

# Primjena strojnog učenja za korekciju artefakata u fluorescencijskoj korelacijskoj spektroskopiji (FCS)

---

**Pavlov, Ivan Emanuel**

**Master's thesis / Diplomski rad**

**2022**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:217:258792>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2025-01-26**



*Repository / Repozitorij:*

[Repository of the Faculty of Science - University of Zagreb](#)



SVEUČILIŠTE U ZAGREBU  
PRIRODOSLOVNO-MATEMATIČKI FAKULTET  
FIZIČKI ODSJEK

Ivan Emanuel Pavlov

Primjena strojnog učenja za korekciju artefakata  
u fluorescencijskoj korelacijskoj spektroskopiji  
(FCS)

Diplomski rad

Zagreb, 2022.

SVEUČILIŠTE U ZAGREBU  
PRIRODOSLOVNO-MATEMATIČKI FAKULTET  
FIZIČKI ODSJEK

INTEGRIRANI PREDDIPLOMSKI I DIPLOMSKI SVEUČILIŠNI STUDIJ  
FIZIKA; SMJER ISTRAŽIVAČKI

**Ivan Emanuel Pavlov**

Diplomski rad

**Primjena strojnog učenja za korekciju  
artefakata u fluorescencijskoj  
korelacijskoj spektroskopiji (FCS)**

Voditelj diplomskog rada: Dr. sc. Goran Zgrablić

Ocjena diplomskog rada: \_\_\_\_\_

Povjerenstvo: 1. \_\_\_\_\_

2. \_\_\_\_\_

3. \_\_\_\_\_

Datum polaganja: \_\_\_\_\_

Zagreb, 2022.

Ovaj diplomski rad napravljen je na Institutu za fiziku pod mentorstvom dr. sc. Gorana Zgrablića. Zahvaljujem se mentoru na ugodnoj suradnji te pozitivnom pristupu za vrijeme pisanja diplomskog rada. Nadam se da će sadržaj rada, eksperimentalni podaci i sve informacije prikupljene za vrijeme pisanja rada biti od koristi znanstvenoj zajednici. Također, ponajviše se zahvaljujem svojoj djevojci Margareti koja mi je bila najveća potpora kroz studij, svojoj i njezinoj obitelji, te užem krugu prijatelja i kolega.

## Sažetak

Fluorescentna korelacijska spektroskopija je etablirana tehnika spektroskopije na razini jedne molekule kojom se promatraju vremenske fluktuacije intenziteta fluorescencije biomolekula markiranih nekom fluorescentnom bojom. Fluktuacije se vremenski autokoreliraju, što potom omogućuje daljnju analizu i mjerenje parametara biomolekule od interesa. Parametri mogu biti koncentracija, difuzijska konstanta i sl. U ovom radu je kao izvor fluorescencije korištena boja Alexa647 u nevezanom obliku. Korištenjem aproksimativne teorijske autokorelacijske funkcije, koja uključuje difuzijsku i tripletnu dinamiku molekule, izračunate su dimenzije detekcijskog volumena u konfokalnom mikroskopu. Zatim su izračunati parametri fiksirani u teorijskom modelu što je omogućilo određivanje vremena difuzije i difuzijske konstante promatrane molekule. Kao alternativna metoda obrade autokorelacijskih krivulja korišteno je strojno učenje. Točnije, korištena je jednostavna neuronska mreža, trenirana na skupu uzoraka poznatih parametara. Potom su provedena predviđanja difuzijske konstante na podskupovima mjerenja koji nisu bili dio treninga modela. Strojno učenje je korišteno kao pokušaj da se odziv optičkog postava odredi izravno iz podataka. Na taj način su se pokušale izbjeći nepouzdanosti mjerenja koje dolaze od optičkih artefakata, a koji nisu predviđeni teorijskim modelima. Vrijednosti dobivene neuronskom mrežom su uspoređeni s vrijednostima dobivenim klasičnim pristupom gdje se parametri sustava dobivaju prilagodbom teorijske funkcije na podatke. Rezultati ukazuju da je pristup korištenjem strojnog učenja i s tako malim brojem podataka koji su se koristili za trening modela, barem usporediv po preciznosti s predviđanjima koje daje klasičan pristup.

Ključne riječi: spektroskopija na razini jedne molekule, fluorescentna korelacijska spektroskopija, difuzijska konstanta, strojno učenje, neuronska mreža, Alexa647, optički artefakti

# Usage of machine learning for artifact correction in Fluorescence Correlation Spectroscopy

## Abstract

Fluorescence correlation spectroscopy is an established technique in single-molecule spectroscopy used for observing temporal fluctuations of the fluorescence intensity of marked biomolecules. Fluctuations are temporally autocorrelated which enables further analysis and measurement of relevant system parameters. Examples of observables are molecule concentrations, diffusion coefficients, etc. In this master's thesis an unbound Alexa647 dye molecule was used as source of fluorescence. By using an approximate theoretical function, which incorporates diffusion dynamics and triplet dynamics of the dye, dimensions of the observational volume of the confocal microscope were calculated. The latter parameters are then fixed in the theoretical model which enabled determination of diffusion time and diffusion constant of the dye. As an alternative method, machine learning was used to process autocorrelation curves. Specifically, a neural network was trained on a dataset that included data of known parameters. After the model was trained, it was used to predict diffusion coefficients of the data in the test set which was not used during the model training. System parameters were measured directly from the data. Doing it that way enables us to potentially avoid systematic errors which come from optical artifacts of the experimental apparatus because approximate theoretical functions do not take optical artifacts into account. The results of the model are compared with the results obtained in the theoretical function fitting procedure. Obtained results signaled that even though the training set used for model training was small, the machine learning model was at least as successful in predicting diffusion coefficients as was the theoretical function fitting procedure.

Keywords: Single-Molecule Spectroscopy, Fluorescence Correlation Spectroscopy, Diffusion coefficient, Machine learning, Neural network, Alexa647, Optical artifacts

# Sadržaj

<b>1</b>	<b>Uvod</b>	<b>1</b>
<b>2</b>	<b>Teorijska razmatranja</b>	<b>3</b>
2.1	Fluorescencija . . . . .	3
2.2	Fluorescencijska korelacijska spektroskopija . . . . .	7
2.2.1	Fluktuacije intenziteta fluorescentnog zračenja . . . . .	7
2.2.2	Osnovni difuzijski model autokorelacijske funkcije . . . . .	8
2.2.3	Difuzija i viskoznost . . . . .	12
2.2.4	Tripletna dinamika . . . . .	13
2.2.5	Odabir fluorofora . . . . .	14
2.2.6	Artefakti i ograničenja . . . . .	15
2.3	Strojno učenje . . . . .	17
2.3.1	O strojnom učenju . . . . .	17
2.3.2	Neuronske mreže . . . . .	18
<b>3</b>	<b>Eksperimentalni postav</b>	<b>22</b>
3.1	Minimizacija šuma . . . . .	22
3.2	Postav . . . . .	23
3.3	Optimizacija postava . . . . .	26
<b>4</b>	<b>Rezultati</b>	<b>28</b>
4.1	Priprema uzoraka . . . . .	28
4.2	Mjerenja . . . . .	30
4.3	Određivanje efektivnog detekcijskog volumena . . . . .	31
4.4	Procjena nepouzdanosti mjerenja . . . . .	34
4.5	Prilagodba teorijske funkcije na podatke . . . . .	35
4.6	Strojno učenje - modeliranje . . . . .	36
4.6.1	Predobrada podataka . . . . .	36
4.6.2	Model i rezultati . . . . .	37
4.7	Usporedba rezultata neuronske mreže i teorijske prilagodbe podacima	39
<b>5</b>	<b>Zaključak</b>	<b>41</b>
	<b>Dodaci</b>	<b>43</b>

<b>A</b>	<b>Vizualni prikaz potpunog skupa mjerenja</b>	<b>43</b>
<b>B</b>	<b>Korišteni programski paketi</b>	<b>44</b>
B.1	NumPy . . . . .	44
B.2	pandas . . . . .	44
B.3	matplotlib . . . . .	44
B.4	scikit-learn . . . . .	45
B.5	TensorFlow . . . . .	45
B.6	Keras . . . . .	45
B.7	random . . . . .	45
B.8	LMFIT . . . . .	45
<b>C</b>	<b>Dostupnost programskog koda</b>	<b>46</b>
	<b>Literatura</b>	<b>47</b>



# 1 Uvod

Eksperimentalna tehnika u fokusu ovog rada je fluorescencijska korelacijska spektroskopija (FCS). FCS je etablirana eksperimentalna tehnika u području spektroskopije na razini jedne molekule (engl. single-molecule spectroscopy) [1]. FCS tehnika se bazira na fenomenu fluorescencije. Ideja je pratiti fluktuacije intenziteta zračenja koje dolazi od malenog broja označenih biomolekula u detekcijskom volumenu (DV-u) koji je reda veličine 1 fL ( $10^{-15}$  L). Konfokalna detekcija, koja se postiže mikroskopskim objektivima velike numeričke aperture (NA) i konfokalnim irisima (engl. confocal pinhole) mikrometarskih dimenzija, najčešći je način kako se postižu mali volumeni potrebni za mjerenja tih fluktuacija. Za označavanje biomolekula od interesa (npr. antitijela, proteina, enzima, peptida, nukleinskih kiselina i sl.) koriste se fluorescentne boje ili fluorofore. To su molekule koje fluoresciraju prilikom svjetlosne pobude. Fluorescencija omogućava utvrđivanje broja i koncentracije označenih biomolekula, njihovo ulaženje i izlaženje iz DV-a procesom difuzije, te još nekih parametara sustava koji nisu relevantni za ovaj rad. Srž FCS tehnike je vremensko autokoreliranje signala što omogućuje određivanje karakterističnog vremena difuzije, kao i određivanje broja označenih biomolekula u DV-u. Navedeni parametri sustava često su od interesa u raznim znanstvenim disciplinama (npr. biofizika, stanična biologija, fotokemija, biomedicina...) stoga je nužno koristiti metode za njihovo određivanje na osnovu mjerenja prikupljenih u laboratoriju.

Jedna od metoda kojom se određuju navedeni parametri je prilagodba teorijske autokorelacijske funkcije (AF) podacima. Teorijska funkcija aproksimira vrijednost AF-e u ovisnosti o vremenskom pomaku. Da bi se modelirala teorijska funkcija koja opisuje FCS mjerenja, potrebno je koristiti pretpostavke o fizici sustava fluorescentnih molekula pod utjecajem pobudnog laserskog zračenja, ali i za pojedini eksperimentalni postav detaljno poznavati intenzitetski profil detekcijskog volumena (engl. molecule detection efficiency function - MDF). MDF je 3D funkcija koja se definira kao umnožak normaliziranog intenzitetskog profila pobudnog snopa i prostorne efikasnosti detekcije fotona, a iznos DV-a se izračuna integracijom  $MDF(\rho, z)$  funkcije u prostoru [2]. Standardni pristup aproksimira MDF gausijanskim profilima okomito (radijalno -  $\rho$ ) i paralelno (aksijalno -  $z$ ) s optičkom osi konfokalnog mikroskopa (smjerom širenja laserskog snopa). No, u realnim primjenama postoje znatna od-

stupanja od gaussijanskih profila, što daje lošije rezultate kod prilagodbe teorijske funkcije podacima [3].

Osim modeliranja MDF-a, minimalna pretpostavka teorijskog modela mora uključiti difuziju biomolekula izvan i unutar DV-a, dok je sljedeća nadogradnja dodavanje člana tripletne fotodinamike fluorofore [4]. Ovaj pristup daje dobre rezultate kod nekih fluorofora, no neke od njih sadrže i druge izražene fotokemijske i fotofizikalne procese koji nisu uključeni u teorijski model uobičajeno korišten kod prilagodbe teorijske krivulje podacima. Također, sam eksperimentalni postav unosi optičke artefakte koji mijenjaju oblik MDF-a na način da gaussijanska aproksimacija više nije zadovoljavajuća što također nije uključeno u teorijski model. Drugim riječima, optički artefakti mjernog postava utjecali bi na vrijednosti autokorelacijske funkcije čak i da je fotodinamika fluorofore savršeno opisana teorijskim modelom.

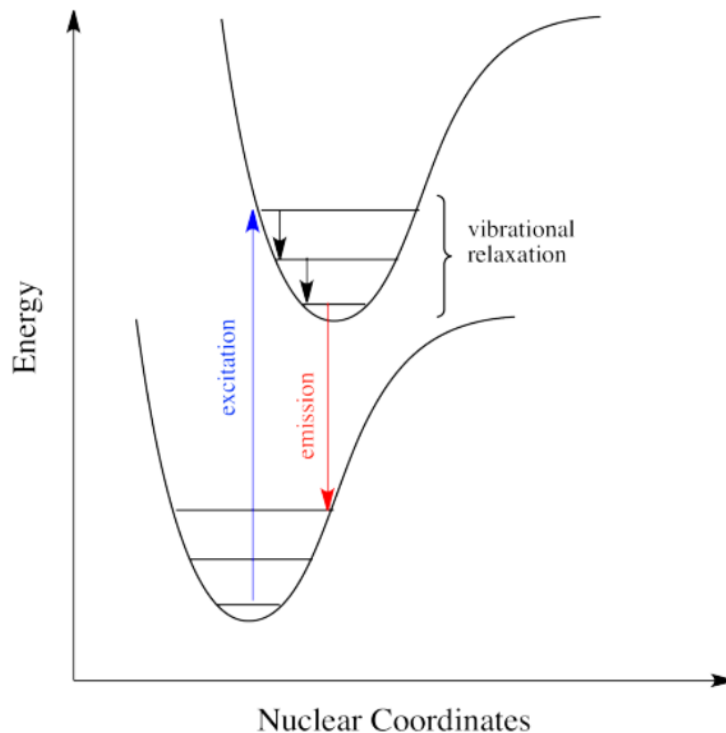
Da bi se pokušalo zaobići probleme koji dolaze od optičkih artefakata, u ovom radu korišten je alternativni pristup obradi podataka koje daje FCS: strojno učenje. Ideja iza modela nadziranog strojnog učenja je da se funkcija koja opisuje mjerene podatke nauči izravno iz podataka. Za razliku od teorijskog modela, modelu strojnog učenja nisu potrebne inicijalne pretpostavke o fotodinamici fluorofore, kao ni o obliku DV-a. Da bi se ovakav pristup isplatio, potrebno je bilo skupiti dovoljno podataka, što je nužan preduvjet za treniranje svakog modela strojnog učenja. Kalibracija mjernog postava, te skupljanje dovoljne količine podataka su se pokazali kao nemala prepreka tako da je ideja ovog rada bila napraviti rudimentarnu ekspoziciju metode. Rezultati se mogu poboljšati prikupljanjem veće količine podataka koji bi se iskoristili za robusniji trening modela.

## 2 Teorijska razmatranja

### 2.1 Fluorescencija

Fizikalni fenomen na kojemu se bazira FCS tehnika je fluorescencija [1]. Fluorescencija pripada u obitelj procesa koji se naziva luminiscencija. Pod luminiscencijom se misli na procese u kojima neka tvar, npr. molekula, emitira foton prilikom relaksiranja iz pobuđenog u osnovno elektronsko stanje. U slučaju fluorescencije molekula prelazi u pobuđeno stanje istog spina nakon čega prolaskom kroz više različitih procesa relaksacije gubi energiju. U konačnici, emitiranjem fotona fluorescencije, koji ima manju energiju u odnosu na prethodno apsorbirani foton, molekula se vraća u osnovno stanje na nanosekundnoj vremenskoj skali. Razliku između energije apsorbiranog fotona i energije emitiranog fotona nazivamo Stokesovim pomakom.

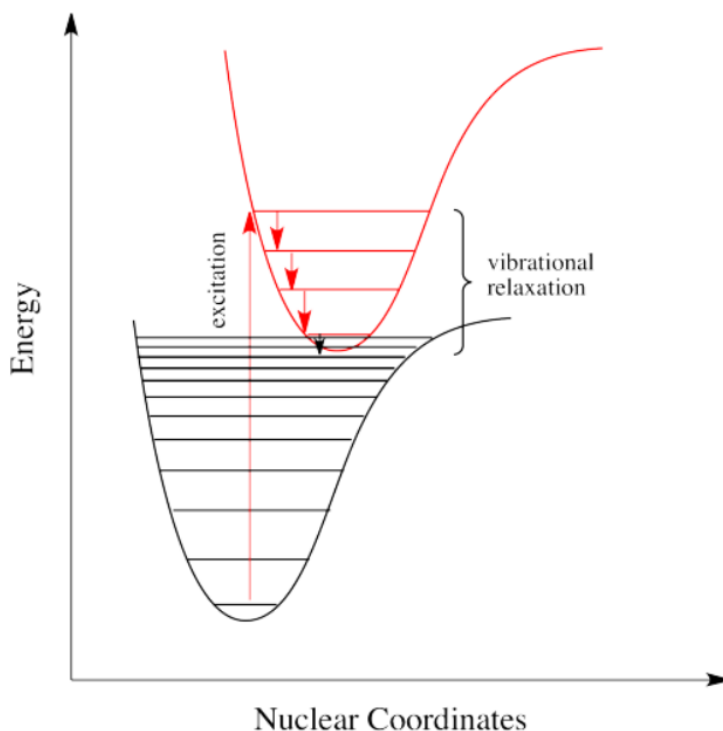
Postoji nekoliko vrsta neradijativnih procesa relaksacije kojima molekula gubi energiju. Prvi proces je proces relaksacije iz viših u niža vibracijska stanja jednog elektronskog stanja. Naime, prilikom npr. laserskog pobuđivanja molekule, molekula će najvjerojatnije doći u jedno od viših vibracijskih stanja pobuđenog elektronskog stanja (plava strelica, slika 2.1). Potom će molekula unutar nekoliko pikosekundi,



Slika 2.1: Primjer relaksiranja vibracijskih nivoa prije emisije fotona [5]

procesom vibracijske relaksacije, završiti u najnižem vibracijskom stanju pobuđenog elektronskog stanja. Nakon toga slijedi emisija fotona niže energije (za Stokesov pomak) od pobudnih fotona. Taj zadnji proces zovemo fluorescencijom (crvena strelica, slika 2.1).

No, može se dogoditi da ipak ne dođe do emisije fotona, te umjesto toga dođe do interne konverzije, što je još jedna vrsta neradijativnih procesa. Na slici 2.2 vidimo primjer slučaja gdje postoji preklapanje nižih vibracijskih stanja pobuđenog elektronskog stanja s višim vibracijskim stanjima osnovnog stanja. Što je veće preklapanje nivoa, veća je vjerojatnost da dođe do interne konverzije. Interna konverzija je ponekad i naziv za samu vibracijsku relaksaciju koja prethodi fluorescenciji, a u kojoj ne dolazi do emisije fotona kao što je i objašnjeno. Kako nema emisije fotona pri-

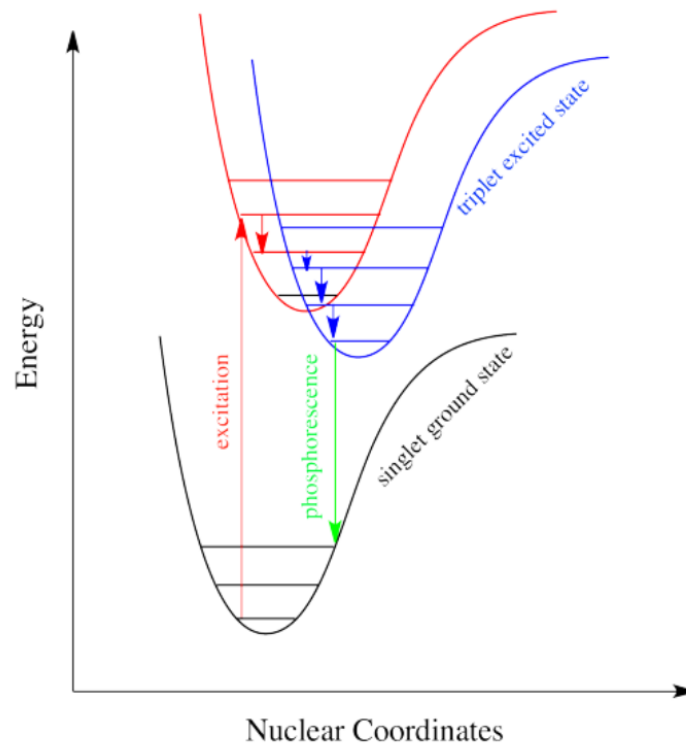


Slika 2.2: Prelazak iz pobuđenog u osnovno elektronsko stanje internom konverzijom [5]

likom ovog procesa, nema ni fluorescencije. To će utjecati na osjetljivost mjerenja tehnikama koje koriste fluorescenciju kao signal, dakle i na FCS.

Još jedna vrsta neradijativnog procesa u kojemu dolazi do promjene elektronskih stanja je međusustavno križanje. Prilikom apsorpcije fotona prijelaz iz osnovnog singletnog stanja u pobuđeno stanje je "zabranjen" i malo vjerojatan (razlika u spinskim brojevima  $\Delta S \neq 0$ ). No, obratan proces, u kojemu se molekula relaksira iz

pobuđenog singletnog stanja i konačno prijeđe u tripletno pobuđeno stanje manje energije, je moguć. Na slici 2.3 vidimo da ako su singletno i tripletno pobuđeno stanje energetski blizu, tada će ponekad doći do prijelaza u tripletno stanje. No, kako je prijelaz iz tripletnog u osnovno singletno emisijom stanje fotona "zabranjen" proces potrebno je dugo vremena da se prijelaz dogodi (red veličine sekunde a može biti i u satima). Proces u kojemu se konačno emitira foton zove se fosforescencija.

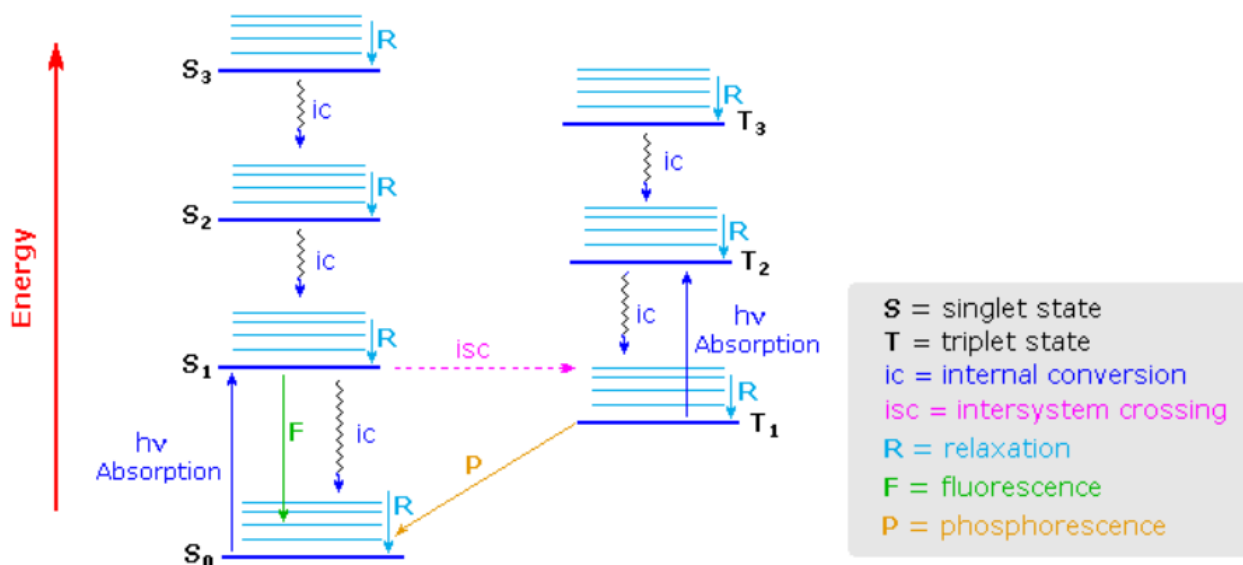


Slika 2.3: Međusustavno križanje koje prethodi fosforescenciji [5]

Neradijativni procesi koji prethode fluorescenciji, radi kojih dolazi do Stokesovog pomaka i kojima molekula gubi energiju, prikazani su na dijagramu Jablonskog (slika 2.4). Ti procesi zaslužni su da broj apsorbiranih fotona nije u 1:1 korespondenciji s brojem emitiranih fotona. Omjer koji govori o tome koliko u sveukupnom procesu relaksacije molekule doprinose radijativni i neradijativni procesi zove se kvantna propusnost  $Q$  (engl. quantum yield)

$$Q = \frac{\Gamma}{\Gamma + k_{nr}} \quad (2.1)$$

Konstante  $\Gamma$  i  $k_{nr}$  su stope radijativnih i neradijativnih procesa. Idealni omjer je 1, što bi značilo da nema neradijativnih procesa. To je naravno nemoguće postići. Stoga se za primjenu u FCS tehnici teži odabrati flourofore sa što većim  $Q$ .



Slika 2.4: Dijagram radijativnih i neradijativnih procesa relaksacije fotoosjetljivih molekula [5]

Još jedan efekt koji je bitno spomenuti a vezan je uz kvantnu propusnost je fotoizbjeljivanje. Naime, nakon nekog vremena pobudno zračenje, kojem su fluoroforescence bile podvrgnute da bi fluorescirale, uzrokuje trajne fotokemijske promjene na fluoroforescencijama čime im onemogućuje fluorescenciju [6]. Uglavnom je to posljedica međusustavnog križanja nakon kojega molekule u tripletnom stanju interagiraju s drugim molekulama. Taj efekt je posebno zamjetan kod velikih pobudnih snaga stoga je potrebno FCS mjerenja raditi u rasponu intenziteta zračenja koji omogućava da se tijekom mjerenja zanemariv dio molekula fotoizbjeli.

## 2.2 Fluorescencijska korelacijska spektroskopija

FCS spada u klasu tehnika za mjerenje jako malih koncentracija ( $nM$  do  $\mu M$ ) i difuzijskih konstanti biomolekula od interesa koje se nalaze bilo *in vivo*, bilo *in vitro* [7]. Za razliku od nekih drugih tehnika koje u principu promatraju intenzitet zračenja s promatranog uzorka, FCS promatra jako male fluktuacije intenziteta fluorescentnog zračenja flouoroforom označenih biomolekula. Moderni eksperimentalni postavi, a najčešće se koristi konfokalni mikroskop, omogućavaju praćenje fluktuacija uzorka u femtolitarskom ( $1 \text{ fL} = 10^{-15} \text{ L}$ ) detekcijskom volumenu (DV-u).

### 2.2.1 Fluktuacije intenziteta fluorescentnog zračenja

Broj molekula u DV-u je dan Poissonovom distribucijom [4]. S obzirom na tu činjenicu te uzevši u obzir da se FCS tehnikom proučavaju fluktuacije intenziteta fluorescencije u DV-u, bitno je znati kako se ponašaju fluktuacije u ovisnosti o broju čestica u DV-u. To možemo vidjeti računajući korijen očekivanja kvadrata fluktuacije broja čestica te uzevši u obzir da je varijanca Poissonove raspodjele jednaka očekivanju. Potom je potrebno normalizirati rezultat da bi vidjeli kako se rezultat ponaša ovisno o prosječnom broju čestica u DV-u.

$$\frac{\sqrt{\langle(\delta N)^2\rangle}}{\langle N \rangle} = \frac{\sqrt{\langle(N - \langle N \rangle)^2\rangle}}{\langle N \rangle} = \frac{\sqrt{\langle N^2 \rangle - \langle N \rangle^2}}{\langle N \rangle} = \frac{\sqrt{\text{Var}(N)}}{\langle N \rangle} = \frac{1}{\sqrt{\langle N \rangle}} \quad (2.2)$$

Iz opisanog računa (izraz 2.2) vidimo da su fluktuacije sve manje za veći broj čestica u volumenu. To ima i smisla s obzirom da će kod malog broja čestica u volumenu svaki ulazak i izlazak iz volumena biti "primjećeniji" nego kada u DV-u imamo puno čestica. Jasno je da je za osjetljiviji eksperiment bitno što više smanjiti broj čestica u DV-u. Moderni FCS postavi to postižu tako da DV dovedu na razinu femtolitra ili manje. Iako fluktuacije koje dolaze od difuzije čestica iz volumena idealno dolaze od malenoga broja čestica, nije idealno ni previše smanjiti DV. U tom slučaju se može dogoditi da u nekom vremenskom intervalu, na koji je eksperiment osjetljiv, niti nema čestica unutar DV-a. S druge strane signal mora prevladati pozadinski šum koji po najviše dolazi od Ramanovog raspršenja što se poistiže kada je volumen u rasponu od 0.1 do 1 fL [1]. Idealno je da očekivani broj čestica u DV-u bude između 0.1 i 1000

čestica [4]. Ako DV iznosi 1 fL, koncentracija čestica mora biti između  $1\mu M$  i  $1nM$ . U principu se koriste nanomolarne koncentracije čestica te je to slučaj i u ovome radu. Fotoni fluorescencije se broje jedan po jedan pomoću jednofotonske lavinske fotodiode (SPAD, engl. Single-Photon Avalanche Diode). Pretpostavljamo konstantnu snagu pobudnog lasera. U tom slučaju fluktuacije fluorescencije definiramo kao

$$\delta F(t) = F(t) - \langle F(t) \rangle \quad (2.3)$$

$$\langle F(t) \rangle = \frac{1}{T} \int F(t) dt \quad (2.4)$$

gdje je  $F(t)$  intenzitet fluorescencije u vremenu  $t$ .

### 2.2.2 Osnovni difuzijski model autokorelacijske funkcije

U slučaju da su sve fluktuacije  $\delta F(t)$  posljedica promjene lokalne koncentracije molekula u DV-u, koje se očituju kao difuzija molekula unutra i izvan DV-a, vrijedi sljedeći izraz [4]

$$\delta F(t) = \kappa \int_V I_{pob}(\rho, z) \cdot S(\rho, z) \cdot \delta(\sigma \cdot q \cdot C(\rho, z, t)) dV \quad (2.5)$$

gdje su  $\kappa$  detekcijska efikasnost,  $\rho$  radijalna koordinata,  $z$  aksijalna koordinata,  $I_{pob}(\rho, z)$  funkcija intenzitetskog profila pobude, te  $S(\rho, z)$  optička funkcija transfera mikroskopski objektiv-iris. Potonja funkcija određuje prostornu efikasnost detekcije fotona. Integracija se proteže duž koordinata diferencijala detekcijskog volumena  $dV$ . Pod irisom se misli na konfokalnu rupicu na koju pada fokusirana svjetlost iz mikroskopskog objektiva. Funkcija irisa je da tu svjetlost prostorno filtrira s ciljem da samo fotoni fluorescencije koji su nastali u žarišnoj ravnini objektiva bivaju detektirani (što je jedna od definicija konfokalnosti). U ovom radu umjesto irisa koristi se ulazna pukotina optičkog vlakna. Delta veličine su  $\delta\sigma$  kao fluktuacija molekularnog apsorpcijskog udarnog presjeka,  $\delta q$  kao fluktuacije kvantne propusnosti (engl. quantum yield) te  $\delta C(\rho, z, t)$  kao fluktuacije lokalne koncentracije fluorofore u DV-u koje su uzrokovane Brownovim gibanjem.

S obzirom da je određivanje svih navedenih parametara praktično nemoguće potrebne su neke aproksimacije koje pojednostavljuju konačni izraz. Prva aproksimacija kombinira dvije prostorne funkcije  $\frac{I_{pob}(\rho, z)}{I_0} \cdot S(\rho, z)$  u jednu funkciju (gdje je  $I_0$  inten-



zitet pobudnog zračenja) koja opisuje prostornu distribuciju emitiranoga zračenja:

$$MDF(\rho, z) = e^{-2\frac{\rho^2}{\rho_0^2}} \cdot e^{-2\frac{z^2}{z_0^2}} \quad (2.6)$$

To je funkcija koja opisuje intenzitetski profil DV-a i u literaturi je poznata po nazivom Molecule Detection efficiency Function - MDF. U gornjoj aproksimaciji distribucija zračenja opada na  $e^{-2}$  početne vrijednosti na  $\rho_0$  u radijalnom smjeru te za isti iznos na  $z_0$  u aksijalnom smjeru. Ovo je uobičajena aproksimacija koja pretpostavlja gausijanski DV. Dalje izvedeni izrazi ovise o ovoj aproksimaciji. S obzirom na ovu aproksimaciju MDF-a, kasnije dobivene funkcije će onoliko dobro opisivati eksperimentalne podatke koliko izraz 2.6 odgovara stvarnom DV-u eksperimentalnog postava.

Kasnije korišteni pristup strojnim učenjem ne pretpostavlja kakvog bi oblika trebali biti određeni parametri modela. Stoga, je jedna od ideja rada bila vidjeti može li se takvim pristupom izračunati funkcija za čiji se izračun koriste isključivo podaci iz eksperimenta dok su sve ostale pretpostavke sadržane isključivo u arhitekturi modela za učenje (primjerice koji tip neuronske mreže se koristi i sl.).

Parametri  $\kappa$ ,  $\sigma$  i  $q$  se mogu kombinirati pomnoženi s  $I_0$  (jer smo u izrazu za MDF dijelili s istom veličinom) da bi se dobila mjera detektiranih fotona po detektiranoj molekuli u sekundi te slijedi:

$$\eta_0 = I_0 \cdot \kappa \cdot \sigma \cdot q \quad (2.7)$$

U literaturi se često za  $\eta_0$  koristi kratica CPM - counts per molecule. Ta veličina mjeri signal-šum omjer mjerenja te se često koristi za podešavanje kvalitete eksperimentalnog postava [4]. Koristeći navedena pojednostavljenja, početna funkcija koja opisuje fluktuaciju se može svesti na

$$\delta F(t) = \int_V MDF(\rho, z) \cdot \delta(\eta C(\rho, z, t)) dV \quad (2.8)$$

Ovaj izraz se sada može iskoristiti za računanje autokorelacije fluktuacija. Autokorelacijska funkcija je definirana kao očekivanje umnoška funkcije, u ovom slučaju fluktuacije, u nekom trenutku  $t$  s tom istom funkcijom u nekom kasnijem trenutku

pomaknutom za  $\tau$  od  $t$  [4]. Dodatno se taj umnožak normalizira te dobivamo

$$G(\tau) = \frac{\langle \delta F(t) \cdot \delta F(t + \tau) \rangle}{\langle F(t) \rangle^2} \quad (2.9)$$

Autokorelacijska funkcija služi da bi se odredila samosličnost signala u trenutku  $t$  i nakon nekog vremena kašnjenja (engl. lag time)  $\tau$ . Ako je  $\tau = 0$  onda vrijedi

$$G(0) = \frac{\langle \delta F(t) \cdot \delta F(t + 0) \rangle}{\langle F(t) \rangle^2} = \frac{\langle \delta F(t)^2 \rangle}{\langle F(t) \rangle^2} = \frac{Var(\delta F(t))}{\langle F(t) \rangle^2} \quad (2.10)$$

Ovaj izraz ćemo kasnije povezati s očekivanom koncentracijom u DV-u.

Izraz 2.8 možemo uvrstiti u 2.9

$$G(\tau) = \frac{\int_V \int_{V'} MDF(\rho, z) \cdot MDF(\rho', z') \cdot \langle \delta(\eta C(\rho, z, t)) \delta(\eta C(\rho', z', t + \tau)) \rangle dV dV'}{(\int_V MDF(\rho, z) \cdot \langle \delta(\eta C(\rho, z, t)) \rangle dV)^2} \quad (2.11)$$

Fluktuacijski delta član se može separirati

$$\delta(\eta C(\rho, z, t)) = C(\rho, z, t) \delta \eta + \eta \delta C(\rho, z, t) \quad (2.12)$$

i uzevši u obzir da se u DV-u ne mijenjaju fluorescencijska svojstva molekule ( $\delta \eta = 0$ ) te da sva promjena dolazi od promjene koncentracije, slijedi:

$$\delta(\eta C(\rho, z, t)) = \eta \delta C(\rho, z, t) \quad (2.13)$$

Koristeći 2.13, izraz 2.9 se svodi na

$$G(\tau) = \frac{\int_V \int_{V'} MDF(\rho, z) \cdot MDF(\rho', z') \cdot \langle \delta C(\rho, z, 0) \delta C(\rho', z', \tau) \rangle dV dV'}{(\langle \delta C \rangle \int_V MDF(\rho, z) \cdot dV)^2} \quad (2.14)$$

Pretpostavljajući da dolazi do difuzije molekula u 3D detekcijskom volumenu te da je difuzijska konstanta fluorofore jednaka  $D$ , može se izračunati [4]

$$\langle \delta C(\rho, z, 0) \delta C(\rho', z', \tau) \rangle = \langle C \rangle \frac{1}{(4\pi D\tau)^{\frac{3}{2}}} e^{-\frac{(\vec{r}-\vec{r}')^2}{4D\tau}} \quad (2.15)$$

gdje je  $\langle C \rangle$  lokalna koncentracija fluorofore u DV-u. Izraz 2.15 se sada može uvrstiti

u 2.14 te se dobiva

$$G(\tau) = \frac{1}{\langle C \rangle (4\pi D\tau)^{-\frac{3}{2}}} \frac{\int_V \int_{V'} MDF(\rho, z) MDF(\rho', z') e^{-\frac{(\vec{r}-\vec{r}')^2}{4D\tau}} dV dV'}{(\int_V MDF(\rho, z) \cdot dV)^2} \quad (2.16)$$

Korisno je definirati izraz koji povezuje difuzijsko vrijeme  $\tau_D$  (u kojemu molekula ostaje unutar DV-a), dimenzije DV-a u radijalnom smjeru  $\rho_0$  te difuzijsku konstantu D:

$$\tau_D = \frac{\rho_0^2}{4D} \quad (2.17)$$

Potom je bitno definirati i izračunati efektivni detekcijski volumen:

$$V_{eff} = \frac{(\int_V MDF(\rho, z) \cdot dV)^2}{\int_V MDF(\rho, z)^2 \cdot dV} = \frac{(\int_V e^{-2\frac{\rho^2}{\rho_0^2}} \cdot e^{-2\frac{z^2}{z_0^2}} \cdot dV)^2}{\int_V (e^{-4\frac{\rho^2}{\rho_0^2}} \cdot e^{-4\frac{z^2}{z_0^2}})^2 \cdot dV} = \pi^{\frac{3}{2}} \cdot \rho_0^2 \cdot z_0 \quad (2.18)$$

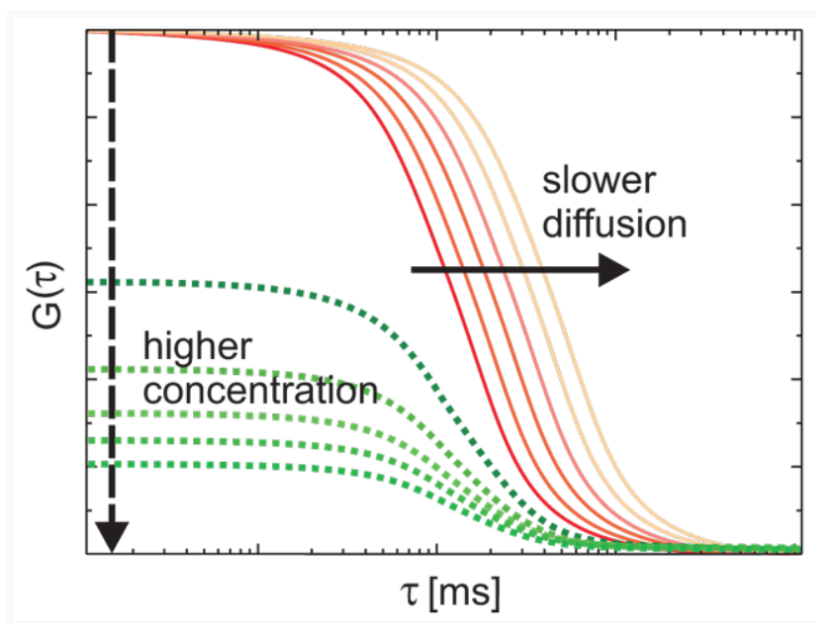
Koristeći rezultate 2.17 i 2.18, iz izraza 2.16 se konačno dobiva [4]:

$$G(\tau) = \frac{1}{V_{eff} \langle C \rangle} \frac{1}{(1 + \frac{\tau}{\tau_D})} \frac{1}{\sqrt{1 + (\frac{\rho_0}{z_0})^2 \cdot \frac{\tau}{\tau_D}}} \quad (2.19)$$

Ako pogledamo prvi član u izrazu 2.19 odmah vidimo da je po definiciji:

$$G(0) = \frac{1}{V_{eff} \langle C \rangle} = \frac{1}{\langle N \rangle} \quad (2.20)$$

Ovo nam govori da će  $G(0)$  davati informaciju o očekivanom broju molekula u DV-u. To je prvi od parametara sustava koje je moguće izračunati pomoću FCS tehnike. U grafovima koje dobivamo FCS tehnikom ta informacija će se vizualno očitovati kao "visina" autokorelacijske krivulje. To znači da će, pod pretpostavkom da su svi ostali parametri sustava nepromijenjeni, povećanje koncentracije spuštati krivulju dolje, dok će smanjivanje koncentracije davati veće maksimalne autokorelacijske vrijednosti (zelene krivulje, sl. 2.5). Ukoliko se koncentracija ne mijenja ali dolazi do usporavanja difuzije, pa molekule sporije izlaze is DV-a, autokorelacijska krivulja zadržava istu "visinu" no vidljiv je pomak na desno prema većim difuzijskim konstantama (crvene krivulje, sl. 2.5).



Slika 2.5: Ilustracija autokorelacijskih krivulja različitih koncentracija i različitih difuzijskih konstanti molekula označenih flouoroforom [8]

### 2.2.3 Difuzija i viskoznost

Stokes-Einstein jednadžba [7] (izraz 2.21) daje nam koeficijent difuzije flouorofore otopljene u fluidu dinamičke viskoznosti  $\eta_v$ :

$$D_{S-E} = \frac{k_b T}{6\pi\eta_v R_v} \quad (2.21)$$

U gornjoj jednadžbi  $k_b$  je Boltzmannova konstanta, T temperatura uzorka,  $\eta_v$  dinamička viskoznost otapala te  $R_v$  hidrodinamički radijus molekule. Viskoznost je veličina koja se lako može kontrolirati u eksperimentu miješajući uzorak fluorescentne boje otopljene u vodi s nekom drugom tvari visoke viskoznosti. Primjerice, mijenjajući omjere vode i glicerola možemo dobiti otopine različitih viskoznosti. Pomoću izraza 2.21 možemo izračunati referentne vrijednosti difuzijske konstante pod uvjetom da je poznat hidrodinamički radijus promatrane molekule, dinamička viskoznost otapala i temperatura uzorka. Hidrodinamički radijus korišten u Stokes-Einstein jednadžbi u slučaju molekule Alexa647 korištene u ovom radu iznosi [9]

$$R_{Alexa647} = (7.52 \pm 0.23) \cdot 10^{-10} m \quad (2.22)$$

Poznavajući gornji hidrodinamički radijus lako je dobiti difuzijske konstante fluorofo-  
 ra u otapalu poznate dinamičke viskoznosti  $\eta$ .

Iz literature [10] je poznata difuzijska konstanta flourofore Alexa647 otopljene u vodi na 25C° pomoću koje se može izračunati difuzijska konstanta na proizvoljnoj temperaturi

$$D(T) = D(25C^\circ) \cdot \frac{T}{298.15K} \frac{8.9 \cdot 10^{-4} Pa \cdot s}{\eta(T)} \quad (2.23)$$

gdje je temperatura T u Kelvinima te  $\eta(T)$  dinamička viskoznost na toj temperaturi. Izraz 2.23 se može koristiti u slučajevima gdje je flourofora otopljena u vodi, no u slučajevima kad se koristi neko drugo otapalo potrebno je koristiti izraz 2.21.

#### 2.2.4 Tripletna dinamika

Izraz 2.19 je teoretska autokorelacijska funkcija koja opisuje difuzijsku dinamiku sustava koji se proučavaju FCS tehnikom. Radeći teorijsku prilagodbu koristeći izraz 2.19, očekuje se da će teorijska krivulja dobro opisivati podatke u slučaju da je jedini značajni efekt radi kojega dolazi do fluktuacija intenziteta fluorescencije upravo difuzija molekula unutar i izvan DV-a. Pretpostavka s kojom je izvedena izraz 2.19 je da vrijedi  $\delta\eta = 0$ . To je u praksi neostvarivo jer postoje dodatni efekti koji utječu na fotodinamiku sustava. Najočitiiji među njima je efekt koji nastaje uslijed prelazaka u zabranjena tripletna stanja iz kojih se fluorofora, u usporedbi s fluorescencijom, par redova veličina sporije relaksira u osnovno stanje.

S obzirom da je vremenska skala na kojoj se događaju tripletni prijelazi puno kraća od skale difuzije molekula, funkciju  $G(\tau)$  koja uključuje i tripletnu dinamiku je moguće separirati [4]

$$G_{ukupno}(\tau) = G_{dif}(\tau) \cdot G_{triplet}(\tau) \quad (2.24)$$

Tripletni član  $G_{triplet}(\tau)$  je moguće modelirati jednadžbom [4]

$$G_{triplet}(\tau) = \frac{1 - T + T e^{-\frac{\tau}{\tau_{trip}}}}{1 - T} \quad (2.25)$$

gdje je  $1-T$  u razlomku normalizacija. Ukupna korelacijska funkcija se tada može

opisati kao:

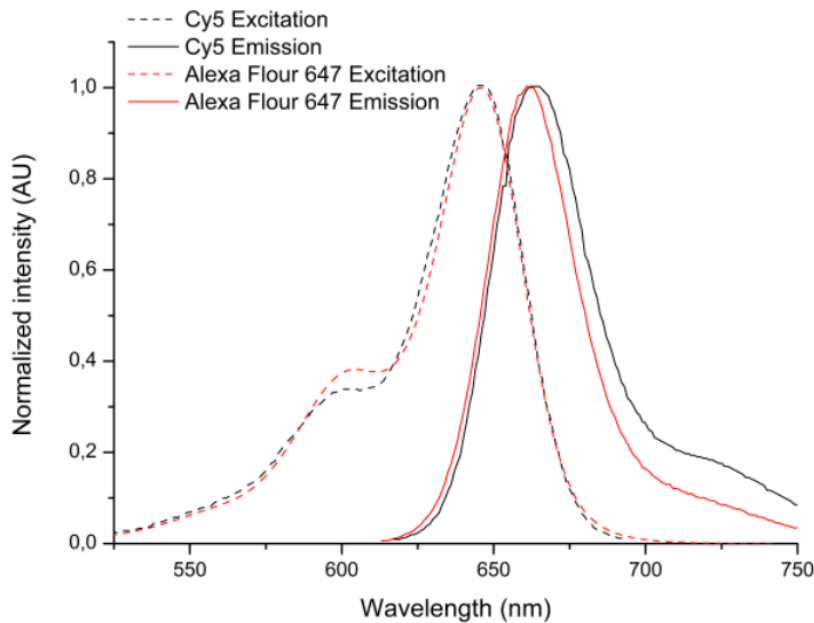
$$G(\tau) = \left( \frac{1 - T + T \cdot e^{-\frac{\tau}{\tau_{trip}}}}{1 - T} \right) \cdot G_0 \cdot \frac{1}{1 + \frac{\tau}{\tau_D}} \cdot \frac{1}{\sqrt{1 + \left(\frac{\rho_0}{z_0}\right)^2 \cdot \frac{\tau}{\tau_D}}} \quad (2.26)$$

Prilikom teorijske prilagodbe na podatke prigodno je gornjoj funkciji, ukoliko podaci nisu normalizirani, dodati konstanti član 1 jer su vrijednosti autokorelacije iznad 1 dok izraz 2.26 trne u 0. Izraz 2.26 se može dodatno generalizirati za slučaj kada je u uzorku prisutno više tipova fluorescentnih molekula (vidi [4]) no to nije bio slučaj u ovome radu. Također, osim tripletne dinamike, koja je jedan od mogućih fotoaktivacijskih procesa u kojemu nema emisije fotona prilikom relaksacije iz viših u niža vibracijska stanja, mogući su i drugi procesi u kojima nema emisije fotona. Izraz 2.25 se može generalizirati za slučaj u kojem je prisutno više "tamnih" stanja (osim tripletnog). I ta diskusija se može pročitati u [4]. U ovom radu je korišten izraz 2.26 kao teorijska krivulja za prilagodbu na eksperimentalne podatke.

### 2.2.5 Odabir fluorofora

FCS tehnika prati male fluktuacije intenziteta fluorescencije. S obzirom da većina biološki interesantnih molekula jako slabo fluorescira, u svrhu praćenja tih biomolekula te se molekule označuju fluoroforima (poznati i pod nazivom fluorescentni markeri). Ti markeri moraju onda imati dobra fluorescencijska svojstva. Svojstva uključuju: dovoljno intenzivno zračenje, visoku kvantnu propusnost te velik broj ciklusa fluoresciranja fluorofora prije nego je fluorescencija uništena fotoizbjeljivanjem, što je obično oko  $10^6$  pobudnih ciklusa [7]. Također, bitno je izabrati fluoroforu čiji se apsorpcijski spektar nalazi u rasponu valnih duljina u kojemu emitira pobudni laser. Bitno je i da su efekti koji se vide u autokorelacijskim krivuljama značajniji od artefakata eksperimenta, posebno za kraće vrijeme kašnjenja  $\tau$ . Ako je difuzija molekula od prvenstvenog interesa onda bi taj proces trebao biti izraženiji i od jake tripletne dinamike i ostalih fotodinamičkih procesa pa se biraju fluorofore u kojima su tripletni prijelazi rjeđi. Takve podatke je lakše opisati teorijskim modelima koji su već izvedeni u prethodnim potpoglavljima. Strojnim učenjem se pokušavaju izbjeći zadnja dva uvjeta.

Prilikom eksperimenata koji su provedeni u svrhu pisanja ovog rada korištene su



Slika 2.6: Apsorpcijski i fluorescentni spektar fluorofora Cy5 i Alexa478 [11]

Cy5 i Alexa647 fluorofore. Oba markera imaju navedena poželjna svojstva te im je maksimum apsorpcijskog spektra blizu valne duljine pobudnog lasera (633 nm). Na slici 2.6 prikazani su apsorpcijski i fluorescencijski spektri obje molekule. Vidimo da postoji veliko preklapanje između spektara tako da su eksperimentalni uvjeti bili praktički identični. Ipak, podaci za Alexa647 su bolje pratili ponašanje modelirano jednačbom 2.26, što je posljedica izraženije tripletne dinamike u slučaju Alexa647. Maksimumi apsorpcijskog spektra su 649 nm za Cy5 te 650 za Alexa647 [12]. U ovom radu su prikazani samo rezultati dobiveni Alexa647 fluoroforom jer je u laboratorijskom radu potrošena Cy5 molekula prije nego su postignuti adekvatni rezultati.

### 2.2.6 Artefakti i ograničenja

Prilikom izvođenja jednačbe 2.26 pretpostavljen je gausijanski DV. Iako taj pristup često daje dobre rezultate, analize eksperimentalnih postava [3] su uočile da je DV gausijanskog profila validan samo u posebnim uvjetima. Oni ovise o promjeru irisa u konfokalnoj ravnini kroz koju svjetlost dolazi do detektora te o svjetlosnoj ispunjenosti stražnje aperture mikroskopskog objektiva. Mjerenja provedena u ovom radu pokazala su da DV korištenog postava nije sasvim gausijanski. Na slici 4.14, kasnije u radu, je vidljivo da za  $\tau = 50\mu s$  teorijska autokorelacijska krivulja malo precjenjuje

vrijednosti dok za  $1000 \mu s$  krivulja malo podcjenjuje vrijednosti. Kod većih odstupanja od gausijanskog profila odstupanje krivulje od vrijednosti bilo bi još izraženije.

Optički artefakti se mogu skoro eliminirati [3] korištenjem irisa malog promjera (25  $\mu m$  ili manje) no time se smanjuju dimenzije DV-a pa je potrebno dulje vrijeme ekspozicije, a smanjuje se i omjer signal-šum. Artefakti se također mogu manifestirati kao dodatan eksponencijalni član vidljiv za mali  $\tau$  koji se u grafovima može zamijeniti za neki drugi fotofizikalni ili fotokemijski proces. Bitno je napomenuti da artefakti mogu izobličiti autokorelacijsku krivulju do te mjere da prilagodbom teorijske krivulje na podatke izračunato difuzijsko vrijeme bude pomaknuto do 80% u odnosu na referentnu vrijednost [3].



## 2.3 *Strojno učenje*

### 2.3.1 O strojnom učenju

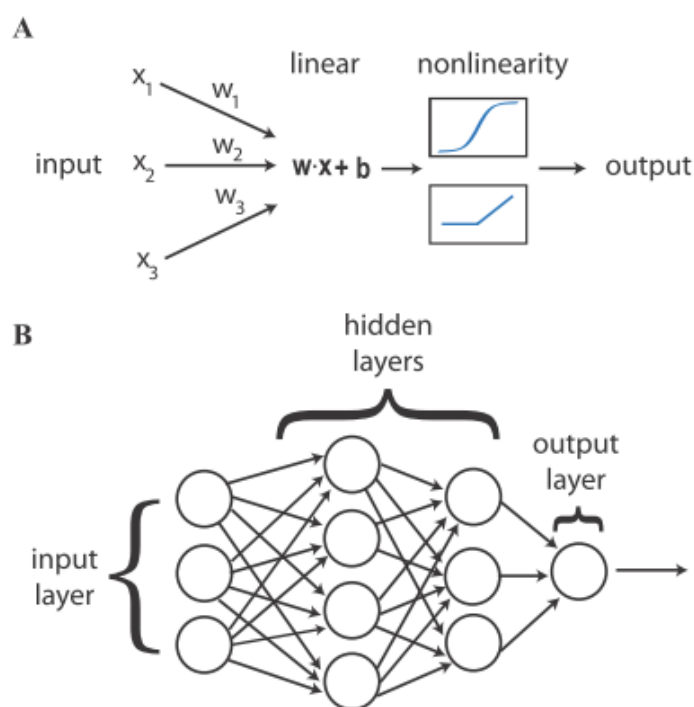
Strojno učenje je programski pristup obradi podataka koji pripada općenitijem polju umjetne inteligencije [13] [14] [15]. Iako algoritmi koji se danas definiraju kao algoritmi strojnog učenja postoje već desetcima godina, poboljšanja implementacija postojećih algoritama, kao i razvoj novih algoritama, učinili su modele strojnog učenja nadmoćnima u raznim primjenama u znanosti kao i industriji. Među poboljšanja spadaju: razvoj novih nelinearnih aktivacijskih funkcija, poboljšanje algoritma kojim se korigiraju parametri modela kroz trening, razvoj novih arhitektura neuronskih mreža itd. Također, razvijanjem same računalne tehnologije, omogućeno je efikasnije i brže treniranje već postojećih kao i novih algoritama.

Ono što strojno učenje čini drugačijim od drugih pristupa učenju iz podataka je to što su obično potrebne minimalne pretpostavke (ili nikakve) o funkcijama koje definiraju podatke. Primjerice, u ovom radu da bi se izvršila prilagodba teorijske autokorelacijske funkcije na podatke koji se dobivaju FCS tehnikom, potrebno je u teorijski opis modela uključiti informaciju o difuziji molekula te tripletnoj dinamici sustava.

Strojno učenje ne pretpostavlja pozadinsko znanje o podacima sustava te pokušava prilagoditi parametre modela tako da na izlazu predviđa vrijednost koja što bolje odgovara "stvarnoj" vrijednosti ciljne varijable. Ciljnu varijablu koju model predviđa biramo iz podskupa varijabli koje se mogu naučiti iz danih podataka. Primjerice, FCS podaci daju informaciju o difuzijskoj konstanti, difuzijskom vremenu, koncentraciji molekula u sustavu itd. U ovom radu ciljna varijabla je bila difuzijska konstanta. Mana takvog pristupa je da su za trening modela potrebne informacije o referentnim vrijednostima, kao primjerice difuzijska konstanta neke molekule u DV-u. Time je teško koristiti ovaj pristup za otkrivanje novih fizikalnih pojava, barem u slučaju nadziranog učenja u koje spada većina popularnih modela strojnog i dubokog učenja. Postoje i druge grane strojnog učenja, kao nenadzirano učenje, koje ne zahtijeva prethodnu informaciju o ciljnoj varijabli no u ovom radu se ne koriste takve metode.

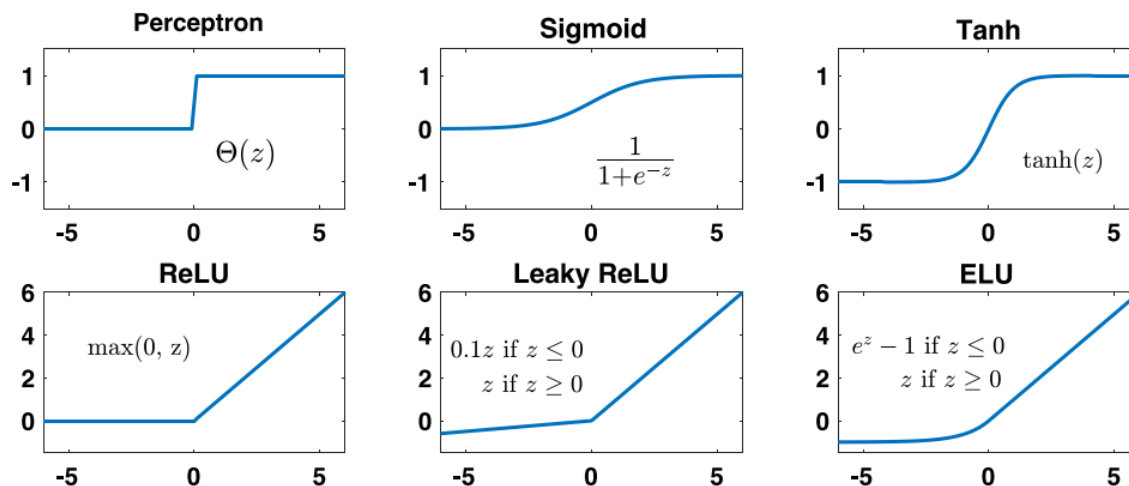
### 2.3.2 Neuronske mreže

Neuronska mreža je nelinearni model strojnog (dubokog) učenja koji je sposoban izravno iz podataka naučiti netrivialna i često nelinearna svojstva sustava pod pretpostavkom da je dostupna adekvatna količina podataka. Adekvatnu količinu podataka uvjetuje tip problema te željena preciznost modela. U principu neuronska mreža sa samo jednim skrivenim slojem može aproksimirati bilo koju nelinearnu kontinuiranu funkciju (engl. "universal approximation theorem") [13]. U praksi je to teško izvedivo, tako da postoje brojne arhitekture neuronskih mreža koje su specijalizirane za obradu i učenje na specifičnim podacima. Primjerice konvolucijske neuronske mreže odlično uče reprezentirati slike kroz niz dubokih slojeva neuronske mreže od kojih svaki sloj uči reprezentaciju na različitom stupnju apstrakcije.



Slika 2.7: Skica jedne ćelije neuronske mreže te obične potpuno povezane neuronske mreže [14]

Na slici 2.7 u gornjem dijelu slike vidimo kako izgleda jedna tipična ćelija neuronske mreže [14]. Ćelija u skrivenom sloju dobiva višestruki ulaz te ih utežava vrijednostima  $w_i (i = 1, 2, \dots, N_{broj\ ulaza})$  te zatim zbraja. Na takvu još uvijek linearnu funkciju se primjenjuje aktivacijska funkcija. Aktivacijska funkcija je nelinearna te služi da bi se mreži omogućilo da učenje nelinearnih svojstava ulaznih podataka. Na slici 2.8 vidimo nekoliko najčeće korištenih aktivacijskih funkcija. U ovom radu se



Slika 2.8: Najčešće korištene aktivacijske funkcije [14]

koristila ReLU aktivacijska funkcija čije djelovanje je vidljivo sa slike u donjem lijevom kutu. Ako je vrijednost ulaza manja od 0 izlaz ćelije je 0, dok u slučaju da je ulaz veći od 0 vrijednost izlaza ćelije linearno ovisi o ulazu.

Na slici 2.7 u donjem dijelu slike vidimo arhitekturu jedne potpuno povezane neuronske mreže s višestrukim ulazom, dva skrivena sloja i jednom izlaznom ćelijom. U slučaju da se radi o regresijskom problemu, kao što je slučaj u ovom radu, zadnji sloj ima samo jednu ćeliju koja kao izlaz vraća kontinuiranu vrijednost. Ta vrijednost predstavlja predviđanje modela u ovisnosti u ulaznim podacima te istreniranim parametrima mreže.

Neuronska mreža potpuno povezanog tipa može imati i veći broj ćelija na izlazu, u slučaju klasifikacijskih problema. Izlazne ćelije onda sadrže neku drugu aktivacijsku funkciju primjereniju klasifikacijskim problemima kao što je primjerice sigmoidna aktivacijska funkcija (druga aktivacijska funkcija na slici 2.8). Cilj klasifikacijskog modela je predvidjeti kojoj klasi pripadaju ulazni podaci. Primjerice ako model na ulazu dobije jednu spinsku konfiguraciju 2d Isingovog modela, a cilj modela je predvidjeti kojoj fazi pripada spinska konfiguracija, izlazni sloj će imati 2 ćelije koje predviđaju vjerojatnost da spinska konfiguracija pripada jednoj od dvije moguće faze.

Mreža uči na velikom broju primjera te kroz trening minimizira funkciju greške (engl. "loss function") na izlaznoj ćeliji [15]. Odabir funkcije greške koja se minimizira ovisi o tipu problema. Na regresijskim problemima, kao funkcija greške se

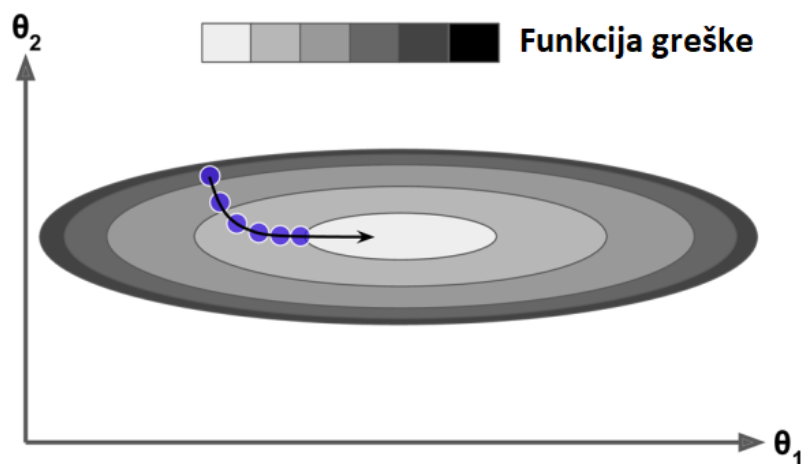
najčešće koristi srednja kvadratna pogreška (engl. "mean squared error" ili MSE)

$$MSE = \frac{1}{n} \sum_i^n (y_i - f(x_i))^2 \quad (2.27)$$

gdje je  $y_i$  vrijednost ciljne varijable koju model koristi za učenje, dok je  $f(x_i)$  predviđanje modela na osnovu podataka  $x_i$ . Cilj modela je što više približiti  $y_i$  i  $f(x_i)$  a da se pritom model ne overfita i postane beskoristan za predviđanje na podacima koji nisu korišteni u treningu. Za minimizaciju funkcije greške zadužen je optimizacijski algoritam. Poznati optimizacijski algoritmi su bazirani na ideji gradijentnog spusta [15]. Parametri ćelije  $w^i$  se korigiraju računajući gradijent funkcije greške

$$w^i \leftarrow w^i - \eta \nabla_{w^i} MSE(w^i) \quad (2.28)$$

gdje je  $\eta$  hiperparametar koji se određuje prije treninga modela i on se naziva stopa učenja (engl. "learning rate") [15]. Na slici 2.9 vidimo ilustraciju gradijentnog spusta



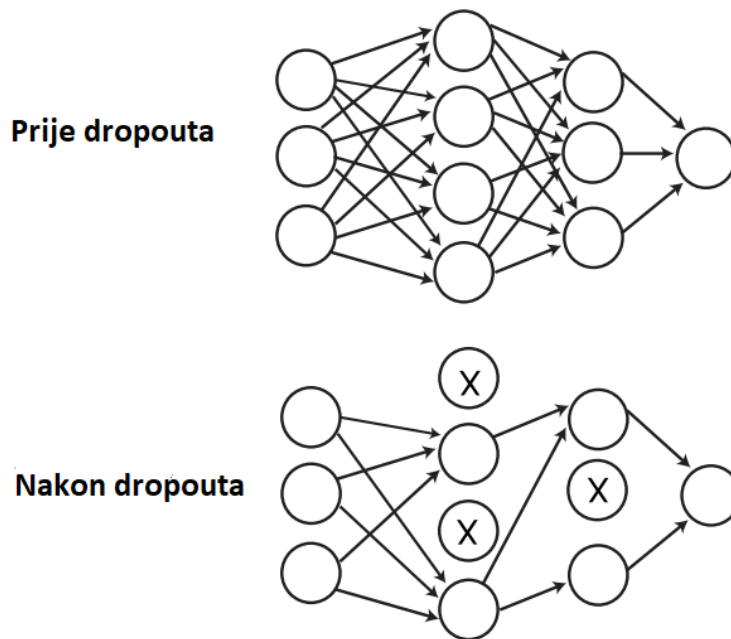
Slika 2.9: Primjer dvodimenzionalnog gradijentnog spusta [15]

u dvije dimenzije. Treba imati u vidu da je prostor parametara obično velik (puno veći od 2) što utječe na duljinu treninga modela. Ideja je kroz svaku iteraciju treninga malo korigirati parametre ćelija mreže dok se ne pronađu optimalne vrijednosti (obično lokalni minimum). Jedna iteracija objašnjene procedure se naziva epoha. Trening obično sadrži desetke ili stotine epoha. Osim gradijentnog spusta definiranog izrazom 2.28 postoje drugi, moderniji optimizacijski algoritmi koji uključuju epohalno ovisnu stopu učenja [14]. Jedan od poznatijih optimizacijskih algoritama

je Adam [14] koji se koristi u ovom radu.

Na kraju svake epohe koristi se validacijski skup podataka. To je dio podataka koji se ne koristi u treningu već služi za računanje funkcije greške na podacima koji se ne koriste u treningu. Nakon treninga model je spreman za predviđanja na dotad neviđenim podacima. Dodatno se može koristiti i testni skup ako su se kroz trening više puta mijenjali hiperparametri modela te je validacijski skup "zagađen", isto kao i trening skup. Na testnom skupu se tada vrši završna provjera kvalitete modela. Hiperparametri modela su broj slojeva mreže, broj ćelija mreže, razni parametri vezani za brzinu i veličinu korekciju težina kroz trening i sl.

Kod treninga algoritama strojnog učenja općenito, ključno je da model bude podešen tako da je sposoban za predviđanje ciljane varijable na podacima koji nisu bili dio treninga, tj. želi se spriječiti overfit modela. Postoji nekoliko načina da se to postigne. Jedan od njih je da se trening odvija dok greška na validacijskom skupu pada. Onda kada funkcija greške počinje rasti, pogotovo ako istovremeno pada na trening skupu, dolazi do overfita modela i trening treba obustaviti. To je lako implementirati tako da se prati greška modela na validacijskom skupu te se trening prekida ako greška nije pala u n epoha, gdje je n hiperparametar koji se određuje prije treninga. Overfit modela se može ublažiti korištenjem dropout sloja za gašenje ćelija (engl. "dropout") koji određuje koliko ćelija u sloju sudjeluje u partikularnoj epohi treninga. Recimo može se podesiti da 30% neurona u sloju ne sudjeluje u treningu, a oni koji ne sudjeluju se biraju nasumično. U sljedećoj epohi se nasumično biraju drugi nesudjelujući neuroni. Postotak privremeno odbačenih neurona je još jedan hiperparametar modela. Na slici 2.10 vidimo prikaz učinka dropout sloja. Ideja dropout sloja je smanjiti potencijalno štetne korelacije između neurona u neuronskoj mreži [14]. Uvođenjem dropout sloja, svaka epoha mreže ustvari kreira malo drugačiju mrežu od prethodne, jer nikada ne sudjeluju isti neuroni. Dropout sloje se također može primjeniti i na ulazne podatke, tj. ulazni sloj u ostatak mreže veže samo neke ulazne ćelije dok su druge privremeno ugašene. Također, moguće je implementirati spremanje najboljeg aktualnog modela. To omogućuje da se u slučaju pogoršanja treninga modela, učita prethodno snimljeni model koji je imao najbolji rezultat na validacijskom skupu.



Slika 2.10: Učinak dropout sloja [14]

### 3 Eksperimentalni postav

#### 3.1 Minimizacija šuma

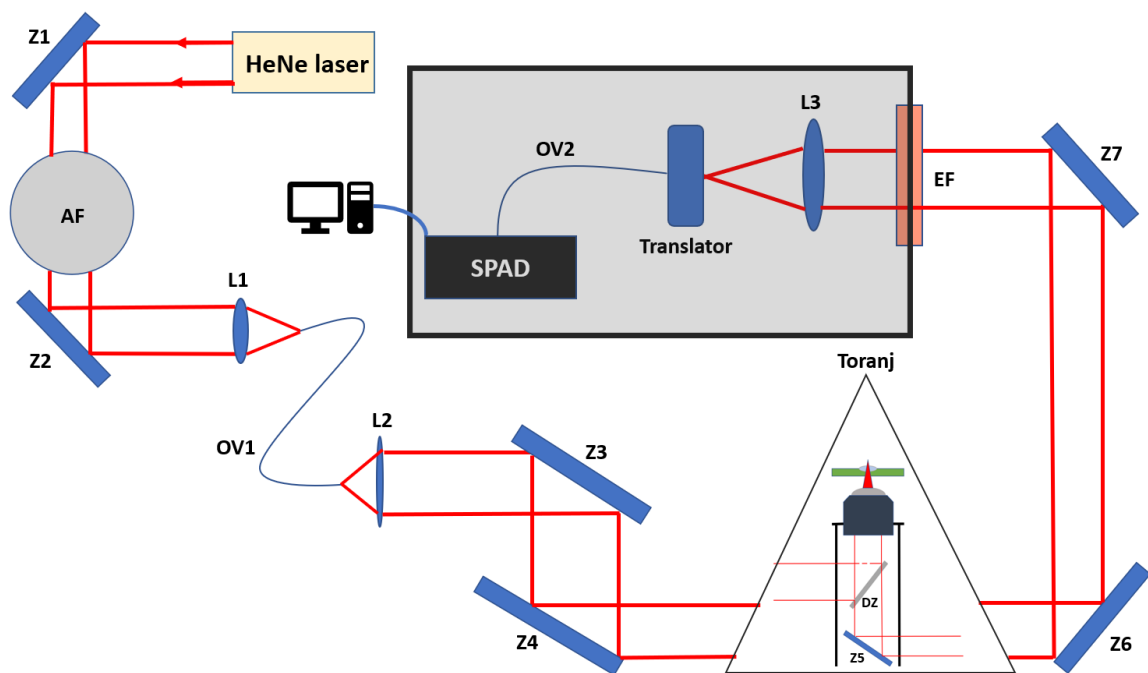
Prvi važan dio eksperimentalnog postava je optički stol na koji se učvršćuju svi ostali elementi postava. Prije svega, stol treba iznivelirati da mu radna površina bude u horizontalnoj ravnini a također je bitno osigurati da bude što otporniji na vanjske vibracije. Stabilnost snage pobudnog snopa mjerila se neposredno prije ulaska snopa u toranj s mikroskopskim objektivom te je uočeno da dolazi do zamjetnog mijenjanja snage kada bi se stol dodirivao. Uočene su čak fluktuacije u snazi uzrokovane zvukovima i vibracijama koje nastaju od primjerice hodanja, pomicanja stvari u laboratoriju i drugih vanjskih faktora. Fluktuacije tog porijekla su dovoljno male da ne utječu bitno na FCS mjerenja. Naime, provedeno je kontrolno mjerenje u kojemu se mijenjala pobudna snaga u rasponu 110-150  $\mu W$  te nisu uočene vidljive promjene autokorelacijskih krivulja.

U eksperimentu je ključno što je više moguće eliminirati i svjetlosni pozadinski šum da bi dobili što bolji omjer signal-šum jer bi šum inače prekrivio fluorescentni signal. Stoga, u laboratoriju su morali biti isključeni svi intenzivni izvori svjetla. Pozadinski šum koji preostaje nastaje radi refleksije i raspršenja laserskog snopa, na

nečistoćama u uzorku, emisije svjetla sa samih optičkih komponenti te tamne struje detektora [1]. U svrhu minimizacije pozadinskog šuma koriste se spektralni filtri o čemu će biti riječi. U eksperimentima izvršenima u ovom diplomskom radu pozadinski šum je uvijek imao stopu brojanja (engl. count rate) ispod 700 fotona u sekundi dok je signal koji dolazi od fluorescencije obično bio 50 ili više puta veći.

### 3.2 Postav

Na slici 3.11 možemo vidjeti shemu cijelog eksperimentalnog postava. Kao izvor svje-



Slika 3.11: Eksperimentalni postav

tla korišten je helij-neonski (HeNe) laser koji daje snop kontinuirane snage (engl. CW - continuous wave) na valnoj duljini od 633 nm. Tijekom rada, provedena su kontrolna mjerenja gdje se mijenjala pobudna snaga u rasponu od 50 do 150  $\mu W$ , dok je maksimum bio 1.5  $mW$ . Cilj je bio naći optimalni raspon u kojemu je izvor daje dovoljno veliku stopu brojanja no mjerenja su također stabilna. Za veće intenzitete lasera brže dolazi do fotoizbjeljivanja flourofore, što nije poželjno, te može doći do izraženijih fotofizikalnih procesa koji nam nisu od interesa u eksperimentu (npr. izraženija tripletna dinamika) [16]. S druge strane, za veće intenzitete potrebno je manje vrijeme integracije signala tako da se mjerenja mogu brže obaviti. Zaključeno je da je optimalna pobudna snaga u rasponu od 50 do 150  $\mu W$ .

Iz HeNe lasera izlazi kolimirani snop približno gausijanskog profila čiji  $e^{-2}$  promjer iznosi 1 mm. Laserski snop nakon refleksije na prvom zrcalu (Z1, sl. 3.11) prolazi kroz filter promjenljive atenuacije (AF, Thorlabs NDC-100S-4 [17]) koji se koristi za regulaciju intenziteta snopa. Atenuacijski filter je disk koji se može okretati  $360^\circ$  oko svoje osi paralelne smjeru širenja snopa (optička os). Veći dio diska je prekriven reflektivnim metalnim premazom čija reflektivnost ovisi o dijelu diska na koji laserski snop nailazi. Dio diska nije prekriven metalnim premazom te za taj raspon kutova svjetlost ne "vidi" filter. Mjera koja govori o tome koliko je intenzitet svjetla smanjen nakon prolaska kroz filter je optička gustoća (OD, engl. Optical density) [16]

$$OD = \log_{10} \frac{I_{in}}{I_{out}} = \log_{10} \frac{1}{T} \quad (3.29)$$

gdje je  $I_{in}$  intenzitet snopa prije filtera,  $I_{out}$  intenzitet poslije filtera a transmitivnost se definira iz izraza  $T = \frac{I_{out}}{I_{in}}$ . Dakle, zakretanjem diska moguće je mijenjati OD te time i intenzitet snopu koji izlazi iz filtera. Filter je zakrenut u položaj koji daje atenuaciju od 20 puta u odnosu na upadni intenzitet snopa. Tako se postigla snaga lasera u rasponu 50-150  $\mu W$ . Snaga lasera se mjeri mjeračem snage (Coherent FieldMaster) čija je sonda sa silicijskim fotodetektorom postavljena neposredno prije nego pobudni snop uđe u toranj s mikroskopskim objektivom. Prije početka akvizicije fotona nužno je maknuti sondu jer blokira snop.

Nakon atenuacijskog filtera snop upada na drugo zrcalo (Z2) koje zakreće snop prema difrakcijski limitiranoj asferičnoj leći koja fokusira snop u monomodno optičko vlakno (OV1). Tiltanjem prvo zrcala Z1, potom zrcala Z2 i konačno mjerenjem intenziteta svjetlosti na izlazu iz optičkog vlakna moguće je optimizirati efikasnost sprege laserskog snopa u vlakno u 3-5 iteracija. Optičko vlakno ima dvojaku funkciju. Prva je da posluži kao prostorni filter koji će proizvesti čisti, prostorno uniforman laserski snop gausijanskog profila. FCS tehnika zahtijeva takav snop jer se na taj način dobija što je moguće manji i geometrijski pravilan DV. Ako je vlakno dovoljno dugo (nekoliko centimetara je sasvim dovoljno) izlaz iz vlakna imat će gotovo savršen gausijanski profil pridružen tom jednom modu koji se propagira kroz optičko vlakno. Na takav način se bilo kakva distorzija snopa može ukloniti pod cijenu gubitka snage na izlazu. Druga funkcija je da vlakno transferira snop do sljedeće asferične leće čija



je uloga da snop ponovno kolimira te da njegov  $e^{-2}$  promjer proširi za tri puta na 3 mm. Sustav L1-OV-L2 (slika 3.11) je istovremeno i prostorni filter i širilo snopa (engl. beam expander) čiji je koeficijent širenja određen omjerom žarišnih duljina  $f_2/f_1$ . Potom se snop pomoću dva zrcala (Z3 i Z4) usmjerava prema tornju s mikroskopskim objektivom.

Toranj je konstrukcijski ekvivalentan invertiranom mikroskopu a sastoji se od dikroičkog zrcala (DZ, Semrock BrightLine FF660-Di01-25x36), te mikroskopskog objektiva (O, Olympus UPLSAPO60xW) i nepropusnog zrcala (Z5). Nakon ulaska u toranj snop se prvo odbija od dikroičkog zrcala te dolazi na objektiv koji, u odnosu na volumen uzorka, svjetlost fokusira na veoma mali dio uzorka (u odnosu na volumen uzorka) femtolitarskog volumena. Mikroskopski objektiv je tzv. "mokri" objektiv jer se, u svrhu postizanja visoke razlučivosti, između leće objektiva i preparata stavlja medij većeg indeksa loma od zraka, u našem slučaju voda. Stoga na objektiv nanosimo kapljicu vode pipetom koja dodiruje donji dio plastične kadice u kojoj se nalazi promatrani uzorak volumena  $300 \mu L$ . Uzorka može biti i puno manje jer je fokalni volumen nekoliko redova veličine manji. No, veći volumeni, uz pokrivanje kadice poklopcem, omogućavaju da se minimizira utjecaj hlapljenja otapala na povećanje koncentracije uzorka tijekom mjerenja.

Mikroskopski objektiv ima žarišnu duljinu  $f_{obj} = 3mm$  i numeričku aperturu  $NA=1.2$ . Objektiv stvara u uzorku pobudni volumen čiji je intenzitetski profil ( $I_{pob}(\rho, z)$  u izrazu 2.5) približno gaussijanskog profila. Snop pobuđuje fluorofore u viša elektronska stanja te molekule fluoresciraju fotone duže valne duljine (niže frekvencije) od onih koji pobuđuju uzorak. Objektiv tada prikuplja generirano fluorescentno zračenje, kolimira ga i usmjerava ga prema dikroičkom zrcalu ispod objektiva. Svjetlost duže valne duljine može proći kroz dugo-propusno (engl. long pass) dikroičko zrcalo te se potom u donjem dijelu tornja još jednom odbija na nepropusnom zrcalu (Z5) i izlazi iz tornja. Dugo-propusno dikroičko zrcalo po definiciji ima vrlo visoku reflektivnost za kraće valne duljine pobudnog zračenja, no veliku propusnost za fluorescenciju duljih valnih duljina [7].

Svjetlost se potom odbija od dva zrcala (Z6 i Z7) na putu ka detekcijskoj kutiji unutar koje se nalazi jednofotonska lavinska fotodioda (SPAD, engl. Single-photon avalanche diode) Excelitas Technology SPCM-AQRH-12-FC. Prije nego uđe u detekcijsku kutiju snop prolazi kroz emisijski uskopojasni filter (EF, Semrock BrighLine

FF01-692/40-25) koji ima visoku propusnost ( $T > 93\%$ ) isključivo u uskom pojasu od 672 – 712 nm a ostale valne duljine blokira. Potonji filter služi za filtriranje nepoželjnog zračenja koje doprinosi pozadinskom šumu. Naime, dikroičko zrcalo nije sposobno u potpunosti eliminirati zračenje koje nije fluorescencija, primjerice Rayleigh raspršenje na vodi. Također dio zračenja dolazi od Raman raspršenja na vodi koje je, kao i fluorescencija, pomaknuto prema dužim valnim duljinama te stoga može proći kroz dikroički filter [16]. EF filtrira dio Raman zračenja te se postiže puno bolji omjer signal-šum [4]. U našim eksperimentima postigli smo omjer signal-šum od barem 50 na više.

Nakon prolaska kroz EF snop se fokusira pomoću akromatskog dubleta žarišne duljine  $f_3 = 300\text{mm}$  (L3, Thorlabs AC508-300-AB) na ulaznu pukotinu multimodnog optičkog vlakna (OV2, Thorlabs M42L01, promjer jezgre 50  $\mu\text{m}$ ) koje vodi fotone fluorescencije u SPAD. Pozicija pukotine na koju se snop fokusira može se pomoću mikrometarskih translatora mijenjati u radijalnom ( $\rho$ ) i aksijalnom ( $z$ ) smjeru te tako ulaz optičkog vlakna postaviti u točku u prostoru na kojoj je sprega fluorescencije u vlakno najefikasnija.

SPAD je spojen na TCSPC karticu (Time Correlated Single Photon Counter, Becker Hickl SPC-130) spoјenu na računalo koja autokorelira signal koji se potom sprema u nekoliko formata na računalu. Budući da pobudni laser nije impulsnog već kontinuiranog tipa, na TCSPC karticu je nužno spojiti i izlaz iz signal generatora koji daje TTL signal na 70 MHz a koji kartici služi kao clock signal za analogno-digitalnu konverziju. Program u realnom vremenu pokazuje graf autokorelacijske krivulje. ASCII zapis autokoreliranog signala je korišten za daljnje modeliranje podataka u Pythonu i programu Origin.

### **3.3 Optimizacija postava**

Prije početka mjerenja cijeli postav nužno je optimizirati prateći broj fotona u vremenu tako da se signal iz SPAD-a spoji na osciloskop kojeg treba postaviti u modalitet brojača (engl. counter mode). Naime, poziciju objektiva je moguće pomicati u aksijalnom ( $z$ ) smjeru te pronaći idealnu poziciju objektiva pri kojoj će broj fotona fluorescencije u sekundi CPS (engl. count per second) biti najveći. Objektiv pomičemo, nekoliko puta na gore pa na dolje, te tako tražimo položaj na  $z$  osi koji daje najjači

mogući signal. Nakon toga slijedi već spomenut proces optimizacije pozicije upadne pukotine multimodnog optičkog vlakna OV2 u detekcijskoj kutiji. Upadna pukotina, koja prije svega mora biti u žarišnoj ravnini leće L3, se polagano pomiče prvo u x smjeru te se traži optimalna pozicija na x osi. Potom se pomiče u y smjeru te se postupak ponavlja. Postupak se iterira nekoliko puta dok nismo sigurni da je signal na osciloskopu maksimiziran.

Optimizirani sustav je podešen za FCS mjerenja za parametre sustava u datom trenutku. No, treba biti oprezan kod rukovanja s opremom (npr. namještanje uzorka iznad objektiva i sl.) jer svaki dodir nekog od dijelova postava sustav može zahtijevati novu optimizaciju (x,y pozicija pukotine OV2, z pozicija objektiva). Ovaj problem bi se mogao eliminirati boljim učvršćivanjem svih dijelova postava no u ovom eksperimentalnom postavu taj problem se rješavao ponovnom optimizacijom u slučaju pogoršanja signala. Moguće je i nakon svakog mjerenja pomoću osciloskopa provjeriti je li sustav još uvijek optimiziran.

## 4 Rezultati

### 4.1 Priprema uzoraka

FCS tehnikom je moguće odrediti parametre sustava kao što su koncentracija fluorofora u uzorku, njihova difuzijska konstanta  $D$  te potom i difuzijsko vrijeme  $\tau_D$ . Kako se kod primjene nadziranog strojnog učenja koriste podaci koji su označeni, tj. svakom podatku u trening skupu je pridružena referentna vrijednost ciljne varijable, potrebno je da je ta varijabla dobro definirana i lako mjerljiva. Idealno bi to bile  $D$  ili  $\tau_D$  no te konstante se obično računaju iz parametara koji proizlaze iz prilagodbe teorijske funkcije na autokorelacijsku krivulju (izraz 2.26). Da bi te parametre dobili strojnim učenjem moramo pribjeći drugom načinu označavanja uzoraka. Difuzijsko vrijeme fluorofore u uzorku možemo proizvoljno mijenjati povećavanjem ili smanjivanjem dinamičke viskoznosti otapala (izraz 2.21). To se postiže mijenjanjem temperature uzorka ali i upotrebom velikog broja otapala čije se viskoznosti monotonno povećavaju u odnosu na baznu viskoznost - viskoznost vode. Takav skup otapala poznatih viskoznosti lako se pripremi miješajući glicerol i vodu u različitim masenim udjelima.

U procesu pripreme uzoraka korištena je analitička vaga. Uzorci su dobiveni miješanjem vode, u kojoj je već bila otopljena Alexa647 fluorofora u koncentraciji od 1 nM, i glicerola. Budući da glicerol ima veliku viskoznost, povećanjem udjela glicerola uzorak postaje sve viskozniji te tako mijenjamo difuzijsku konstantu  $D$  a time i difuzijsko vrijeme  $\tau_D$  fluorofore (izraz 2.17). Prije pripreme uzoraka, izračunate su potrebne mase glicerola i vode da bi se postigla željena viskoznost uzorka. Za dohvat otapala korištene su dvije pipete maksimalnih volumena  $1\text{mL}$  te  $200\mu\text{L}$ . Prilikom prenošenja glicerola u spremnik uzorka, vagom je u realnom vremenu praćena unešena masa glicerola te ukupna masa smjese u spremniku.

Dobiveni maseni udjeli su obično malo odstupali od onih planiranih prije vaganja. To je posljedica visoke viskoznosti glicerola, radi čega bi glicerol prijanjao uz stijenke pipete pa je bilo nemoguće postići željene masene udjele. To je ujedno i razlog zašto je u pripremi uzoraka korištena analitička vaga. Na taj način, masa glicerola i ukupna masa uzorka su mogli biti vagani s nepouzdanošću od 1 mg što je omogućilo puno finiju podjelu uzoraka po viskoznosti, a posljedično i po difuzijskoj konstanti, nego bi to bilo moguće isključivo pipetiranjem.

U tablici 4.1 su prikazane viskoznosti skupa uzoraka kojeg smo pripravili u laboratoriju. Nulti uzorak sadrži samo fluoroforu Alexa647 otopljenu u vodi a potom slijedi 15 uzoraka koji sadrže sve veći maseni udio glicerola  $w_m$ . Iz vrijednosti  $w_m$  i

Uzorak	$w_m$ [%]	$\eta$ [ $10^{-2} Poise$ ]	$D$ [ $10^{-10} \frac{m^2}{s}$ ]	$T$ [ $^{\circ}C$ ]
0	0	0.95	3.03	22.3
1	3.0	1.02	2.82	22.4
2	12.0	1.28	2.25	22.5
3	24.0	1.84	1.57	22.6
4	32.5	2.49	1.16	22.7
5	37.0	2.91	0.99	23.4
6	42.7	3.79	0.76	22.8
7	47.1	4.59	0.63	23.3
8	50.4	5.49	0.53	23.0
9	54.0	6.65	0.43	23.1
10	57.8	8.83	0.35	23.1
11	60.8	10.08	0.29	23.1
12	62.8	11.52	0.25	23.1
13	65.0	13.46	0.21	23.1
14	67.0	15.54	0.19	23.2
15	69.1	18.30	0.16	23.2

Tablica 4.1: Maseni udjeli glicerola  $w_m$ , dinamičke viskoznosti  $\eta$  te dinamičke konstante  $D$  pripremljenih smjesa glicerol-voda na temperaturi laboratorija  $T$ . Mjerenja su obavljena u istom danu.

temperature uzorka  $T$  u trenutku mjerenja, mogu se izračunati dinamičke viskoznosti otapala  $\eta$  uz pomoć online kalkulatora viskoznosti smjese glicerol-voda [18] [19].

Svi uzorci stajali su u laboratoriju minimalno sat vremena u svrhu termalizacije te je pretpostavljeno da temperatura uzorka jednaka temperaturi laboratorija. Temperatura se mjerila digitalnim termometrom s nepouzdanošću od 0.1 K. Temperature u laboratoriju su se uvijek povećavale tijekom dana kako bi se vršila mjerenja. Minimalna temperatura u laboratoriju za vrijeme bilo kojeg mjerenja je bila 20.8°C dok je najveća bila 23.4°C. Koristeći izraz 2.21 izračunate su difuzijske konstante i pripadajuću nepouzdanost svakog uzorka gdje su ulazni parametri bili dinamička viskoznost, temperatura laboratorija te hidrodinamički radijus Alexa647 molekule (izraz 2.22).

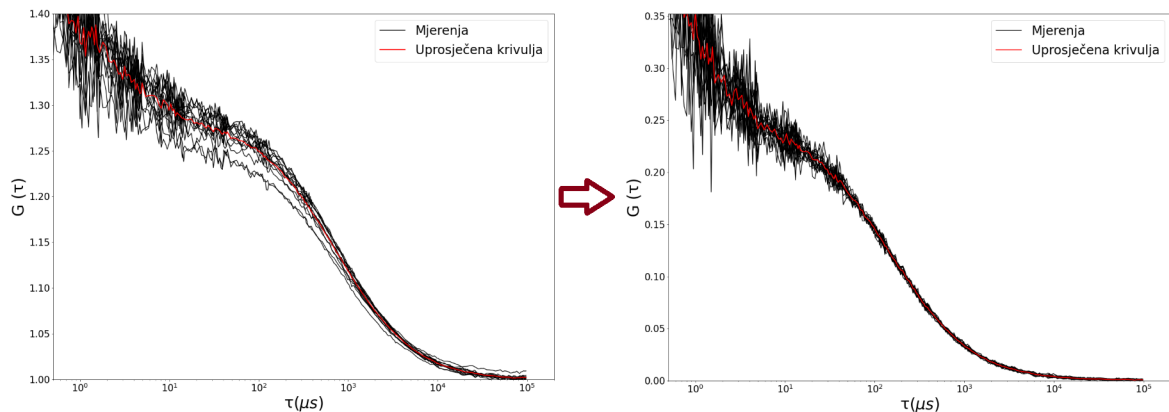
## 4.2 Mjerenja

Svakoga dana izmjereno je svih 15 uzoraka iz tablice 4.1 a posebna pozornost dana je mjerenju temperature T neposredno pred pokretanje FCS mjerenja. Definiramo da mjerenja nastala u jednom danu čine jedan skup mjerenja. Svaki uzorak mjeren je 15 puta, s vremenom integracije od 60 s po mjerenju, što definiramo da čini jedan podskup mjerenja. Broj mjerenja po podskupu može se zadati u programu za akviziciju koji ih konsektivno pokreće.

Autokorelacijske krivulje koje su drastično odstupale od prosječnog oblika krivulje u pojedinom podskupu su izbačene iz podskupa. Time su eliminirana mjerenja koja nastaju zbog raspršenja na nečistoćama u uzorku, povremenih fluktuacija u snazi lasera, ambijentalnih vibracija optičkog stola i sl. Zbog potrebe za normalizacijom krivulja, autokorelacijske vrijednosti su korigirane tako da im vrijednost asimptomatski teži nuli (umjesto jedinici kako je u originalnim mjerenjima). Također, za veliki  $\tau$  krivulje su normalizirane u odnosu na jedno referentno mjerenje unutar podskupa. Na slici 4.12 desno vidimo primjer jednog podskupa mjerenja gdje su lijevo prikazana mjerenja prije normalizacije.

Razlog za ovakvu predobradu podataka je postavljeni cilj da naš model strojnog učenja nauči predviđati difuzijske karakteristike sustava koje se očituju kao pomak krivulje u lijevo ili desno, ovisno o iznosu  $\tau_D$  za pojedini uzorak. Ostale karakteristike sustava, primjerice promjena broja fluorofora u DV-u što se očituje kao pomicanje krivulje gore ili dolje (sl. 2.5), su šum koji algoritmu smeta da iz podataka nauči predviđati  $\tau_D$ . Iako su svi uzorci sadržavali približno 1 nM fluorofore, eksperimentalni postav unosi grešku mjerenja kojoj najviše doprinose male promjene u veličini DV-a od mjerenja do mjerenja što prividno mijenja koncentraciju fluorofore. Normalizacijom mjerenja unutar podskupa brišemo informaciju o prividnim malim razlikama u koncentracijama fluorofore među uzorcima. Također kod usporedbe podskupova mjerenja izvedenih na različite dane dolazi do istog efekta zato što nije moguće mjerni postav postaviti u potpuno identično stanje i dobiti isti DV. U slučaju da koristimo nenormalizirane podatke, potrebno bi bilo više mjerenja da algoritam jednako uspješno nauči predviđati ciljnu varijablu.

Za manje vrijednosti  $\tau$  primjetne su veće oscilacije autokorelacijskih vrijednosti koje su uzrokovane odzivom mjernog uređaja. Za veće vremenske pomake ( $\tau > 10\mu s$ ) vidimo karakterističan oblik krivulje koji opisuje difuziju molekula u i izvan



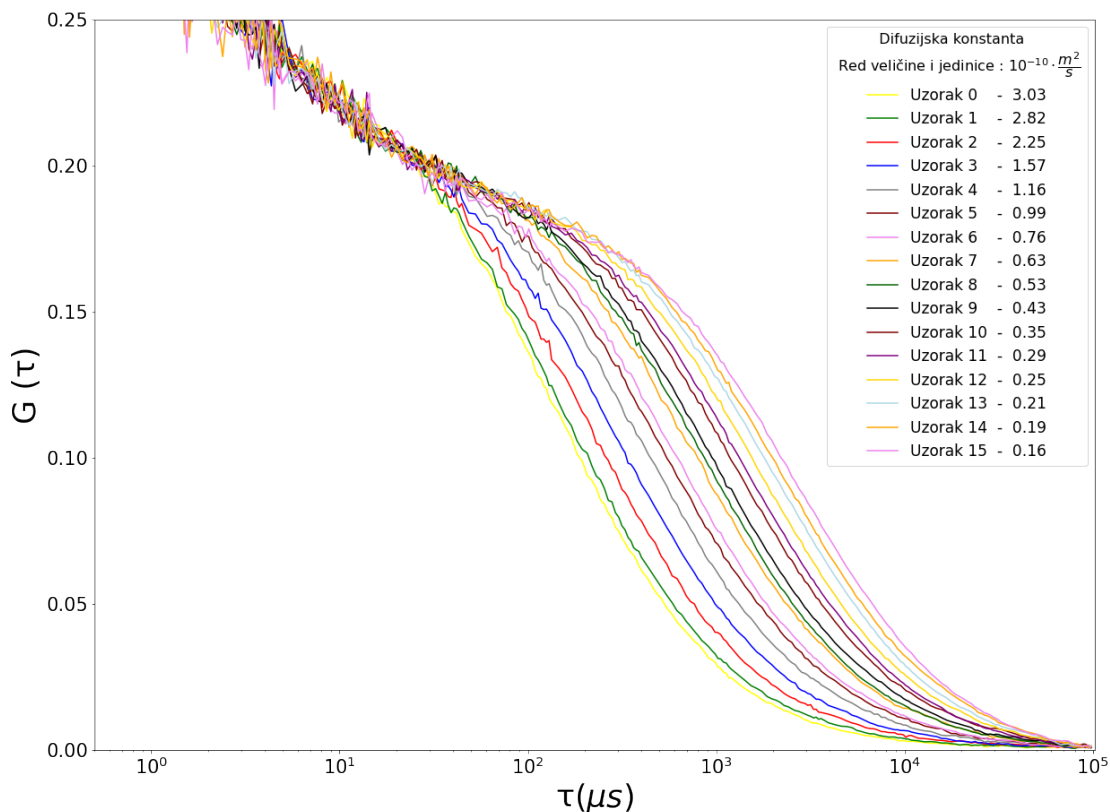
Slika 4.12: Mjerenja u jednom podskupu prije i poslije normalizacije

DV-a.

Na slici 4.13 možemo vidjeti grafove uprosječenih autokorelacijskih krivulja (u smislu da je uprosječen cijeli podskup od oko 15 mjerenja po uzorku) za cijeli skup uzoraka navedenih u tablici 4.1. Svaka uprosječena krivulja je ponovno normalizirana u odnosu na uprosječenu krivulju referentnog uzorka (najčešće je to uzorak s fluoroforom u čistoj vodi). Uzorci su poredani u silaznom poretku iznosa difuzijske konstante. Već vizualnom analizom razvidno je da se difuzijsko vrijeme konzistentno povećava kako pada difuzijska konstanta. Treba imati u vidu da su mjerenja sa slike 4.13 obavljena u istom danu. Naime, radi razlika u temperaturi laboratorija za vrijeme provođenja mjerenja na istim uzorcima na različite dane, difuzijske konstante različitih podskupova mjerenja istog uzorka nisu nužno bile iste. Ta se razlika očituje kao mali pomak jedne krivulje u odnosu na drugu. Vizualni prikaz cijelog skupa podataka, koji uključuje 56 podskupova mjerenja prikupljenih u 5 dana, može se vidjeti u dodatku A.1.

### 4.3 Određivanje efektivnog detekcijskog volumena

U slučaju kad modeliramo MDF gausijanskim funkcijama (izraz 2.6), efektivni detekcijski volumen se može izračunati iz izraza 2.18. Dakle, potrebno je znati dimenzije DV-a u radijalnom ( $\rho$ ) i aksijalnom ( $z$ ) smjeru -  $\rho_0$  i  $z_0$ . Prilagodbom teorijskog modela 2.26 na referentna mjerenja uzorka dobije se difuzijsko vrijeme  $\tau_D$  i omjer  $K = \frac{z_0}{\rho_0}$ . Difuzijsku konstantu na temperaturi uzorka  $T$  računamo iz izraza 2.23 u kojem je ulazni parametar i viskoznost otapala  $\eta(T)$ . Postupak pripreme smjese glicerol-voda unosi bi veću nepouzdanost viskoznosti u izraz 2.23 nego u slučaju da



Slika 4.13: Prikaz uprosječenih autokorelacijskih krivulja za sve uzorke iz tablice 4.1 u jednom danu mjerenja

se za otapalo koristi čista voda. Stoga, za referentna mjerenja odabiremo Alexa647 fluoroforu otoplenu u čistoj vodi. Sad kad je difuzijska konstanta fluorofore  $D_{Alexa647}$  određena, iz izraza 2.17 može se izračunati  $\rho_0$ , a potom iz omjera K se dobije i  $z_0$ . Primjećeno je da, neovisno o korištenom uzorku, teorijska prilagodba ne daje razumne vrijednosti za  $\rho_0$  i  $z_0$ . To znači da prilagodba treba konvergirati te da  $\rho_0$  bude ispod  $1 \mu m$ , dok  $z_0$  mora biti 3-5x veći. Stoga je potrebno fiksirati K na vrijednost koja daje najstabilnije rezultate. Budući da  $\rho_0$  i  $z_0$  divergiraju kada K nije fiksiran, možemo zaključiti da DV eksperimentalnog postava koji smo koristili nema gausijanski profil u aksijalnom smjeru [2]. Fiksiranjem K se omogućava korištenje aproksimacije gausijanskog profila u aksijalnom smjeru, a time omogućuje i korištenje izraza 2.26. Omjer K smo fiksirali na vrijednost koja se najčešće koristi u literaturi [2]:

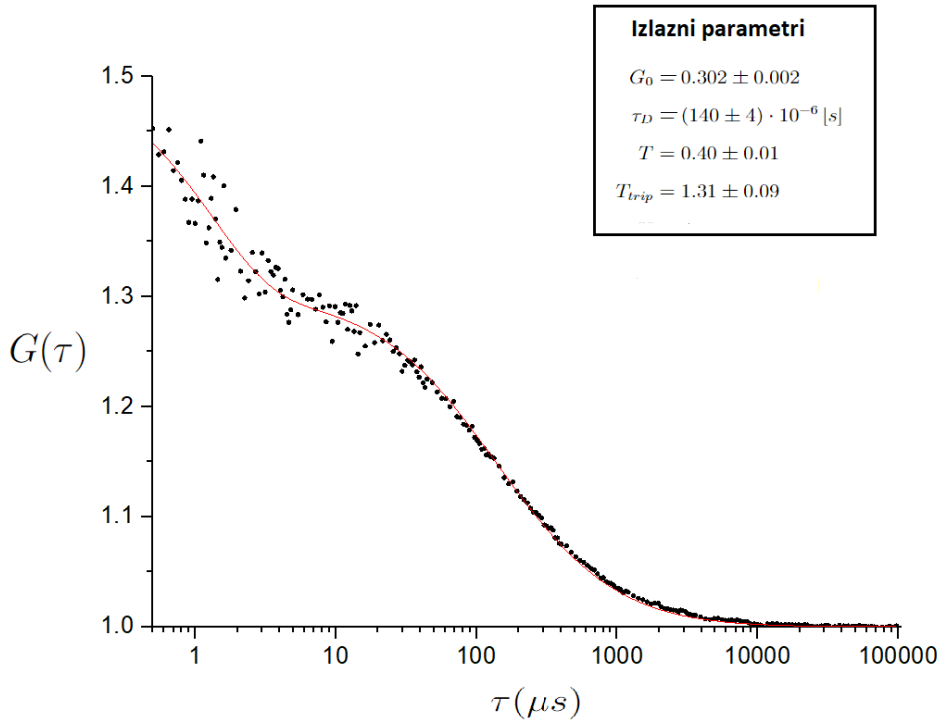
$$K = \frac{z_0}{\rho_0} = 5 \quad (4.30)$$



Također, koristili smo izraz 2.23 da bi izračunali difuzijsku konstantu fluorofore Alexa647 u čistoj vodi na 21.9°C. Dobili smo

$$D(21.9) = (3.0 \pm 0.1) \cdot 10^{-10} \frac{m^2}{s} \quad (4.31)$$

Na slici 4.14 vidimo graf koji sadrži usrednjene podatke (15 mjerenja na 21.9°C) prilagođene na funkciju 2.26.



Slika 4.14: Autokorelacijska krivulja za fluoroforu Alexa647 u vodi na 21.9°C (crne točkice) i teorijska prilagodba na nju (crvena linija)

Iz parametara prilagodbe modela i izračunate difuzijske konstante na 21.9°C °(izraz 4.31) te koristeći izraz 2.17 dobivamo

$$\rho_0 = (0.407 \pm 0.003) \mu m \quad (4.32)$$

$$z_0 = (2.04 \pm 0.01) \mu m \quad (4.33)$$

$$V_{eff} = (1.88 \pm 0.04) fL \quad (4.34)$$

Dobivena vrijednost DV-a je u femtolitarskom rasponu što je standardni iznos za FCS mjerenja [4]. Izračunate vrijednosti  $\rho_0$  i  $z_0$  su bile fiksirane u svim daljnim teorijskim prilagodbama u ostatku rada. Vrijednosti autokorelacijske krivulje nisu korigirane s

minimalne vrijednosti 1 na 0 jer nije bilo potrebe za normalizacijom krivulja kao na slici 4.13.

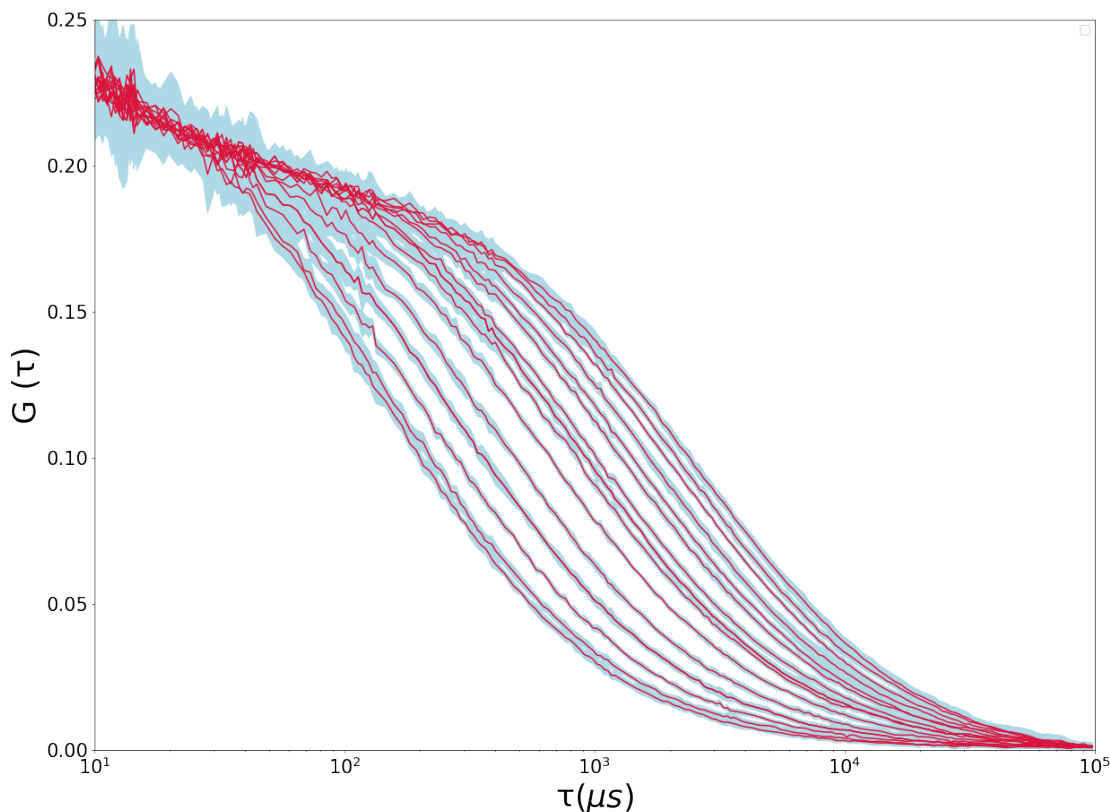
#### 4.4 Procjena nepouzdanosti mjerenja

Prilikom pripreme uzoraka koji su se koristili u mjerenjima bilo je bitno pripremiti uzorke koji su dovoljno različiti jedan od drugoga, u smislu da im se autokorelacijske krivulje ne preklapaju, kako bi algoritam strojnog učenja dobio dovoljno različitih krivulja za trening modela.

Na slici 4.15 možemo vidjeti iste krivulje kao na slici 4.13 zajedno s izračunatim nepouzdanostima po podskupu mjerenja i za svaku točku vremenskog pomaka (plava boja). Vidimo da su za male vrijednosti vremenskog pomaka  $\tau$ , do otprilike 100  $\mu s$ , nepouzdanosti velike. No to nije značajniji problem jer taj dio krivulje ne nosi informaciju o difuzijskom vremenu. U rasponu  $\tau$  vrijednosti u kojemu dolazi do difuzije ( $\tau > 100 \mu s$ ) nepouzdanosti su manje od razmaka između krivulja osim kod prve dvije krivulje, što je logično jer su to krivulje vode i uzorka s najmanjim udjelom glicerola.

Razmaci u difuzijskim konstantama među uzorcima (tablica 4.1) te prethodna analiza nam govore da bi algoritam strojnog učenja trebao biti u stanju razlikovati krivulje te, uz dovoljno podataka za trening, naučiti predviđati  $\tau_D$  i  $D$  za uzorke čije se difuzijske konstante razlikuju za barem:

$$\Delta D \approx 0.05 \cdot 10^{-10} \frac{m^2}{s} \quad (4.35)$$



Slika 4.15: Uprosječene autokorelacijske krivulje iz slike 4.1 (crveno) te nepouzdanost vrijednosti krivulje po točkama vremenskog pomaka  $\tau$  (plavo)

#### 4.5 Prilagodba teorijske funkcije na podatke

Prilikom testiranja modela strojnog učenja korišten je testni skup koji je sadržavao 7 podskupova mjerenja. Svaki podskup označavao je jedan uzorak određene difuzijske konstante te je sadržavao oko 15 snimljenih autokorelacijskih krivulja (nije točno 15 po uzorku jer su neke uklonjene, vidi potpoglavlje 4.3). Za usporedbu modela strojnog učenja s klasičnim pristupom prilagodbe teorijske krivulje 2.26 na autokorelacijske krivulje, provedena je prilagodba za svih petnaestak mjerenja iz pojedinog podskupa. Iz tih prilagodbi su kao izlazni parametri dobivene difuzijske konstante i pripadajuće nepouzdanosti za svaku krivulju iz podskupa. Potom su izračunate opća srednja vrijednost i opća nepouzdanost difuzijske konstante podskupa. Postupak je ponovljen za svaki podskup (tj. uzorak A-G) iz testnog skupa. Dobiveni parametri su dani u tablici 4.2.

<i>Uzorci</i>	$D_{fit}$	$M_{D_{fit}}$	$D_{S-E}$	$ D_{fit} - D_{S-E} $
E	2.67	0.03	2.82	0.15
G	1.15	0.01	1.16	0.01
F	1.01	0.01	0.985	0.025
A	0.56	0.006	0.525	0.032
C	0.48	0.005	0.405	0.075
D	0.41	0.003	0.324	0.085
B	0.33	0.004	0.267	0.059

Tablica 4.2: Rezultati prilagodbe analitičke funkcije 2.26 na podatke iz testnog skupa korištenog za testiranje neuronske mreže poredani po veličini difuzijske konstante u jedinicama  $10^{-10} \frac{m^2}{s}$ .

## 4.6 Strojno učenje - modeliranje

### 4.6.1 Predobrada podataka

Podaci su originalno bili u dvostupčanom prikazu (vrijeme vremenskog pomaka i vrijednost autokorelacijske funkcije). Podaci koji su korišteni kao ulaz u mrežu su učitani u Dataframe objekt u programskom jeziku Python [20] pomoću Pandas biblioteke [21]. Dataframe objekt sadrži 254 stupaca - po jedan stupac za svaku vrijednost autokorelacijske funkcije za određeni vremenski pomak  $\tau$  (ukupno 255 vremenskih pomaka). Također, provedeno je skaliranje i normalizacija kako je objašnjeno u potpoglavlju 4.2. Na slici 4.16 je prikaz dijela ulaznih podataka gdje brojke iznad stupaca označavaju  $\tau$ , a indeksi u svakom retku označavaju broj podskupa te broj mjerenja u podskupu.

Time[ $\mu$ s]	0.0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	
G_14_15	-0.977094	-0.977094	0.675606	3.074910	1.579390	1.006146	0.689475	0.622442	0.456016	0.444459	0.462951	0.430590	0.509180	0.451394	0.479131	0.428279	0.3
G_7_10	-0.918234	-0.918234	1.090358	3.929223	1.840959	1.014878	0.817793	0.700381	0.532649	0.524262	0.624901	0.285244	0.444589	0.402656	0.411043	0.415236	0.5
G_15_14	-0.926486	-0.926486	0.217546	1.921059	0.967320	0.659662	0.451138	0.346307	0.313262	0.264264	0.250591	0.267683	0.201593	0.289333	0.246033	0.211849	0.1
G_8_12	-0.975587	-0.975587	0.329890	2.322459	1.316451	0.726589	0.678622	0.516572	0.473790	0.447862	0.386931	0.420638	0.547685	0.425823	0.432305	0.386931	0.3
G_5_03	-0.960905	-0.960905	0.342508	2.271112	1.310996	0.853267	0.671849	0.504387	0.424843	0.445776	0.489037	0.466709	0.505783	0.384373	0.362045	0.369022	0.4
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
G_6_13	-0.918157	-0.918157	0.052227	1.530731	0.817505	0.546714	0.386842	0.305048	0.311864	0.281501	0.284599	0.218296	0.272825	0.304428	0.227590	0.232548	0.2
G_13_10	-0.946809	-0.946809	0.128450	1.908381	1.071980	0.731548	0.563315	0.441109	0.428412	0.463328	0.459360	0.410160	0.440315	0.381592	0.463328	0.383180	0.4
G_9_06	-0.935871	-0.935871	0.284933	2.152115	1.233030	0.737513	0.606381	0.410263	0.499619	0.529791	0.443917	0.447398	0.394017	0.380091	0.355722	0.423028	0.3
G_20_04	-0.991237	-0.991237	0.147702	1.897340	1.065384	0.718396	0.533063	0.431008	0.441622	0.447337	0.370591	0.377939	0.284864	0.359161	0.315889	0.382021	0.3
G_14_02	-1.008663	-1.008663	0.646700	3.084267	1.545845	0.890976	0.625910	0.493377	0.456996	0.405022	0.282884	0.298476	0.345252	0.342653	0.358246	0.256897	0.3

Slika 4.16: Prikaz dijela ulaznih podataka modela

Ukupan skup podataka čini 56 podskupova mjerenja gdje, kao što smo definirali u potpoglavlju 4.5, jedan podskup označava jedan uzorak određene difuzijske konstante i sadrži petnaestak snimljenih autokorelacijskih krivulja. Iako smo pripremili samo 16 uzoraka različitih masenih udjela glicerola (vidi tablicu 4.1), podskupova je

bilo 56 jer je isti uzorak obično mjereno više puta u različite dane. Budući da je temperatura laboratorija varirala kroz dane, difuzijske konstante istih uzoraka su se malo razlikovale. Za potrebe treniranja modela, izračunata je difuzijska konstanta za svaki od 56 podskupova mjerenja, te su njome označeni podskupovi mjerenja. Posljedično, difuzijska konstanta je igrala ulogu ciljne varijable modela.

Podaci su podijeljeni u tri skupa. Trening skup, validacijski skup i testni skup. Trening skup sadrži 524 mjerenja, validacijski skup 131 mjerenja dok testni skup sadrži 97 mjerenja. Mjerenja koja su dio testnog skupa se ne mijenjaju. Točnije, radi malog broja uzoraka na kojima su obavljena mjerenja, testni skup je kreiran ručno. Izabrano je 7 različitih uzoraka (od njih ukupno 16) te od svakog izabranog uzorka jedan podskup mjerenja. Odabrani podskupovi mjerenja nisu nužno bili prikupljeni u isti dan. Nije bilo moguće kreirati veći testni skup jer bi se ostavilo premalo podataka za učenje modela (za trening i validacijski skup). Trening skup i validacijski skup su se kreirali tako da je od preostalih mjerenja, nasumično 80% mjerenja razdvojeno u trening skup, dok je preostalih 20% smješteno u validacijski skup. Prilikom svakog pokretanja treninga modela trening skup i validacijski skup se nanovo inicijaliziraju, što znači da trening skup i validacijski skup nisu nikada isti za dva različita istrenirana modela. Time se postiže da se kroz svaki trening modela generira malo drugačiji model, s obzirom da je trening skup uvijek malo drugačiji. To u konačnici omogućuje kreiranje ansambla modela. U praksi, ansambli modela daju bolja predviđanja [15].

#### 4.6.2 Model i rezultati

Korištena je jednostavna potpuno povezana neuronska mreža. Ulazni sloj se sastoji od 244 neurona. On prima jedan redak Pandas Dataframe objekta, koji sadrži jedno mjerenje (244 stupaca po retku). Svaki redak inače ima 254 stupaca, no prvih 10 nije korišteno jer sadrže šum (mali vremenski pomaci). Vremenski pomaci se kreću od  $0.5 \mu s$  do  $97431 \mu s$ . Nakon ulaznog sloja, nalaze se dva duboka sloja neuronske mreže. Prvi sadrži 50 neurona, a drugi 20 neurona. Oba sloja su aktivirana ReLu aktivacijskom funkcijom. Izlazni sloj je linearna funkcija koja prikuplja ulaz iz posljednjeg dubokog sloja te kao rezultat daje predviđenu vrijednost difuzijske konstante za ulazne podatke. Nakon oba duboka sloja nalaze se dropout slojevi koji odbacuju 30% neurona u oba sloja prilikom svake epohe, kao što je objašnjeno u teorijskom razmatranju. Korišten je Adam optimizacijski algoritam. Broj epoha treninga je bio

varijabilan jer se koristila metoda ranijeg zaustavljanja na osnovu promjene funkcije greške na validacijskom skupu. Na slici 4.17 vidimo broj parametara modela kao i arhitekturu mreže.

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 50)	12250
dense_1 (Dense)	(None, 20)	1020
dense_2 (Dense)	(None, 1)	21

=====  
 Total params: 13,291  
 Trainable params: 13,291  
 Non-trainable params: 0

Slika 4.17: Arhitektura mreže s brojem parametara mreže

Istrenirano je 10 modela iste arhitekture, no kako su prije inicijalizacije treninga svakog modela nanovo inicijalizirani trening i validacijski skup, svaki model je malo drugačiji. Nakon treninga, svaki model radi predviđanje difuzijske konstante na cijelom testnom skupu, tj. predviđa koja difuzijska konstanta pripada svakoj krivulji u skupu. Predviđanja svakog modela se uprosječuju po podskupu mjerenja (njih 7). Svakom modelu tada pripada lista duljine 7, u koju su spremljene vrijednosti srednjih difuzijskih konstanti 7 podskupova iz testnog skupa. Analogno se sprema lista s nepouzdanostima difuzijske konstante. Nakon treninga svih 10 modela, vrijednosti su spremljene u  $10 \times 7$  matricu, te se računa opća srednja vrijednost  $D_{NN}$  i opća nepouzdanost  $M_{D_{NN}}$  podskupova mjerenja (stupci matrice) na osnovu vrijednosti iz matrice.  $D_{NN}$  i  $M_{D_{NN}}$  možemo shvatiti kao predviđanja ansambla modela. Ovom procedurom se dobivaju robusniji rezultati nego u slučaju da se koristi samo jedan model, pogotovo uzevši u obzir da je podataka jako malo. Rezultati su predstavljeni u tablici 4.3.

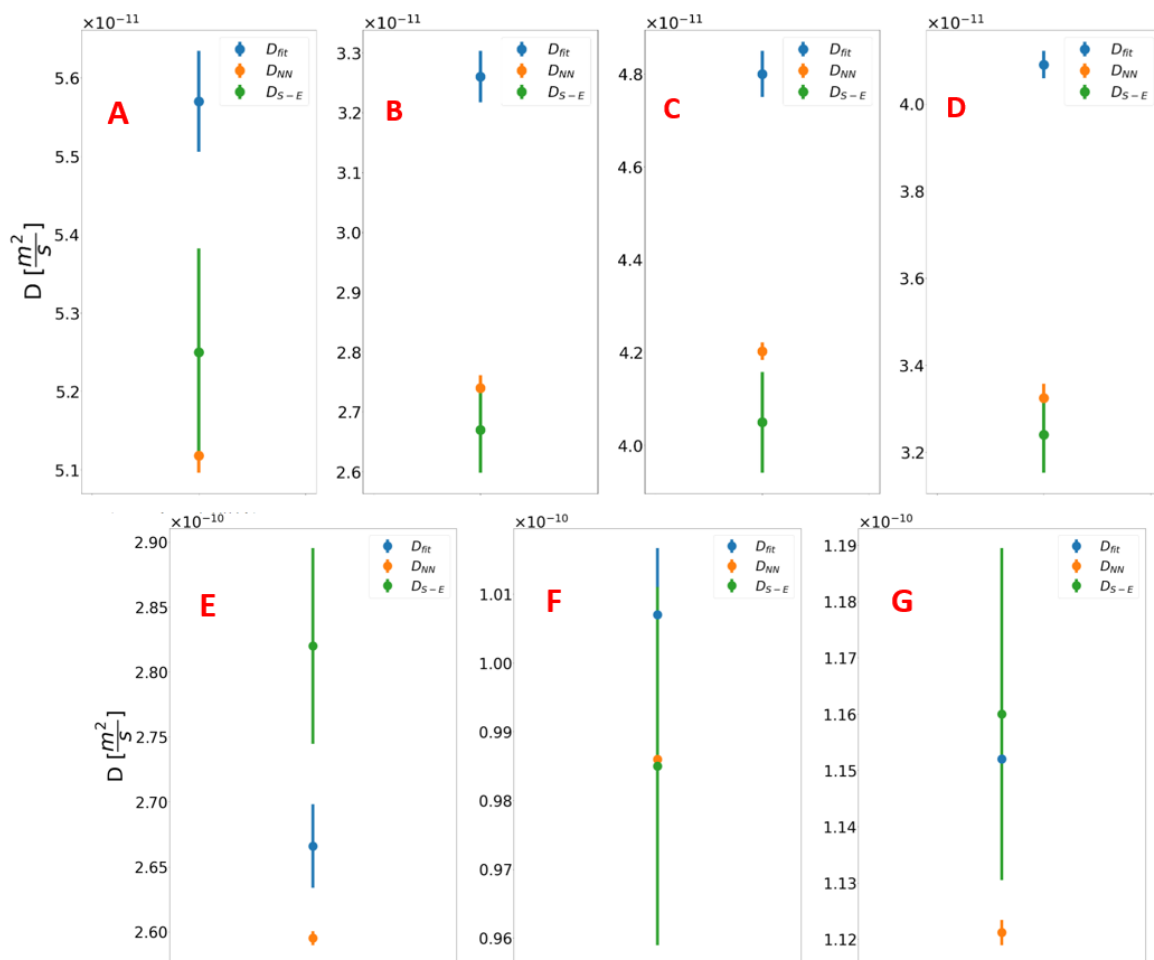
<i>Uzorci</i>	$D_{NN}$	$M_{D_{NN}}$	$D_{S-E}$	$ D_{NN} - D_{S-E} $
E	2.595	0.005	2.82	0.23
G	1.121	0.002	1.16	0.04
F	0.986	0.004	0.985	0.001
A	0.512	0.002	0.525	0.013
C	0.420	0.002	0.405	0.015
D	0.332	0.002	0.324	0.08
B	0.274	0.002	0.267	0.007

Tablica 4.3: Rezultati neuronske mreže poredani po veličini difuzijske konstante u jedinicama  $10^{-10} \frac{m^2}{s}$ .  $D_{NN}$  opća srednja vrijednost ansambla modela,  $M_{D_{NN}}$  opća nepouzdanost ansambla modela,  $D_{S-E}$  vrijednost difuzijske konstante izračunate iz Stokes-Einstein jednadžbe 2.21 te  $|D_{NN} - D_{S-E}|$  apsolutna razlika vrijednosti  $D_{NN}$  dobivenih neuronskom mrežom te one izračunate iz 2.21.

#### 4.7 Usporedba rezultata neuronske mreže i teorijske prilagodbe podacima

Na slici 4.18 vidimo vizualni prikaz rezultata dobivenih prilagodbom teorijske funkcije 2.26 na autokorelacijske krivulje i rezultata dobivenih neuronskom mrežom. Za usporedbu rezultata korišten je prethodno opisani testni set. Vidimo da je u svih 7 slučajeva srednje predviđanje neuronske mreže unutar 10% od vrijednosti dobivenih iz izraza 2.21 što je uspjeh s obzirom na količinu podataka korištenih u treningu modela. Razvidno je da neuronska mreža za uzorke A, B, C, D i F daje vrijednosti bliže od vrijednosti dobivenih teorijskom prilagodbom podacima, a tek u dva slučaja (uzorak E i G) je obratno.

S obzirom da je podataka za trening mreže i testiranje pristupa premalo, nije moguće zaključiti koji pristup je bolji. No, s obzirom da teorijska prilagodba podacima nije ovisna o broju dostupnih mjerenja (zato što prilagodba uvijek konvergira na iste parametre), dok moć predviđanja neuronske mreže jako ovisi o količini podataka koji se koriste u treningu mreže, za pretpostaviti je da bi skupljanjem većeg skupa podataka za trening neuronska mreža popravila preciznost predviđanja difuzijske konstante na cijelom testnom skupu. Proširenje skupa podataka bi se ponajprije odnosilo na povećanje broja različitih uzoraka u smislu pripreme većeg broja smjese glicerol-voda različitih viskoznosti. Isto tako, povećanje ukupnog broja podskupova mjerenja po uzorku moglo bi se s lakoćom izvesti ukoliko bi se spremnik uzorka mogao dovesti i održavati na proizvoljnoj temperaturi (npr. termoelektričnim hladnjakom ili grijačem), neovisnoj od temperature laboratorija.



Slika 4.18: Srednje vrijednosti i nepouzdanosti difuzijske konstante na testnom skupu



## 5 Zaključak

U ovom radu testirana je alternativna metoda obrade podataka koje generira fluorescentna korelacijska spektroskopija (FCS). Metoda koristi strojno učenje, točnije neuronsku mrežu, da bi se iz podataka odredila difuzijska konstanta fluorescentne molekule (fluorofore). Izveden je teorijski model autokorelacijske funkcije s uključenom tripletnom i difuzijskom dinamikom koristeći gaussijansku aproksimaciju detekcijskog volumena. Mjerenja pokazala da ta aproksimacija nije sasvim opravdana barem u aksijalnom smjeru. Motivacija za pronalaženje alternativne metode je neprikladnost korištenja gaussijanske aproksimacije u većini eksperimentalnih postava pa i ovom kojeg smo mi koristili.

Strategija za validaciju alternativne metode sastoji se od sljedećih koraka. Prvo, prikupljanje mjerenja na skupu uzoraka čije se viskoznosti otapala, i posljedično difuzijske konstante fluorofore Alexa647, mijenjaju u rasponu od interesa za opažanje biokemijskih reakcija.

Drugo, koristeći izvedeni teorijski model autokorelacijske funkcije, određen je detekcijski volumen FCS postava te njegove dimenzije u radijalnom i aksijalnom smjeru ( $\rho_0, z_0$ ) što je bilo nužno za određivanje difuzijske konstante pomoću teorijske prilagodbe. Potom se podaci obrađuju te se radi prilagodba teorijskih krivulja na dobivena mjerenja. Iz prilagodbi se kao izlazni parametri dobivaju difuzijske konstante fluorofore u uzorcima različitih dinamičkih viskoznosti.

Treći korak je treniranje neuronske mreže na prikupljenim mjerenjima koja predviđa difuzijsku konstantu fluorofore u uzorku. U programskom jeziku Python su obrađeni podaci iz FCS mjerenja, te je kreiran jednostavan model potpuno povezane neuronske mreže. S obzirom na ograničenja u resursima nije bilo moguće skupiti veći skup mjerenja na kojemu bi se neuronska mreža mogla još bolje optimizirati.

U zadnjem koraku, rezultati dobiveni teorijskom prilagodbom uspoređuju se s onima dobivenima neuronskom mrežom. Možemo zaključiti da je model neuronske mreže korišten u ovom radu u stanju predvidjeti difuzijsku konstantu uzoraka unutar 10% od referentne vrijednosti.

Potrebno je naglasiti da je veliki dio vremena proveden u laboratoriju, te većina mjerenja nije bila uporabljiva što je ograničilo primjenu kompleksnijih modela strojnog učenja. No, dobiveni rezultati ukazuju da je pristup korištenjem strojnog učenja

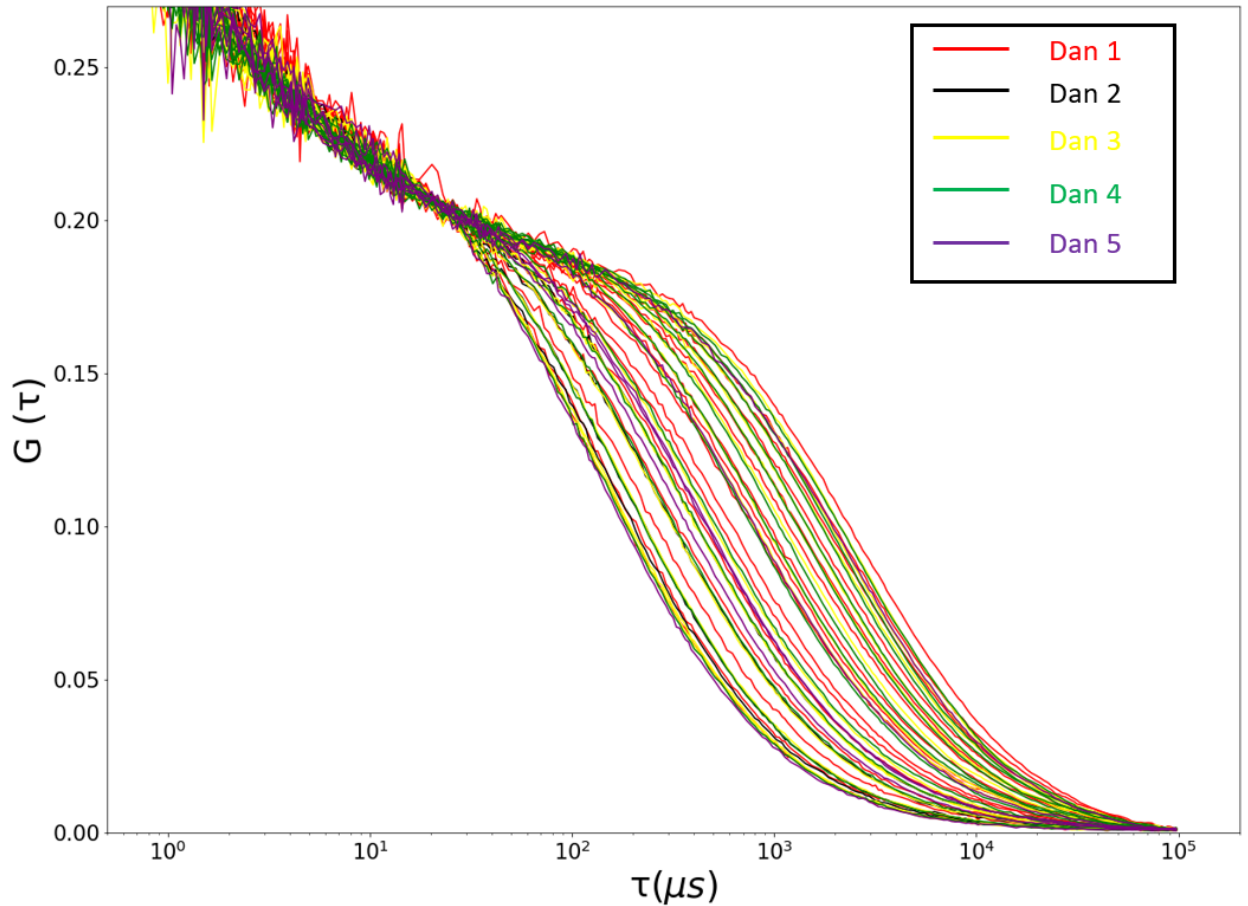
i s tako malim brojem podataka koji su se koristili za trening modela, barem usporediv po preciznosti s predviđanjima koje daje prilagodba teorijske funkcije korištene u ovom radu. Također, kako se dinamika Alexa647 molekule može dobro opisati teorijskim modelom s uključenom difuzijskom i tripletnom komponentom, rezultati dobiveni prilagodbom teorijske funkcije su usporedivi s rezultatima dobivenima neuronskom mrežom. No, lako je zamisliti da bi u slučaju nekih drugih molekula s kompliciranijom fotofizikom i fotokemijom rezultati bili lošiji. Potrebno bi bilo i te efekte uključiti u teorijski model.

U tom slučaju, pristup korištenjem strojnog učenja bi potencijalno mogao izravno iz podataka naučiti difuzijska svojstva molekule i davati drastično bolje rezultate jer model ne zahtijeva dodatne pretpostavke. Također, u slučaju mjerenja provedenih u sklopu ovog rada eksperimentalni artefakti, u obliku odstupanja detekcijskog volumena od gaussijanskog profila, nisu bili jako izraženi. U slučaju izraženijih artefakata bi pristup strojnim učenjem opet mogao dati robusnije rezultate.

Rezultati ukazuju da bi se strojno učenje moglo koristiti i u drugačijim tipovima problema. Primjerice, moguće je promatrati više različitih molekula u mediju fiksirane viskoznosti, te istrenirati model na skupu podataka autokorelacijskih krivulja koje pripadaju tim molekulama. Nakon treninga, model bi mogao primiti novo mjerenje, tj. jednu novu autokorelacijsku krivulju, te bi na osnovu predviđanja difuzijske konstante bilo moguće procijeniti kojoj molekuli pripada ta autokorelacijska krivulja. Drugi mogući pristup je treniranje klasifikatora molekula u otopini. Slično kao u prethodnom slučaju, na osnovu velikog broja mjerenja na  $n$  različitih molekula te autokorelacijskih krivulja iz tih mjerenja, može se istrenirati model koji predviđa vjerojatnost da ulazna autokorelacijska krivulja pripada svakoj od  $n$  molekula.

# Dodaci

## Dodatak A Vizualni prikaz potpunog skupa mjerenja



Slika A.1: Prikaz uprosjećenih autokorelacijskih krivulja koje čine potpuni skup mjerenja (56 podskupova) provedenih kroz pet dana. Samo u prvom danu (crvene krivulje) su provedena mjerenja na svim uzorcima, te je to razlog zašto između nekih crvenih krivulja nema interpoliranih krivulja drugih boja.

## **Dodatak B Korišteni programski paketi**

U ovom diplomskom radu korišten je programski jezik Python te biblioteke Python okruženja. Za pokretanje koda korišten je Google Colaboratory [22]. To je okruženje koje nudi mogućnost pokretanja koda na oblaku (engl. cloud). Korištenjem vanjskih resursa, omogućeno je treniranje modela onima koji ne posjeduju potrebu infrastrukture.

### ***B.1 NumPy***

NumPy [23] je biblioteka koja se koristi za znanstveno računanje i manipulaciju višedimenzionalnih reprezentacija podataka. Te reprezentacije se mogu interpretirati kao matrice i tenzori. Objekti kao, numpy array (tenzor), često se, a nekad i nužno, koriste kao ulaz u razne algoritme strojnog učenja.

### ***B.2 pandas***

Pandas je biblioteka koja služi za učitavanje podataka u raznim formatima u Dataframe ili Series objekte implementirane u biblioteci. Series reprezentacija služi za obradu i vizualizaciju vremenskih nizova. Dataframe reprezentacija služi za prikaz višedimenzionalnih podataka, gdje svaka dimenzija igra ulogu jedne varijable (stupca), a svaki podatak sadržan je u jednom retku. Pandas sadrži brojne metode za manipulaciju forme podataka kao i provođenje deskriptivne statistike. Također, moguće je lako transformirati podatke pomoću proizvoljne funkcije. To je korisno jer se omogućava primjerice transformiranje samo jednog stupca, dok ostali ostaju nepromijenjeni. Time se lako kreiraju nove varijable koje mogu biti od koristi pri treningu modela. U pandasu su također implementirane rudimentarne metode vizualizacije podataka, ponajviše korisne pri analizi vremenskih serija.

### ***B.3 matplotlib***

Matplotlib [24] je najpoznatija biblioteka za vizualizaciju podataka. Daje puno veći spektar mogućnosti od metoda implementiranih u pandasu. Lako se kombinira s NumPy bibliotekom.

## ***B.4 scikit-learn***

Scikit-learn [25] je biblioteka koja sadrži vrhunske implementacije nekih od najpoznatijih algoritama strojnog učenja. U pozadini scikit-learn koristi druge poznate biblioteke kao NumPy i matplotlib. Ova biblioteka ne podržava mijenjanje arhitekture modela izvan dostupnih metoda i hiperparametara modela.

## ***B.5 TensorFlow***

TensorFlow [26] je biblioteka specijalizirana za strojno, posebice duboko učenje. Posjeduje brojne metode za kreaciju tenzorskih objekata te njihovu manipulaciju. U TensorFlowu je moguće implementirati nove algoritme. To je i bitna razlika između ove biblioteke i scikit-learna.

## ***B.6 Keras***

Keras [27] je biblioteka koja u pozadini koristi TensorFlow, a specijalizirana je za duboko učenje. Omogućava provođenje tenzorskih operacija na CPU, GPU i TPU računalnim jedinicama. Tenzorske operacije su većinom skrivene u pozadini te ne zahtijevaju od korisnika duboko poznavanje implementacije tih operacija. Ideja iza Kerasa je "sakriti" niži stupanj apstrakcije rada algoritma, a u fokus staviti viši stupanj apstrakcije. U Kerasu je lako definirati razne arhitekture neuronskih mreža pozivanjem već predefiniраниh slojeva mreže implementiranih u biblioteci. Također, Keras sadrži brojne optimizacijske algoritme, aktivacijske funkcije i sl.

## ***B.7 random***

Random [28] je modul koji je sastavni dio implementacije programskog jezika Python. Modul sadrži implementirane pseudo-slučajne generatore raznih statističkih distribucija.

## ***B.8 LMFIT***

LMFIT [29] je biblioteka koja služi za provođenje prilagodbi teorijskih funkcija na podatke. U biblioteci su već definirane brojne često korištene funkcije, no moguće je definirati vlastitu funkciju te ju koristiti za prilagodbu.

## **Dodatak C Dostupnost programskog koda**

Programski kod je dostupan na GitHubu [30]. Korišteni podaci su dostupni na zahtjev.

## Literatura

- [1] Lakowicz, J. R. Principles of Fluorescence Spectroscopy. 3rd ed. Boston : Springer, 2006.
- [2] Blom, H.; Björk, G. Lorentzian spatial intensity distribution in one-photon fluorescence correlation spectroscopy. // Applied Optics. Vol. 48, 31(2009), str. 6050-6058.
- [3] Hess, S.; Webb, W. Focal Volume Optics and Experimental Artifacts in Confocal Fluorescence Correlation Spectroscopy. // Biophysical Journal. Vol. 83, 4(2002), str. 2300-2317.
- [4] Schwille, P.; Haustein, E. Fluorescence Correlation Spectroscopy - An Introduction to its Concepts and Applications. // Spectroscopy. Vol.94, 3(2009), str. 1-33.
- [5] Fluorescence and Phosphorescence, (1.9.2020.), Chemistry LibreTexts, <https://chem.libretexts.org/@go/page/41400>, 20.10.2021
- [6] Fluorescence Microscopy Interactive Tutorials: Photobleaching, (12.9.2016.), Molecular Expressions, <https://micro.magnet.fsu.edu/primer/java/fluorescence/photobleaching/>, 21.10.2021.
- [7] Bacia K.; Haustein E.; Schwille P. Fluorescence correlation spectroscopy: principles and applications. // Cold Spring Harb Protoc. Vol. 2014, 7(2014), str. 709-25.
- [8] Fluorescence Correlation Spectroscopy (FCS): Correlation analysis of temporal fluctuations in fluorescence intensity of fluorophores, PicoQuant, <https://www.picoquant.com/applications/category/life-science/fluorescence-correlation-spectroscopy-fcs>, 15.11.2021.
- [9] Loman, A. Molecular Sizing using Fluorescence Correlation Spectroscopy. Doktorski rad. Göttingen : Universität zu Göttingen, 2010.
- [10] Absolute Diffusion Coefficients: Compilation of Reference Data for FCS Calibration, PicoQuant, [https://www.picoquant.com/images/uploads/page/files/7353/appnote\\_diffusioncoefficients.pdf](https://www.picoquant.com/images/uploads/page/files/7353/appnote_diffusioncoefficients.pdf), 16.1.2022.

- [11] Gregersen, S. Fluorescent Peptide-Stabilized Silver-Nanoclusters. Doktorski rad. Københavns : Københavns Universitet, 2014.
- [12] Fluorophore Selection, ThermoFisher, <https://www.thermofisher.com/hr/en/home/life-science/cell-analysis/fluorophores.html>, 16.11.2021.
- [13] Nielsen, M. Neural networks and deep learning. Determination Press, 2015., <http://neuralnetworksanddeeplearning.com/>, 16.1.2022.
- [14] Mehta, P.; Bukov, M.; Wang, C.; Day, A.; Richardson, C.; Fisher, C.; Schwab, D. A high-bias, low-variance introduction to Machine Learning for physicists. // Physics Reports, Vol. 210 (2018), str. 1-124.
- [15] Géron, A. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensor-Flow. 2nd ed. New York : O'Reilly Media, 2019
- [16] Hac, A. Diffusion Processes in Membranes Containing Coexisting Domains Investigated by Fluorescence Spectroscopy. Doktorski rad. Göttingen : Universität zu Göttingen, 2003.
- [17] Round Continuously Variable Metallic Neutral Density Filters, Thorlabs, [https://www.thorlabs.com/NewGroupPage9.cfm?ObjectGroup\\_ID=1393&pn=NDC-25C-2&RiLink=1](https://www.thorlabs.com/NewGroupPage9.cfm?ObjectGroup_ID=1393&pn=NDC-25C-2&RiLink=1), 12.11.2021
- [18] Calculator to estimate the density and viscosity of glycerine / water mixtures, Chris Westbrook, [https://www.met.reading.ac.uk/~sws04cdw/viscosity\\_calc.html](https://www.met.reading.ac.uk/~sws04cdw/viscosity_calc.html), 2.1.2021
- [19] Volk, A.; Kähler, C.J. Density model for aqueous glycerol solutions. // Experiments in Fluids volume 59, 75(2018), str. 75-79.
- [20] Python, <https://www.python.org>, 1.10.2021.
- [21] pandas, <https://pandas.pydata.org/>, 1.10.2021.
- [22] Google Colaboratory, <https://colab.research.google.com/>, 1.10.2021.
- [23] NumPy, <https://numpy.org/>, 1.10.2021.
- [24] matplotlib, <https://matplotlib.org/>, 1.10.2021.



- [25] scikit-learn, <https://scikit-learn.org/>, 1.10.2021.
- [26] TensorFlow, <https://www.tensorflow.org/>, 1.10.2021.
- [27] Keras, <https://keras.io/>, 1.10.2021.
- [28] random, <https://docs.python.org/3/library/random.html>, 1.10.2021.
- [29] LMFIT, <https://lmfit.github.io/lmfit-py/>, 1.10.2021.
- [30] Primjena strojnog učenja za korekciju artefakata u fluorescencijskoj korelacijskoj spektroskopiji, GitHub, [https://github.com/Paarmenid/Primjena\\_strojnog\\_ucenja\\_za\\_korekciju\\_artefakata\\_u\\_fluorescencijskoj\\_korelacijskoj\\_spektroskopiji](https://github.com/Paarmenid/Primjena_strojnog_ucenja_za_korekciju_artefakata_u_fluorescencijskoj_korelacijskoj_spektroskopiji), 15.2.2022.