boston*, 2008.
- [120] J. F. Wakerly, *Digital Design: Principles And Practices, 4/E*. Pearson Education India, 2008.
- [121] D. Punj, M. Mivelle, S. B. Moparthi, T. S. van Zanten, H. Rigneault, N. F. van Hulst, M. F. García-Parajó, and J. Wenger, “A plasmonic ‘antenna-in-box’ platform for enhanced single-molecule analysis at micromolar concentrations,” *Nature Nanotechnology*, vol. 8, pp. 512–516, July 2013.
- [122] D. Gérard, J. Wenger, N. Bonod, E. Popov, H. Rigneault, F. Mahdavi, S. Blair, J. Dintinger, and T. W. Ebbesen, “Nanoaperture-enhanced fluorescence: Towards higher detection rates with plasmonic metals,” *Physical Review B*, vol. 77, p. 045413, Jan. 2008. Publisher: American Physical Society.
- [123] H. Aouani, J. Wenger, D. Gérard, H. Rigneault, E. Devaux, T. W. Ebbesen, F. Mahdavi, T. Xu, and S. Blair, “Crucial Role of the Adhesion Layer on the Plasmonic Fluorescence Enhancement,” *ACS Nano*, vol. 3, pp. 2043–2048, July 2009. Publisher: American Chemical Society.
- [124] M. U. Raza, S. S. S. Peri, L.-C. Ma, S. M. Iqbal, and G. Alexandrakis, “Self-induced back action actuated nanopore electrophoresis (SANE),” *Nanotechnology*, vol. 29, no. 43, p. 435501, 2018. ISBN: 0957-4484 Publisher: IOP Publishing.
- [125] S. Jones, A. A. Al Balushi, and R. Gordon, “Raman spectroscopy of single nanoparticles in a double-nanohole optical tweezer system,” *Journal of Optics*, vol. 17, p. 102001, Sept. 2015. Publisher: IOP Publishing.

- [126] J.-A. Huang, M. Z. Mousavi, Y. Zhao, A. Hubarevich, F. Omeis, G. Giovannini, M. Schütte, D. Garoli, and F. De Angelis, “SERS discrimination of single DNA bases in single oligonucleotides by electro-plasmonic trapping,” *Nature Communications*, vol. 10, p. 5321, Nov. 2019.
- [127] J. Wenger, F. Conchonaud, J. Dintinger, L. Wawrezynieck, T. W. Ebbesen, H. Rigneault, D. Marguet, and P.-F. Lenne, “Diffusion Analysis within Single Nanometric Apertures Reveals the Ultrafine Cell Membrane Organization,” *Biophysical Journal*, vol. 92, pp. 913–919, Feb. 2007.
- [128] Y. Gorodetski, N. Shitrit, I. Bretner, V. Kleiner, and E. Hasman, “Observation of Optical Spin Symmetry Breaking in Nanoapertures,” *Nano Letters*, vol. 9, pp. 3016–3019, Aug. 2009. Publisher: American Chemical Society.
- [129] K. Li, J. Wang, W. Cai, H. He, M. Cen, J. Liu, D. Luo, Q. Mu, D. Gérard, and Y. J. Liu, “Electrically Switchable, Polarization-Sensitive Encryption Based on Aluminum Nanoaperture Arrays Integrated with Polymer-Dispersed Liquid Crystals,” *Nano Letters*, vol. 21, pp. 7183–7190, Sept. 2021. Publisher: American Chemical Society.
- [130] P. Fürjes, Z. Fekete, L. Illés, A. Tóth, G. Battistig, and R. Gyurcsányi, “Effects of the Focused Ion Beam Parameters on Nanopore Milling in Solid State Membranes,” *26th European Conference on Solid-State Transducers, EUROSENSOR 2012*, vol. 47, pp. 684–687, Jan. 2012.
- [131] A. J. Storm, J. H. Chen, X. S. Ling, H. W. Zandbergen, and C. Dekker, “Fabrication of solid-state nanopores with single-nanometre precision,” *Nature Materials*, vol. 2, pp. 537–540, Aug. 2003.
- [132] Y. Wang, Q. Chen, T. Deng, and Z. Liu, “Nanopore fabricated in pyramidal HfO<sub>2</sub> film by dielectric breakdown method,” *Applied Physics Letters*, vol. 111, p. 143103, Oct. 2017. Publisher: American Institute of Physics.
- [133] S. Pud, D. Verschueren, N. Vukovic, C. Plesa, M. P. Jonsson, and C. Dekker, “Self-Aligned Plasmonic Nanopores by Optically Controlled Dielectric Breakdown,” *Nano Letters*, vol. 15, pp. 7112–7117, Oct. 2015. Publisher: American Chemical Society.
- [134] A. L. Ravindranath, M. S. Shariatdoust, S. Mathew, and R. Gordon, “Colloidal lithography double-nanohole optical trapping of nanoparticles and proteins,” *Optics Express*, vol. 27, pp. 16184–16194, May 2019. Publisher: OSA.
- [135] T. Park, S. J. Lee, J. H. Cha, and W. Choi, “Scalable fabrication of nanopores in membranes via thermal annealing of Au nanoparticles,” *Nanoscale*, vol. 10, no. 47, pp. 22623–22634, 2018. Publisher: Royal Society of Chemistry.
- [136] Q. Chen and Z. Liu, “Fabrication and Applications of Solid-State Nanopores,” *Sensors*, vol. 19, no. 8, 2019.
- [137] P. Roy, C. Badie, J.-B. Claude, A. Barulin, A. Moreau, J. Lumeau, M. Abbarchi, L. Santinacci, and J. Wenger, “Preventing corrosion of aluminum metal with nanometer-thick films of Al<sub>2</sub>O<sub>3</sub> capped with TiO<sub>2</sub> for ultraviolet plasmonics,” 2021. arXiv: 2106.09392 [physics.optics].

## References

---

- [138] V. Tabard-Cossa, D. Trivedi, M. Wiggin, N. N. Jetha, and A. Marziali, “Noise analysis and reduction in solid-state nanopores,” *Nanotechnology*, vol. 18, p. 305505, June 2007. Publisher: IOP Publishing.
- [139] S. Liang, F. Xiang, Z. Tang, R. Nouri, X. He, M. Dong, and W. Guan, “Noise in nanopore sensors: Sources, models, reduction, and benchmarking,” *Nanotechnology and Precision Engineering*, vol. 3, pp. 9–17, Mar. 2020. Publisher: American Institute of Physics.
- [140] D. V. Verschueren, S. Pud, X. Shi, L. De Angelis, L. Kuipers, and C. Dekker, “Label-Free Optical Detection of DNA Translocations through Plasmonic Nanopores,” *ACS Nano*, vol. 13, pp. 61–70, Jan. 2019. Publisher: American Chemical Society.
- [141] P. E. Sheehan and L. J. Whitman, “Detection Limits for Nanoscale Biosensors,” *Nano Letters*, vol. 5, pp. 803–807, Apr. 2005. Publisher: American Chemical Society.
- [142] P. Chen, J. Gu, E. Brandin, Y.-R. Kim, Q. Wang, and D. Branton, “Probing Single DNA Molecule Transport Using Fabricated Nanopores,” *Nano Letters*, vol. 4, pp. 2293–2298, Nov. 2004. Publisher: American Chemical Society.
- [143] L. Qiao, M. Ignacio, and G. W. Slater, “Voltage-driven translocation: Defining a capture radius,” *The Journal of Chemical Physics*, vol. 151, p. 244902, Dec. 2019. Publisher: American Institute of Physics.
- [144] D. Verschueren, X. Shi, and C. Dekker, “Nano-Optical Tweezing of Single Proteins in Plasmonic Nanopores,” *Small Methods*, vol. 3, p. 1800465, May 2019. Publisher: John Wiley & Sons, Ltd.
- [145] S. S. S. Peri, M. K. Sabnani, M. U. Raza, S. Ghaffari, S. Gimlin, D. D. Wawro, J. S. Lee, M. J. Kim, J. Weidanz, and G. Alexandrakis, “Detection of specific antibody-ligand interactions with a self-induced back-action actuated nanopore electrophoresis sensor,” *Nanotechnology*, vol. 31, p. 085502, Nov. 2019. Publisher: IOP Publishing.
- [146] R. M. M. Smeets, U. F. Keyser, N. H. Dekker, and C. Dekker, “Noise in solid-state nanopores,” *Proceedings of the National Academy of Sciences*, vol. 105, p. 417, Jan. 2008.
- [147] C.-C. Chien, S. Shekar, D. J. Niedzwiecki, K. L. Shepard, and M. Drndić, “Single-Stranded DNA Translocation Recordings through Solid-State Nanopores on Glass Chips at 10 MHz Measurement Bandwidth,” *ACS Nano*, vol. 13, pp. 10545–10554, Sept. 2019. Publisher: American Chemical Society.
- [148] U. Joshi, V. Venkatesh, and H. Yamamoto, “Progress Report I: Fabrication of Nanopores in Silicon Nitride Membranes using Self-Assembly of PS-b-PMMA,” 2019.
- [149] W. Zhang, J. Li, L. Cao, Y. Wang, W. Guo, K. Liu, and J. Xue, “Fabrication of nanoporous silicon dioxide/silicon nitride membranes using etched ion track technique,” *Radiation Effects in Insulators*, vol. 266, pp. 3166–3169, June 2008.

- [150] A. Amar, R. L. Lozes, Y. Sasaki, J. C. Davis, and R. E. Packard, "Fabrication of sub-micron apertures in thin membranes of silicon nitride," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, vol. 11, pp. 259–262, Mar. 1993. Publisher: American Institute of Physics.
- [151] S. Roberts, A. Cole, K. C. Heasman, S. Devenport, M. D. Brown, and T. M. Bruton, "Process development of coloured LGBC solar cells for BIPV applications," in *Proc. 22nd Eur. Photovolt. Sol. Energy Conf.*, pp. 4276–4279, 2007.
- [152] D. K. Kaushik, S. K. Chattopadhyaya, and N. Nath, "Thin film thickness monitoring using a doubly oscillating quartz crystal and measurement of growth rate," *Journal of Physics E: Scientific Instruments*, vol. 14, pp. 345–348, Mar. 1981. Publisher: IOP Publishing.
- [153] P. H. Holloway, "Gold/chromium metallizations for electronic devices," *Gold Bulletin*, vol. 12, pp. 99–106, Sept. 1979.
- [154] J. Yang and O. Paul, "Fracture properties of LPCVD silicon nitride thin films from the load–deflection of long membranes," *Selected papers from Eurosenors XV*, vol. 97-98, pp. 520–526, Apr. 2002.
- [155] T. W. Orent and R. A. Wagner, "Investigation of the chemical bonding of Cr and Ti to silicon nitride," *Journal of Vacuum Science & Technology B: Microelectronics Processing and Phenomena*, vol. 1, pp. 844–849, July 1983. Publisher: American Institute of Physics.
- [156] M. Najiminaini, F. Vasefi, B. Kaminska, and J. J. Carson, "Optical resonance transmission properties of nano-hole arrays in a gold film: effect of adhesion layer," *Optics Express*, vol. 19, pp. 26186–26197, Dec. 2011. Publisher: OSA.
- [157] Q. Jiang, B. Rogez, J.-B. Claude, A. Moreau, J. Lumeau, G. Baffou, and J. Wenger, "Adhesion layer influence on controlling the local temperature in plasmonic gold nanoholes," *Nanoscale*, vol. 12, no. 4, pp. 2524–2531, 2020. Publisher: Royal Society of Chemistry.
- [158] T. C. Tisone and J. Drobek, "Diffusion in Thin Film Ti–Au, Ti–Pd, and Ti–Pt Couples," *Journal of Vacuum Science and Technology*, vol. 9, pp. 271–275, Jan. 1972. Publisher: American Vacuum Society.
- [159] I. V. Treshin, V. V. Klimov, P. N. Melentiev, and V. I. Balykin, "Optical Tamm state and extraordinary light transmission through a nanoaperture," *Physical Review A*, vol. 88, p. 023832, Aug. 2013. Publisher: American Physical Society.
- [160] M. Manocchio, M. Esposito, A. Passaseo, M. Cuscunà, and V. Tasco, "Focused Ion Beam Processing for 3D Chiral Photonics Nanostructures," *Micromachines*, vol. 12, no. 1, 2021.
- [161] J. F. Ziegler, M. Ziegler, and J. Biersack, "SRIM – The stopping and range of ions in matter (2010)," *19th International Conference on Ion Beam Analysis*, vol. 268, pp. 1818–1823, June 2010.

## References

---

- [162] C. Sadatnajafi, E. Balaur, and B. Abbey, “Bimodal Plasmonic Color Filters Enable Direct Optical Imaging of Ion Implantation in Thin Films,” *Advanced Functional Materials*, vol. n/a, p. 2009419, Feb. 2021. Publisher: John Wiley & Sons, Ltd.
- [163] J. Mao, Y. Wang, K. Appusamy, S. Guruswamy, and S. Blair, “Effect of Ga Implantation and Hole Geometry on Light Transmission through Nanohole Arrays in Al and Mg,” *The Journal of Physical Chemistry C*, vol. 122, pp. 10535–10544, May 2018. Publisher: American Chemical Society.
- [164] C. Zhang, J. Li, A. Belianinov, Z. Ma, C. K. Renshaw, and R. M. Gelfand, “Nanoaperture fabrication in ultra-smooth single-grain gold films with helium ion beam lithography,” *Nanotechnology*, vol. 31, p. 465302, Aug. 2020. Publisher: IOP Publishing.
- [165] P. Fürjes, “Precisely tailored solid state nanopores for molecule recognition,” in *2016 Symposium on Design, Test, Integration and Packaging of MEMS/MOEMS (DTIP)*, pp. 1–5, June 2016. Journal Abbreviation: 2016 Symposium on Design, Test, Integration and Packaging of MEMS/MOEMS (DTIP).
- [166] FEI, “Helios NanoLab 400/400S/400ML/600 User Operation Manual,” 2010. Publisher: FEI.
- [167] P. Fürjes, “Controlled Focused Ion Beam Milling of Composite Solid State Nanopore Arrays for Molecule Sensing,” *Micromachines*, vol. 10, no. 11, 2019.
- [168] P. Hovington, D. Drouin, and R. Gauvin, “CASINO: A new monte carlo code in C language for electron beam interaction —part I: Description of the program,” *Scanning*, vol. 19, pp. 1–14, Jan. 1997. Publisher: John Wiley & Sons, Ltd.
- [169] S. Valkealahti and R. M. Nieminen, “Monte-Carlo calculations of keV electron and positron slowing down in solids,” *Applied Physics A*, vol. 32, pp. 95–106, Oct. 1983.
- [170] H. Gümüş and A. Bentabet, “CSDA range, stopping power and mean penetration depth energy relationships in some hydrocarbons and biologic materials for 10 eV to 100 MeV with the modified Rohrlich–Carlson model,” *Applied Physics A*, vol. 123, p. 334, Apr. 2017.
- [171] M. J. Walker, D. Berman, C. Nordquist, and J. Krim, “Electrical Contact Resistance and Device Lifetime Measurements of Au-RuO<sub>2</sub>-Based RF MEMS Exposed to Hydrocarbons in Vacuum and Nitrogen Environments,” *Tribology Letters*, vol. 44, p. 305, Sept. 2011.
- [172] C. Haverkamp, K. Höflich, S. Jäckle, A. Manzoni, and S. Christiansen, “Plasmonic gold helices for the visible range fabricated by oxygen plasma purification of electron beam induced deposits,” *Nanotechnology*, vol. 28, p. 055303, Dec. 2016. Publisher: IOP Publishing.
- [173] A. Thedsakhulwong and W. Thowladda, “Removal of carbon contamination on silicon wafer surfaces by microwave oxygen plasma,” 2008.
- [174] M. P. Brenner and D. Lohse, “Dynamic Equilibrium Mechanism for Surface Nanobubble Stabilization,” *Physical Review Letters*, vol. 101, p. 214505, Nov. 2008. Publisher: American Physical Society.

- [175] H. Conrads and M. Schmidt, "Plasma generation and plasma sources," *Plasma Sources Science and Technology*, vol. 9, pp. 441–454, Oct. 2000. Publisher: IOP Publishing.
- [176] D. Berman and J. Krim, "Impact of oxygen and argon plasma exposure on the roughness of gold film surfaces," *Thin Solid Films*, vol. 520, pp. 6201–6206, July 2012.
- [177] J. W. Coburn and H. F. Winters, "Plasma etching—A discussion of mechanisms," *Journal of Vacuum Science and Technology*, vol. 16, pp. 391–403, Mar. 1979. Publisher: American Vacuum Society.
- [178] M. Yamamoto, T. Matsumae, Y. Kurashima, H. Takagi, T. Suga, T. Itoh, and E. Higurashi, "Comparison of Argon and Oxygen Plasma Treatments for Ambient Room-Temperature Wafer-Scale Au–Au Bonding Using Ultrathin Au Films," *Micromachines*, vol. 10, no. 2, 2019.
- [179] D. Yang, N. W. Phillips, K. Song, R. J. Harder, W. Cha, and F. Hofmann, "Annealing of focused ion beam damage in gold microcrystals: an in situ Bragg coherent X-ray diffraction imaging study," *Journal of Synchrotron Radiation*, vol. 28, no. 2, pp. 550–565, 2021. ISBN: 1600-5775 Type: doi:10.1107/S1600577520016264.
- [180] M. D. Abràmoff, P. J. Magalhães, and S. J. Ram, "Image processing with ImageJ," *Biophotonics international*, vol. 11, no. 7, pp. 36–42, 2004. Publisher: Laurin Publishing.
- [181] A. Zehtabi-Oskuie, J. G. Bergeron, and R. Gordon, "Flow-dependent double-nanohole optical trapping of 20nm polystyrene nanospheres," *Scientific Reports*, vol. 2, p. 966, Dec. 2012.
- [182] C. Escobedo, A. G. Brolo, R. Gordon, and D. Sinton, "Optofluidic Concentration: Plasmonic Nanostructure as Concentrator and Sensor," *Nano Letters*, vol. 12, pp. 1592–1596, Mar. 2012. Publisher: American Chemical Society.
- [183] C.-W. Tsao and D. L. DeVoe, "Bonding of thermoplastic polymer microfluidics," *Microfluidics and Nanofluidics*, vol. 6, pp. 1–16, Jan. 2009.
- [184] R. Mukhopadhyay, *When PDMS isn't the best*. ACS Publications, 2007.
- [185] P. I. Norland, "NOA 68 Datasheet," 2021. Publisher: Norland Products Inc.
- [186] P. Wägli, A. Homsy, and N. de Rooij, "Norland optical adhesive (NOA81) microchannels with adjustable wetting behavior and high chemical resistance against a range of mid-infrared-transparent organic solvents," *Sensors and Actuators B: Chemical*, vol. 156, pp. 994–1001, Aug. 2011.
- [187] A. Fragasso, S. Pud, and C. Dekker, "1/f noise in solid-state nanopores is governed by access and surface regions," *Nanotechnology*, vol. 30, no. 39, p. 395202, 2019. ISBN: 0957-4484 Publisher: IOP Publishing.
- [188] R. B. Schoch, H. van Lintel, and P. Renaud, "Effect of the surface charge on ion transport through nanoslits," *Physics of Fluids*, vol. 17, p. 100604, Oct. 2005. Publisher: American Institute of Physics.

## References

---

- [189] R. M. M. Smeets, U. F. Keyser, D. Krapf, M.-Y. Wu, N. H. Dekker, and C. Dekker, “Salt Dependence of Ion Transport and DNA Translocation through Solid-State Nanopores,” *Nano Letters*, vol. 6, pp. 89–95, Jan. 2006. Publisher: American Chemical Society.
- [190] K. Lin, Z. Li, Y. Tao, K. Li, H. Yang, J. Ma, T. Li, J. Sha, and Y. Chen, “Surface Charge Density Inside a Silicon Nitride Nanopore,” *Langmuir*, vol. 37, pp. 10521–10528, Sept. 2021. Publisher: American Chemical Society.
- [191] S. W. Kowalczyk, A. Y. Grosberg, Y. Rabin, and C. Dekker, “Modeling the conductance and DNA blockade of solid-state nanopores,” *Nanotechnology*, vol. 22, no. 31, p. 315101, 2011. ISBN: 0957-4484 Publisher: IOP Publishing.
- [192] D. R. Lide, *CRC handbook of chemistry and physics*, vol. 85. CRC press, 2004.
- [193] V. Gun’ko, L. Andriyko, V. Zarko, A. Marynin, V. Olishevskyi, and W. Janusz, “Effects of dissolved metal chlorides on the behavior of silica nanoparticles in aqueous media,” *Open Chemistry*, vol. 12, no. 4, pp. 480–491, 2014. ISBN: 2391-5420 Publisher: Versita.
- [194] W. Han and X. Chen, “A review: applications of ion transport in micro-nanofluidic systems based on ion concentration polarization,” *Journal of Chemical Technology & Biotechnology*, vol. 95, pp. 1622–1631, June 2020. Publisher: John Wiley & Sons, Ltd.
- [195] J. Yuan and X. Long, “CCD-area-based autocollimator for precision small-angle measurement,” *Review of Scientific Instruments*, vol. 74, pp. 1362–1365, Mar. 2003. Publisher: American Institute of Physics.
- [196] F. L. Fernández, L. Martino, V. Elvira, D. Delgado, and J. López-Santiago, “Adaptive Quadrature Schemes for Bayesian Inference via Active Learning,” *IEEE Access*, vol. 8, pp. 208462–208483, 2020.
- [197] S. Schmid and C. Dekker, “The NEOtrap – en route with a new single-molecule technique,” *iScience*, p. 103007, Sept. 2021.
- [198] T. Grigaitis, A. Naujokaitis, S. Tumėnas, G. Juška, and K. Arlauskas, “Characterization of silicon nitride layers deposited in three-electrode plasma-enhanced CVD chamber,” *Lithuanian Journal of Physics*, vol. 55, no. 1, 2015. ISBN: 2424-3647.
- [199] S. Mariani, L. M. Strambini, and G. Barillaro, “Electrical double layer-induced ion surface accumulation for ultrasensitive refractive index sensing with nanostructured porous silicon interferometers,” *ACS sensors*, vol. 3, no. 3, pp. 595–605, 2018. ISBN: 2379-3694 Publisher: ACS Publications.
- [200] J. Serrano, M. Lipinski, T. Wlostowski, E. Gousiou, E. van der Bij, M. Cattin, and G. Daniluk, “The white rabbit project,” 2013.
- [201] Y. Zhao, M. Iarossi, A. F. De Fazio, J.-A. Huang, and F. De Angelis, “Label-Free Optical Analysis of Biomolecules in Solid-State Nanopores: Toward Single-Molecule Protein Sequencing,” *ACS photonics*, vol. 9, no. 3, pp. 730–742, 2022. Publisher: ACS Publications.

- [202] Y. Zhao, A. Hubarevich, J.-A. Huang, and F. De Angelis, “Nanofluidic trapping and enhanced Raman detection of single biomolecules in plasmonic bowl-shaped nanopore,” *arXiv preprint arXiv:2203.06666*, 2022.

