

**ADVERTIMENT.** L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  <https://creativecommons.org/licenses/?lang=ca>

**ADVERTENCIA.** El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <https://creativecommons.org/licenses/?lang=es>

**WARNING.** The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>

**Decision-making in quantum environments:**  
from model-free to model-aware learning of quantum controls

by  
MATIAS BILKIS

*This page is intentionally left blank*

Universitat Autònoma de Barcelona

**Decision-making in quantum environments:**  
from model-free to model-aware learning of quantum controls

by  
MATIAS BILKIS

under the supervision of  
Prof. JOHN CALSAMIGLIA COSTA,

A thesis submitted in partial fulfillment for the degree of  
Doctor of Philosophy in Physics

Física Teòrica: Informació i Fenòmens Quàntics  
Departament de Física, Facultat de Ciències



Bellaterra, Defense date

This work is licensed under a [Creative Commons](https://creativecommons.org/licenses/by-nc-sa/3.0/) “Attribution-NonCommercial-ShareAlike 3.0 Unported” license.



*To my family and friends,  
from here, there and beyond,  
who supported me  
during this amazing adventure.*

*This page is intentionally left blank*

---

# Abstract

---

The success in the experimental control of quantum systems, along with the advent of artificial intelligence, raises the question of where do these worlds meet. In this thesis, a series of frameworks is proposed, in which a machine-learning agent sequentially interacts with a quantum system in order to learn control strategies. Such frameworks are interconnected with each other by the degree of knowledge the agent possesses on the setting at hand.

Our journey begins by model-free learning of long-distance quantum communication protocols, continues towards an intermediate-stage of awareness in which NISQ-computing problems are tackled, and finally lands on the fully-aware terrain, where statistical-inference problems are considered for continuously-monitored systems.

Specifically, we depart from a completely agnostic assumption, where a reinforcement learning agent is posed to the task of calibrating a quantum receiver that decodes information out of a continuous-variable system. From here, some light is shed into the learning problem, and in the second framework we consider how to design useful quantum circuits for currently-available quantum computers. To this end, a semi-agnostic algorithm that jointly optimizes both the structure and the parameters of the circuit is introduced. Finally, we consider how open quantum dynamics can be inferred by an agent which is continuously-monitoring a system. Here, the structure of the quantum evolution is assumed to be known, and the agent needs to learn which is the underlying hypothesis that describes systems' dynamics out of a noisy measurement signal.

The methods developed in this thesis can potentially be applied to real-life scenarios, and represent a step forward towards the constitution of quantum machine learning.





---

# Resum

---

L'èxit en el control experimental dels sistemes quàntics, juntament amb la presència quotidiana de diferents intel·ligències artificials, planteja la pregunta d'on es creuen aquests mons. En aquesta tesi es proposa una sèrie de marcs en què un agent d'aprenentatge automàtic interactua seqüencialment amb un sistema quàntic per tal d'aprendre estratègies de control. Aquests marcs estan interconnectats entre si pel grau de coneixement l'agent té sobre l'escenari en qüestió.

El nostre viatge comença amb l'aprenentatge sense model de protocols de comunicació quàntica sobre llarga distància, continua cap a una etapa intermèdia de consciència en què s'aborden problemes de computació quàntica tipus NISQ i finalment aterra al terreny totalment conscient, on es consideren problemes d'inferència estadística en sistemes quàntics contínuament monitoritzats.

Específicament, partim d'un escenari completament agnòstic on un agent d'aprenentatge per reforç s'enfronta a la tasca de calibrar un receptor quàntic que descodifica informació clàssica d'un sistema quàntic de variable contínua. Partint d'aquest context, llancem una mica de llum sobre el problema de l'aprenentatge, i el segon escenari consisteix a dissenyar circuits quàntics útils per als ordinadors quàntics actualment disponibles. Per això, introduïm un algorisme semi-agnòstic que optimitza de manera conjunta tant l'estructura com els paràmetres del circuit. Finalment, considerem com la dinàmica d'un sistema quàntic pot ser inferida per un agent que està monitoritzant contínuament aquest sistema. Aquí, assumim que l'estructura de la dinàmica és coneguda per l'agent, que necessita aprendre quina és la hipòtesi subjacent que descriu l'evolució a partir d'un senyal de mesura sorollós.

Els mètodes desenvolupats en aquesta tesi estan preparats per poder ser aplicats a escenaris de la vida real i representen un pas endavant cap a la constitució de l'aprenentatge automàtic quàntic.



---

# Resumen

---

El control experimental exitoso de ciertos sistemas cuánticos, junto con la presencia de la inteligencia artificial, plantea la pregunta de dónde estos mundos se encuentran. En esta tesis proponemos una serie de marcos donde agente de aprendizaje automático interactúa secuencialmente con un sistema cuántico para aprender estrategias de control. Los marcos están interconectados entre sí por el grado de conocimiento el agente posee sobre el escenario en cuestión.

Nuestro viaje comienza con el aprendizaje sin modelo para protocolos de comunicación cuántica, continúa hacia una etapa intermedia de conciencia donde se abordan problemas de computación cuántica, y finalmente aterriza en el terreno totalmente consciente, considerando problemas de inferencia estadística en sistemas continuamente monitoreados.

En particular, partimos de una escenario agnóstico, donde un agente de aprendizaje por refuerzo se enfrenta a la tarea de calibrar un receptor cuántico que decodifica estados coherentes. Continuamos arrojando algo de luz sobre el grado de conciencia, y nuestro segundo escenario consiste en el diseño de circuitos cuánticos útiles para dispositivos actualmente disponibles. Para ello, introducimos un algoritmo semi-agnóstico que optimiza conjuntamente la estructura y los parámetros del circuito. Finalmente, estudiamos cómo inferir la dinámica de un sistema a partir de su monitoreo continuo. Aquí, asumimos que la estructura de la dinámica es conocida por el agente, quien necesita aprender cuál es la hipótesis subyacente que describe la evolución, a partir de la señal de medidas ruidosas.

Los métodos desarrollados en esta tesis están listos para ser aplicados a escenarios de la vida real y representan un paso adelante hacia la constitución del aprendizaje automático cuántico.

*This page is intentionally left blank*

---

# Declaration

---

I declare that the thesis has been composed by myself and that the work has not been submitted for any other degree or professional qualification. I confirm that the work submitted is my own, except where work which has formed part of jointly-authored publications has been included. My contribution and those of the other authors to this work have been explicitly indicated below. I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others.



---

## Related works

---

- [BRY+20] M. Bilkis et al. “Real-time calibration of coherent-state receivers: Learning by trial and error”. In: *Phys. Rev. Research* 2 (3 2020), p. 033295. DOI: [10.1103/PhysRevResearch.2.033295](https://doi.org/10.1103/PhysRevResearch.2.033295).
- [BCV+21] M Bilkis et al. “A semi-agnostic ansatz with variable structure for quantum machine learning”. In: *arXiv preprint arXiv:2103.06712* (2021).
- [BRC21] M. Bilkis, Matteo Rosati, and John Calsamiglia. “Reinforcement-learning calibration of coherent-state receivers on variable-loss optical channels”. In: *2021 IEEE Information Theory Workshop (ITW)*. 2021, pp. 1–6. DOI: [10.1109/ITW48936.2021.9611396](https://doi.org/10.1109/ITW48936.2021.9611396).
- [GGC] M. Bilkis G. Gasbarri and J. Calsamiglia. *Sequential testing for continuously monitored quantum systems - appearing soon*.



*This page is intentionally left blank*

---

## Acknowledgments & Cariños

---

Many (so many!) helped me during this beautifully intense process called PhD. Among all of them, John Calsamiglia has undoubtedly been my super-hero, being available in stormy moments when the only clear thing was confusion. Furthermore, this thesis would not exist if it wasn't for the two extraordinary post-docs I have worked with: Matteo Rosatti and Giulio Gasbarri. To the amazing combination of PI & post-docs I will be forever grateful: they taught me an incredible amount of things, and not only physics but also life. During these years, the research group has been some sort of family to me, and it has been a pleasure (and of course, sometimes a pain, that's family!)<sup>1</sup> to share the office with Matt, María, Minglai, Mani, Giulio, Matteo, Andreas, Anna, Ramón, Emili, Philipp, Gael, Suzana, Joe, Carlo, Naga, Illaria, Marco, Jennifer, Yago, Alonso, Jessica, Andreu, Antoni, Elisabeth, Esteban, Farzin, Zahra, and so many more!

Tracing back a bit, I thank to all the people who encouraged me to move so far away from home to do this PhD. Above all, my family, who was an essential support during these times, even from so far away: mis viejos, mi hermana, Elena, Mariano, Flavia, Mandi, Paloma, Ruth, Tomi, Gonza, Marcelo, los abuelos Yaco y Scheine, Fanny y Abraham, y todo el resto.

It was José Edelstein who first suggested me to move here (Barcelona); the Toni Acín, who has always been super receptive, was my first contact with this amazing city and got me in touch with GIQ. So big thanks to them!

Also, big thanks to all of my teachers and collaborators from whom I have learnt and still learn so much; among Raúl Rossignoli, Norma Canosa, Mauricio Matera, Lorena Rebón, José Esté, Raul Morral, Marco Cerrezo, Lukasz Cincio, Patrick Coles

---

<sup>1</sup>This is a joke, if someone is reading this, I expect her/him to laugh :)

and Guillaume Verdon. Big thanks to the research institutions Instituto de Física La Plata and Los Alamos National Laboratory for hosting me during this years.

Also, I an enormous help to my psychologist, who was certainly necessary to help me to take so many difficult decisions. Por el infinito apoyo durante estos años, muchas gracias Maren!

Son muchos años y por lo tanto mucha gente a la que agradecer, sin la cual toda esta aventura tampoco hubiera sido posible. Son al menos dos los lados del océano que toca agradecer, pero las fronteras se suelen confundir un poco en estas épocas de tanta conexión. Sin más orden que el desorden, millones de gracias por todo el aguante a Valen Caso Rosendi, Caro Durián, Joaco Gil, Pedro Pablo, Javi Argüello, Mecchi, Eric, Amy, Sara, Jero, Seba Cabuli, Jose Giugno, Ines, Lucas, Nico (y toda la banda de natación platense!). Al maravishoso equipo Mashonesero que formamos con Cris y Bea, a los vecinos de La Floresta y Vallvidrera: Carol, Ale, Taba, Nati, Iña, Lucho y compañía. También Teresa, Marta, que les quiero mucho! Al Max Nordau en La Plata que siempre estará en mis raíces, a Lauti Krol, Pau Fioramonti, Eli Butinof, Juli Coloma, Caro Bikis que es mi hermana pero ya que estamos la nombro de nuevo. ¡Cuánto aguante! También a Raulo, Manu, Juli, Caro y Alicia Golijov (y Custom!). También Benja, los Martines (Lagares y Tomé), Poema, Floppy, Nela, Nacho Bruvera, Fiorela, Cecchi. A la Unió Ciclista de Sant Cugat y todo el equipo de triatlón! Cada persona que nombro me ha dado —y también las muchas que me olvido nombrar en el apuro de enviar esta tesis!—, durante estos años, un apoyo inigualable.

I also want to thank the unique support from people at Port d'Informació Científica (PIC), who lended me their computational resources which turned extremely useful when getting the numerical results that we will later analyze in the thesis.

Ah! Before closing, special thanks to my dog that lives in La Plata, Dam, who has always been there — even virtually — for the routinary hugs :)

*This page is intentionally left blank*



---

# Contents

---

|          |                                                     |          |
|----------|-----------------------------------------------------|----------|
| <b>1</b> | <b>Motivation</b>                                   | <b>1</b> |
| <b>2</b> | <b>Preliminaries</b>                                | <b>5</b> |
| 2.1      | Some basic notions of quantum information . . . . . | 5        |
| 2.1.1    | Closed-system dynamics . . . . .                    | 8        |
| 2.1.2    | Quantum channels . . . . .                          | 9        |
| 2.1.3    | Generalized quantum measurements . . . . .          | 10       |
| 2.1.4    | Quantum instruments . . . . .                       | 11       |
| 2.1.5    | Quantum master equation . . . . .                   | 12       |
| 2.2      | NISQ & the qubits . . . . .                         | 14       |
| 2.2.1    | Variational Quantum algorithms . . . . .            | 14       |
| 2.2.2    | The Ansatz . . . . .                                | 18       |
| 2.2.3    | The optimization procedure . . . . .                | 22       |
| 2.2.4    | Barren-plateaus . . . . .                           | 28       |
| 2.2.5    | Discussion . . . . .                                | 31       |
| 2.3      | Continuous-variable systems . . . . .               | 31       |
| 2.3.1    | Gaussian systems . . . . .                          | 33       |
| 2.3.2    | Some methods in phase space . . . . .               | 36       |
| 2.3.3    | Gaussian operations and beyond . . . . .            | 40       |
| 2.3.4    | Measurements . . . . .                              | 41       |
| 2.4      | Continuously-monitored systems . . . . .            | 42       |
| 2.4.1    | Cavity emission and photon-detection . . . . .      | 43       |
| 2.4.2    | Continuous photon-detection . . . . .               | 46       |
| 2.4.3    | Continuous homodyne detection . . . . .             | 49       |

|          |                                                     |            |
|----------|-----------------------------------------------------|------------|
| 2.4.4    | Intermezzo II: Ito calculus                         | 52         |
| 2.4.5    | Imperfect detection                                 | 56         |
| 2.4.6    | The Gaussian case                                   | 57         |
| 2.4.7    | Optomechanical systems                              | 60         |
| 2.5      | Statistical Inference                               | 63         |
| 2.5.1    | Hypothesis testing                                  | 64         |
| 2.5.2    | Single-shot quantum state discrimination            | 65         |
| 2.5.3    | Back to the classics: asymmetric hypothesis testing | 70         |
| 2.5.4    | Sequential hypothesis testing                       | 75         |
| 2.5.5    | Parameter estimation                                | 81         |
| 2.6      | Reinforcement learning                              | 83         |
| 2.6.1    | Bandit problems                                     | 85         |
| 2.6.2    | Value functions and the Bellman equation            | 91         |
| 2.6.3    | Q-learning                                          | 94         |
| 2.6.4    | Model aware methods                                 | 96         |
| 2.6.5    | Discussion                                          | 98         |
| <b>3</b> | <b>Learning in the darkness</b>                     | <b>99</b>  |
| 3.1      | Coherent-state quantum receivers                    | 102        |
| 3.1.1    | Gaussian receivers                                  | 104        |
| 3.1.2    | Kennedy receivers                                   | 105        |
| 3.1.3    | Dolinar receivers                                   | 109        |
| 3.1.4    | Time-domain Dolinar receiver & optimality           | 112        |
| 3.1.5    | The sequential structure of Dolinar receiver        | 114        |
| 3.2      | The model-aware approach                            | 117        |
| 3.3      | The model-free approach                             | 122        |
| 3.3.1    | Reinforcement learning the Dolinar receiver         | 122        |
| 3.3.2    | Q-learning the Dolinar receiver                     | 124        |
| 3.3.3    | Optimal state-action values & convergence           | 128        |
| 3.3.4    | A little help from my bandit friends                | 133        |
| 3.3.5    | Noise robustness                                    | 136        |
| 3.4      | Discussion & Future work                            | 143        |
| 3.4.1    | Code                                                | 144        |
| <b>4</b> | <b>Learning in the twilight</b>                     | <b>145</b> |
| 4.1      | The Variable Ansatz (VAns) Algorithm                | 148        |
| 4.1.1    | Insertion method                                    | 151        |
| 4.1.2    | Simplification method                               | 151        |
| 4.1.3    | Scaling of VAns                                     | 156        |

|          |                                                             |            |
|----------|-------------------------------------------------------------|------------|
| 4.2      | Using VAns                                                  | 156        |
| 4.2.1    | Transverse Field Ising model (TFIM)                         | 157        |
| 4.2.2    | $XXZ$ Heisenberg Model                                      | 161        |
| 4.2.3    | Molecular Hamiltonians                                      | 162        |
| 4.2.4    | Quantum Autoencoder                                         | 166        |
| 4.2.5    | Unitary compilation                                         | 169        |
| 4.2.6    | Noise and the $\lambda$ -model                              | 172        |
| 4.3      | Discussion and future perspectives                          | 175        |
| 4.3.1    | Mitigating the effect of barren plateaus                    | 176        |
| 4.3.2    | Future directions                                           | 177        |
| 4.3.3    | Code                                                        | 179        |
| <b>5</b> | <b>Learning in the daylight</b>                             | <b>181</b> |
| 5.1      | Introduction                                                | 181        |
| 5.2      | Hypothesis testing in continuously-monitored systems        | 182        |
| 5.3      | Sequential testing & Gaussian systems                       | 185        |
| 5.3.1    | The log-likelihood ratio                                    | 190        |
| 5.3.2    | To be or not to be Gaussian                                 | 196        |
| 5.3.3    | Joint system evolution as an OU-process                     | 203        |
| 5.3.4    | Deterministic vs Sequential                                 | 206        |
| 5.4      | Parameter estimation & continuously-monitoring              | 209        |
| 5.4.1    | The spectral power and the Lorentizan fit                   | 209        |
| 5.4.2    | Evolution of the Fisher information                         | 211        |
| 5.4.3    | Maximum-likelihood estimation via automatic-differentiation | 212        |
| 5.4.4    | Learning a linear force                                     | 214        |
| 5.5      | Discussion & future perspectives                            | 215        |
| 5.5.1    | Code                                                        | 216        |





*It is imperfection - not perfection - that is the end result of the program written into that formidably complex engine that is the human brain, and of the influences exerted upon us by the environment and whoever takes care of us during the long years of our physical, psychological and intellectual development.*

Rita Levi-Montalcini

# 1

---

## Motivation

---

Motivating this thesis should be quite easy, a friend told me a couple of days ago: just ask ChatGPT. However, I will not do this, and the main reason for that is because I have actually found joy during the writing of my thesis, and I think this is a good starting point. If we find pleasure in doing certain things, like research for example, why would we care in finding protocols to *automatize* them, as done in this thesis? *Ups*: tricky question.

Well, I can think of two reasons. The first one is for the fun in doing so: it has been (and still is) an amazing experience! Yes, definitely fun... but... it does not sound very convincing that my only motivation to do a PhD in quantum machine learning was only for the fun of it, right?

Thus, we need to think on a second reason. Think, think... come on... four and a half years thinking on this PhD and now nothing... Well, at least we can try to imagine the *shape* of the answer. It should be something like... mmmm... a bit harder to swallow... but not too much, since we really want to keep the reader up to (at least) the end of this Section. Exactly! that's the concept I need: "Sections!"

While Sections help us structure things (and I can assure you, my Dear Reader that we *will use them* later in the thesis), they do also place a distance between things. For us scientists, such distance is necessary – and often essential – in order to focus on a specific research topic. But more often than not, scientific communities (and we can well generalize this to the *human knowledge*) fall into an abuse of *knowledge*

*compartmentalization*. And this sounds a more satisfying motivation for why doing this PhD: work on connecting Science back (by contribution with our little grain of sand). In turn, this thesis brings two fields that (back to 2018, when my PhD began) were quite far away from each other. Moreover, and as it happens in many situations when bringing two different worlds together, many beautiful things come up. While I will not use the card that quantum AI could lead to a major breakthrough —I could sell it, but I don't buy it— I do think that research directions that focus on intersecting two communities with each other is *healthy* for science, and this is some kind of *enough motivation* to me.

So what about *quantum*? and what about *AI*? Falling (perhaps not too much!) in contradiction with the previous paragraph, let's go by parts.

Quantum theory is a route to understand physical reality, and I want to put emphasis on its fundamental aspects. Those have to do with describing the reality by means of quantum superpositions and quantum correlations. Moreover, the theory is intrinsically stochastic in nature. Yet the physical reality that we inhabit has — a priori — nothing to do with this wierd phenomena of superpositions and non-local correlations. Furthermore, classical physics is intrinsically deterministic, in the sense that with precise knowledge of a given system, we can perfectly predict its behaviour. Some authors describe this gap as a desert of ignorance occupying many decimals in the scale: how our classical & deterministic world can be reconciled with the quantum theory? We simply do not know the answer yet, and this is perhaps one of the biggest misteries that we currently need to deal with.

On the other hand we have artificial intelligence (AI), and a huge (classical) computing power available —at least to some— right now. While ChatGPT and all the advent of AI is truly impressive, we are certainly not living in an Asimov's tale, and many questions remain open on how can current AIs become *truly intelligent* as we humans are, if there even exists a clear definition for such concept. As of now, the possibility of an AI having the scientific *aha!*s that we scientists are often illuminated with (call it *intuition*) sounds like a fantasy. Perhaps this will never happen, and our human condition hindered us to completely understand the logic behind our reasoning (be it concious or unconcious).

However, a PhD thesis consists in leaving almost all of the unresolved misteries of science, technology and humanity aside for a while, but focusing on only *some*. And I guess this is what motivates the current thesis: on the one hand there is clear evidence that we can control quantum systems up to an unprecedented level<sup>1</sup>. On the other hand, while current AIs might be a game-changer in a volatile society like

---

<sup>1</sup>By this, we mean preparing and storing a system in a quantum superposition for a sufficiently long time.

ours, they do not represent a change-of-paradigm on how intelligence is understood, as they do not even get close to it. This might sound polemic, I know, since for example ChatGPT has recently passed Scott Aaronson’s quantum-mechanics exam with a 7/10 score. However, I do refer to intelligence as something beyond passing an exam. Again, I would be truly impressed (and of course chances are that I am wrong) that an AI *aha!*, and come up with a new theory of knowledge on its own, *i.e.* without human intervention. Thus, AI should be understood as a tool, and not as a stand-alone system.

In particular, we will pay special attention to the currently available quantum technologies, by considering three paradigmatic devices. First we will study continuous-variable quantum receivers, which are aimed to decode information out of a quantum state of light. Secondly, we focus on NISQ computers, which are relatively small and extremely noisy version of quantum computers, but have able provide to provide a quantum advantage with respect to any classical computer paradigm. Thirdly, we will consider the case of continuous-time quantum sensors, where a quantum system stored in an optical cavity is being monitored, which provides an effective to infer information of its environment.

In this thesis we tackle each of the aforementioned devices –also called quantum environments– from the machine-learning perspective. This is done by introducing an agent equipped with a varying level of awareness on the quantum environment she is faced to.

The physical systems that we consider, along with the methods and machine-learning algorithms, are widely explained in the *Preliminaries* Chapter. This (considerably long) section serves as a reference for the subsequent ones. Thus, the reader can straightforwardly begin from its preferred Chapter, and consult the *Preliminaries* when necessary (appropriate reference are included).

The learning journey begins *in the darkness*, where the agent is asked to optimally calibrate a quantum receiver without any knowledge of the setting at hand. Here, the only information given to the agent is a binary reward (a bit, valued either 0 or 1), which stands for the correctness of the guess made for the underlying signal that is aimed to be decoded.

We continue *in the twilight*, with an agent facing the problem of building noisy quantum circuits, in the context of NISQ computing. Here, the solutions the agent is asked to reach are even unknown to us, and we can only aid her by providing specifically-tailored circuit compression rules. To this end, we introduce the VAns algorithm, which sequentially modifies the quantum circuit by following a variable-ansatz structure. Our method is thoroughly tested in paradigmatic quantum machine learning applications, ranging from VQE to quantum autoencoder

---

tasks.

Finally, we *learn during daylight*, where a fully model-aware agent is asked to tackle statistical inference problems. Here, our attention is put on continuously-monitored quantum systems, where the performance of sequential discrimination protocols is studied. Here, we analytically study the evolution of the log-likelihood ratio, by providing insight on error rates and stopping time probability distribution for a sequential test. In addition, the case of parameter estimation is considered, and a maximum-likelihood estimation is carried out by an automatic-differentiation method; such recurrent structure allow us to infer parameters of external signals out of measurement signal.

The main contribution of this thesis is that of “bringing the quantum scientist closer to the AIs, in such a way that they can collaborate with each other”. This thesis, as most of the scientific research, provides a (small) step further towards the constitution of such collaboration. Having this in mind, the reader shall not understand each of the Chapters as a closed-one, but rather as an invitation to continue writing it. In turn, our results do not constitute a milestone in quantum AI, but we rather see them as blueprints towards more elaborate quantum AI framework.

*Yo no se qué pasó // en mil calles la calle se abrió. //*  
*Me perdí y en verdad // no sabía que rumbo tomar//.*  
*Pregunté, porque al fin // preguntando se llega a Pekín.*

— Lo que me costó el Amor de Laura  
Alejandro Dolina

# 2

---

## Preliminaries

---

This chapter is aimed to explain both physics and algorithms studied in the thesis. We begin with quantum physics; after recapping the basic ingredients of quantum information in Sec. 2.1.5, we discuss about (discrete-variable) NISQ computing in Sec. 2.2. We then review continuous-variable quantum systems in Sec. 2.3, with a particular focus on Gaussian states and operations. Introducing such a formalism allows us to discuss continuously-monitored systems in Sec. 2.4, and this constitutes the end of our tour through quantum systems. In Sec. 2.5.2 we introduce the problem of quantum state discrimination, and we then softly turn to the classical world by discussing statistical inference problems in Sec. 2.5. Finally, we introduce the reinforcement learning framework in Sec. 2.6, with a particular focus on model-free Q-learning agents.

### 2.1 Some basic notions of quantum information

A system  $\mathcal{S}$  is described by quantum theory using a  $d$ -dimensional Hilbert space  $\mathcal{H}$ ; the dimension  $d$  might be infinite, *e.g.* if we want to describe systems in the continuous-variable (CV) ones, such as optical or mechanical ones. The state of the system is represented by a self-adjoint, positive semi-definite and unit-trace operator  $\rho \in \mathcal{H}$ .

## 2.1. SOME BASIC NOTIONS OF QUANTUM INFORMATION

---

Physical quantities such as position, angular momenta or energy are described by self-adjoints operators, known as observables. As such they have an associated eigenbasis, each eigenstate being a unit-vector in  $\mathcal{H}$  and deemed *pure state*. Note however that a quantum state is generally found in a *mixed state*, e.g. a probabilistic mixture of pure states.

The value that physical quantities can take is obtained by measuring the system, and this makes quantum theory a *physical* one, since we can readily contrast measurement statistics with theoretical predictions.

Given an orthonormal basis  $\{|i\rangle\}_{i=1}^d$ , a quantum state<sup>1</sup> can be written as

$$\rho = \sum_{i,j} \rho_{ij} |i\rangle\langle j|, \quad (2.1)$$

where  $\rho_{ij}$  are the matrix components of  $\rho$ . Since  $\rho \geq 0$ , its spectral decomposition is always available, with an associated eigenbasis  $\{|k\rangle\}$

$$\rho = \sum_k p_k |k\rangle\langle k|, \quad 0 \leq p_k \leq 1, \quad \sum_k p_k = 1, \quad (2.2)$$

where the conditions on  $p_k$  arise from the fact that  $\rho$  is positive semi-definite and unit-trace operator. The purity of the state is defined as  $\text{Tr}[\rho^2]$ ; while pure-states are represented by rank-one operators (of purity 1), mixed-states' purity is strictly less than 1.

Composite systems are described in a Hilbert space formed by the tensor-product of the individual components. As an example, let us consider bipartite systems with parties  $\mathcal{S}$  and  $\mathcal{A}$ ; the composite quantum state  $\rho_{SA}$  is an operator in  $\mathcal{H}_{SA} = \mathcal{H}_S \otimes \mathcal{H}_A$ . Taking a basis for  $\mathcal{H}_{SA}$  formed by a tensor-product of basis  $|i_S\rangle$  and  $|j_A\rangle$  in  $\mathcal{S}$  and  $\mathcal{A}$  respectively, we obtain that

$$\rho_{AS} = \sum_{ijkl} \rho_{ij}^{kl} |i\rangle_S \langle j| \otimes |k\rangle_A \langle l|. \quad (2.3)$$

A state is *separable* if it can be written as a convex combination of product-like states, i.e.

$$\rho_{AS}^{sep} = \sum_k p_k \rho_S^{(k)} \otimes \rho_A^{(k)}, \quad (2.4)$$

where  $\rho_i^{(k)}$  is a *local* quantum state in subsystem  $i = S, A$ . If  $\rho_{AS}$  cannot be written in this form, then the state exhibits quantum correlations.

---

<sup>1</sup>While much of the following discussion applies to CV systems as well, we will restrict in this Section to finite  $d$ , and study CV-systems in Sec. 2.3.

The reduced state of each subsystem is defined via the partial-trace [NC00]; e.g.  $\rho_{S(A)} = \text{Tr}_{A(S)} [\rho_{AS}]$ . For instance, the local state of system  $S$  reads

$$\rho_S = \text{Tr}[\rho_{AS}] = \sum_m \langle m_A | \left( \sum_{ijkl} \rho_{ij}^{kl} |i\rangle_S \langle j| \otimes |k\rangle_A \langle l| \right) |m_A\rangle = \sum_{ijm} \rho_{ij}^{mm} |i\rangle_S \langle j|. \quad (2.5)$$

### Projective measurements

In order to extract information out of a quantum state, we need to measure it. Quantum measurements deliver a classical outcome, and provide a route to map quantum information to classical one. While generalized measurements will be discussed in Sec. 2.1.3, we will now discuss a particular type of measurement.

Projective measurements consist on orthogonal projectors  $\{E_k\}$  such that  $E_k E_j = \delta_{kj} E_j$ , with  $\sum_k E_k = \mathbb{I}$ , and  $E_k = E_k^\dagger$ . Here, the possible measurement outcomes are  $\{k\}_{k=0}^{d-1}$ , and the probability of outcome  $k$  given the state of the system is  $\rho$  reads

$$p(k) = \text{Tr}[\rho E_k]. \quad (2.6)$$

Moreover, the state of the system after this outcome occurs, known as the post-measurement state, is given by

$$\rho \rightarrow \rho|_k = \frac{E_k \rho E_k}{\text{Tr}[\rho E_k]}, \quad (2.7)$$

where we note that the normalization is given by the probability of such measurement outcome. On the contrary, if the measurement is performed but the outcome is not known, the *unconditional post-measurement state* is given by

$$\rho \rightarrow \sum_k E_k \rho E_k. \quad (2.8)$$

Here, we observe that (i) the unconditional state is the average of all possible conditional-states, as weighted by the corresponding probability  $p(k)$ , and (ii) if the original state to be measured is pure, then not recording the measurement outcome will generally convert it to a mixed state.

As an example, let us consider an observable  $O$ . We can readily construct a projective measurement via its spectral decomposition, i.e.  $\hat{O} = \sum_k o_k |k\rangle \langle k|$ , leading to  $E_k = |k\rangle \langle k|$  and

$$\langle O \rangle_\rho = \text{Tr}[\hat{O} \rho] = \sum_k p_k o_k, \quad (2.9)$$



where  $p_k = \text{Tr}[\rho |k\rangle\langle k|]$  is the probability of finding the state  $\rho$  in the  $k^{\text{th}}$  eigenstate of  $\hat{O}$ . Here, let us recall that if  $\hat{H}$  is the system's Hamiltonian, then *the variational theorem* is obtained by noting that  $p_k \geq 0 \forall k$ ; it follows that  $\langle H \rangle_\rho \geq E_0$  where  $E_0$  is the ground-state energy, *i.e.* the lowest eigenvalue of  $\hat{H}$ . The inequality is saturated in the case that  $\rho$  is the ground-state of  $\hat{H}$ ; this is the basis of variational quantum algorithms, where  $\rho$  is evolved in such a way to get as close as possible to the ground-state, as discussed Sec. 2.2.

### 2.1.1 Closed-system dynamics

The evolution of a closed quantum system is given by an unitary transformation  $\hat{U}_t$ , generated by a Hamiltonian  $\hat{H}$ , *e.g.*  $\hat{U}_t = e^{-i\hat{H}t}$ , where the Hamiltonian will be assumed time-independent. Here, the Schrodinger equation describes quantum-state evolution according to

$$\partial_t \rho(t) = -i[\hat{H}, \rho(t)]. \quad (2.10)$$

The solution to such equation can be expressed as

$$\rho(t) = \hat{U}_t \rho(0) \hat{U}_t^\dagger, \quad (2.11)$$

and this constitutes the *Schrodinger picture*. However, the time-dependence of the quantum state can be entirely translated to the observables, which gives rise to the *Heisenberg picture*. In turn, by defining

$$\hat{O}(t) = \hat{U}_t^\dagger \hat{O}(0) \hat{U}_t, \quad (2.12)$$

we observe that the expected value of  $O$  reads

$$\langle O \rangle(t) = \text{Tr}[O(0)\rho(t)] = \text{Tr}[O(t)\rho(0)]. \quad (2.13)$$

The corresponding equation of motion for the observable reads

$$\partial_t O(t) = i[H, O(t)]. \quad (2.14)$$

Moreover, the *interaction picture* —or interaction frame— is obtained by decomposing the Hamiltonian in terms of the *free* Hamiltonian  $H_0$  and an interaction term  $V$  according to  $H = H_0 + V$ . In this picture, both states and observables depend on time, according to

$$\rho_{IF}(t) = e^{-iH_0 t} \rho(0) e^{iH_0 t}, \quad (2.15)$$

$$\hat{O}_{IF}(t) = e^{-iH_0 t} \hat{O}(0) e^{iH_0 t}, \quad (2.16)$$

and the evolution of  $\rho_{IF}(t)$  reads

$$\partial_t \rho_{IF}(t) = -i[V, \rho_{IF}(t)], \quad (2.17)$$

where a similar evolution applies for observables as well.

Changes in quantum-states can be generalized to consider the case in which the quantum system is interacting with an environment, *i.e.* an open-quantum system. While our discussion of the continuous-time case will be postponed until Sec. 2.1.5, we will now introduce a tool to represent discrete changes, such as those modeling the noise that a quantum-receiver suffers, as studied in Sec. 3.3.5, or the noise injected into a quantum state by an imperfect NISQ-component, as we study in Sec. 4.2.6.

### 2.1.2 Quantum channels

A quantum channel  $\Phi$  is a lineal map  $\Phi : L(\mathcal{H}_1) \rightarrow L(\mathcal{H}_2)$ , where  $L(\mathcal{H})$  denotes the set of operators in  $\mathcal{H}$  and the dimensions of the input-space  $\mathcal{H}_1$  and output-space  $\mathcal{H}_2$  might differ.

The lineal map is required to be:

- Completely positive (CP): let  $\mathcal{H}_{1'}$  be of dimension  $d'$ , then

$$(\mathbb{I}_{d'} \otimes \phi)[\sigma] \geq 0 \quad \forall d', \quad \forall \sigma \in \mathcal{H}_{1'} \otimes \mathcal{H}_1. \quad (2.18)$$

- Trace-preserving (TP).

The first condition is required such that when  $\Phi$  acts locally (*i.e.* on one out of many subsystems), the resulting global state is a positive state. The second condition is required such that the resulting state is unit-traced.

Choi-Krauss theorem [Man19] guarantees that any CPTP channel admits a *Krauss decomposition*:

$$\mathcal{M}(\rho) = \sum_{k=1}^N M_k \rho M_k^\dagger, \quad (2.19)$$

with

$$\sum_{k=1}^N M_k^\dagger M_k = \mathbb{I}_d. \quad (2.20)$$

The operators  $\{M_k\}_{k=1}^N$  are known as *Krauss operators*, and it is important to note that this decomposition is not unique.

The non-uniqueness of Krauss decomposition is explicated by the *Stinespring dillation*. This states that any CTPT channel can be obtained by unitarily evolving the input-system with an ancilla and tracing the latter out. To see this, let the initial state of the system and ancilla be  $\rho$  and  $|0_A\rangle\langle 0_A|$  respectively, and the closed (unitary) evolution of the joint system given by  $\hat{U} = \hat{U}_{SA}$ . The resulting state of the system is

$$\rho \rightarrow \text{Tr}_A \left[ \hat{U} \rho \otimes |0_A\rangle\langle 0_A| \hat{U}^\dagger \right] \quad (2.21)$$

$$= \sum_{k=1}^N \langle k_A | \left( \hat{U} |0_A\rangle \rho \langle 0_A| \hat{U}^\dagger \right) |k_A\rangle \quad (2.22)$$

$$= \sum_{k=1}^N M_k \rho M_k^\dagger, \quad (2.23)$$

where we defined  $M_k = \langle k_A | \hat{U} |0_A\rangle$  by considering a basis  $\{|k_A\rangle\}_{k=1}^N$  of the ancilla. However, we can readily note that a different basis  $\{|\nu_A\rangle\}_{k=1}^N$  (of a possibly enlarged) ancilla could be considered, where the basis-elements are related to each other by an isometry  $\hat{V}$  such that  $|\nu\rangle = \sum_{l=1}^N V_{\nu l} |l\rangle$ . Using the latter basis, the two sets of Krauss operators relate to each other as  $N_\nu = \sum_{l=1}^N V_{\nu l}^\dagger M_l$ .

Moreover, note that the map  $\Phi$  will be trace-preserving iff  $\sum_{k=1}^N M_k^\dagger M_k = \mathbb{I}$ . Also, we remark that as long as Eq. 2.20 is satisfied, the Krauss operators can also be time-dependent [Man19].

As an example, we can consider an unitary channel as in Eq. 2.11, where the Krauss representation reduces to  $M_k = \hat{U}_t$ , with  $N = 1$ .

### 2.1.3 Generalized quantum measurements

A generalized quantum measurement given by a set of lineal operators  $\{E_k\}_{k=1}^n$ , each associated to a possible outcome  $k \in \{1, \dots, n\}$ . The probability that outcome  $k$  happens, conditioned on the quantum state being  $\rho$ , is given by the Born rule:

$$p(k|\rho) = \text{Tr}[\rho E_k]. \quad (2.24)$$

Since such probability is required to be positive for any quantum state, it follows that  $E_k \geq 0$  (which imply self-adjointness). Moreover, since  $\sum_k p(k|\rho) = 1$  for any quantum state, then  $\sum_k E_k = \mathbb{1}_d$ . This constitute a *Positive Operator-Valued Measure (POVM)*, which we will often denote by  $\mathcal{M}$ .

POVMs arise when a projective measurement is carried out on an ancilla that the system is coupled with. Similarly to our previous discussion on quantum channels<sup>2</sup>, we can readily define an ancilla system  $A$ , with an associated eigen-basis  $\{|k_A\rangle\}_{k=1}^n$ . Now, considering the initial-state of the system and ancilla to be  $\rho_S$  and  $|0_A\rangle\langle 0_A|$  respectively, and the joint unitary evolution to be  $\hat{U}$ , a projective measurement  $\{|k_A\rangle\langle k_A|\}$  carried out on sub-system  $A$  will occur with probability

$$p(k) = \text{Tr}[(\mathbb{1}_d \otimes |k_A\rangle\langle k_A|)\rho_S \otimes |0_A\rangle\langle 0_A|] = \text{Tr}[\rho M_k^\dagger M_k] = \text{Tr}[\rho E_k], \quad (2.25)$$

where similarly to the previous Section we have  $M_k = \langle k_A|\hat{U}|0\rangle$ . Thus, under this Krauss decomposition we have that  $E_k = M_k^\dagger M_k$ , and the conditional state of  $\rho$  upon obtaining measurement outcome  $k$  reads

$$\rho \rightarrow \rho_{|k} = \frac{M_k \rho M_k^\dagger}{\text{Tr}[\rho M_k^\dagger M_k]}. \quad (2.26)$$

Similarly, the unconditional post-measurement state under such Krauss decomposition reads

$$\rho \rightarrow \sum_k^n M_k \rho M_k^\dagger. \quad (2.27)$$

It is important to recall that this decomposition is not unique, since different Krauss decompositions are equivalent by an isometry, as discussed in the previous Section. In turn, Eq. 2.27 can be understood as a quantum channel acting on  $\rho$ , similar to Eq. 2.20. In this direction, it is useful to understand different Krauss decompositions as different *unravelings* of the open-quantum system dynamics, all averaging up to the same unconditional evolution.

The notion of quantum measurements and channels can be generalize to take into account stochastic channels, using the concept of a quantum instrument [Wat18], or quantum operation [Pre98], which we now turn to introduce.

### 2.1.4 Quantum instruments

A quantum instrument is described by a collection of CP maps  $\{\phi_k\}_{k=1}^n$ , which sum up to a channel, *i.e.*  $\Phi = \sum_k \phi_k$ . Such object is to be considered as a generalization of a POVM: after applying a quantum instrument on a state  $\rho$ , two things happen:

- A measurement outcome  $k$  is obtained with probability  $\text{Tr}[\phi_k(\rho)]$ ,

---

<sup>2</sup>In fact, measurements can be regarded as quantum-to-classical channels [Wat18]

- The conditional state of  $\rho$  upon observing measurement outcome  $k$  is

$$\rho \rightarrow \frac{\phi_k(\rho)}{\text{Tr}[\phi_k(\rho)]}. \quad (2.28)$$

We can readily consider a Krauss decomposition of  $\phi_k$ , given by  $\{M_{k\mu}\}$ . Such Krauss operators should obey the completeness relation derived above, namely  $\sum_{k\mu} M_{k\mu}^\dagger M_{k\mu} = \mathbb{1}$ . Here, we can interpret the quantum instrument as an object describing an stochastic application of  $\phi_k$ : whereas  $k$  is recorded, the label  $\mu$  is ignored. As such, the post-measurement state reads

$$\phi_k(\rho) = \sum_{\mu} M_{k\mu} \rho M_{k\mu}^\dagger, \quad (2.29)$$

with outcome  $k$  occurring with probability  $p(k) = \text{Tr}[\phi_k(\rho)]$ . We note that since outcome  $\mu$  is ignored, the channel  $\phi_k$  is generally trace-decreasing, and the normalized post-measurement state coincides with the post-measurement state defined above as  $\rho|_k = \frac{\phi_k(\rho)}{\text{Tr}[\phi_k(\rho)]}$ .

Moreover, we might think of a process that consists on  $M$  consecutive measurements, leading to an stochastic sequence  $\mathbf{k} = \{k_1, \dots, k_j, \dots, k_M\}$ . The conditional state of the system upon observing outcome  $k_j$  is obtained via  $\phi_{k_j}$ . Note that when keeping the system's state unnormalized up to the  $M$ -outcome, the trace of the unnormalized state, *i.e.*  $\phi_{k_M} \circ \dots \circ \phi_{k_1}(\rho)$ , represents the likelihood of observing such sequence of outcomes, and the normalized state reads:

$$\rho|_{\mathbf{k}} = \frac{\phi_M \circ \dots \circ \phi_1(\rho)}{\text{Tr}[\phi_{k_M} \circ \dots \circ \phi_{k_1}(\rho)]}. \quad (2.30)$$

Our discussion has so-far limited to discrete changes of the quantum state. In the following we will discuss how a dynamical equation is obtained for the (unconditional) evolution of an open-quantum system.

### 2.1.5 Quantum master equation

While we have previously introduced the notion of a quantum channel given by a CPTP map between quantum states, no notion of continuous-time has entered so-far in the discussion. Thus, we will here consider a quantum channel  $\mathcal{E}_t$ , and discuss a dynamical equation accounting for  $\rho(t) = \mathcal{E}_t[\rho(0)]$ . In the Markov approximation, this can be expressed as a differential equation of the form

$$d\rho = \mathcal{L}[\rho]dt, \quad (2.31)$$

where  $\mathcal{L}$  is a superoperator that generates the *quantum dynamical semi-group*<sup>3</sup> associated to  $\mathcal{E}_t$  and can be shown to admit the so-called *Libland form* [Lin76; WM09; Man19]:

$$\mathcal{L}(\rho) = -i[\hat{H}, \rho] + \sum_{k=1}^M \mathcal{D}[\hat{L}_k]\rho, \quad \mathcal{D}[\hat{L}_k]\rho = \hat{L}_k\rho + \rho\hat{L}_k^\dagger - \frac{1}{2}\{\hat{L}_k\hat{L}_k^\dagger, \rho\}, \quad (2.32)$$

where  $\hat{H}$  is the system Hamiltonian (generator of the closed-system dynamics) and the linearly-independent operators  $\{\hat{L}_k\}$  are known as the *Libland operators*, or *jump operators*. We note that the Libland form is invariant under (i) unitary transformations of the jump operators, i.e.  $\hat{L}_k \rightarrow \sum_l \hat{V}_{kl}\hat{L}_l$  with  $\hat{V}$  unitary (which is a consequence of the unitary invariance of the Krauss decomposition), and (ii) shifts of the Libland operators accompanied by a new term in the Hamiltonian:

$$\hat{L}_k \rightarrow \hat{L}_k + \xi_k, \quad \hat{H} \rightarrow \hat{H} - \frac{i}{2} \sum_{k=1}^M \xi_k^* \hat{L}_k - \xi_k \hat{L}_k^\dagger. \quad (2.33)$$

The dynamical equation introduced above is known as the *quantum master equation*, and its generator is given by Eq. 2.32. We remark that this evolution can also be obtained by suitable time-dependent Krauss operators:

$$\rho(t + dt) = \mathcal{E}_{dt}[\rho(t)] = \rho + \sum_{k=1}^M K_k(dt)\rho K_k^\dagger(dt) \quad (2.34)$$

$$= \rho + \mathcal{L}[\rho]dt. \quad (2.35)$$

As discussed in Sec. 2.1.3, the choice of Krauss and jump operators  $K_k(dt)$  in Eq. 2.34 plays an important role when studying the *conditional* dynamics, arising as a consequence of recording the measurement outcome at each time-step. This can be understood as a stochastic sequence of channels, as introduced in Sec. 2.1.4, which give rise to the notion of a *quantum-trajectory*, as discussed in Sec. 2.4.

To conclude, in this Section we have revisited some of the basics in quantum information theory. Many of the concepts discussed here will be recaped later in

---

<sup>3</sup>A quantum dynamical semi-group consists on a family of CPTP maps  $\{\mathcal{E}_t : t \geq 0\}$  such that (i)  $\mathcal{E}_t\mathcal{E}_s = \mathcal{E}_{t+s}$  and (ii)  $\text{Tr}[\mathcal{E}_t[\rho]\hat{O}]$  is continuous over  $t$ , for any state-observable pair  $(\rho, \hat{O})$ . Note that the first restriction is quite restrictive, since an arbitrary evolution might not admit such decomposition.

the thesis. We will now turn to introduce some of the physics and concepts that set the basis for Chapter 4.

## 2.2 NISQ & the qubits

In this Section we will focus on discrete-variable systems. Special attention will be paid to Noisy Intermediate-Scale Quantum (NISQ) computing, a concept standing for currently available technologies, where quantum circuits of about 50 – 100 qubits are already being implemented and controlled despite the strong presence of noise [Pre18].

While error-correcting codes cannot be built with such a low amount of qubits –it is estimated that nearly a thousand *physical qubits* are required in order to get a single *logical qubit* [FMM+12]– special-purpose quantum devices can readily be built, exhibiting a genuinely quantum behaviour. Whether such technologies will be able to provide a game-changer quantum advantage in industry applications, or they are just a stepping stone towards fault-tolerant architectures is a matter of vivid debate between the community [AAB+19; PCZ22; LLZ+22]. However, we remark that the sole feasibility of experimental preparation and control of quantum systems already constitute a game-changer paradigm in our scientific community. In turn, not only a plethora of new questions regarding scalability and overall implementation of such technologies arise [ABB+21], but also the demand for new theoretical tools, particularly to better analyze performance issues that we will later discuss [MBS+18; CSV+21; PCW+21a; HAY+21; ZG21; TWN+21; SCC+22; PNG+21; MKW21; CC21; ACC+21; BK21].

In the following we will concentrate on an specific topic that falls under the umbrella of NISQ computing, called Variational Quantum Algorithms.

### 2.2.1 Variational Quantum algorithms

Recently, a superposition of hype and hope has been put on the field of quantum machine learning [BWP+17; AAB+19; SK22; Aar15; Pre18; PMS+14; PCW+21b; SCH+22], particularly when *training* Parametrized Quantum Circuits (PQCs).

PQCs consists on a series of quantum gates, *e.g.* CNOTs and rotations, where the real parameters of the latter are modified in order to *optimize* the circuit. Here, and in analogy to the classical neural networks scenario [GBC16], we aim to minimize a given cost function (and thereby, these circuits are also known as *quantum neural networks*). The overall training of PQCs is known as a Variational Quantum Algorithm (VQA), and consists on an hybrid classical-quantum scheme.

In turn, a quantum device is only used to estimate cost-function values, whereas a classical algorithm is in charge of controlling that quantum device by optimizing over accessible degrees of freedom, *e.g.* rotation parameter values. Recently, many efforts have been carried out in order to analyze the performance, trainability and overall feasibility of VQAs [CAB+21a; BCLK+22].

Training a VQA means to solve an optimization problem encoded into a cost function of the form

$$C(\mathbf{k}, \boldsymbol{\theta}) = \sum_i^M f_i \left( \text{Tr}[\hat{O}_i U(\mathbf{k}, \boldsymbol{\theta}) \rho_i U^\dagger(\mathbf{k}, \boldsymbol{\theta})] \right). \quad (2.36)$$

Here,  $\{\rho_i\}$  are  $n$ -qubit states forming a *training set* of cardinality  $M$ , and  $U(\mathbf{k}, \boldsymbol{\theta})$  is our NISQ circuit parametrized by continuous parameters  $\boldsymbol{\theta}$  (*e.g.*, rotation angles) and by discrete parameters  $\mathbf{k}$  (*e.g.*, gate placements). Moreover,  $\hat{O}_i$  are observables and  $f_i$  are functions that encode the optimization task at hand.

For example, when employing the Variational Quantum Eigensolver (VQE) algorithm [PMS+14] we have  $f_i(x) = x$  and the cost function reduces to

$$C_{\text{VQE}}(\mathbf{k}, \boldsymbol{\theta}) = \text{Tr}[\hat{H}U(\mathbf{k}, \boldsymbol{\theta})\rho U^\dagger(\mathbf{k}, \boldsymbol{\theta})], \quad (2.37)$$

where  $\rho$  is the input state (and the only state in the training set) and  $H$  is the Hamiltonian whose ground state we seek to prepare. Alternatively, in a classification problem where the training set is of the form  $\{\rho_i, y_i\}$ , with  $y_i \in \{1, \dots, n\}$  being the true label, the choice  $f_i(x) = (x - y_i)^2$  leads to the least square error cost [LLD22; FN18; SBS+20; BBF+20; CCL19a]. In this regard, and as it happens in the reinforcement-learning scenario discussed in Sec. 2.6, the possibility of encoding a given problem into a cost function is a matter of creativity.

Given the cost function, a NISQ circuit is employed to estimate each term in Eq. (2.36). Usually, the available measurements consist on local projections over the eigenbasis of  $\sigma_z^i$  for each qubit  $i = 1, \dots, n$ . Here, in order to estimate the cost function, we decompose each operator appearing in the cost-function in terms of *Pauli-strings*. Recall that Pauli-string operators form a basis in the space of  $n$ -dimensional hermitian operators and hence

$$\hat{O}_i = \sum_{k=1}^{4^n} c_k P_k, \quad (2.38)$$

with  $P_k = \bigotimes_{i=1}^n \sigma_j$ , where  $\sigma_j \in \{\mathbb{I}, \sigma_x, \sigma_y, \sigma_z\}$ . Note that Pauli-string operators are orthogonal to each other, *i.e.*  $\text{Tr}[P_k P_{k'}] = 2^n \delta_{kk'}$  and thus  $c_k = \text{Tr}[\hat{O}_i P_k] / 2^n$ .



Thus, the cost-function can be estimated by computing averages of some pre-determined observables, which are in turn a weighted sum of expected values of Pauli-strings. Moreover, we can estimate each expected value  $\langle P_k \rangle$  by performing the corresponding single-qubit transformation beforehand and measuring  $\sigma_z$ <sup>4</sup>. In this regard, if an  $\epsilon$ -accuracy is required in order to estimate this expected value, an order of  $\sim \frac{1}{\epsilon^2}$  measurements needs to be done (this will have consequences when trying to resolve between very small values of the cost-function in Sec. 2.2.4).

We also note that the Pauli-string procedure outlined above consists on an exponentially large sum of expected values. However, we are often interested in local hamiltonians, where many of these terms vanish (for example, TFIM and XXZ models considered in Ch. 4). In this case, techniques such as randomized measurements (RMs) have recently been developed, in which the cost function can be computed more efficiently. While we do not consider such approaches in this thesis, we refer the interested reader to Ref. [EFH+22], and we note that a combination between our methods (discussed in Chapter 4) and classical shadows (a sub-class of RMs) is discussed in Sec. 4.3.

Once the cost-function has been estimated, the power of classical optimization algorithms is leveraged to solve the optimization task

$$\operatorname{argmin}_{\mathbf{k}, \boldsymbol{\theta}} C(\mathbf{k}, \boldsymbol{\theta}). \quad (2.39)$$

The estimation of a cost function via a quantum circuit and its modification (hopefully towards a cost-minimizing direction) by a classical optimization algorithm defines a cycle, which is repeated until convergence, as depicted in Fig. 2.1. The success of such scheme hinges on several factors. As a matter of fact, the classical optimizer must be able to efficiently train the parameters, and in the past few years, there has been a tremendous effort in developing quantum-aware optimizers [VPB18; KAC+20; ACS+20; SIK+20; KB19; NFT20; FCA+20]. In particular, severe trainability issues has been pointed out, which forbid the success of VQAs and as constitute the main barrer (along with hardware-noise) for making use of current NISQ devices in the VQA framework, and we will review some of them in Sec. 2.2.2.

However, we note that even if such trainability issues could be overcome, the performance of a VQA will intrinsically be linked to the specific quantum circuit used for the training. Here, the choice of an ansatz for  $U(\mathbf{k}, \boldsymbol{\theta})$  plays a crucial role in determining the success of a VQA scheme; for instance, for a given Hamiltonian

---

<sup>4</sup>This can be understood as a change-of-basis.

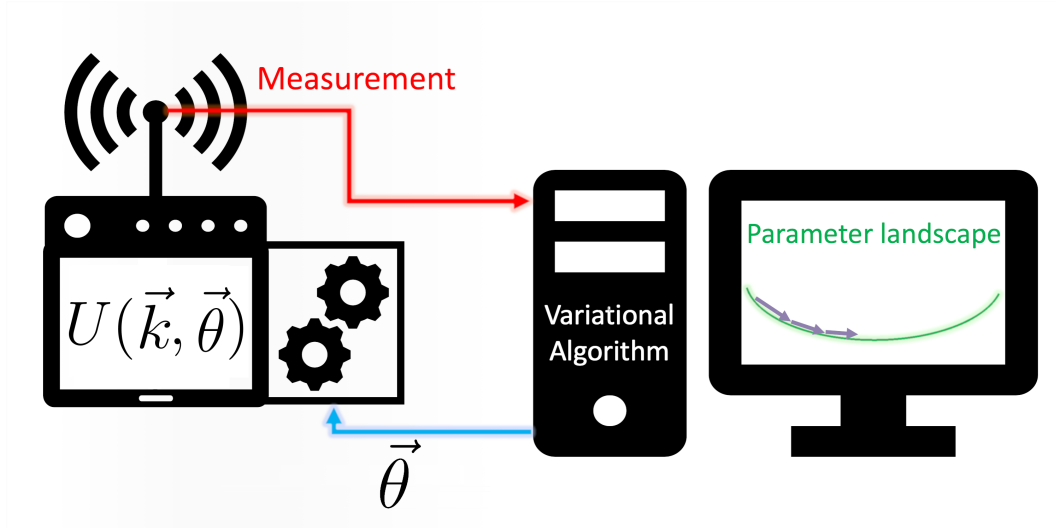


Figure 2.1: Figure adapted from Ref. [YEZ+19]: a generic VQA algorithm is sketched. Here, a quantum device is used to estimate a cost-function value, whereas the device itself is controlled by a classical optimization algorithm. This cycle is iterated until convergence.

in the VQE algorithm, circuits that are close to the ground-state preparing one might lie outside the subspace  $\mathcal{U}$  of the unitary group  $U(n)$  which is generated by varying  $\theta$  in  $U(\mathbf{k}, \theta)$ , with a fixed circuit layout given by  $\mathbf{k}$  [HSC+22].

In this respect, not every unitary transformation in  $U(n)$  can straightforwardly be implemented, since the availability of quantum gates is often restricted; we generally need to deal with a *gate-alphabet*  $\mathcal{D}$ , which in this thesis will be composed of one-qubit rotations and of CNOTs entangling any pair of qubits present in the circuit. We note that such an alphabet is universal [NC00]: any quantum gate can be implemented provided enough gates of the alphabet are used. Nevertheless, such a *compilation* is a tricky one: noise accumulates with circuit's depth, and hence only a non-trivial set of unitaries can be reached. Moreover, we remark that the availability of fully-connected qubits is a slight simplification, and connectivity constraints should in principle be considered. In the latter case, entangling two qubits which are not straightforwardly connected represents an overhead in circuit's depth, and many efforts have been carried out recently in order to find novel strategies to tackle this issue [Gri19].

For these reasons, it is important to discuss different strategies usually considered when constructing the quantum-circuit to be used in a VQA, also known as an *ansatz*.

### 2.2.2 The Ansatz

We refer to the *ansatz* as the quantum-circuit structure, *i.e.* the layout defining quantum-gates placements. While such a word stand for *an educated guess or an additional assumption made to help solve a problem* [Wik], let us remark that trainability and noise-related issues make it non-trivial to define what an *educated guess* for a given problem is in the NISQ-era.

We will make a distinction between *fixed*-structured ansatzes (*i.e.* where  $\mathbf{k}$  is fixed), and *variable*-structure ones, in which also the structure of the circuit is optimized. In the latter case, the optimization is arguably more complex, since it consists on finding the right circuit structure on top of the continuous parameter optimization.

A key parameter that characterizes the ansatz is the number of gates present in the circuit. It is thus essential to construct ansatzes that maintain this number as low as possible to mitigate noise and trainability-related issues, but also that have enough expressibility to contain the problem solution (or at least an approximate version of it).

In the following we will first discuss some commonly-used fixed-structured ansatzes, and we will then turn to the variable-structure case.

Let us now discuss on some commonly-used fixed-structure ansatzes. We first consider a *separable* ansatz as depicted in Fig. 2.2. Here, the structure is so trivial that no entanglement is generated by the circuit, since only local operations are performed on each qubit. Assuming that the initial state is separable —we often consider it to be  $|0\rangle^{\otimes n}$ —, only a very small fraction of the Hilbert space can be reached under this circuit’s choice [ZHS+98], and the overall performance is expected to be poor since the presence of entanglement is generally required. Also, note that the separable case can efficiently be simulated in a classical computer, and thus we do would not expect them to provide any quantum advantage.

This issue can be addressed with the layered Hardware Efficient Ansatz [KMT+17], which we refer as HEA. Here, the gates are arranged in a brick-like fashion and act on alternating pairs of qubits, as depicted in Fig. 2.2. We define a  $L$ -HEA as a circuit in which  $L$  layers are stacked next to each other: a single layer consists on  $n - 1$  two-qubit gates (in the figure, rotations around  $x$  and  $z$  axis, followed by a CNOT) that correlate two neighbour qubits in the circuit. We here will consider the qubits as a cyclic chain, where the bottom one is also conected to the upper one, although this definition can be adapted to specific qubit-connection constraints of the quantum computer at hand. As  $L$  increases, it can be expected that HEA

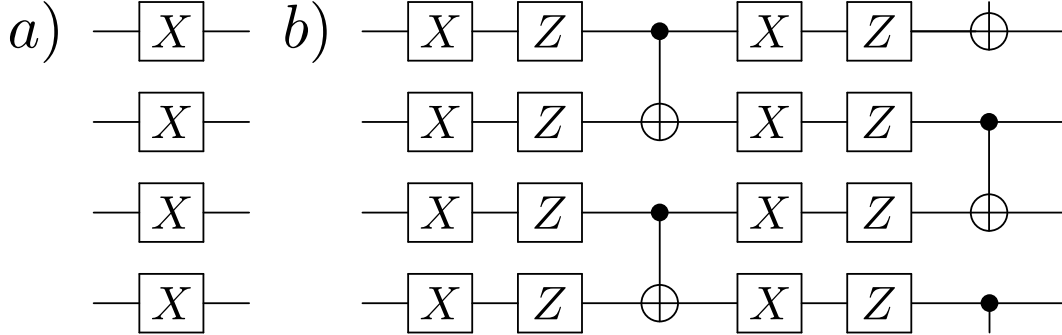


Figure 2.2: We show examples of commonly-used quantum circuits. In (a) we depict a separable product ansatz which generates no entanglement between the qubits. On the other hand, (b) shows two layers of a shallow alternating Hardware Efficient Ansatz where neighboring qubits are initially entangled. Here  $Z$  ( $X$ ) indicates a parametrized rotation about the  $z$  ( $x$ ) axis (the angles  $\theta$  are not explicitly shown, but such are the degree of freedom to optimize over).

becomes more expressible. as opposed to the separable circuit, whose expressibility is very poor.

In turn, HEA turns out to be *too* expressible [HSC+22], in the sense that any unitary can be mimicked using enough HEA-layers, and this can lead to trainability issues. For a sufficiently-expressible PQC, the landscape of possible unitaries  $\mathcal{U}$  that can be reached by varying its parameters turns to be hard to navigate, in the sense that the cost-function landscape becomes very flat, a phenomenon known as a *barren plateau*. The latter fact constitute a big challenge for an optimization procedure that needs to navigate such landscape and reach the lowest cost-function value, as discussed in Sec. 2.2.4.

One of the main advantages of HEA is that it employs gates native to the specific device used, hence avoiding an unnecessary overhead in the number of gates present in the circuit, arising from compiling non-native unitaries into native gates (for instance, the building-blocks in the figure could be replaced with gates coming from another dictionary). This type of ansatz is *problem-agnostic*, in the sense that it is expressible enough so that it can be generically employed for any task; in Chapter 4 we will often use this ansatz for benchmarking purposes.

On the other hand, a hint on the solution for the target problem to be solved by a quantum computer can, in principle, be helpful when building the ansatz. This claim should certainly holds in the ideal (*e.g.* noise-less) scenario, and in VQAs, this is reflected by the so-called *problem-inspired* ansatzes. Here the goal is to encode information of the problem into the ansatz’s structure so the optimal solution of

Eq. (2.39) exists within the parameter space without requiring high expressibility.

An example of such fixed-structure ansatzes is the Quantum Alternating Operator Ansatz (QAOA) for optimization problems, which is aimed to adiabatically prepare the ground-state of a *problem*-hamiltonian, and thus alternates between a *mixing* unitary and a *problem* one. We will not discuss these ansatzes here, and the interested reader can find more details in Refs. [FGG14; HWO+19].

On a different note, considerable effort has been made in order to construct physically-inspired ansatzes in the field of quantum chemistry, and the Unitary Coupled Cluster (UCC) Ansatz [CRO+19; BM07] is an example of them. We will discuss the quantum-chemistry problem in greater detail in Sec. 4.2.3. Essentially, we here seek to prepare the ground state of a molecular Hamiltonian, which (under the Born-Oppenheimer approximation) translates to solve a many-body fermionic system, known as the *electronic-structure problem*. While much insight from the classical methods that tackle this problem can be brought to the NISQ scenario, some issues raise up. Namely, we need to map the [UCC] ansatz from fermions to qubits, as well as the molecular Hamiltonian, since we are here targeting applications on digital quantum computers. Such mapping can be done by the Jordan-Wigner transform (or alternatively Bravi-Kitaev). Unfortunately, it usually comes with an overhead in the number of quantum gates required to build the specific circuit (and also a large number of Pauli observables to represent the molecular Hamiltonian). This fact constitutes a big challenge for NISQ applications, since deep circuits are severely affected by noise.

We might think of quantum chemistry as a subset of problems that quantum computing can potentially address. Generally, for a given physical problem, a family of quantum-circuit structures reflecting certain symmetries of potential solutions could in principle be constructed. However, quantum-hardware noise makes it highly non-trivial to directly translate such physical intuition into the structure of the quantum circuit.

Under the presence of limited resources, each component of the VQA that could in principle be optimized over, should be optimized over. This constitutes the spirit of several efforts [GEB+19; TSB+21; ZKK+21; RHP+19; CSU+20; CRS+21; CSS+18; DHY+20; ZHZ+20] to adapt the circuit layout to the specific scenario at hand (e.g. a noise model, or the restricted availability of quantum hardware). In the following we will review some of such efforts, which are known as *variable-structure ansatzes*. We remark that this is not the end of the story regarding optimization of VQA's components. As a matter of fact, we could also optimize the way cost-function

estimates are obtained [GPRS+21; Alg], though we will only focus to optimizing circuit's structure.

The overall strategy of variable-structure ansatzes consists of iteratively changing the quantum circuit by placing (or removing) gates that empirically lower the cost-function value after the continuous-parameter optimization. The first proposal for variable ansatzes for quantum chemistry was introduced in [GEB+19] under the name of ADAPT-VQE. Here, the authors follow a circuit structure similar to that used in the UCC ansatz introduced above, and propose to iteratively grow the circuit by appending gates that implement fermionic operators chosen from a pool of single and double excitation operators. At each iteration, one decides which operator in the pool is to be appended, which can lead to a considerable overhead if the number of operators in the pool is large. Similarly to the UCC case, the mapping from fermions to qubits can lead to prohibitively deep circuits. This issue can be overcome using the qubit-ADAPT-VQE [TSB+21] algorithm, where the pool of operators is modified in such a way that only easily implementable gates are considered. However, the size of the pool still grows with the number of qubits. We refer the reader to [CWM+20] for a detailed comparison between ADAPT-VQE and UCC ansatzes.

A different approach to variable-structure ansatzes that has gained considerable attention are machine-learning-aided evolutionary algorithms (EA) that upgrade individuals (quantum circuits) from a population. Noticeably, the presence of quantum correlations makes it so that it is not straightforward to combine features between circuits during the evolution, as simply merging two promising circuits does not necessarily lead to low cost-function values. Thus, only random mutations have been considered so far. An example of this method is found in the Evolutionary VQE (EVQE) [RHP+19], where one explores the Hilbert space *smoothly* by growing the circuit with identity-initialized blocks of gates and randomly removing sequences of gates. Another example of an evolutionary algorithm is the Multi-objective Genetic VQE (MoG-VQE) [CSU+20], where one uses building blocks that are randomly placed along the circuit, and simultaneously optimizes both the energy and number of entangling gates. Evolutionary algorithms constitute a promising approach to ansatzes design, they nevertheless come at the cost of high quantum-computational resources to evolve populations of quantum circuits.

In addition, in Refs. [DHY+20; ZHZ+20; PT20; DHY+20] tools from auto-machine were employed in order to learn how to build ansatzes. The overall idea behind these methods is that of employing neural network (supernet) that suggests the quantum circuit structure; supernet-training can be done using policy-gradient methods, a variant of the reinforcement-learning algorithms that we consider

in Sec. 2.6. While the idea of artificial-intelligence improving (quantum) neural-network architectures is exciting, it is extremely resource-consuming: recall that the amount of data required to train a neural network is very often prohibitively high. Similarly to the EA case, the high amount of resources consumed by this method is a challenge that needs to be addressed.

Finally, in Refs. [CSS+18; CRS+21] a methodology, formalized by the VAns algorithm [BCV+21] – *i.e.* our main contribution to the field, and explained and showcased in great detail in Chapter 4 – was used to obtain a short-depth version of a given unitary under specific quantum hardware constraints (such as connectivity, noise-model as represented by quantum channels or available gates).

Having reviewed some commonly-used quantum circuit structures, we will now turn to discuss the continuous optimization.

### 2.2.3 The optimization procedure

With a quantum circuit parametrized by  $U(\mathbf{k}, \boldsymbol{\theta})$ , VQAs employ a classical optimization algorithm in order to minimize the cost-function in Eq. (2.36). Here, we will focus on derivative-based methods, which at each cycle  $\ell$  of the algorithm, the gradient  $\nabla_{\boldsymbol{\theta}} C(\mathbf{k}, \boldsymbol{\theta})$  is computed, and a cost-minimizing direction is followed. Usually, the initial parameters  $\boldsymbol{\theta}^\ell$  are randomly chosen (this random initialization should be understood according to the Haar measure, as we discuss in the next Section). In the following we will first discuss about optimization algorithms, and then on how such gradients can be computed.

The most straightforward approach in gradient-based optimization is that of following the gradient; as such we are guaranteed to attain (at least) a local minima. Thus, in Gradient Descent algorithm, the continuous parameters are updated according to

$$\boldsymbol{\theta}^{(\ell+1)} = \boldsymbol{\theta}^{(\ell)} - \alpha \nabla_{\boldsymbol{\theta}} C(\mathbf{k}, \boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(\ell)}}, \quad (2.40)$$

where  $\alpha$  is the *learning-rate*, accounting for the update-step size, and this is guaranteed to reach, at least, a local-minima.

In optimization problems where the dataset is too large, computing the gradients can be prohibitively expensive (for instance, if the entire dataset does not fit in the memory). In this case, the training set is splitted into chunks, and an stochasticity arises when estimating gradient; such is the case of Stochastic Gradient Descent (SGD), where the gradient is estimated out of  $b$ -dimensional data subsets known as *batches*. At each cycle, the data is randomly splitted and the update rule of Eq. 2.40 is sequentially applied using the gradient  $\nabla_{\boldsymbol{\theta}} C(\mathbf{k}, \boldsymbol{\theta})$  computed using each batch (note that this injects stochasticity); an *epoch* is consequently defined as an entire

pass of the original training set. We note that this stochasticity, which might in principle appear undesired, it can actually be helpful since taking random directions often help in escaping local minima.

While a large learning-rate value in Eq. 2.40 can potentially forbid the optimizer to reach the actual minima (since the update becomes simply *too large*), we note that small learning-rate values will potentially dampen the convergence rate. More elaborate optimization algorithms can be considered, which adapt the magnitude of the learning-rate according to the gradient landscape at hand. Among them, a very popular optimizer is the Adaptive Moment Estimation (Adam) algorithm [KB15], which sequentially adapts the learning rate for each parameter, by specifically tailoring it out from moving averages of first and second moments for each gradient component. In practice, this allows to reach a better optimization performance, although we shall ultimately adapt the optimization algorithm to the specific problem at hand. In the context of VQA, many issues need to be addressed, such as shot-noise, sampling-complexity and optimal choice of learning-rate values to take into account the quantum nature of the optimization landscape; in this regard there has been a tremendous effort in developing quantum-aware optimizers [VPB18; KAC+20; ACS+20; SIK+20; KB19; NFT20; FCA+20; GLD+21].

With a basic understanding on how optimization algorithms proceed, let us now discuss how the essential ingredient in the update-rule can be obtained under the VQA framework. For this, we distinguish between two approaches: (i) classical simulation of quantum circuits, and (ii) experimental implementations on quantum hardware.

### Classical simulation of quantum circuits & Automatic differentiation

If no particular symmetries are imposed, we are able to simulate quantum systems classically for up to  $\sim 30$  qubits<sup>5</sup>. State-of-the-art techniques used to classically compute VQA-cost-function gradients rely on automatic differentiation (AD), in which cost-function derivatives are obtained by tracking each intermediate-operation derivative, and following the chain-rule [BVM+20; LLZ+20]. This is done by decomposing a generic program in terms of elementary operations whose derivatives can be computed, and is the spirit behind the paradigm of *differentiable*

---

<sup>5</sup>On the contrary, many systems can be approximated efficiently by using clever ansatzes; this lies at the heart of Tensor Network (TN) methods such as Matrix Product States (MPS), Density Matrix Renormalization Group (DMRG) or Multiscale Entanglement Renormalization Ansatz (MERA). There is currently much excitement about TNs: while the space of all possible quantum states is large, only a portion of it is *physically relevant* and it is believed that such portion can be simulated efficiently using TNs [BB17; Orú19].



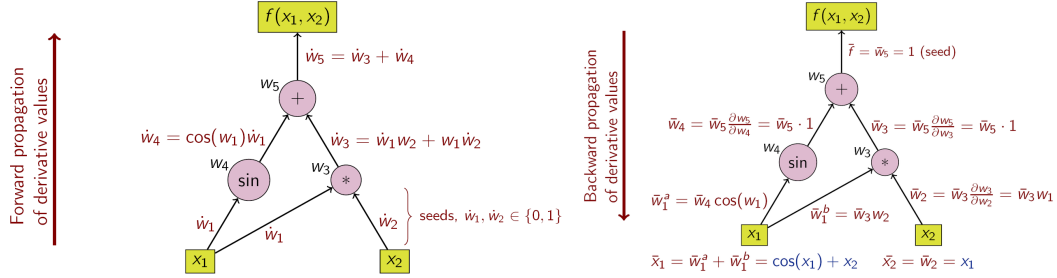


Figure 2.3: We show how the computational graph is travelled forward and backwards in each of the possible modes of AD, for the example of an scalar function  $f(x_1, x_2) = x_1 x_2 + \sin(x_1)$

programming [REF+20; LLW+19; SSK+21].

We remark that AD differs from numerical differentiation (such as finite differences) since it is an exact method. Similarly to symbolic differentiation, each operation involved in the computation provides a rule for its derivative with respect to the corresponding input. However, the output of AD is a numerical value of the derivative, and not a mathematical expression for it.

In turn, AD is carried out by constructing a *computational graph*, where nodes are associated to the elementary operations present in the computation (and whose derivatives are known), and where the dependence on the parameters is explicated. In order to compute the function derivative, the graph is transversed either forward or backwards; the latter case is known as *backpropagation*<sup>6</sup>. Thanks to efficient matrix multiplication sub-routines and special-purpose hardware such as GPUs and TPUs, this can be done considerably fast, and in turn AD stands as one of the main ideas behind the advent of Artificial Intelligence.

Let us consider an example<sup>7</sup> of an algorithm taking as input a real-value  $x$ , and outputing  $y = f(g(h(x)))$  where  $f, g$  and  $h$  are elementary functions whose derivatives are well-known and can be recorded when building the computational graph. We aim to compute  $\partial_x y$ , and to this end we define  $w_0 = x$ ,  $w_1 = h(w_0)$ ,  $w_2 = g(w_1)$  and  $w_3 = f(w_2) = y$ . Now, the chain rule implies that

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial w_2} \frac{\partial w_2}{\partial w_1} \frac{\partial w_1}{\partial x}. \quad (2.41)$$

Thus, the computational graph is built by recording each of the elementary functions

<sup>6</sup>The optimal way to transverse the computational graph generally depends on the setting at hand. For high-dimensional inputs (as customary when working with neural networks), backpropagation is preferred at the cost of an increase in memory usage.

<sup>7</sup>Taken from Ref. [Wik23]

that need to be computed *and* each of the elementary functions derivatives, which are known in advance. Then, forward-AD transverses the chain-rule from inside to outside, by first computing  $\frac{\partial w_1}{\partial x}$ , then  $\frac{\partial w_2}{\partial w_1}$  and finally  $\frac{\partial y}{\partial w_2}$ . On the contrary, backwards-AD transverses the chain-rule from outside to inside, by first computing  $\frac{\partial y}{\partial w_2}$ , then  $\frac{w_2}{w_1}$  and finally  $\frac{\partial w_1}{\partial x}$ . As a concrete example, we can consider the function

$$y = f(x_1, x_2) \tag{2.42}$$

$$= x_1 x_2 + \sin(x_1) \tag{2.43}$$

$$= w_1 w_2 + \sin(w_1) \tag{2.44}$$

$$= w_3 + w_4 \tag{2.45}$$

$$= w_5, \tag{2.46}$$

where at each elementary operation we define a new variable  $w_i$ , which represents a node in the computational-graph. In the table we show the operations needed to compute the function, forward-AD, and backward-AD (this is also schematised in Fig. 2.3).

| Function computation  | Forward-AD                                              | Backward-AD                                             |
|-----------------------|---------------------------------------------------------|---------------------------------------------------------|
| $w_1 = x_1$           | $\dot{w}_1 = 1$ (seed)                                  | $\bar{w}_5 = 1$ seed                                    |
| $w_2 = x_2$           | $\dot{w}_1 = 0$ (seed)                                  | $\bar{w}_4 = \bar{w}_5 \cdot 1$                         |
| $w_3 = w_1 \cdot w_2$ | $\dot{w}_3 = \dot{w}_1 \cdot w_2 + w_1 \cdot \dot{w}_2$ | $\bar{w}_3 = \bar{w}_5 \cdot 1$                         |
| $w_4 = \sin(w_1)$     | $\dot{w}_4 = \cos(w_1) \cdot \dot{w}_1$                 | $\bar{w}_2 = \bar{w}_3 \cdot w_1$                       |
| $w_5 = w_3 + w_4$     | $\dot{w}_5 = \dot{w}_3 + \dot{w}_4$                     | $\bar{w}_1 = \bar{w}_3 \cdot w_2 + \bar{w}_4 \cos(w_1)$ |

Here, we used the notation of  $\bar{w}_i = \frac{\partial y}{\partial w_i}$  for the backward-AD, and note that the *seed value* in this example is trivial for such case (since it stands for possible multidimensional outputs, e.g. cases where  $y \in \mathbb{R}^n$ ). On the contrary, since the example function has two inputs  $(x_1, x_2)$ , forward-AD needs to be specified which derivative we are computing —either  $\partial_{x_1} y$ , as set by the seed value in the table, or  $\partial_{x_2} y$ —. Note that to compute the gradient, we would need two passes of the graph for this example<sup>8</sup>.

While this discussion only provides an introduction to the meaning of AD, we remark that much work has been done in this field during the last years, and we refer the interested reader to Refs. [Wik23; ABC+16; MDA15; Mac16; BPR+18].

---

<sup>8</sup>For a vector field  $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ , then computing the gradient requires  $m$  computational graphs sweeps for the backward-AD, and  $n$  sweeps for the forward-AD.

### Experimental implementations on quantum hardware

In NISQ circuits, backpropagating cost-function derivatives imply that the quantum state at each step in the circuit should be kept in memory, a fact essentially forbidden because of exponentially large Hilbert-spaces. Alternatively, we can use quantum hardware to compute such gradients.

In this scenario, cost-function derivatives can be obtained by the so-called *parameter-shift rules* where cost-function derivatives are expressed as linear combinations of cost-function values, each obtained by using the very same circuit, but shifting the parameters by a finite amount. We remark that this is an analytic result, unrelated to numerical differentiation techniques such as finite differences<sup>9</sup>.

In the following we will detail the idea behind parameter-shift rules. Let us consider a single parameter  $\alpha \in \boldsymbol{\theta}$ , and a VQE problem, whose Hamiltonian is  $\hat{H}$  and thus the cost-function becomes

$$C(\boldsymbol{\theta}) = C_{\text{VQE}}(\boldsymbol{\theta}) = \langle 0 | U^\dagger(\boldsymbol{\theta}) \hat{H} U(\boldsymbol{\theta}) | 0 \rangle, \quad (2.47)$$

where we momentarily simplified our notation by dropping the dependence on  $\mathbf{k}$ , and used the notation  $|0\rangle^{\otimes n} \equiv |0\rangle$ .

Moreover, we will assume that  $U(\boldsymbol{\theta}) = U_L G(\alpha) U_R$ , where  $U_L$  and  $U_R$  are unitary transformations parametrized by  $\boldsymbol{\theta} - \{\alpha\}$ . Upon defining  $|\psi\rangle = U_R |0\rangle$ , and  $\hat{Q} = U_L^\dagger \hat{H} U_L$ , it follows that

$$\partial_\alpha C(\boldsymbol{\theta}) = \langle \psi | G^\dagger \hat{Q} (\partial_\alpha G) | \psi \rangle + \langle \psi | (\partial_\alpha G^\dagger) \hat{Q} G | \psi \rangle, \quad (2.48)$$

where we used  $G$  as a shorthand for  $G(\alpha)$ . Now, let us consider  $G(\alpha) = e^{-i\frac{\alpha}{2}P}$ , with  $P \in \{\sigma_x, \sigma_y, \sigma_z\}$ . In this case, we obtain

$$\partial_\alpha C(\boldsymbol{\theta}) = \frac{i}{2} \langle \psi | G^\dagger [P, \hat{Q}] G | \psi \rangle. \quad (2.49)$$

Now, using the following identity, which holds for any operator  $\sigma$  [MNK+18]:

$$[P, \sigma] = i \left( G\left(\frac{\pi}{2}\right) \sigma G\left(-\frac{\pi}{2}\right) - G\left(-\frac{\pi}{2}\right) \sigma G\left(\frac{\pi}{2}\right) \right), \quad (2.50)$$

we get

$$\partial_\alpha C(\boldsymbol{\theta}) = -\frac{1}{2} \left( \langle \psi | G(\alpha)^\dagger G^\dagger\left(-\frac{\pi}{2}\right) \hat{Q} G\left(-\frac{\pi}{2}\right) G(\alpha) | \psi \rangle + \langle \psi | G(\alpha)^\dagger G^\dagger\left(\frac{\pi}{2}\right) \hat{Q} G\left(\frac{\pi}{2}\right) G(\alpha) | \psi \rangle \right)$$

---

<sup>9</sup>In turn, finite-differences do not get along with gradient-based optimizers, and often leads to instabilities due to approximation errors; this issue intensifies in the NISQ framework, since we generally need to resolve between cost-function values that are small. Estimating such difference in cost-function values turns difficult, since not only a high amount of samples is required, but also the strong presence of hardware-noise can potentially forbid to resolve it.

By noting that  $G(\alpha)G(\beta) = G(\alpha + \beta)$ , and recalling that  $\hat{Q} = U_L^\dagger \hat{H} U_L$ , it follows that

$$\partial_\alpha C(\boldsymbol{\theta}) = \frac{1}{2} (C(\boldsymbol{\theta})|_{\alpha \leftarrow \alpha + \frac{\pi}{2}} - C(\boldsymbol{\theta})|_{\alpha \leftarrow \alpha - \frac{\pi}{2}}) \quad (2.51)$$

This procedure can be extended to consider arbitrary Pauli strings [MNK+18], and unitary transformations for which the derivative is not unitary. In the latter case, the derivative can be written as a linear combination of unitary operations, resulting in a generalized parameter-shift rule which – as opposed to Eq 2.51 – involves more than two contributions [SBG+19; WIW+22].

Thus, the cost-function derivatives can be obtained by estimating each term in Eq. 2.51 using a quantum computer parametrized by  $U(\boldsymbol{\theta})$ , and repeating this procedure for each parameter  $\alpha \in \boldsymbol{\theta}$ . As mentioned, this result is exact (contrary to numerical differentiation methods); it nevertheless relies on cost-function estimates, thus statistical fluctuations arising from measurement outcomes will always be present. This plays an important role if the value of such derivatives is small, in which case many measurements will be required in order to deal with sufficient accuracy.

Thus, VQAs' optimizers will generally need to deal with an intrinsically stochastic component, and convergence properties of SGD has been analyzed in Ref. [SWM+20]. Following such reference, we can readily distinguish between different sources of *stochasticity*, the first one being shot-noise. Alternatively, in Eq. 2.38 we have decomposed the Hamiltonian as a sum of Pauli-string, and at each cycle of the optimization algorithm, the gradient could be estimated out of a subset of all the Pauli strings present in such decomposition. In this case, another source of stochasticity arises, and is associated to estimating cost-function gradient out of a subset of terms appearing in the cost-function. Methods combining both techniques are known as *doubly-stochastic* optimizers [SWM+20; HN21].

Overall, parameter-shift rules can be understood as a *recipe* for computing gradients of cost functions; we remark that the approach presented above naturally generalizes to cost-functions of the form in Eq. 2.36. Under this recipe, circuit's structure needs not to be modified, but rather the parameter in question is shifted, and the derivative is obtained as linear combination of the corresponding cost-function estimates. For classical simulation purposes, the *best of both worlds* are usually combined, and backpropagation methods are used along with the parameter-shift rules (provided that the systems under consideration are not too large); this is one among the many efforts behind projects like TensorFlow-Quantum or PennyLane [BIS+18; BVM+20].

By now, we have introduced the basic ingredients required to run a VQA. Namely, a strategy that defines the structure of our quantum circuit, a cost-function

that evaluates its performance, and a (classical) optimization algorithm that takes control of free parameters such as rotation values. The ultimate goal is to find the global minima of the cost function by navigating the optimization landscape. In this regard many issues have been recently raised up, which put into question the overall feasibility of VQA framework, and that is the matter of our next section.

### 2.2.4 Barren-plateaus

The barren plateau (BP) phenomenon has recently received considerable attention as one of the main challenges to overcome for the VQA framework to provide a quantum speedup. BPs refer to the exponentially vanishing (in the number of qubits) for a random circuit to present a gradient component which is slightly higher than zero. In turn, this constitutes a big challenge when navigating the cost-function landscape, since such landscape (*e.g.* the set of unitaries that are generated when varying the parameters of a sufficiently random PQC) becomes flat in the presence of a BP.

In the following we will first outline how the BP phenomenon emerges from a 2-design, and then discuss several extensions of this phenomena to other classes of circuits.

The notion of  $t$ -design quantifies how similar the measure obtained from varying parameters  $\theta$  in  $U(\mathbf{k}, \theta)$  differs from that of the Haar measure, which is the uniform measure in  $U(n)$ . A random event refers to randomly sampling an unitary transformation from  $\mathcal{U}$ , *i.e.* the set of unitaries that are generated from  $U(\mathbf{k}, \theta)$  by modifying the parameters  $\theta$ . Thus, we can define an expressibility super-operator as

$$\mathcal{A}_{\mathcal{U}}^{(t)}(\cdot) = \int_{\mathcal{U}} d_{\mu}(U) U^{\otimes t}(\cdot)U^{\dagger \otimes t} - \int_{U(n)} d_{\mu_H}(U) U^{\otimes t}(\cdot)U^{\dagger \otimes t}, \quad (2.52)$$

where  $d_{\mu_H}(U)$  denotes the volume element of the Haar measure, and  $d_{\mu}(U)$  is the volume element corresponding to the uniform distribution over  $\mathcal{U}$  [HSC+22]; for PQCs such set is generated by uniformly sampling over the parameters  $\theta$ .

Thus,  $\mathcal{U}$  forms a  $t$ -design if  $\mathcal{A}^{(i)}(X) = 0$  for every operator  $X$  and every  $i = 1, \dots, t$ , meaning that the uniform distribution over  $\mathcal{U}$  matches the uniform distribution over  $U(n)$  up to the first  $t$ -moments. Using some useful identities from

integrating over the Haar measure, namely that

$$\int_{U(n)} d_\mu(U) U_{ij} U_{mk}^* = \frac{\delta_{im} \delta_{jk}}{n} \quad (2.53)$$

$$\begin{aligned} \int_{U(n)} d_\mu(U) U_{ij} U_{kl} U_{mn}^* U_{op}^* &= \frac{1}{n^2 - 1} (\delta_{im} \delta_{jn} \delta_{ko} \delta_{lp} + \delta_{io} \delta_{km} \delta_{jp} \delta_{ln}) \\ &\quad - \frac{1}{n(n^2 - 1)} (\delta_{im} \delta_{ko} \delta_{jp} \delta_{ln} + \delta_{io} \delta_{km} \delta_{jn} \delta_{lp}), \end{aligned} \quad (2.54)$$

the landscape of 2-designs was explored in Ref. [MBS+18]. In particular, using the identities we can show that 2-designs have gradients which are on average (over the Haar measure) zero-valued, which indicate that the landscape is not *biased* towards any particular direction. Moreover, the identities can also be used to compute the variance of the cost-function gradient, and the result (for 2-designs) is that it exponentially vanishes with the number of qubits present in the circuit:

$$\text{Var} [\partial_\alpha C_{\text{VQE}}(\mathbf{k}, \boldsymbol{\theta})] \leq F(n), \quad \text{with} \quad F(n) = \mathcal{O}\left(\frac{1}{2^n}\right), \quad (2.55)$$

where  $\alpha \in \boldsymbol{\theta}$ . From Chebyshev's inequality we have that  $\text{Var} [\partial_\alpha C(\mathbf{k}, \boldsymbol{\theta})]$  bounds the probability that the cost-function partial derivative deviates from its mean value (of zero) as

$$\Pr \left[ \left| \partial_\alpha C(\mathbf{k}, \vec{\theta}) \right| \geq c \right] \leq \frac{\text{Var}[\partial_\alpha C(\mathbf{k}, \vec{\theta})]}{c^2} \sim \mathcal{O}(2^{-n}), \quad (2.56)$$

for any  $c > 0$ . This indicates that when navigating the set  $\mathcal{U}$  by varying parameters  $\boldsymbol{\theta}$ , an optimizer will observe a considerably flat landscape with a high probability. Here, such probability is associated to the chances that, after a random circuit initialization, the cost-function value associated to such circuit presents a gradient whose value is larger than  $c$ , as per Eq. 2.56.

Thus, two difficulties arise when using a quantum computer in this context.

On the one hand, we will need to deal with shot-noise arising from measurements done to estimate the cost-function value. Here, the accuracy of the estimations is proportional to the squared-root-inverse of the number of shots  $N$ .

On the other hand, navigating through the set  $\mathcal{U}$ , *i.e.* the set of unitary transformations that can be reached from the parametrized quantum circuit under consideration  $U(\mathbf{k}, \boldsymbol{\theta})$ , when varying the continuous parameters  $\boldsymbol{\theta}$ . Here, the probability of reaching a higher-than- $\epsilon$  gradient in Eq. 2.56 should be understood in terms of the parameter landscape: high gradients are exponentially unlikely to

be reached, when randomly varying the parameters. Nevertheless, the optimizer will ultimately need to deal with such small gradients, and for the optimization procedure not to behave like a random-walk, we will need to resolve between this gradients, and the accuracy required to accomplish that will generally require a high number of measurements [MBS+18]. In this sense, exponentially many resources will be required to navigate such flat landscape.

While barren plateaus were first identified for large circuits, they were shown to be present in shallow circuits as well. Here, cost-function locality plays a crucial role: while several choices of observables  $\{O_i\}$  and functions  $\{f_i\}$  can lead to different faithful cost functions (*i.e.*, cost functions whose global optima correspond to the solution of the problem), global cost functions can lead to trainability issues for large problem sizes [CSV+21; SCC+22]. Here we recall that global cost functions are defined as ones where  $O_i$  acts non-trivially on all  $n$  qubits. On a different note, the BP-phenomena discussed above has so-far considered 2-designs. In this sense, Refs. [HSC+22; LCS+22] study the presence of BPs when the circuits are  $\epsilon$ -close to a 2-design. In particular, the expressibility of a quantum circuit (*i.e.*, which sample large regions of the unitary group [SJAG19]), can be linked to the amount of entanglement it generates [SCC+22; PNG+21; MKW21], which serves as a tool to diagnosticate the presence of a BP. For example, the Hardware-Efficient-Ansatz is a highly expressible circuit, and consequently suffers from an expressibility-induced BP. Several promising strategies have been proposed to *mitigate barren plateaus*, such as correlating parameters [VC21], layerwise training [SMM+21], and clever parameter initialization [GWO+19; VBM+19]. Nonetheless, constructing smart ansatzes that do not present BPs seem to be the most promising route to avoid them; an example of such are the Quantum Convolutional Neural Networks [CCL19b; PCW+21a].

However, this is not the end of the barren-plateau story: there exists a second effect that leads to barren plateaus which can even affect smart ansatzes with no randomness or entanglement-induced barren plateaus. As shown in Ref. [WFC+21], the presence of certain noise models acting throughout the circuit maps the input state toward the fixed point of the noise model (*i.e.*, the maximally mixed state) [WFC+21; SFGP21], which effectively implies that the cost function value concentrates exponentially around its average as the circuit depth increases. Explicitly, in a noise-induced barren plateau (NIBP) we now find that

$$|\partial_\alpha C(\mathbf{k}, \boldsymbol{\theta})| \leq g(n), \quad \text{with} \quad g(L) = \mathcal{O}\left(\frac{1}{q^L}\right), \quad (2.57)$$

where  $q > 1$  is a noise parameter and  $L$  the number of ansatz's layers. From Eq. 2.57 we see that noise-induced barren plateaus will be critical for circuits whose depth

scales (at least linearly) with the number of qubits. It is worth remarking that Eq. 2.57 is no longer probabilistic as the whole landscape flattens. Here, we note that strategies aimed at reducing the expressibility of the circuit cannot generally prevent the cost from having a noise-induced barren plateau, since here reducing the circuit noise (e.g. improving the quantum hardware) and employing shallow circuits seem to be the only viable and promising strategies to prevent these barren plateaus at the moment.

### 2.2.5 Discussion

In this Section we have discussed about NISQ computing, and strategies to train parametrized quantum circuits. We have introduced the quantum machine learning gargon, namely a cost-function, an ansatz and an optimizer. Then, we have discussed about trainability issues that appear due to the fact that cost-function derivatives concentrates around its mean-value (e.g. zero); such value of the gradient does not correspond to a local minima, but rather to a plateau, known as barren plateau. Such landscape represents an important issue for optimizers, since finding cost-minimizing directions is forbiddenly resource-consuming. Moreover, we note that the barren-plateau phenomena has also been observed in gradient-free optimizers [ACC+21]. Under the vast literature studying this trainability issue, we particularly emphasized that neither strategies that mitigate randomness nor entanglement in ansatzes [VBM+19; VC21; SMM+21; GWO+19; PCW+21a; ZHL+20; BH21; CSA+22], which cause the appearance of barren plateaus will be effective in mitigating a different source of landscape-flatness, called noise-induced barren plateau.

In this regard, it is widely accepted that designing smart ansatzes which prevent altogether barren plateaus is one of the most promising applications. In Chapter 4 we move a step-forward in this direction, providing a method to train quantum circuits on both parameter and structure-wise.

## 2.3 Continuous-variable systems

Let us now introduce the formalism describing continuous variables quantum systems, which sets the playground to work with optical and optomechanical quantum systems, as we do in Chapters 3 and 5. This section is intended as a brief introduction to the formalism of continuous-variable quantum systems, with an emphasis on Gaussian states and operations, and thus several topics remain



out of the scope. The interested reader is referred to Refs. [Ser17a; FOP05] for a comprehensive review of these topics and more.

A quantum continuous variable system is defined as a system whose  $n$  degrees of freedom, as described by  $2n$  canonical operators  $\hat{\mathbf{r}} = (\hat{q}_1, \hat{p}_1, \dots, \hat{q}_n, \hat{p}_n)^T$ , obey the Canonical Commutation Relations (CCR):

$$[\hat{q}_j, \hat{p}_k] = i\hbar\delta_{jk}, \quad j, k = 1, \dots, n. \quad (2.58)$$

Note that we can also define annihilation and creation operators as  $\hat{a}_j = \frac{\hat{q}_j + i\hat{p}_j}{\sqrt{2}}$ , and  $\hat{\mathbf{a}} = (\hat{a}_1, \hat{a}_1^\dagger, \dots, \hat{a}_n, \hat{a}_n^\dagger)^T$ , where an unitary transformation  $\bar{U} = \bigoplus_{j=1}^n \bar{u}$ , where  $\bar{u} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix}$  relates the former operators with the later as  $\hat{\mathbf{a}} = \bar{U}\hat{\mathbf{r}}$ , and the CCR can be compactly written as

$$[\hat{\mathbf{r}}, \hat{\mathbf{r}}^\dagger] = i\Omega, \quad \Omega = \bigoplus_{i=1}^n \Omega_i, \quad \Omega_i = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad (2.59)$$

$$[\hat{\mathbf{a}}, \hat{\mathbf{a}}^\dagger] = \bar{\sigma}_z = \bigoplus_{i=1}^n \sigma_z^{(i)} \quad (2.60)$$

where the notation  $[\hat{\mathbf{b}}, \hat{\mathbf{c}}^\dagger] = \hat{\mathbf{b}}\hat{\mathbf{c}}^\dagger - (\hat{\mathbf{b}}\hat{\mathbf{c}}^\dagger)^\dagger$  should be understood as an outer product which in components reads  $[\hat{\mathbf{b}}, \hat{\mathbf{c}}^\dagger]_{jk} = \hat{b}_j\hat{c}_k^\dagger - \hat{b}_k\hat{c}_j^\dagger$ .

Each canonical degree of freedom  $j$  defines a *mode* equipped with an infinite-dimensional Hilbert space  $\mathcal{H}_j$ , and thus we talk of a  $n$ -mode system whose Hilbert space is  $\otimes_{j=1}^n \mathcal{H}_j$ . Moreover, the  $\Omega$  matrix in Eq. 2.59 is known as the *symplectic* form, and it can be checked that holds: (i): antisymmetry  $\Omega = -\Omega^T$ , (ii) its inverse is its negative  $\Omega^2 = -\mathbb{1}_{2n}$ , and (iii) it corresponds to a real orthogonal transformation  $\Omega^T\Omega = -\Omega^2 = \mathbb{1}_{2n}$ . From here, we are define Weyl transformations as

$$\hat{D}_{\vec{r}} := e^{i\vec{r}\Omega\hat{\mathbf{r}}}, \quad (2.61)$$

where  $\vec{r} = (q_1, p_1, \dots, q_n, p_n)$  is a constant vector in the phase space. A composition law of Weyl operators and its action on canonical operators can readily be derived, and reads<sup>10</sup>

$$\hat{D}_{\vec{r}_1 + \vec{r}_2} = \hat{D}_{\vec{r}_1} \hat{D}_{\vec{r}_2} e^{\frac{i\vec{r}_1^T \Omega \vec{r}_2}{2}} \quad (2.64)$$

<sup>10</sup>Here, we have used the CCR and the BCH formular, which relates the exponential of two operators  $X, Y$  as

$$e^X e^Y = e^Z, \quad Z = X + Y + \frac{[X, Y]}{2} + \frac{[X, [X, Y]]}{12} + \frac{[Y, [X, Y]]}{12} + \dots \quad (2.62)$$

$$\hat{D}_{\vec{r}} \hat{r} \hat{D}_{\vec{r}}^\dagger = \hat{r} + \vec{r}. \quad (2.65)$$

This last equation indicates that Weyl operators act as *displacement* transformations. Equivalently, we can express these transformations with respect to annihilation and creation operators; let  $\vec{\alpha} = (\text{Re}[\alpha_1], \text{Im}[\alpha_1], \dots, \text{Re}[\alpha_n], \text{Im}[\alpha_n])$ , with  $\alpha_j = \frac{q_j + ip_j}{\sqrt{2}}$ , e.g.  $\vec{r} = \bar{U}^\dagger \vec{\alpha}$ , and define

$$\hat{D}_{\vec{\alpha}} := \hat{D}_{\vec{r}}|_{\vec{r}=-\vec{\alpha}} \quad (2.66)$$

$$= e^{\vec{\alpha} \bar{\sigma}_z \hat{a}}, \quad (2.67)$$

and thus Eq. 2.64 reads

$$\hat{D}_{\vec{\alpha}+\vec{\beta}} = \hat{D}_{\vec{\alpha}} \hat{D}_{\vec{\beta}} e^{-\alpha^T \bar{\beta} + \bar{\alpha}^T \beta} \quad (2.68)$$

We will now introduce an important set of quantum states, which in turn constitute our battle-horse when dealing with continuous degrees of freedom.

### 2.3.1 Gaussian systems

We define Gaussian quantum states as thermal states of a quadratic hamiltonians

$$\rho_G = \frac{e^{-\beta \hat{H}}}{\text{Tr}[e^{-\beta \hat{H}}]}, \quad \beta \in \mathbb{R}^+, \quad (2.69)$$

$$\hat{H} = \frac{\hat{r}^T H \hat{r}}{2} + \hat{r}^T \vec{r}, \quad (2.70)$$

where  $\vec{r}$  is a constant vector in phase space, and  $H$  is the *Hamiltonian matrix*, which is positive-defined and symmetric. Note this definition includes the limit  $\beta \rightarrow \infty$ , which corresponds to the case of pure states.

Such construction allows us to find the so-called normal mode form of  $\rho_G$ , in which the system behaves as a set of  $n$  decoupled harmonic oscillators. In particular, studying the transformation that takes  $\rho_G$  into its normal form is instructive, and we will comment on it nextly. There are at least two important reasons to study Gaussian states. Firstly, many problems become analytically tractable when dealing with such a class. Secondly, manipulating Gaussian states is experimentally feasible.

---

Moreover, if we take a *central* commutator  $[X, Y]$ , i.e.  $[X, [X, Y]] = [Y, [X, Y]] = 0$ , then the action of  $e^X$  by similarity reads

$$e^X Y e^{-X} = Y + [X, Y] \quad (2.63)$$

Moreover, light (*e.g.*) is a natural candidate for quantum communication scenarios, and coherent states – a particular type of Gaussian states – describe very well the behaviour of lasers.

We note that, up to a constant shift in the energy landscape,  $\hat{H}$  is equivalent to

$$\hat{H}' = \frac{(\hat{\mathbf{r}} - \bar{\mathbf{r}})^T H (\hat{\mathbf{r}} - \bar{\mathbf{r}})}{2}, \quad \bar{\mathbf{r}} = -H^{-1} \vec{\mathbf{r}}. \quad (2.71)$$

Note that  $H^{-1}$  does always exist since  $H > 0$ . It follows that

$$\hat{H}' = \frac{1}{2} \hat{D}_{-\bar{\mathbf{r}}} \hat{\mathbf{r}}^T H \hat{\mathbf{r}} \hat{D}_{\bar{\mathbf{r}}}, \quad (2.72)$$

we can thus focus on the non-displaced hamiltonian  $\hat{\mathbf{r}}^T H \hat{\mathbf{r}}$ . In the Heisenberg picture, the evolution of canonical operators corresponds to  $\dot{\hat{\mathbf{r}}}_j = i [\hat{H}, \hat{\mathbf{r}}_j]$ , and by using the CCR, it follows that  $\dot{\hat{\mathbf{r}}} = \Omega H \hat{\mathbf{r}}$ . Thus, we conclude that  $\hat{\mathbf{r}}(t) = e^{\Omega H t} \hat{\mathbf{r}}(0)$ . Since any system transformation should preserve the CCR, it follows that

$$i\Omega = [\hat{\mathbf{r}}(t), \hat{\mathbf{r}}^T(t)] = e^{\Omega H} [\hat{\mathbf{r}}(0), \hat{\mathbf{r}}(0)^T] (e^{\Omega H})^T = i e^{\Omega H} \Omega (e^{\Omega H})^T, \quad (2.73)$$

and thus

$$\Omega = S \Omega S^T, \quad S := e^{\Omega H t}. \quad (2.74)$$

The last line is the condition for  $S$  to belong to the real symplectic group  $\text{Sp}_{2n, \mathbb{R}}$ , which is the reason for calling  $\Omega$  the symplectic form. If we now explicit the unitary character of canonical operators' time-evolution, we obtain

$$\dot{\hat{\mathbf{r}}} = \hat{U}^\dagger(t) \hat{\mathbf{r}}(0) \hat{U}(t) = e^{-i\hat{H}t} \hat{\mathbf{r}}(0) e^{i\hat{H}t} = e^{\Omega H t} \hat{\mathbf{r}}(0) = S \hat{\mathbf{r}}(0), \quad (2.75)$$

we see that there is a clear correspondence between unitary operators and symplectic transformations, and in fact  $\text{Sp}_{2n, \mathbb{R}} \cap SO(2n)$  is isomorphic to  $U(n)$ . Thus, there is a correspondence between symplectic transformations  $S = e^{\Omega H}$  and unitary operators generated by second-order hamiltonians,  $\hat{S} = e^{\frac{i\hat{\mathbf{r}}^T H \hat{\mathbf{r}}}{2}}$ :

$$\hat{S} \hat{\mathbf{r}} \hat{S}^\dagger = S \hat{\mathbf{r}}. \quad (2.76)$$

In particular, said symplectic transformation  $S$  can be written in its singular value decomposition as

$$S = O_1 Z O_2, \quad (2.77)$$

where  $O_1, O_2 \in \text{Sp}_{2n, \mathbb{R}} \cap SO(2n)$  are orthogonal symplectic transformations and  $Z = \bigoplus_{j=1}^n \begin{pmatrix} z_j & 0 \\ 0 & z_j^{-1} \end{pmatrix}$  stands for the *squeezing* transformation. The energy of

the system  $(\hat{\mathbf{r}}^T \hat{\mathbf{r}})$  is invariant under the action of  $O_i$ , and thus we refer to those transformations as *passive* ones. Importantly, the isomorphism with the unitary group guarantees that such symplectic transformations can be constructed out from *phase-shifters* and *beam-splitters* [RZB+94; WPGP+12].

In particular, phase-shifters  $R_\phi$  and beam-splitters  $B_{\theta,\psi}$  have the following matrix representations

$$R_\phi = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix}, \quad B_{\theta,\psi} = \begin{pmatrix} \cos \theta & e^{i\psi} \sin \theta \\ -e^{-i\psi} \sin \theta & \cos \theta \end{pmatrix} \otimes \mathbb{I}_2. \quad (2.78)$$

In passing, we note that if we move the annihilation and creation operators, the action of phase-shifters is that of adding a phase  $e^{i\phi}$ , whereas beam-splitters mix two modes  $a_1$  and  $a_2$  (which correspond to the input modes):

$$B_{\theta,\psi}^\dagger \hat{a}_1 B_{\theta,\psi} = \cos \theta \hat{a}_1 + e^{i\psi} \sin \theta \hat{a}_2 \quad (2.79)$$

$$B_{\theta,\psi}^\dagger \hat{a}_2 B_{\theta,\psi} = -e^{-i\psi} \sin \theta \hat{a}_2 + \cos \theta \hat{a}_1, \quad (2.80)$$

where the added phase  $\psi$  can, for instance, be corrected via a proper phase-shift.

To sum up, any symplectic transformation can be implemented out of squeezing operations and interferometers (*e.g.* a combination of phase-shifters and beam-splitters, which can in turn mimic any passive symplectic transformation).

Let us now focus on the structure of the Hamiltonian matrix  $H$ . In particular, we recall that given  $M \in \mathbb{R}^{2n \times 2n}$ , with  $M > 0$ , there exists a transformation  $S \in \text{Sp}_{2n, \mathbb{R}}$  taking  $M$  into its normal form [Wil36], *e.g.*  $SMST^T = D$  with  $D = \bigoplus_j d_j \mathbb{I}_2$  containing the symplectic eigenvalues  $\{d_j\}_{j=1}^n$ ; recall that such transformation can be challenging to find [JPP21; PSL09]. In this regard,  $S$  is related to the transformation  $L$  that puts  $i\Omega M$  into its diagonal form, *e.g.*  $L\Omega ML^{-1}$  is diagonal, with  $S = (L^{-1}\bar{U})^\dagger$ , and the symplectic eigenvalues of  $M$  are the (absolute values of) the eigenvalues of  $i\Omega M$ , which come into pairs  $\{\pm d_j\}_{j=1}^n$ . We can readily apply such decomposition to our quadratic Hamiltonian matrix  $H$ :

$$H = S^T \left( \bigoplus_{j=1}^n \omega_j \mathbb{I}_2 \right) S, \quad (2.81)$$

where by denoting the symplectic eigenvalues by  $\omega_j$  we highlight its role as frequencies of  $n$  non-interacting free modes. By inserting such normal mode decomposition into  $\hat{H}$ :

$$\hat{H} = \frac{1}{2} \hat{\mathbf{r}}^T H \hat{\mathbf{r}} \quad (2.82)$$

$$= \frac{1}{2} \hat{\mathbf{r}}^T S^T \left( \bigoplus_{j=1}^n \omega_j \mathbb{I}_2^{(j)} \right) S \hat{\mathbf{r}} \quad (2.83)$$

We now use that the symplectic transformation  $S$  has a unitary representation  $\hat{S}$ :

$$\hat{H} = \frac{1}{2} \hat{\mathbf{r}}^\dagger H \hat{\mathbf{r}} \quad (2.84)$$

$$= \frac{1}{2} \hat{S} \hat{\mathbf{r}}^T \left( \bigoplus_{j=1}^n \omega_j \mathbb{1}_2^{(j)} \right) \hat{\mathbf{r}} \hat{S}^\dagger \quad (2.85)$$

Upon defining  $\hat{H}_{\omega_j} := \frac{\omega_j}{2} (q_j^2 + p_j^2)$ , we note that any second-order Hamiltonian with zero displacement is equivalent (up to a unitary transformation) to a set of decoupled oscillators. The decomposition of the original quadratic hamiltonian reads

$$\hat{H} = \frac{1}{2} (\hat{\mathbf{r}} - \bar{\mathbf{r}})^T H (\hat{\mathbf{r}} - \bar{\mathbf{r}}) \quad (2.86)$$

$$= \hat{D}_{-\bar{\mathbf{r}}} \hat{S} \left( \sum_{j=1}^n \hat{H}_{\omega_j} \right) \hat{S}^\dagger \hat{D}_{\bar{\mathbf{r}}}, \quad (2.87)$$

which implies that gaussian states have the following structure:

$$\rho_G = \frac{\hat{D}_{-\bar{\mathbf{r}}} \hat{S} \left( \bigotimes_{j=1}^n e^{-\beta \hat{H}_{\omega_j}} \right) \hat{S}^\dagger \hat{D}_{\bar{\mathbf{r}}}}{\prod_j \text{Tr}[e^{-\beta \hat{H}_{\omega_j}}]} \quad (2.88)$$

$$= \mathcal{N}_\beta \hat{D}_{-\bar{\mathbf{r}}} \hat{S}^\dagger \left( \bigotimes_{j=1}^n \sum_{k_j=0}^{\infty} e^{-\beta k_j \omega_j} |k_j\rangle \langle k_j| \right) \hat{S} \hat{D}_{\bar{\mathbf{r}}}, \quad (2.89)$$

where in the second line we have expanded  $\hat{H}_{\omega_j}$  in the Fock basis  $\{|k_j\rangle \langle k_j|\}$  of each mode  $j$ , and defined  $\mathcal{N}_\beta := \prod_j (1 - e^{-\beta \omega_j})$ . We observe that for  $\beta \rightarrow \infty$  a pure state is obtained, since only vacuum terms survive. Hence, pure gaussian states are generated by applying gaussian unitary operations to the n-mode vacuum state. Quite remarkably, either Gaussian states have rank 1 (pure states), or infinite rank (mixed states).

We thus conclude that Gaussian states are obtained from orthogonal symplectic transformations (and displacements) applied to a set of non-interacting harmonic oscillators. In particular, the action of such transformations is linear in phase space, a fact that becomes clear when inspecting its action on the statistical moments.

### 2.3.2 Some methods in phase space

Another parametrization of Gaussian states can be obtained, by departing from Eq. 2.88; in particular we note that the first two moments of  $\rho_G$  encode all its

information, as

$$\begin{aligned}\bar{\mathbf{r}} &:= \text{Tr}[\rho_G \hat{\mathbf{r}}] \\ \Sigma &:= \text{Tr}[\{\hat{\mathbf{r}} - \bar{\mathbf{r}}, (\hat{\mathbf{r}} - \bar{\mathbf{r}})^\dagger\} \rho] = S \left( \bigoplus_j \nu_j \mathbb{1}_2^{(j)} \right) S^T,\end{aligned}\tag{2.90}$$

where  $\nu_j = \frac{1 + e^{-\beta\omega_j}}{1 - e^{-\beta\omega_j}}$  stand for the symplectic eigenvalues of  $\Sigma$ <sup>11</sup>. All the information about the Gaussian state is encoded in both the expectation value and covariance. On the one hand, the displacement value  $\bar{\mathbf{r}}$  stand for the first statistical moment of the Gaussian state. On the other hand, the symplectic transformation  $S$  (associated to the gaussian unitary  $\hat{S}$ ), as well as the symplectic eigenvalues of  $H$  can be obtained through  $\Sigma$ . In this regard, we note that the spectrum of the Gaussian state can be obtained from the covariance matrix, and thus correlation measures between the modes, as well as the states' purities, can be inferred from the second statistical moment.

Importantly, any covariance matrix  $\Sigma$  of a quantum state should satisfy the Robertson-Schrödinger uncertainty relation

$$\Sigma + i \Omega \geq 0,\tag{2.91}$$

which is a direct consequence of the CCR and the positivity of the quantum state  $\rho$ . The singular value decomposition of symplectic transformations in (2.77) turns useful to understand the structure of the covariance matrix. For example, if we restrict to pure Gaussian states, then it follows that

$$\Sigma = O D O^T,\tag{2.92}$$

where  $D = \bigoplus_{j=1}^n \begin{pmatrix} d_j & 0 \\ 0 & \frac{1}{d_j} \end{pmatrix}$ ,  $d_j > 0$  and  $O$  stands for a passive, orthogonal transformation. Moreover, if we restrict to single-mode systems, we observe that covariance matrices are generated by squeezing and rotating the identity; this is aligned with the fact that all pure gaussian states can be obtained through symplectic transformations applied to the vacuum state  $|0\rangle^{\otimes n}$ .

While preparation of pure Gaussian states has such intuitive interpretation, this picture can be complemented with a collection of quasiprobability distributions that depict the quantum state in phase space which we now turn to present.

---

<sup>11</sup>Note that for the single-mode case, the covariance reads  $\Sigma_{nm} = \langle \Delta \hat{\mathbf{r}}_n \Delta \hat{\mathbf{r}}_m \rangle + \langle \Delta \hat{\mathbf{r}}_m \Delta \hat{\mathbf{r}}_n \rangle$ , with  $\Delta \hat{\mathbf{r}}_i = \hat{\mathbf{r}}_i - \langle \hat{\mathbf{r}}_i \rangle$ , and  $\hat{\mathbf{r}} = (q, p)$ .

In analogy to classical mechanics, we can study the quantum system in phase space. In doing so, we arrive to the so-called quasiprobability density functions, which are equivalent to the quantum state and provides yet another operational approach to its characterisation. Unsurprisingly, such analogy finds frontiers, since such quasiprobability distributions can take negative values (for instance, for non-Gaussian states). In this regard, we find that the set of quantum Gaussian states behave classically, since their associated quasiprobabilities are – spoiler alert – Gaussian distributions.

The phase-space formalism hinges on the overcompleteness of Weyl operators, which serves as a bridge between Hilbert and phase spaces. In the following we will restrict to the *single*-mode case; generalizations to  $n$ -mode systems are straightforward.

Recall that a coherent state is defined an eigenstate of  $\hat{a}$

$$\begin{cases} \hat{a}\hat{D}_\alpha|0\rangle = \alpha\hat{D}_\alpha|0\rangle \\ |\alpha\rangle = \hat{D}_\alpha|0\rangle \end{cases}. \quad (2.93)$$

In particular, note that  $|\alpha\rangle\langle\alpha| = \hat{D}_\alpha|0\rangle\langle 0|\hat{D}_\alpha^\dagger$ , meaning that coherent states are – in line with the formalism revisited in the previous section – just displaced vacuum states, with idle symplectic transformations and thus an identity as covariance matrix. In particular, by writing such states in the Fock basis, we obtain

$$|\alpha\rangle = e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle, \quad (2.94)$$

from which it is not hard to prove that coherent states resolve to identity:

$$\frac{\int_{\mathbb{C}} |\alpha\rangle\langle\alpha| d^2\alpha}{\pi} = \mathbb{1}. \quad (2.95)$$

This shows that coherent states resolve to a measurement, called *heterodyne* measurement. Moreover, given any (bounded) operator  $\hat{O}$ , we can express it as

$$\hat{O} = \frac{1}{\pi} \int_{\mathbb{C}} d^2\alpha \text{Tr}[\hat{O}\hat{D}_\alpha] \hat{D}_{-\alpha}, \quad (2.96)$$

$$\text{Tr}[\hat{O}] = \frac{1}{\pi} \int_{\mathbb{C}} d^2\alpha \langle\alpha|\hat{O}|\alpha\rangle, \quad (2.97)$$

which is known as the *Fourier-Weyl relation*. This allows us to write any quantum state  $\rho$  in terms of coherent states as

$$\rho = \frac{1}{\pi} \int_{\mathbb{C}} \text{Tr}[\rho\hat{D}_\alpha] \hat{D}_{-\alpha}. \quad (2.98)$$

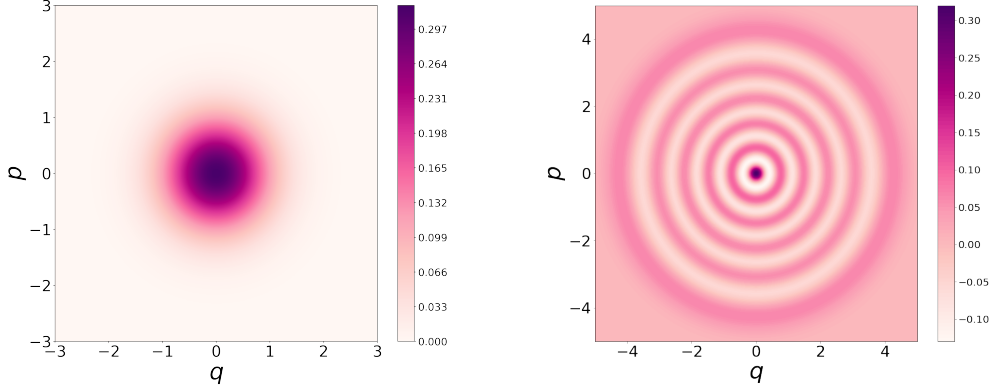


Figure 2.4: We show the Wigner functions of the vacuum state  $|0\rangle$  (left) and of the Fock state  $|5\rangle$  (right). While in the former case the probability distribution is a Gaussian, the latter represents a non-Gaussian state (e.g. is the fifth eigenstate of the harmonic oscillator), which here translates into negative values of the Wigner function.

In particular, the *characteristic function*  $\chi(\alpha)$  is defined as the coefficient accompanying  $\hat{D}_{-\alpha}$  in the integral, e.g.

$$\chi(\alpha) = \text{Tr}[\rho \hat{D}_\alpha] \tag{2.99}$$

$$\rho = \frac{1}{\pi} \int_{\mathbb{C}} \chi(\alpha) \hat{D}_{-\alpha}. \tag{2.100}$$

Such characteristic function allows us to define the Wigner function as its (complex) Fourier transform:

$$W(\alpha) = \frac{1}{\pi^2} \int_{\mathbb{C}} d^2\beta \chi(\beta) e^{\alpha\bar{\beta} - \bar{\alpha}\beta}, \tag{2.101}$$

which satisfies that  $\int W(\alpha) d^2\alpha = 1$ . By expanding  $\alpha = \frac{q+ip}{\sqrt{2}}$  and  $\beta = \frac{\tilde{q}+i\tilde{p}}{\sqrt{2}}$ , we find that  $W(q, p) = \int d\tilde{q} e^{ip\tilde{q}} \langle q + \tilde{q} | \rho | q - \tilde{q} \rangle$ . Such parametrization provides an operational interpretation of Wigner functions, since  $\frac{1}{2} \int_{-\infty}^{\infty} dp W(q, p) = \langle q | \rho | q \rangle$ , e.g. marginal distributions predict probability measurement outcomes over conjugate variables and if varying the direction of  $p$  in the integrand one can perform a tomographic reconstruction of the quantum state. Finally, for Gaussian states, the



characteristic and Wigner functions are Gaussians, *e.g.*

$$\chi(\vec{r}) = e^{-\frac{(\Omega\vec{r})^T \Sigma \Omega\vec{r}}{2} - i(\Omega\vec{r})^T \vec{r}} \quad (2.102)$$

$$W(\vec{r}) = \frac{2^n}{\pi^n \sqrt{|\Sigma|}} e^{-(\vec{r}-\bar{r})^T \Sigma^{-1} (\vec{r}-\bar{r})}. \quad (2.103)$$

In Fig. 2.4 we show two examples of such Wigner functions, one for a coherent (and thus Gaussian) state, and other one for a Fock number state; we observe in the latter case that the Wigner function can take negative values: such negativity play a role of a classical-quantum frontier in continuous-variable systems [Wal21].

We have thus identified the set of Gaussian states as thermal (and ground) states of quadratic hamiltonians. Equivalently, Gaussian states can be defined as those whose characteristic function is Gaussian. On the one hand, we have observed that Gaussian states can be understood as a set of  $n$  non-interacting harmonic oscillators, up to a symplectic transformation  $\hat{S}$ . Such symplectic transformation is ultimately linked to a Gaussian unitary, *e.g.* generated by a quadratic hamiltonian. Moreover, Eq. 2.90 explicits the action of such transformations on the first two statistical moments of the Gaussian state, namely displacing the canonical coordinates and transforming the covariance matrix via a symplectic matrix that acts by similarity. In turn, such operations act linearly in phase space and that is the reason for which Gaussian operations are deemed so in literature [WPGP+12; Oli12].

### 2.3.3 Gaussian operations and beyond

A quantum channel is deemed Gaussian if it preserves the Gaussian character of the state. Such definition applies to unitary transformations, but also to CP-maps (which can suitably be obtained via a Stinespring dilation with an ancilla).

A handful parametrization of Gaussian CP-maps can readily be obtained by studying the evolution of system's state after a Gaussian unitary interaction with a (Gaussian) auxiliary system. The result is an update rule for the first two moments, parametrized by matrices  $X$  and  $Y$  in  $\mathbb{R}^{2n \times 2n}$ :

$$\bar{r} \rightarrow X\bar{r} \quad (2.104)$$

$$\Sigma \rightarrow X\Sigma X^T + Y, \quad (2.105)$$

with  $Y + i\Omega \geq iX\Omega X^T$  (such condition guarantees that Robertson-Schrödinger uncertainty relations in Eq. (2.91) are preserved). We note that such evolution is different than the resulting one when continuously monitoring the system, as studied in Sec. 2.4.

In particular an attenuator – or *lossy channel* – is obtained by mixing a quantum state with the vacuum by a beam-splitter, and we denote it by  $\mathcal{L}_\eta$ , where  $\eta$  is the transmissivity coefficient. For such a channel, and restricting to the single-mode case, one finds that  $X = \eta\mathbb{I}_2$  and  $Y = \sqrt{1 - \eta^2}\mathbb{I}_2$ , with  $\eta \in [0, 1]$ . We note that if the input state is a coherent state, the covariance matrix remains unchanged (since  $\Sigma = \mathbb{I}_2$ ), whereas the first moment gets attenuated as  $\alpha \rightarrow \eta\alpha$ . This channel turns to accurately model the behaviour of a laser beam transmitted through the atmosphere, though the transmissivity  $\eta$  non-trivially depends on parameters such as height, temperature and more [DVR+19; AP05; UHP+12; Pir21a; Pir21b; VSV11; VSV+17].

### 2.3.4 Measurements

A measurement is deemed gaussian if (i) it preserves the Gaussian character of the state upon post-conditioning, and (ii) when measuring a Gaussian state, the outcome probability distribution is Gaussian. For instance, counting excitations correspond to projections over Fock states: recall that  $\sum_{n=0}^{\infty} |n\rangle\langle n| = \mathbb{I}$ , and are obviously non-gaussian since number states (asside from vacuum) are not; moreover, the underlying distribution is Poissonian. Nevertheless, it is experimentally possible (though expensive) to resolve excitations in quantum optics laboratories, thorough an apparatus known as *photon-counter* [LMN08]. In this regard, one may be interested in only resolving between “no photons” or “one or more photons”, an apparatus known as *on/off photodetector*, which correspond to jut partitioning the Hilbert space differently, e.g.  $\mathcal{M}_{on/off} = \{|0\rangle\langle 0|, \mathbb{I} - |0\rangle\langle 0|\}$ , where the second measurement operator is non-gaussian.

Let us now restrict to Gaussian projections. In particular, we have mentioned homodyne detection corresponds to projecting over a quadrature direction in phase space, as defined by the operator  $\hat{q}_\theta = \cos\theta\hat{q} + \sin\theta\hat{p}$ , where  $\hat{q}_\theta |q_\theta\rangle = q_\theta |q_\theta\rangle$  and  $\mathbb{I} = \int_{\mathbb{R}} |q_\theta\rangle\langle q_\theta| dq_\theta$ , and  $p(q_\theta) = \text{Tr}[|q_\theta\rangle\langle q_\theta| \rho]$ . Such measurement can be realized, for instance, by the so-called *balanced* homodyne detection. This consists in mixing  $\rho$  with a local oscillator ( $|\alpha\rangle$  with  $|\alpha| \gg 1$ ) by a balanced beam-splitter ( $\theta = \frac{\pi}{4}$  and  $\psi = 0$  in Eq 2.79), and subtracting the detected intensities at the two outputs of the beam-splitter (each measured by a photodetector). An alternative way to realize a homodyne detection is the *direct* homodyne detection, which consists on mixing the incoming signal with a local oscillator by a low-reflectivity BS (which transmits most of the original signal, adding only a small amount of the local oscillator) and measuring the intensity of the reflected port.

On the other hand, we can consider heterodyne measurements, which in ac-

cordance to Eq. (2.95) are projective measurements on coherent state of varying intensity and phase. They can be realized by combining the state  $\rho$  with a vacuum into a balanced beam splitter, and homodyning in the direction of  $\hat{q}$  and  $\hat{p}$  (e.g.  $\theta = 0$  and  $\theta = \frac{\pi}{2}$  in Eq. 2.78 respectively) in the paragraph of above); in this regard heterodyne measurements correspond to joint measurements of position and momenta, and its probability measurement outcome is  $p(\alpha) = \langle \alpha | \rho | \alpha \rangle \pi^{-1}$ .

Such a mechanism of implementing a measurement by homodyning and having access to ancillas can actually be generalized to any Gaussian operation via Gaussian ancillas, Gaussian unitary interactions and homodyning/discarding the reduced state of the ancilla system [GIC02].

## 2.4 Continuously-monitored systems

Our focus in this Section will be put on continuously-monitored quantum systems. Opposite to the previously discussed single-shot scenarios, here we are interested in measuring the system repeatedly, which in the continuous limit leads to a *measurement signal*. Here, we will give an introduction to the topic, and the interested reader is encouraged to complement our discussion with Refs. [WM09; Jac14; GC85; DJ99; WM93a].

The most direct way to implement a continuous measurement on a quantum system is to couple it an electromagnetic auxiliary mode that is subsequently measured via photon-counting or homodyning measurements as shown in Fig. 2.5. While several physical scenarios allow for this kind of detection — among them atomic sensors [JMnT+18] and mesoscopic electrometers using quantum dots [LJP+03]—, we will focus on optomechanical cavities [AKM14].

To this end, we begin in Sec. 2.4.1 by providing some insight on optical cavities and their interplay with measurements done on the light that leaks out of the cavity. Since such measurement outcomes are stochastic, the evolution of the system will be given by a *stochastic master equation* (SME). Thus, in Sec. 2.4.2 we derive an SME for the case of a single optical mode stored in the cavity and photon-detection done on the outside modes, leading to a SME driven by point processes, e.g. the measurement record is a sequence of photon clicks. We then consider homodyne detection in Sec. 2.4.3, where the measurement signal becomes continuous, and a Wiener noise appears in the play. Next, a general equation for a continuously-monitored Gaussian system is discussed in Sec. 2.4.6. Finally, in Sec. 2.4.7 we introduce the optomechanical model that we consider in Chapter 5, which is used as a testbed for our approaches to the statistical inference problems considered in that Chapter.

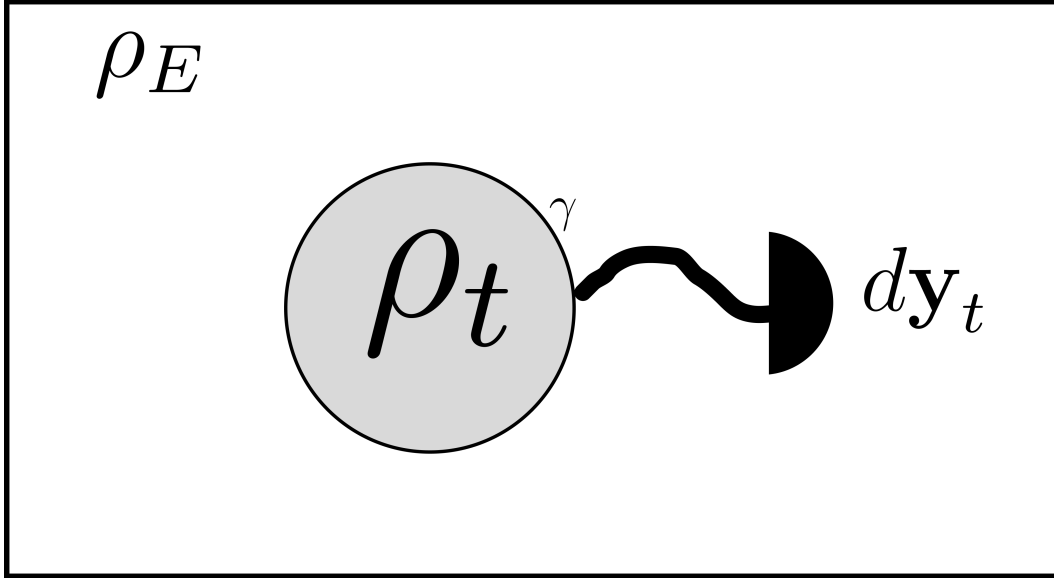


Figure 2.5: We depict the continuously-monitored setting under consideration. This consists on an optomechanical cavity that stores a quantum system  $\rho_t$ , and whose decay rate is  $\gamma$ , and is coupled to a bosonic bath  $\rho_E$ , the latter being constantly measured. This gives rise to the measurement signal, whose (stochastic) value at time  $t$  is denoted by  $dy_t$ .

### 2.4.1 Cavity emission and photon-detection

Optical cavities consist on convex mirrors faced together at some distance, where light can be stored in a standing wave mode for long periods of time. In order to access the cavity mode, one of the mirrors can be made slightly transmissive, as in Fig. 2.6. This allows to retrieve information by measuring the leaked photons and to pump the cavity field by an external laser. Since light is trapped inside the cavity for long periods of times, placing a quantum system inside it, is one of the most effective means of coupling quantum systems to light.

The leaked light occupies travelling modes (free propagating), which means that such photons will never interact again with the system. This fact allows one to describe the evolution of the cavity mode as a Markovian dynamics. This corresponds to an electromagnetic (bosonic) cavity mode coupled to a bosonic bath, which in turn consists in the free-propagating modes. For optical modes, this bath is taken to be at zero temperature, *i.e.* the vacuum state, but the formalism also allows us to consider non-zero temperatures, that are present for example in the baths of optomechanical systems.

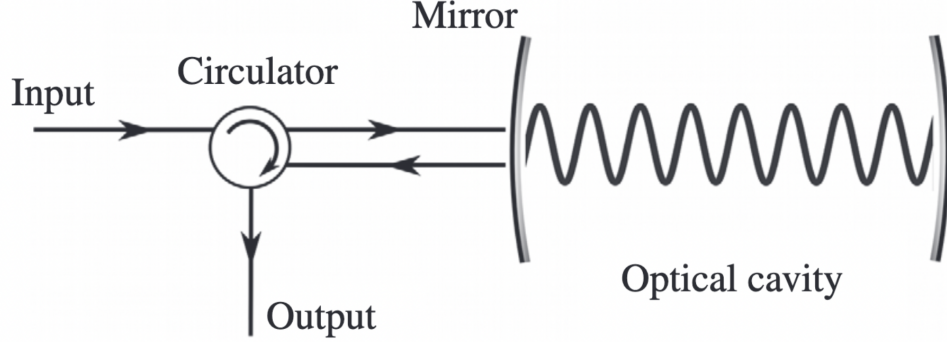


Figure 2.6: An optical cavity is depicted (image taken from Ref. [Jac14], Ch.3). Here, light is input (and output) to (and from) the cavity by the same mirror, and such two modes are splitted by the circulator.

We denote the cavity's decay rate by  $\gamma$  (depending on the transmissivity of the mirror), the annihilation operator for the cavity mode by  $a$ , and the anhilliation operator at detection time  $t$ , by  $b_t$ . By approximating the bath's autocorrelation function as a delta-function, which makes the bath Markovian<sup>12</sup> we obtain the following commutation for the bosonic bath's operators in the interaction picture w.r.t. system's Hamiltonian:

$$[b_t, b_{t'}^\dagger] = \delta(t - t'). \quad (2.106)$$

Under the rotating-wave approximation, the interaction potential between cavity and bath is modelled to be

$$V_{IF} = -i\sqrt{\gamma}(b_t a^\dagger - b_t^\dagger a), \quad (2.107)$$

and describes the exchange of excitations between each other. However, due to the singularity appearing in Eq. 2.106, we must be careful when studying the dynamics generated by  $V_{IF}$ . A convinient approach is that of interpreting  $b_t$  a la Ito [WM93a]<sup>13</sup>. Thus, an infinitesimal operator is defined as

$$dB_t = b_t dt, \quad (2.108)$$

<sup>12</sup>This matter is discussed further in Sec.3.3. and Sec.3.11 of Ref. [WM09]

<sup>13</sup>Ito calculus is a tool of stochastic calculus, that allows to compactly write differential equations for random variables subject to Wiener noise, and also integrals. While working with Ito calculus, one proceeds similarly than traditional calculus but taking into account a few important rules. We briefly discuss this topic in Sec. 2.4.4.

where  $dt$  accounts for an infinitesimal time-interval leading to  $\delta(0)dt = 1$ . With the new operators, we then get the following commutation relations:

$$[dB_t, dB_t^\dagger] = dt, \quad (2.109)$$

which indicates that the operator  $dB_t$  can be understood as being order  $\sqrt{dt}$ . With the re-defined bath's operators, we can now study how an infinitesimal evolution looks like. Here, we must be careful when performing expansions, since the  $\sqrt{dt}$  scaling requires to go up to second order terms. To do this, we consider the infinitesimal evolution generated by  $V_{IF}$  in Eq. 2.107, and we get

$$\begin{aligned} U(t + dt, t) &= e^{\sqrt{\gamma}(a dB_t^\dagger - a^\dagger dB_t)} \\ &= \mathbb{I} + \sqrt{\gamma}(a dB_t^\dagger - a^\dagger dB_t) - \frac{1}{2}\gamma a^\dagger a dt - \frac{\gamma}{2}\{a^\dagger, a\}dB_t^\dagger dB_t \\ &\quad + \frac{\gamma}{2}(a^2 dB_t^{\dagger 2} + a^{\dagger 2} dB_t^2), \end{aligned} \quad (2.110)$$

which describes an infinitesimal evolution of the joint system, in the interaction picture.

We will now assume that the bosonic bath is always found in the ground-state  $|0\rangle$ , and that at each time-step  $t$ , the system interacts with a new copy of such ground-state, *i.e.* with a new mode  $b_t$ . Let us also assume that system's state is given by  $|\psi(t)\rangle$ . With this, the global state at time  $t + dt$  is given by

$$U(t + dt, t) |0\rangle |\psi(t)\rangle = \left( \mathbb{I} - \frac{\gamma}{2} a^\dagger a dt \right) |0\rangle |\psi(t)\rangle + \gamma dB_t^\dagger |0\rangle a |\psi(t)\rangle. \quad (2.111)$$

Under these assumptions, the probability of finding an excitation in the bath is given by

$$p_1(dt) = \gamma \langle 0 | dB_t dB_t^\dagger | 0 \rangle \langle \psi(t) | a^\dagger a | \psi(t) \rangle, \quad (2.112)$$

which turns out to be small if we recall that  $dB_t dB_t^\dagger = dt$ .<sup>14</sup>

We also note that if we *do measure* the state of the leaked mode, and detect a photon, the state of the cavity becomes  $|\psi_1(t + dt)\rangle \propto a |\psi(t)\rangle$  (a photon-subtracted state), while if no photon is detected  $|\psi_0(t + dt)\rangle \propto (\mathbb{I} + \frac{\gamma}{2} a^\dagger a) |\psi(t)\rangle$ . In the next section we formalize this (stochastic) conditional state dynamics using a sequence of quantum instruments applied one infinitesimal step after the other.

---

<sup>14</sup>This relationship holds since we only keep the non-normally ordered operators in bath's commutation relations, as such is assumed to be in the ground-state

### 2.4.2 Continuous photon-detection

To begin our study of conditional post-measurement states, we will consider the photon-detection example outlined in the previous section, but we will also include the action of the system's hamiltonian. By direct identification of the post-measurement (unnormalized) conditional states are given by  $|\psi_0(t + dt)\rangle = M_0 |\psi(t)\rangle$  and  $|\psi_1(t + dt)\rangle = M_1 |\psi(t)\rangle$  where the two Krauss operators are:

$$M_0 = \mathbb{I} - \left( \frac{R}{2} + iH \right) dt \quad (2.113)$$

$$M_1 = c\sqrt{dt}, \quad (2.114)$$

where  $H$  stands for the system's Hamiltonian<sup>15</sup>,  $dt$  is an infinitesimal amount of time,  $c = \sqrt{\gamma}a$ , with  $\gamma$  being the cavity's decay rate and  $R = c^\dagger c$ .

By dropping terms of  $\mathcal{O}(dt^2)$ , it is straightforward to check that  $\{M_0, M_1\}$  constitute a valid quantum instrument ( $M_0^\dagger M_0 + M_1^\dagger M_1 = \mathbb{I}$ ), and the unconditional post-measurement state, *e.g.* the state of the system at time  $t + dt$  when not recording the measurement outcome, recovers the Libland form in Eq. 2.32 for the cavity emission model:

$$\rho(t + dt) = M_0 \rho M_0^\dagger + M_1 \rho M_1^\dagger \quad (2.115)$$

$$= \rho - i[H, \rho] dt + c^\dagger \rho c - \frac{1}{2} \{c^\dagger c, \rho\}, \quad (2.116)$$

If we do keep track of the measurement record, at each (infinitesimal) time step we can get a detection “click” (outcome  $i = 1$ ) with probability

$$p_1(dt) = \text{Tr}[\rho M_1^\dagger M_1] = \text{Tr}[\rho c^\dagger c] dt. \quad (2.117)$$

and the absence of a click (outcome  $i = 0$ ) with probability

$$p_0(dt) = \text{Tr}[\rho M_0^\dagger M_0] = 1 - \text{Tr}[\rho c^\dagger c] dt. \quad (2.118)$$

---

<sup>15</sup>We will drop the hat notation in this Section.

and the corresponding normalized states will be given by 2.113 we have

$$|\psi_1(t + dt)\rangle = \frac{M_1 |\psi(t)\rangle}{\|M_1 |\psi(t)\rangle\|} = \frac{c_1 |\psi(t)\rangle}{\sqrt{\langle \psi(t) | c^\dagger c | \psi(t) \rangle}}, \quad (2.119)$$

$$|\psi_0(t + dt)\rangle = \frac{M_0 |\psi(t)\rangle}{\|M_0 |\psi(t)\rangle\|} = \frac{1}{\sqrt{\langle M_0^\dagger M_0 \rangle}} \left[ \mathbb{I} - \left( \frac{R}{2} + iH \right) |\psi(t)\rangle \right] \quad (2.120)$$

$$\begin{aligned} &\simeq \left( 1 + \frac{\langle c^\dagger c \rangle}{2} dt \right) \left[ \mathbb{I} - \left( \frac{R}{2} + iH dt \right) \right] \\ &= \left( \mathbb{I} - dt \left[ iH + \frac{R - \langle R \rangle}{2} \right] \right) |\psi\rangle, \end{aligned} \quad (2.121)$$

where we have Taylor-expanded the normalization term in  $|\psi_0(t + dt)\rangle$  up to the first order, and dismissed terms  $\mathcal{O}(dt^2)$ . We that while most of the time the system will evolve continuously with  $(M_0 = \mathbb{I} + \mathcal{O}(dt))$ , with very small probability,  $p_1 = \mathcal{O}(dt)$ , the state will suffer a very drastic (finite) change, the so-called *quantum jump*. Note that in both cases, upon a click or a no-click, the act of measuring has always a *back-action* on the system dynamics.

Now we wish to express the above results describing the stochastic evolution of the system in a single stochastic differential equation. For this purpose, we define the stochastic variable  $N(t)$ , corresponding to the total number of clicks until time  $t$ , *i.e.* the number of occurrences of measurement outcome  $i = 1$  registered up to time  $t$ . Since such quantity can only increase by a single unit, the following properties hold:

$$dN_t^2 = dN_t \quad (2.122)$$

$$\mathbb{E}[dN_t] = \text{Tr}[M_1^\dagger(t)M_1(t)\rho(t)] = dt \langle \psi(t) | c^\dagger c | \psi(t) \rangle, \quad (2.123)$$

A variable  $N(t)$  with these properties defines what is called a *point process* [Dol73; WM09]. Since such variable  $dN$  can only take the binary values of zero and one, we can compactly write the conditional evolution of the (normalized) pure state as

$$|\psi(t + dt)\rangle = (1 - dN) \left[ \mathbb{I} - \left( iH + \frac{c^\dagger c - \langle c^\dagger c \rangle}{2} \right) dt \right] |\psi(t)\rangle + dN \frac{c |\psi\rangle}{\sqrt{\langle c^\dagger c \rangle}}, \quad (2.124)$$

where the time-dependence of the variable  $N$  will be left implicit from now on. This a non-linear, stochastic equation, known as a *Stochastic Schrödinger Equation* since it preserves the purity of the state. Any solution to this equation defines a quantum trajectory, and describes the possible paths that the quantum state can



take under the stochastic outcomes obtained from the point process. By dropping the terms of order  $dNdt$ <sup>16</sup>, we get

$$|\psi(t + dt)\rangle = |\psi\rangle + dN\left(\frac{c}{\sqrt{\langle c^\dagger c \rangle}} - 1\right)|\psi\rangle + \frac{1}{2}dt\left(c^\dagger c - \langle c^\dagger c \rangle - 2iH\right)|\psi\rangle. \quad (2.125)$$

We can now extend this to density matrices by witting  $\rho(t) = |\psi(t)\rangle\langle\psi(t)|$  and dropping high order terms in  $dt$ :

$$d\rho = dN\left(\frac{c\rho c^\dagger}{\langle c^\dagger c \rangle}\right) - dt\left(i[H, \rho] + \frac{1}{2}\{c^\dagger c, \rho\} - \frac{1}{2}\langle c^\dagger c \rangle\rho\right). \quad (2.126)$$

which by linearity extends to arbitrary mixed initial states. Note that averaging this equation out over all possible measurement outcomes and recalling  $\mathbb{E}[dN_t] = \langle c^\dagger c \rangle$  one recovers again the Lindbladian dynamics.

This can also be written as

$$d\rho(t) = dN\mathcal{G}[c]\rho(t) - dt\mathcal{H}\left[iH + \frac{c^\dagger c}{2}\right]\rho(t), \quad (2.127)$$

in terms of the super-operators  $\mathcal{G}$  and  $\mathcal{H}$  as

$$\mathcal{G}[c]\rho = \frac{c\rho c^\dagger}{\langle c^\dagger c \rangle} \quad (2.128)$$

$$\mathcal{H}[c]\rho = c\rho + \rho c^\dagger - \text{Tr}[\rho(c + c^\dagger)]\rho \quad (2.129)$$

We note that one could easily extend this approach to include consider more *jump operators*,  $\{c_i\}_{i=1}^M$  by taking  $M_i = c_i\sqrt{dt}$ ,  $M_0 = \mathbb{I} - (\frac{R}{2} + iH)dt$  with  $R = \sum_{i=1}^M c_i^\dagger c_i$ . For example, one could include losses the other mirror. Most interestingly, this approach lets us easily describe situations in which only some types of jumps are monitored or accounted for. One then obtains a stochastic equation with jump terms of the form  $G[c]$  for the monitored modes and the averaged Lindbladian counterpart for the unobserved modes.

In this section we have studied the conditional evolution of system under the particular measurement choice in Eq. 2.113 (corresponding to photon-counting), and we have obtained the stochastic equation describing the continuously monitored system. As discussed in the introduction to quantum channels and quantum instruments, for a fixed interaction between the system and the auxiliary system, different choices of measurement on the auxiliary system lead to different Kraus

<sup>16</sup>this can be justified by the fact that the probability of  $dN = 1$  is of order  $dt$

representations of the same quantum map. In the next section we will study the conditional dynamics when one performs a different a different type of measurement on the auxiliary (bath) modes. Of course, this will lead to a completely different type of stochastic equation, but if we average over all its realization we will recover the same Lindblad equations.

The different stochastic differential equations corresponding to the same average open system Lindbladian equations are called *unravelings*. We remark here that while in many works these unravellings are presented as just convenient interpretations or as a numerical tool to simulate (Montecarlo) open system dynamics —in many occasions it is much simpler to propagate the pure state, than the full density matrix—, in this thesis the particular unravelling will have a direct physical meaning and we will be concerned about inference methods based the single-trajectories, single (possibly long) measurement records rather than average quantities of the state.

### 2.4.3 Continuous homodyne detection

We have previously study an stochastic master equation for an specific *unraveling* of unconditional dynamics, namely photon-detection. In such case, a detection induces a *quantum jump* on the system's density matrix, with a probability that is proportional to  $dt$ , as given by Eq. 2.122. Here, we will study a different unraveling of the dynamics.

The existence of different unravelings arises from the invariance of the master equation — *i.e.* the undonditional open quantum system dynamics revisited in Sec. 2.1.5 — under certain transformations (see also Eq. 2.33) of the form

$$c \rightarrow c + \alpha \tag{2.130}$$

$$H \rightarrow H - \frac{i}{2}(\bar{\alpha}c - \alpha c^\dagger), \tag{2.131}$$

with  $\alpha \in \mathbb{C}$ . This is (unitarily equivalent) to a different set of Krauss representation:

$$M_0 \rightarrow \tilde{M}_0(t) = \mathbb{I} - dt \left( iH + \frac{\bar{\alpha}c - \alpha c^\dagger - (c + \alpha)^\dagger(c + \alpha)}{2} \right) \tag{2.132}$$

$$M_1 \rightarrow \tilde{M}_1(t) = \sqrt{dt}(c + \alpha). \tag{2.133}$$

This transformation can physically be achieved by mixing the input field with a strong local oscillator (L.O.) by a low transmissivity beam-splitter (BS)<sup>17</sup>, as shown

---

<sup>17</sup>To see this, let the transmissivity of the BS be  $\eta \sim 1$ , and let the intensity of the L.O. be  $\frac{\alpha}{\sqrt{1-\eta}}$ ;

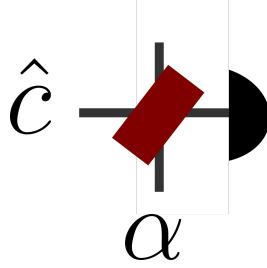


Figure 2.7: We sketch a direct homodyne detection. The beam-splitter is chosen such that the reflected part (which is subsequently measured) has only a small contribution of the local oscillator, and is dominated by the original signal associated to the system we are measuring.

in Fig. 2.7, which corresponds to direct homodyne detection. In turn, let us restrict to  $\alpha \in \mathbb{R}$ , then a detection (as defined in the previous section) happens with probability

$$p_1(dt) = \mathbb{E}[dN] = \text{Tr}[\rho(c + \alpha)^\dagger(c + \alpha)]dt \quad (2.135)$$

$$= \text{Tr}[\rho(c^\dagger c + \alpha(c + c^\dagger) + \alpha^2)] \quad (2.136)$$

$$= \text{Tr}[\rho(c^\dagger c + \alpha q + \alpha^2)], \quad (2.137)$$

where  $q = c + c^\dagger$  stands for the position quadrature<sup>18</sup>. Thus, if  $\alpha \gg \langle c^\dagger c \rangle$ , photon-detection at the reflected port is equivalent (up to a constant) to measuring the  $q$  quadrature. This corresponds to the case of direct *homodyne* detection, which was introduced in Sec. 2.3.4; we will here consider the direct homodyne detection, which is realized by mixing the system with a strong local oscillator by a low transmissivity beam-splitter, as shown in Fig. 2.7.

Repeating the same procedure outlined by the end Sec. 2.4.2, but now with the transformed set of Krauss operators  $\{\tilde{M}_0(t), \tilde{M}_1(t)\}$ , we get a stochastic master equation that reads

$$d\rho = \mathcal{G}[c + \alpha]\rho dN + dt\mathcal{H}\left[iH + \alpha c + \frac{c^\dagger c}{2}\right]. \quad (2.138)$$

the reflected mode will then be

$$\sqrt{\eta}c + \alpha, \quad (2.134)$$

assuming we treat the local oscillator classically, an approximation valid when the intensity of the L.O. is sufficiently high [Ser17a]. From here, since  $\sqrt{\eta} \sim 1$ , we get the desired transformation.

<sup>18</sup>We will momentarily drop the  $\sqrt{2}$  factor to ease the notation

From here, we aim to take the continuous limit, in which the photocurrent is approximated by a continuous function of time. This corresponds to the ideal limit of homodyne detection, in which the intensity of the local oscillator goes to infinity. In particular, we are interested in a regime under which the number of detections per time is high and changes in the system are small. In such regime, we obtain a *Belavkin-Zakai* equation, which is yet another quantum stochastic master equation, that describes the continuous homodyne-monitoring of a quantum system. We will now derive such equation by making expansions in terms of the local oscillator intensity, following Ref. [WM09]<sup>19</sup>.

We begin by considering a small time-interval  $\delta t = \mathcal{O}(\alpha^{-\frac{3}{2}}) \ll 1$ <sup>20</sup>, that results in an expected number of detections,

$$\mathbb{E}[\delta N] = \delta t \operatorname{Tr}[\rho(\alpha^2 + \alpha q + c^\dagger c)] = \mathcal{O}(\alpha^{\frac{1}{2}}), \quad (2.139)$$

which is large in comparison to system's evolution (since  $\delta t = \mathcal{O}(\alpha^{-\frac{3}{2}})$ ). Under this choice of scaling, the detection probability is also large  $p_1(\delta t) = \mathcal{O}(\sqrt{\alpha})$ .

As pointed out in Sec. 2.3.4,  $\delta N$  is a Poissonian random variable. However, it is possible — since the number of detections happening per  $\delta t$  is large — to approximate such probability distribution by a normal distribution [WM09; WM93b] (intuitively we can understand this as a central-limit theorem) with a mean-value given by Eq. 2.139 and variance

$$\sigma^2 = \delta t(\alpha^2 + \mathcal{O}(\alpha^{\frac{3}{2}})). \quad (2.140)$$

By recurring to Ito calculus, Eq. (2.140) can be written compactly as

$$\delta N = \alpha^2(\delta t) \left[ 1 + \frac{\langle q \rangle(t)}{\alpha} \right] + \alpha \delta W, \quad (2.141)$$

where  $\delta W$  describes a differential white noise, e.g. a zero-mean normal random variable whose variance is  $\mathbb{E}[\delta W^2] = \delta t$ .

It is now time to make a brief pause, and jump to revise Ito calculus. The interested reader is referred to Ref. [Gar04], where further details about this topic can be found.

---

<sup>19</sup>Note that there are alternative ways to derive this equation. For instance by straightforwardly considering the action of a Gaussian measurement on the quantum state and Taylor-expanding it in terms of  $dt$  [Jac14; JS06], or considering the interaction between a Gaussian system, bath and measurement [Ser17b; GLS16], although this formalism is focused on Gaussian systems, and directly leads to the set of stochastic linear equations for mean and covariance of the quantum state, which we discuss later on, bypassing stochastic master equations.

<sup>20</sup>The only purpose of considering  $\delta$  rather than  $d$  is that we take a limit in terms of local oscillator intensity by the end of this discussion

### 2.4.4 Intermezzo II: Ito calculus

In a nutshell, Ito provides a tool to tackle problems involving stochastic differential equations, where the infinitesimal change  $dX_t$  of a variable  $X_t$  occurring in an infinitesimal time  $dt$  interval is written in terms of deterministic part (proportional to  $dt$ ) and a stochastic infinitesimal increment  $dW$ . Here we will focus on processes driven by Wiener noise and Ito calculus will give us tools to make sense of and compute integrals such as  $\int X_t dW$ .

A Wiener process  $W(t) = \int_0^t dW$  describes stochastic curve with zero mean and a variance that grows as  $t$ , and is defined in such a way that it is *continuous* in time. The infinitesimal Wiener increments  $dW$  have zero mean and variance  $\mathbb{E}(dW^2) = dt$ , but in order to guarantee that  $W(t)$  is a continuous function, the increments have the (surprising) property that  $dW^2 = dt$ , *i.e.* it is deterministic (the distribution of  $\Delta W^2$  become more and more peaked as the time intervals  $\Delta t$  smaller). A Wiener process can be thought of as the continuum limit of a random walk or brownian motion

Consider the general (1D) Ito stochastic differential equation

$$dX_t = a(X_t, t)X_t dt + b(X_t, t)dW_t. \quad (2.142)$$

While the coefficients  $a(X_t, t)$  and  $b(X_t, t)$  might be arbitrary functions, in case they are state-independent — known as a drift-diffusion process — an integral expression can be defined as

$$X_t = \int_{t_0}^t a(t')dt' + \int_{t_0}^t b(t')dW_{t'}, \quad (2.143)$$

where the second integral is an *stochastic Ito integral* (whose value depends on the specific realization of the Wiener process up to time  $t$ ).

Contrary to Riemman integral, when dealing with stochastic integrals it is of utmost importance to choose the value at which to evaluate the function for on each partition, due to the stochastic nature of the resulting function. Ito integral is defined by considering the value that the function takes at the beginning of each partition. For example, in a different stochastic calculus named by Stratonovich, the integral is defined by averaging up the values at the two extremes of the interval, and the rules of calculus are completely different. In this thesis, we will work with Ito integrals.

Here, we observe that  $X_t$  clearly does *not* depend on the *future* realisations of the Wiener process (that is, on values  $dW_s$  for  $s > t$ ); this captures the notion of a *non-anticipating function of  $t$* , which we denote by  $NA(t)$ . That is, a stochastic

function  $G(t)$  is  $NA(t)$  if  $\forall t$  and  $\forall s$  such that for  $t < s$ , the values that  $G(t)$  take are statistically independent of the future increments  $W(s) - W(t)$ , with  $W(T) = \int_0^T dW_t$ .

Given  $G(t)$  and  $H(t)$  non-anticipating functions, and when integration *a la Ito*, the following holds:

1.  $\langle \int_{t_0}^t G(t') dW_{t'} \rangle = 0$ , which can be seen by the construction of Ito's integral.
2.  $\langle \int_{t_0}^t G(t') dW_{t'} \left[ \int_{t_0}^{t'} H(s') dW_{s'} \right] \rangle = \int_{t_0}^t \langle G(t') H(t') \rangle dt'$ , which can be understood as a correlator formula, and where  $\langle dW_t dW_s \rangle = \delta_{t,s} dt$  rules the stochastic integration out.
3.  $\langle \int_{t_0}^t G(t') dW_{t'} \left[ \int_{t_0}^{t'} H(s') dW_{s'} \right] \rangle = 0$ , which can be understood from the fact that  $\tilde{H}(t') = \int_{t_0}^{t'} H(s') dW_{s'}$  is  $NA(t')$  and thus the function  $G(t') H(t')$  is  $NA(t')$ , whose stochastic integral averages to zero (see first item).
4. Let  $\tilde{G}$  and  $\tilde{H}$  be deterministic functions of time<sup>21</sup>, then

$$\left\langle \int_{t_0}^t \tilde{G}(t') dW_{t'} \int_{t_0}^s \tilde{H}(s') dW_{s'} \right\rangle = \int_{t_0}^{\min(t,s)} dt' \tilde{G}(t') \tilde{H}(t'). \quad (2.144)$$

Where we use the  $\langle \cdot \rangle$  notation instead  $\mathbb{E}(\cdot)$  for the expected values whenever there's no ambiguity with the expectation over quantum states.

When considering the derivative of a (twice-differentiable) function  $f(X_t, t)$ , one has to be carefull in not using blindly the chain rule. One needs to take Taylor expand  $f$  up to second-order, taking into account that  $dW \sim O(\sqrt{dt})$ :

$$df = \left( \frac{\partial f}{\partial t} + a(t) \frac{\partial f}{\partial x} + \frac{b^2(t)}{2} \frac{\partial^2 f}{\partial x^2} \right) dt + b(t) \frac{\partial f}{\partial x} dW_t. \quad (2.145)$$

The latter formula is known as the *Ito lemma*, and we can understand it as a way to take into account a change of variables when studying the dynamics of a stochastic system.

When the drift  $a(t)$  and diffusion  $b(t)$  coefficients in Eq. 2.142 are constant, the process is known as an *Orstein-Uhlenbeck process*:

$$dX_t = \kappa(\mu - X_t)dt + \sigma dW_t, \quad (2.146)$$

---

<sup>21</sup>This expression finds use when considering, for example,  $\tilde{G}$  and  $\tilde{H}$  to be diffusion coefficients of respectively different stochastic processes.

whose solution is easily obtained by a change of variables  $Y_t = X_t e^{\kappa t}$ . Using Ito's lemma:

$$dY = \kappa X_t e^{\kappa t} dt + e^{\kappa t} dX_t = \kappa \mu e^{\kappa t} dt + \sigma e^{\kappa t} dW_t$$

from where

$$Y_t = Y_0 + \int_0^t \kappa \mu e^{\kappa t'} dt' + \int_0^t \sigma e^{\kappa t'} dW_{t'},$$

and using  $X_t = Y_t e^{-\kappa t}$ :

$$X_t = X_0 + \mu(1 - e^{-\kappa t}) + \sigma \int_0^t e^{-\kappa(t-t')} dW_{t'} \quad (2.147)$$

we get

$$X_t = \mu(1 - e^{-\kappa t}) + \sqrt{D} \int_0^t e^{-\kappa(t-t')} dW_{t'}. \quad (2.148)$$

Using the aforementioned rules, we can readily check that, for  $t \rightarrow \infty$ ,  $\mu = \mathbb{E}[X_t]$  and  $\text{Var}[X_t] = Dt$ , which indicates that the distribution diffuses in time; in turn, this solution corresponds to a Gaussian probability distribution. The mean and variance of  $X_t$  can now be readily computed using the integration rules above,

$$\langle X_t \rangle = X_0 e^{-\kappa t} + \mu(1 - e^{-\kappa t}) \xrightarrow{t \rightarrow \infty} \mu \quad (2.149)$$

$$\text{Var}[X_t] = \frac{\sigma^2}{2\kappa} (1 - e^{-2\kappa t}) \xrightarrow{t \rightarrow \infty} \frac{\sigma^2}{2\kappa} \quad (2.150)$$

That is, unlike the Wiener process, the mean and variance converge to constant values for large times. Moreover, since the process is Gaussian, the probability distribution of  $X_t$  at a given time is a Gaussian with the mean and variance above. In turn, the probability of whole trajectory is fully determined by the first and second moments (including  $\langle X_t X_{t'} \rangle$ , which can also be readily computed).

The connection between SDE and probability distribution of the variable  $X_t$  at a given time can be established for more general processes. In turn, considering any function  $f(X_t)$  where the dynamics of  $X_t$  is given by Eq. (2.142), we have that

$$\frac{\langle df(X_t) \rangle}{dt} = \left\langle \frac{df(X_t)}{dt} \right\rangle = \frac{d}{dt} \langle f(X_t) \rangle \quad (2.151)$$

$$= \langle a(X_t, t) \partial_x f + \frac{1}{2} b(X_t, t)^2 \partial_x^2 f \rangle \quad (2.152)$$

Considering that  $X_t$  has a conditional probability density  $p(X_t, t|X_0, t_0)$  then

$$\frac{d}{dt} \langle f(X_t) \rangle = \int dX f(X) \partial_t p(X, t|X_0, t_0) \quad (2.153)$$

$$= \int dX \left[ a(X_t, t) \partial_X f + \frac{1}{2} b(X_t, t)^2 \partial_X^2 f \right] p(X_t, t|X_0, t_0). \quad (2.154)$$

Integrating by parts and using the fact that  $f(x)$  is arbitrary, we arrive to the *Fokker-Planck* equation

$$\partial_t p(X_t, t|X_0, t_0) = -\partial_X [a(X, t)p(X, t|X_0, t_0)] + \frac{1}{2} \partial_X^2 [b(X, t)^2 p(X, t|X_0, t_0)], \quad (2.155)$$

resulting in an analogous description of the process in terms of an differential equation for its probability distribution.

*Fin del intermezzo.*

Recalling that the evolution of cavity's system under the transformed measurement operators of Eq. (2.132) reads

$$\delta \rho = \delta N(t) \mathcal{G}[c + \alpha] \rho - \delta t \mathcal{H} \left[ iH + \alpha c + \frac{c^\dagger c}{2} \right] \rho, \quad (2.156)$$

we can now work the  $\mathcal{G}[c + \alpha]$  term and expand it in powers of  $\alpha^{-1}$ ; this means keeping the terms up to  $\frac{1}{\alpha^2}$  arising by Taylor expanding up to the second order the normalization in  $\mathcal{G}$  (see Eq. 2.128). After such expansion, we obtain

$$\delta \rho = \delta N(t) \left[ \frac{\mathcal{H}[c]}{\alpha} + \frac{\mathcal{G}[c] - \langle q \rangle \mathcal{H}[c]}{\langle c^\dagger c \rangle} \right] \rho - \delta t \mathcal{H} \left[ iH + \alpha c + \frac{c^\dagger c}{2} \right] \rho \quad (2.157)$$

Now, plugging the Ito form of  $\delta N$  into Eq. 2.157, we get

$$\delta \rho = \left( \alpha^2 \delta t \left[ 1 + \frac{\langle q \rangle}{\alpha} \right] + \alpha \delta W \right) \left( \frac{\mathcal{H}[c]}{\alpha} + \frac{\mathcal{G}[c] \langle c^\dagger c \rangle - \langle q \rangle}{\alpha^2} \right) \rho + \delta t \mathcal{H} \left[ -iH - \alpha c - \frac{c^\dagger c}{2} \right] \rho.$$

From here, we can expand all the products on the above equation and keep terms of order  $\frac{1}{\sqrt{\alpha}}$  or higher. Finally, taking the limit  $\alpha \rightarrow \infty$  we get the desired Belavkin-Zakai equation for homodyne measurement:

$$d\rho = dt \left( -i [H, \rho] + \mathcal{D}[c] \rho \right) + \mathcal{H}[c] \rho dW \quad (2.158)$$

$$\mathcal{D}[c] \rho = c \rho c^\dagger - \left\{ \frac{c^\dagger c}{2}, \rho \right\} \quad (2.159)$$

$$\mathcal{H}[c] \rho = c \rho + \rho c^\dagger - \text{Tr}[(c + c^\dagger) \rho],$$



where to facilitate reading we included the definitions of the super-operators  $\mathcal{D}[c]$  and  $\mathcal{H}[c]$  introduced before.

The last equation describes the evolution of a Markovian system which is subject to a continuous monitoring under homodyne detection. As a consequence, the so-called *back-action* term arises (the one accompanying the Wiener noise), which takes into account the effect measurement result recording. Similar to photon-detection situation, when averaging out over possible measurement outcomes, the Libland form is recovered.

On the other hand, a measurement *signal* arises, which is also a continuous function of time. To see this, let us return to the Ito expression for  $\delta N(t)$ , namely

$$\delta N(t) = \alpha^2 \delta t \left[ 1 + \frac{\langle q \rangle (t)}{\alpha} \right] + \alpha \delta W, \quad (2.160)$$

which after subtracting the intensity of the local oscillator we get

$$dy(t) = \lim_{\alpha \rightarrow \infty} \frac{\delta N(t) - \alpha^2 \delta t}{\alpha} = \langle q \rangle dt + dW, \quad (2.161)$$

where clearly  $\langle q \rangle$  is time-dependent and also the Wiener noise. As we will later see one can use this equation to replace de Wiener term in stochastic equation (2.158) by  $dW = dy - \langle q \rangle dt$  to obtain the conditional quantum state of the system given a homodyne measurement record.

Overall, we have seen one gets very different kind of information about the cavity mode from the information that has leaked from it, when we mix the leaked light with a local oscillator before the detector instead of measuring leaked light directly with a photo-detector (in-phase quadrature vs. photon number). We also note that the nature of the measured signal is very different: for homodyning we obtain essentially a continuous noisy signal while for photo-counting one gets a sequence of sparse sequence of “clicks”. Similarly, the stochastic equation describing the system is of a very different nature: in the first case it’s a smooth diffusion-type equation, while in the later exhibits abrupt quantum jumps. Of course, both lead to the same master equation when the leaked photons are traced out (or equivalently when all conditional dynamics is averaged out).

In the next Section we briefly explain how to take into account detection inefficiencies.

### 2.4.5 Imperfect detection

First, the case of inefficient detection, in which the efficiency of the photodetector  $\eta$  is less than unity. In this case, the master equation describing cavity emission of

Eq. 2.115 can be understood as

$$d\rho = -i[H, \rho]dt + (1 - \eta)\mathcal{D}[c]\rho dt + \eta\mathcal{D}[c]\rho, \quad (2.162)$$

and only unravel the last term<sup>22</sup>, resulting in a modified stochastic master equation that for the homodyne case Eq. 2.158 is slightly modified to read

$$d\rho = -i[H, \rho]dt + \mathcal{D}[c]\rho dt + \sqrt{\eta}\mathcal{H}[c]\rho dW. \quad (2.163)$$

Moreover, the homodyne measurement signal for the inefficient detection case reads

$$dy = \sqrt{\eta}\langle q \rangle dt + dW. \quad (2.164)$$

The remaining refinement that we consider deals with the bosonic bath, which so far has been assumed to be in the ground-state. However, we might assume that the cavity is in contact with a thermal bath. Deriving stochastic master equations when measuring such thermal bath turns to be an obscure subject [WM09; Jac14]. However, we can assume that the thermal bath is just an extra reservoir which we do not measure. In this case, we recall that the master equation (for the cavity plus thermal bath only) reads:

$$\frac{d\rho}{dt} = -i[H + i(\bar{\beta}c + \beta c^\dagger), \rho] (n_{th} + 1) + \mathcal{D}[c]\rho + n_{th}\mathcal{D}[c^\dagger]\rho, \quad (2.165)$$

where  $n_{th}$  is the average number of photons in the cavity and depends on bath's temperature, and we observe that the mean value of the incoming field,  $\beta dt = \langle dB \rangle$ , has a driving effect on the cavity [WM09]. This driving term induces a constant shift in cavity's quadratures that in the following will be dismissed.

As stressed in Sec. 2.3, measuring system's quadrature belongs to the set of Gaussian measurements, and as such preserves the gaussian character of (an initially gaussian) state. This motivates the following Section, in which we study the quadratures dynamics of a quantum system under continuous homodyne detection.

## 2.4.6 The Gaussian case

In the previous Sections we have studied the dynamics of a cavity when its bath is being continuously-monitored, and also discussed a model for imperfect detections. We will now discuss the dynamics structure when the system is assumed Gaussian.

---

<sup>22</sup>note that  $\mathcal{D}[\sqrt{\eta}c] = \eta\mathcal{D}[c]$ .

We have introduced Gaussian systems in Sec. 2.3.1, and also studied how a Gaussian map modifies the first moments of a Gaussian state in Sec. 2.3.3. Opposite to such Section, we are here interested in studying the *conditional* diffusive dynamics of the Gaussian system, arising as a consequence of continuously-monitoring the bath coupled to it.

Our discussion follows *the Wiseman approach* [WM09; WD05; WM93a], which consists on deriving dynamical equations for the moments of a Gaussian state *from* the stochastic master equation of Eq. 2.158. However, we remark that for the Gaussian case, the same equations can be derived by straightforwardly working with the moments, and assuming an specific (Gaussian) structure for the interaction between bath's and system's modes. This constitutes the *Serafini approach* one, and while it will not be discussed here, the interested reader can consult Refs. [Ser17a; GLS16].

For the sake of simplicity, we will restrict to a single-mode quantum system. We consider an equation of the form in Eq. 2.158:<sup>23</sup>

$$d\rho = \mathcal{L}_0\rho dt + \mathcal{H}[c]\rho dW_t, \quad (2.166)$$

where  $c$  is an arbitrary system operator ( $c = \sqrt{\gamma}a$  in the cavity model), and  $\mathcal{L}_0$  is the Liblandian describing the non-diffusive open-quantum dynamics (that takes into account interaction with additional modes).

We recall that Gaussian states are generated by Hamiltonians which are at most quadratic in system's quadratures:

$$H = \frac{1}{2}\hat{\mathbf{r}}G\hat{\mathbf{r}} - \hat{\mathbf{r}}^T\Omega Bu(t), \quad (2.167)$$

where  $G$  is the Hamiltonian matrix (*i.e.*  $G$  is real and symmetric)<sup>24</sup> and included a possible time-dependent drive  $u(t)$  that enters in the Hamiltonian as a linear term (*e.g.* a displacement due by a driving force). Also, we recall that  $\Omega$  is the symplectic form defined in Eq. 2.59, and that system's mean and covariance are defined as  $\bar{\mathbf{r}} = \langle \hat{\mathbf{r}} \rangle = \text{Tr}[\rho\hat{\mathbf{r}}]$  and  $\Sigma_{nm} = \langle \Delta\hat{\mathbf{r}}_n\Delta\hat{\mathbf{r}}_m \rangle + \langle \Delta\hat{\mathbf{r}}_m\Delta\hat{\mathbf{r}}_n \rangle$ , with  $\Delta\hat{\mathbf{r}}_i = \hat{\mathbf{r}}_i - \langle \hat{\mathbf{r}}_i \rangle$ , and  $\hat{\mathbf{r}} = (q, p)$ . Moreover, the transformation  $\tilde{C}$  mapping system's quadratures with the operator  $c$  in Eq. (2.166) is defined via  $c = \tilde{C}\hat{\mathbf{r}}$ , which for the cavity model reads  $\tilde{C} = \sqrt{\gamma}\begin{pmatrix} 1, i \end{pmatrix}$ .

<sup>23</sup>Let us remark that while we derived such equation for an optical cavity, it can be used to describe more complex systems, such as quantum dots coupled to nano-mechanical resonators [WM09]. Moreover, by the end of this Section we will consider a slightly more complex model for the cavity, and discuss an scenario where a mechanical mode is also considered.

<sup>24</sup>Note our change in notation with respect to Sec. 2.3, *e.g.*  $G \equiv H$  in Eq. 2.70.

If we only consider the unconditional dynamics, given by the Liblandian  $\mathcal{L}_0$ , and use the canonical commutation relations in Eq. 2.59, we can derive a dynamical equation for the first two moments  $\bar{\mathbf{r}}, \Sigma$ :

$$\begin{aligned} d\bar{\mathbf{r}} &= \left( A\bar{\mathbf{r}} + Bu(t) \right) dt \\ d\Sigma &= \left( A\Sigma + \Sigma A^T + D \right) dt, \end{aligned} \quad (2.168)$$

where  $A = \Omega(G + \text{Im}[C])$  and  $D = \Omega \text{Re}[C] \Omega^T$  and  $C = \tilde{C}^\dagger \tilde{C}$  [WD05]. We note that Eq. 2.168 is just a different parametrization of the Gaussian evolution described in Eq. 2.104. However, if we now consider the *conditional evolution*, that arises from continuously-monitoring via homodyne detection and is captured by Eq. 2.166, an stochastic term is expected to appear in moments' dynamics.

To this end, let us re-write the measurement signal in Eq. 2.164 in terms of system's quadratures<sup>25</sup> as

$$dy_t = C(\bar{\mathbf{r}}dt + C^{-1})d\mathbf{W}_t, \quad (2.169)$$

with  $C = \sqrt{4\eta\gamma} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ . Here, our notation  $dy_t$  accounts for the fact that two components are in principle obtained when monitoring the first moment of the quantum state. For homodyne detection this implies that the second component is uninformative and always zero-valued, a fact also stressed by including the pseudo-inverse of  $C$  in the above equation.

Computing the first moments — *e.g.* by taking the corresponding expectation values in Eq. 2.166, where we take into account the interaction with a thermal bath in  $\mathcal{L}_0$  —, and using Isserlis' theorem to simplify the covariance evolution [JS06], we get the following system of lineal stochastic equations:

$$\begin{aligned} d\bar{\mathbf{r}}_t &= \left( A - \chi(\Sigma_t)C \right) \bar{\mathbf{r}}_t dt + \chi(\Sigma_t) dy_t = A\bar{\mathbf{r}}_t dt + \chi(\Sigma_t) dW_t \\ d\Sigma_t &= A\Sigma_t + \Sigma_t A^T + D - \chi(\Sigma_t)^T \chi(\Sigma_t), \end{aligned} \quad (2.170)$$

where  $\chi(\Sigma) = \Sigma C^T + \Gamma^T$ <sup>26</sup>. Here, we have made explicit the time-dependence of the moments, and also explicitated the *back-action* term in the evolution for the first moment, *i.e.* by including the measurement outcome  $dy_t$  in the evolution, and replacing the innovation  $d\mathbf{W}_t$  by it. We note that the measurement outcome  $dy_t$  is

<sup>25</sup>We will now recover the factor 2 dropped when defining  $q = a + a^\dagger$  in the previous Section

<sup>26</sup>The expression for  $\Gamma$  is not relevant for our discussion and numerics, where it is always zero-valued, but the interested reader is referred to Ref. [WD05].

the only experimentally-accessible object that can be used to update our knowledge of the system's state, *e.g.* we do not have access to the value of the Wiener noises. The evolution for the covariance matrix is known as a Riccati equation, and admits a stationary solution if certain stability conditions, known as Hurwitz conditions, are satisfied [WD05]: the eigenvalues of  $A$  are required to have a positive-semidefinite value. Noticeably, such evolution for the second moment is deterministic.

The equations Eq. 5.13 are mathematically equivalent to the *Kalman-Bucy equations* for solving the classical filtering problem of estimating the hidden-state of a linear dynamic system out of a series of noisy measurements [DJ99]. In turn, we can understand the equations as those governing the evolution of the first two moments of the normalized probability distribution, and obtained through the Bayesian update of a classical linear system, using the information acquired through the noisy measurement. Quantumly, this fact is not surprising, and is a direct consequence of the bayesian structure of the theory.

Finally, the measurement statistics are described by Eq. 2.169. We remark that such equation provides the value of the measurement outcome  $dy_t$  obtained at time  $t$ , and it is not meant to be understood as a differential equation but rather as a compact formula that describes outcomes statistics. Thus, the probability distribution of  $dy_t$  is a Gaussian, centered at  $C\bar{r}_t dt$ , and with a variance proportional to  $dt$ :

$$p(dy_t|\bar{r}_t) = \mathcal{N} e^{-\frac{\|dy_t - C\bar{r}_t dt\|^2}{2dt}}, \quad (2.171)$$

with  $\mathcal{N} = \frac{1}{\sqrt{2\pi dt}}$ .

To sum up, in this Section we have studied how the stochastic master equation, which describing a quantum system under continuous homodyne monitoring of the field it is coupled with, reduces to a system of lineal stochastic equations under the Gaussian assumption.

Before ending our discussion on continuously-monitored systems, we will discuss the motion of a mechanical-mode stored in the cavity, and in contact with the optical mode. This constitutes yet another refinement of the cavity emission example discussed along this Section, and will fix the structure of the matrices  $A$ ,  $C$  and  $D$  of Eq. 5.13.

### 2.4.7 Optomechanical systems

We have previously discussed the dynamical behaviour of a quantum state of light stored in a cavity, that is being continuously-monitored by constantly measuring the bosonic bath which is coupled with. In this section, we will include the mechanical

mode, and discuss how its equations of motion, which we anticipate have the same structure than those of Eq. 5.13— are obtained. While here we provide an overview of the optomechanical model, the interested reader can find a rigorous justification of it in Refs. [DJ99; WM93b].

We consider a single mechanical mode describing the motion of a moving mirror that forms the end of an optical cavity [AKM14], e.g. the one not leaking light in Fig. 2.6. Moreover, the cavity stores an optical mode, and the Hamiltonian of the full system — in the interaction picture with respect to cavity’s optical-mode Hamiltonian— reads

$$H = H_m - ga^\dagger ax + H_d \quad (2.172)$$

where  $a$  is the optical-mode annihilation operator,  $H_m$  and  $x$  stand for the Hamiltonian and position of the mechanical mode, and  $g$  is the coupling constant. Moreover, we included a driving term in the cavity, of the form  $H_d = iE(a - a^\dagger)$ , where  $E \sim \alpha\gamma$  and with  $\alpha$  related to the power of the driving laser. Note that this interaction relates the number of photons with the position of the mechanical mode. Intuitively, since  $N$  is the generator of phase-shifts in the optical mode, the phase of the output light will contain information of mirror’s position.

If we now monitor the cavity via homodyne-detection, we get a Belavkin-Zakai equation for the optomechanical system, similarly to our discussion in Sec. 2.4.3:

$$d\rho_{c,m} = \mathcal{L}_0[\rho_{c,m}]dt + \mathcal{L}_{mon}[\rho_{c,m}]dt + \sqrt{\eta\gamma}\mathcal{H}[a]\rho_{c,m}. \quad (2.173)$$

Let us now focus on the mechanical mode, and perform some assumptions. Namely, we will assume that the cavity mode is *slaved* to the mirror’s dynamics, which correspond to the *bad-cavity-limit* (high values of  $\gamma$ ). In this regime, the cavity mode can be *adiabatically eliminated*; this consists on expanding the optomechanical state  $\rho_{c,m}$  in terms of a parameter that captures the difference in time-scales between the cavity and motional dynamics [DJ99]. Intuitively, the time-scale at which the optical mode changes is much shorter than that of the mechanical one. After performing such expansion and focusing on the reduced state of the mirror, i.e.  $\rho_m = \text{Tr}_c \rho_{c,m}$ , we get

$$d\rho_m = \mathcal{L}[\rho_m]dt + \sqrt{\eta\kappa}[x]\rho_m dW, \quad (2.174)$$

that describes the evolution of the system following a similar structure than that if we only monitored the optical mode. Here, we defined  $\mathcal{L}[\rho] = \mathcal{L}_0[\rho] + \mathcal{L}_{mon}[\rho]$ <sup>27</sup> and some constants were introduced, such as the *measurement constant*  $\kappa = \frac{g^2|\alpha|^2}{\gamma}$ ,

---

<sup>27</sup>In particular,  $\mathcal{L}_{mon}[\rho] = -i[H_m - g|\alpha|^2 x, \rho_m]dt + 2\kappa\mathcal{D}[x]\rho_a dt$ , and  $\mathcal{L}_0$  describes the interaction with a thermal bath as in Eq. 2.165.

describing the rate at which information on mirror's position is obtained and fed back into the first momenta (*i.e.* the strength of the back-action term).

Recapitulating our discussion of Sec. 2.4.6, the equations describing the mechanical-mode dynamics can be simplified considerably by the Gaussian assumption (*i.e.* the state is Gaussian, and all channels acting on it are Gaussian as well). In this case, the evolution of the first two moments – of the mechanical mode – is given by Eqs. 5.13, *e.g.*

$$\begin{aligned} d\bar{\mathbf{r}}_t &= (A - \chi(\boldsymbol{\Sigma}_t)C)\bar{\mathbf{r}}_t dt + \chi(\boldsymbol{\Sigma}_t)d\mathbf{y}_t = A\bar{\mathbf{r}}_t dt + \chi(\boldsymbol{\Sigma}_t)d\mathbf{W}_t \\ d\boldsymbol{\Sigma}_t &= A\boldsymbol{\Sigma}_t + \boldsymbol{\Sigma}_t A^T + D - \chi(\boldsymbol{\Sigma}_t)^T \chi(\boldsymbol{\Sigma}_t), \quad \chi(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma}C^T + \Gamma, \end{aligned} \quad (2.175)$$

with the matrices having the following structure:

$$A = \begin{pmatrix} -\frac{\gamma}{2} & -\omega \\ \omega & -\frac{\gamma}{2} \end{pmatrix}, \quad D = \left[ \gamma(n + \frac{1}{2}) + \kappa \right] \mathbb{I}_2, \quad C = \sqrt{4\eta\kappa} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad (2.176)$$

and  $\Gamma = 0$ . Here, we recall that  $\gamma$  is cavity decay rate,  $\omega$  is the mechanical-mode frequency,  $\kappa$  the measurement strength and  $\eta$  the measurement efficiency. In this context, the measurement signal reads

$$d\mathbf{y}_t = C \langle \bar{\mathbf{r}} \rangle dt + d\mathbf{W}_t. \quad (2.177)$$

This measurement model can be further simplified, when the mechanical-mode frequency is known. This simplification, introduced in Ref. [SDH+11]<sup>28</sup>, consists on demodulating the measurement signal, resulting in an heterodyne-like measurement of a (rotating) system quadratures<sup>29</sup>. Intuitively, for a “sufficiently informative” measurement, *e.g.* after monitoring the system for enough time, the

<sup>28</sup>see also Ref. [DSH+12] for clearer explanation

<sup>29</sup>Here, the measurement current is obtained by the transformations

$$\Delta y_t^{(x)} = \int_t^{t+\Delta t} \cos(\omega_m t') dy_{t'} \sim \sqrt{4\mu\kappa} \langle X \rangle \Delta t + \Delta W_x \quad (2.178)$$

$$\Delta y_t^{(p)} = \int_t^{t+\Delta t} \sin(\omega_m t') dy_{t'} \sim \sqrt{4\mu\kappa} \langle P \rangle \Delta t + \Delta W_p, \quad (2.179)$$

where the rotating quadratures are defined as

$$\begin{aligned} X &= \frac{1}{\sqrt{2}} (ae^{i\omega_m t} + a^\dagger e^{-i\omega t}) \\ P &= \frac{-i}{\sqrt{2}} (ae^{i\omega_m t} - a^\dagger e^{-i\omega t}), \end{aligned}$$

and satisfy  $[X, P] = i$ , and  $\Delta W_x \sim \int_t^{t+\Delta t} \cos \omega t' dW(t')$  (and the same for  $p$ ).

mirror's behaviour should easily be inferred from the signal. As such, its spectral power should be peaked around  $\omega_m$ , and performing such demodulation can be understood as moving to a rotating frame. We will further discuss how to infer the value of the mechanical-mode frequency in Sec. 5.4. When moving to such rotating frame, the matrices in Eq. 2.180 are modified and read

$$A = \begin{pmatrix} -\frac{\gamma}{2} & 0 \\ 0 & -\frac{\gamma}{2} \end{pmatrix}, \quad D = \left[ \gamma \left( n + \frac{1}{2} \right) + \kappa \right] \mathbb{I}_2, \quad C = \sqrt{4\eta\kappa} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad (2.180)$$

namely  $A$  has only a damping term, and  $C$  is proportional to the identity.

The aforementioned model will be considered in Chapter 5, when analyzing sequential hypothesis testing strategies and parameter estimation problems. Overall, the physical model consists on continuously-monitoring the position of a mechanical mode; a quantum-mechanical treatment of this results in a stochastic master equation that comes with back-action term due to the measurement result; note that such back-action term is un-present in the classical scenario. The dynamical equations of the moving mirror are described, in the Gaussian regime, by Eqs. 2.180. Moreover, if the frequency of the mechanical mode is known, then the model is simplified even further. In such case, some analytical insight can be gained, as further discussed in Chapter 5.

## 2.5 Statistical Inference

It is time to put the attention on our (classical) daily lives. Arguably, the highlight of this century is *data* (classical, at least for the first two decades), and motivating this Section by writing that *we are surrounded by data* is, by now, a cliché, though an useful one. In turn, we are constantly faced towards new information and we usually need to take decisions based on it (either conciously or unconsciously)<sup>30</sup>.

While many of our daily decisions turns to be hard to model, some others can be placed on the scientific stage (*i.e.* an experiment, where we acquire data in a systematic way). In this situation, we can rigourously study the model and increase the possibilities of taking correct decisions in the face of uncertainty. In this context, we will now turn study (binary) hypothesis discrimination in Sec 2.5.1, and parameter estimation in Sec. 2.5.5.

---

<sup>30</sup>The number of such decisions is obviously a random variable, but neuro-science community estimate it to be around 35K per day [SL13]



### 2.5.1 Hypothesis testing

The canonical example to discuss hypothesis testing problems deals with medical trials, where the effectiveness of a new development (for instance, a new test that tells whether a person has COVID or not) is to be tested. We will restrict to two hypothesis, or models, consisting on the null hypothesis  $H_0$  (the patient is healthy), and the alternate hypothesis  $H_1$  (the patient has COVID). In this context, we are interested in minimizing the chances that a sick patient is said to be healthy<sup>31</sup>. This notion is captured by the overall probability that the test retrieves a *false negative*, and is known as the  $\beta$ -error or type-II error. Denoting by  $\hat{H}_i$  the result of the test, *i.e.*  $\hat{H}_1$  implies that the test result is *positive* (diagnosing the patient as sick) and  $\hat{H}_0$  *negative* (telling the patient is healthy), have

$$\beta = p(\hat{H}_0|H_1). \quad (2.181)$$

It is easy to come up with a test makes this quantity zero: always diagnose the patient as sick irrespective of the result obtained in the test. Such test is obviously useless, and we are often interested in keeping the *false positive* rate bounded. This quantity is given by the  $\alpha$ -error, or type-II error:

$$\alpha = p(\hat{H}_1|H_0). \quad (2.182)$$

We refer to  $\alpha$  and  $\beta$  as *weak* errors, since they certify the average performance of the test; on the contrary, when discussing sequential strategies we will also deal with the notion of *strong certification*, in the sense that we will be interested in providing guarantees for *each* realisation of the stochastic processes.

In general it is not possible to minimize both errors and a tradeoff appears: reducing one type of error typically increases the other. Conventionally, we are interested in minimizing  $\beta$  for a fixed value of  $\alpha$ , and this is known as *asymmetric hypothesis testing*. On the contrary, there are situations where both type of errors need to be minimized, and thus we consider a linear combination of the errors, leading to the *symmetric* error probability. In particular, the total or mean error probability is given by

$$P_e = p_0p(\hat{H}_1|H_0) + p_1p(\hat{H}_0|H_1), \quad (2.183)$$

where  $p_k$  stands for the prior probability of having the  $k^{\text{th}}$  model (and  $k = 0, 1$ ), and  $\sum_k p_k = 1$ . Analogously, the success probability is defined as

$$P_s = 1 - P_e = p_0p(\hat{H}_0|H_0) + p_1p(\hat{H}_1|H_1), \quad (2.184)$$

---

<sup>31</sup>By the 2020s, there was pandemics, and before having vaccines, we were particularly cautious when testing negative for COVID. For instance, we used to do double checks: if someone contracted COVID chances were extremely high to spread it among people in close contact with.

Our notation emphasizes that the symmetric error can be interpreted Bayesianly, where prior probabilities are assigned to each hypothesis and the total error is obtained by multiplying each prior with the posterior probability of having such hypothesis given the observed data.

Let us consider the example of a discrete random variable  $X$  (say a dice) which takes values  $x \in \{0, 1, \dots, n\}$ , with probability distribution  $p(x)$  under hypothesis  $H_0$  or  $q(x)$  under hypothesis  $H_1$ . Upon receiving a sample, i.e. a value  $x$ , we have to make a guess  $\hat{H}_{g(x)}$  about the true hypothesis according to a *decision or guessing rule*  $g(x) \in \{0, 1\}$ . The total success probability can be written as

$$\begin{aligned} P_s &= \sum_{x=1}^n (P(g(x) = 0, x|H_0)\pi_0 + P(g(x) = 1, x|H_1)\pi_1) \\ &\leq \sum_{x=1}^n \max\{P(g(x) = 0, x|H_0)\pi_0, P(g(x) = 1, x|H_1)\pi_1\} \end{aligned} \quad (2.185)$$

This upper bound can be attained by always guessing for the most likely hypothesis: i.e.

$$g(x) = \begin{cases} 0 & \text{if } P(x|H_0)\pi_0 \geq P(x|H_1)\pi_1 \\ 1 & \text{if } P(x|H_0)\pi_0 \leq P(x|H_1)\pi_1 \end{cases} \quad (2.186)$$

That is the *maximum-likelihood* criterium is optimal for the minimum-error case (symmetric error probability optimization).

Before further discussing hypothesis testing and the assymetric scenario, we will discuss a situation in which the underlying models are quantum states. That is, given a copy of a quantum state we are asked to correctly label it. However, this situation is subtle: the maximum-likelihood criterium can only be applied once a particular measurement (POVM) is implemented, and this is of course also subject to optimization.

## 2.5.2 Single-shot quantum state discrimination

Analogies and similarities do certainly allow us to build intuition, but we do also rely on differences to asses the *meaning* of physical phenomena. We do not need to go very far: discriminating between these letters allows us to be reading this thesis. Suppose, on the other hand, that your glasses drop all of a sudden: you will definitely find difficulties to distinguish between letters. This means that some kind of *noise* appears, and hinders—at least partially—the information to be perfectible discernable. At zero-noise level we are able, in principle, to perfectly tell which are

the letters. Thus, in classical scenarios, external noise-sources are the only cause for which letters might not be perfectly distinguishable.

However, in the quantum realm, letters might well be in a superposition. As such, we usually encounter situations in which even at zero-noise level, not perfect distinguishability can be achieved when looking measuring the (quantum) information, a phenomena unrelated to the transmission channel, but intrinsically linked to the quantum nature of information.

While no glasses can help to achieve perfect distinguishability of quantum states, we are here interested in studying optimal ways to distinguish between the symbols, *e.g.* given now by quantum states. As a matter of fact, among all the possibilities allowed by quantum mechanics to extract information out of a system, some measurements are more helpful than others when it comes to discriminate between a given set of quantum states. Moreover, quantum physics provides an ultimate limit to the distinguishability of states, a fact that we will now turn to discuss.

### The Helstrom bound

We focus on the one-shot discrimination problem between two quantum states  $\rho_0$  and  $\rho_1$ . Given a single copy of a state  $\rho$ , the task is to tell if either hypothesis  $H_0 : \rho = \rho_0$  holds, or hypothesis  $H_1 : \rho = \rho_1$  does it, each having a prior probability of occurring  $p_k$ , with  $\sum_k p_k = 1$ . To this end, we introduce a two-outcome POVM  $\mathcal{M} = \{M_0, M_1\}$  such that  $\sum_{k=0}^1 M_k = \mathbb{I}$  and  $M_k \geq 0$ . Based on the measurement outcome  $k$ , the decision rule reads  $\hat{k} = k$ <sup>32</sup> while the outcome probability is given by Eq. 2.24, *e.g.*  $p(k|\rho) = \text{Tr}[M_k \rho]$ . The success and error probabilities of such discrimination protocol is given by

$$\begin{aligned}
 P_s(\mathcal{M}) &= \sum_{k=0,1} p_k p(\hat{k}|k) = \sum_{k=0,1} p_k \text{Tr}[\rho_k M_k], \\
 P_e(\mathcal{M}) &= \sum_{k=0,1} p_{\bar{k}} p(\hat{k}|\bar{k}) = \sum_{k=0,1} p_{\bar{k}} \text{Tr}[\rho_{\bar{k}}] M_k = 1 - P_s(\mathcal{M}),
 \end{aligned}
 \tag{2.187}$$

where we define the complementary hypothesis  $\bar{k} = k + 1$  (modulus 2). Under these definitions, we will now derive a lower bound for the error probability, known

---

<sup>32</sup>Note that we do not lose generality by restricting to two outcomes POVM, since any POVM with more outcomes,  $\{E_i\}_{i=1}^n$  can be regrouped in an effective POVM with the same performance:  $M_0 = \sum_{i \in S_0} E_i$ , and  $M_1 = \sum_{i \in S_1} E_i$  where  $S_k$  is the set of outcomes for which we guess hypothesis  $k$ .

as *the Helstrom Bound*. To this end, let us write the error probability as per

$$\begin{aligned} P_e(\mathcal{M}) &= p_0 \text{Tr}[M_1 \rho_0] + p_1 \text{Tr}[M_0 \rho_1] \\ &= \text{Tr}[(\mathbb{I} - M) \tilde{\rho}_0] + \text{Tr}[M \tilde{\rho}_1] \end{aligned} \quad (2.188)$$

where we define  $M = M_0$  and  $\tilde{\rho}_k = p_k \rho_k$  as a shorthand. Moreover, we let  $\Delta := \tilde{\rho}_0 - \tilde{\rho}_1$ , it follows that while  $\tilde{\rho}_k$  is a positive semi-definite operator,  $\Delta$  might not be so, although is Hermitian and thus admits a diagonal representation. Having this in mind, we define  $\Delta_{\pm}$  as the positive and negative part of  $\Delta$  respectively, each being a positive-definite operator, *i.e.*

$$\Delta_{\pm} = \sum_{\lambda: \lambda > 0} \lambda |\lambda\rangle \langle \lambda|, \quad \Delta = \Delta_+ - \Delta_- \quad (2.189)$$

Then, we have that

$$\begin{aligned} P_e(\mathcal{M}) &= \text{Tr}[\tilde{\rho}_0] - \text{Tr}[M \Delta] = \text{Tr}[\tilde{\rho}_0] - (\text{Tr}[M \Delta_+] - \text{Tr}[M \Delta_-]) \\ &\geq \text{Tr}[\tilde{\rho}_0] - \text{Tr}[M \Delta_+] \\ &\geq \text{Tr}[\tilde{\rho}_0] - \text{Tr}[\Delta_+] \\ &= \text{Tr}[\tilde{\rho}_0] - \frac{1}{2} (\|\Delta\|_1 + \text{Tr}[\Delta]) \\ &= \frac{1}{2} (1 - \|p_0 \rho_0 - p_1 \rho_1\|_1), \end{aligned} \quad (2.190)$$

where we used that the operators  $M$  and  $\Delta_-$  are positive semi-definite, and that  $\text{Tr}[\Delta_+] = \frac{1}{2} (\|\Delta\|_1 + \text{Tr}[\Delta])$ , which follows from the fact that  $\|\Delta\|_1 = \text{Tr}[\Delta] = \text{Tr}[\Delta_+] + \text{Tr}[\Delta_-]$ .

The structure of the optimal POVM  $\mathcal{M}^*$  can be obtained as follows. Assuming  $\Delta$  has no vanishing eigenvalues, then  $M = \Pi_+$ , where  $\Pi_+$  is the projector over the positive eigenspace  $\Delta_+$ , implying that  $M_1 = \mathbb{I} - M_0$  projects onto the negative part of  $\Delta$ . For this case, the error probability reads

$$\begin{aligned} P_e(\mathcal{M}^*) &= \text{Tr}[\tilde{\rho}_0] - \text{Tr}[\Pi_+ \Delta] \\ &= \text{Tr}[\tilde{\rho}_0] - (\text{Tr}[\Pi_+ \Delta_+] - \text{Tr}[\Pi_+ \Delta_-]) \\ &= \text{Tr}[\tilde{\rho}_0] - \frac{1}{2} (\|\Delta\|_1 + \text{Tr}[\Delta]) \\ &= \frac{1}{2} (1 - \|p_0 \rho_0 - p_1 \rho_1\|_1). \end{aligned} \quad (2.191)$$

While the optimizing measurement  $\mathcal{M}^*$  is a projector, it is a quite particular one, which requires access to the positive and negative subspace of  $\Delta$ . As such, obtaining its expression can be challenging, since it is not always possible to diagonalize  $\Delta$ . Moreover, we are often concerned with a physical implementation of that measurement, which is a subtle matter that will be discussed in Chapter 3.

### The pure-states case

As an example, let us consider the case of two arbitrary pure states  $|\psi\rangle$  and  $|\phi\rangle$ . Here, the operator  $\Delta$  can be written in a two-dimensional basis spanned by  $|\psi\rangle$  and its orthogonal complement  $|\psi^\perp\rangle$ , and the Helstrom bound reduces to:

$$P_e(\mathcal{M}^*, |\psi_0\rangle, |\psi_1\rangle) = \frac{1}{2} \left( 1 - \sqrt{1 - 4p_0p_1|c|^2} \right), \quad (2.192)$$

where  $c = \langle\psi|\phi\rangle$  is the overlap between the pure states; as seen in Fig. 2.8, the bound tends to  $\frac{1}{2}$  (i.e. randomly guessing) as the overlap between the states approaches to unity, whereas it goes to zero when the states become orthogonal, which can be understood as a the classical limit. The optimal measurement is a projector onto a superposition of  $|\psi_0\rangle$  and  $|\psi_0^\perp\rangle$ .

To see this, we can readily construct an orthogonal basis that spans the two-dimensional subspace of the Hilbert space needed to represent the states  $\{|\psi\rangle, |\phi\rangle\}$ . This can be done, for example, via Gram-Schmidt decomposition, where we consider

$$|\psi^\perp\rangle = \frac{|\phi\rangle - c|\psi\rangle}{\sqrt{1 - |c|^2}}, \quad (2.193)$$

allowing us to write

$$|\psi\rangle = c|\alpha\rangle + s|\alpha^\perp\rangle, \quad (2.194)$$

where we defined  $s = \sqrt{1 - |c|^2}$ . Now, the operator  $\Delta$  can be represented in the two-dimensional space spanned by  $\{|\alpha\rangle, |\alpha^\perp\rangle\}$  as per

$$\Delta = \pi_0 |\psi\rangle\langle\psi| - \pi_1 |\phi\rangle\langle\phi| \equiv \begin{pmatrix} \pi_0 - \pi_1 c^2 & -\pi_1 c s \\ -\pi_1 \bar{c} s & -\pi_1 s^2 \end{pmatrix}. \quad (2.195)$$

An straightforward diagonalization leads to the eigenvalues of  $\Delta$ :

$$\lambda_\pm = \frac{1}{2} \left( 1 - 2\pi_1 \pm \sqrt{1 + 4\pi_0\pi_1|c|^2} \right), \quad (2.196)$$

where we can readily see that expression in Eq. 2.192 is obtained by inserting the trace-norm of  $\Delta$  in Eq. 2.191. Moreover, we can also compute the eigenstates (each associated to the positive and negative parts of  $\Delta$ ) as per

$$|\lambda_+\rangle = \alpha_1 |\psi\rangle + \alpha_2 |\psi^\perp\rangle \quad (2.197)$$

$$|\lambda_-\rangle = \alpha_3 |\psi\rangle + \alpha_4 |\psi^\perp\rangle, \quad (2.198)$$

where the coefficients are included in the footnote<sup>33</sup>. Recalling that the optimal POVM is constructed by projecting over  $|\lambda_\pm\rangle$ , we observe that generally the state we project over is a superposition between  $|\psi\rangle$  and  $|\phi\rangle$ , e.g. the hypothesis we aim to distinguish. This will play an important role when discussing the discrimination between two coherent states, since we can readily see that the optimal POVM in this case is a projection onto a cat-like state, e.g. a superposition of coherent-states.

To sum up, we have discussed the basics of single-shot quantum state discrimination; long is the road that continues this discussion. For instance, we omitted unambiguous state discrimination, where we demand that every time we guess for one hypothesis no errors are committed. This requires that we allow an extra *I don't know outcome*, that deems that data inconclusive; in such case we are interested in minimizing the probability associated to such inconclusive outcome.

Also, we have not discussed multiple-hypothesis problems. In particular, closed-form solutions are known if the states are generated by a symmetry group; here, the optimal POVM is the *pretty-good* (or square-root) measurement, and can be obtained explicitly in terms of the candidate states. In the non-symmetric case, such measurement often provides a *pretty-good* success probability [HW94; BC09]. Also it is worth mentioning that semidefinite programming [BV04] will not be discussed here. The latter is a technique that allows to efficiently convex optimization problems (and relates to state discrimination when numerically optimizing over measurements).

---

33

$$\alpha_1 = -\frac{1}{2}\sqrt{2\sqrt{1-c^2}+2} \quad (2.199)$$

$$\alpha_2 = \frac{c}{\sqrt{2}\sqrt{\sqrt{1-c^2}+1}} \quad (2.200)$$

$$\alpha_3 = \frac{c^2 + \sqrt{1-c^2} - 1}{\sqrt{2}\sqrt{(c^2-1)(\sqrt{1-c^2}-1)}} \quad (2.201)$$

$$\alpha_4 = \frac{c}{\sqrt{2-2\sqrt{1-c^2}}} \quad (2.202)$$

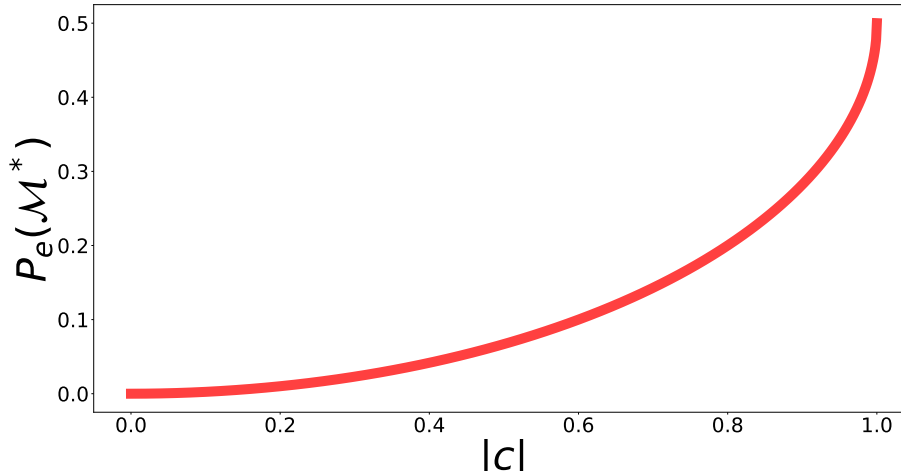


Figure 2.8: We show Helstrom bound for the error probability when discriminating between two pure states as a function of their overlap  $c$  (absolute value).

### 2.5.3 Back to the classics: asymmetric hypothesis testing

Error probabilities non-trivially depend on the amount of data available, e.g. the number of samples. Intuitively, the more data we have (e.g. the more we have sampled our system), the smaller should the errors be on average. Quite generically, it can be seen that the error decays exponentially with the number of samples, with a rate that is an entropic function of the two underlying distributions; we will discuss some characterizations of such rates later in this Section.

In this context, a very powerful lemma known as the *Neyman-Pearson lemma*, provides the statistic that is optimal for a wide range of settings.

*Neyman-Pearson Lemma.* Given a binary hypothesis testing problem where we need to tell if data  $\mathbf{x} = (x_1, \dots, x_n)$  has been sampled from either  $H = H_0$  or  $H = H_1$ , for  $T > 0$  we define the *acceptance region* as

$$A(T) = \left\{ \frac{p(\mathbf{x}|H_1)}{p(\mathbf{x}|H_0)} \geq T \right\}, \quad (2.203)$$

where  $\mathbf{x}$  are the samples acquired, and  $T$  is the *decision-boundary* (note that the point  $T = 1$  is where the hypothesis are equally likely). The decision region plays a prominent role in the values that the resulting errors take, and in this case read

$\alpha^* = p(A(T)|H_0)$  and  $\beta^* = p(\overline{A(T)}|H_1)$ <sup>34</sup>. Now, consider any other decision region  $B$ , whose error probabilities are  $\alpha_B$  and  $\beta_B$ . Then, the lemma states that if  $\alpha_B \leq \alpha^*$ , it follows that  $\beta_B \geq \beta^*$  (and also if  $\beta_B \leq \beta^*$  then  $\alpha_B \geq \alpha^*$ )<sup>35</sup>.

As discussed previously, we are generally interested in minimizing  $\beta$  error when keeping  $\alpha$  bounded. Considering the set of all possible statistics that can be defined, the power Neyman-Pearson lemma is that of providing the optimal one, which is given by the *log-likelihood ratio*.

### Large deviations and asymptotic error rates

The dependence of the error probabilities is non-trivial with respect to the number of samples available, and we intuitively expect that the error probabilities become smaller and smaller when more data is acquired. This can be formally studied, and we here provided a short outline of the topic.

We will begin by discussing the notion of *random variable concentration*. We consider the random sequence  $\mathbf{x}$  with  $x_k \in \mathcal{A} = \{a_1, \dots, a_M\}$ , and consisting on  $n$  *i.i.d* samples, whose underlying probability will be denoted by  $Q$ , *i.e.*  $Q^n(\mathbf{x}) = \prod_{k=1}^n Q(x_k)$

We define a *type*  $P_{\mathbf{x}}(\mathcal{A})$ —for simplicity we will denote it as  $P_{\mathbf{x}}$ —which is the empirical probability distribution of each letter in alphabet  $\mathcal{A}$ , associated to the sequence  $\mathbf{x}$ , *i.e.*  $P_{\mathbf{x}}(a) = \frac{N_{\mathbf{x}}(a)}{n}$ ,  $\forall a \in \mathcal{A}$ , where  $N_{\mathbf{x}}(a)$  is the number of times that letter  $a$  appears in  $\mathbf{x}$ . Next, we define as  $\mathcal{P}_n$  as the set comprising all possible types  $P$  generated by a denominator  $n$ . For example taking a binary alphabet  $\mathcal{A} = \{0, 1\}$ , we get

$$\mathcal{P}_n = \{(P(0), P(1)) : \left(\frac{0}{n}, \frac{n}{n}\right), \left(\frac{1}{n}, \frac{n-1}{n}\right), \dots, \left(\frac{n}{n}, \frac{0}{n}\right)\}.$$

Since many sequences can lead to the same type, we define the type class  $T(P)$  as

$$T(P) = \{\mathbf{x} : P_{\mathbf{x}} = P\}. \quad (2.204)$$

Now, the probability of a given sequence  $\mathbf{x}$ , which is *i.i.d.* and sampled from  $Q$  can

---

<sup>34</sup> $\overline{A(T)}$  denotes the complement of  $A(T)$ , *e.g.*  $\overline{A(T)} = \left\{ \mathbf{x} : \frac{p(\mathbf{x}|H_1)}{p(\mathbf{x}|H_0)} < T \right\}$

<sup>35</sup>The proof of the lemma is straightforward, and consists on exploiting decision functions associated to  $A(T)$  and  $B$ , and playing a bit with the definition of  $A(T)$ .



be written in terms of its type  $P_{\mathbf{x}}$  as<sup>36</sup>

$$Q(\mathbf{x}) = e^{-n \left( D(P_{\mathbf{x}} \| Q) + H(P_{\mathbf{x}}) \right)}, \quad (2.205)$$

where the relative entropy between two (discrete) random-variable distributions is defined as  $D(p \| q) = \sum_k p_k \log \frac{p_k}{q_k}$  and  $H$  is the Shannon entropy.

From Eq. 2.205 we can readily see that for typical sequences, i.e. sequences  $\mathbf{x}$  with  $P_{\mathbf{x}} = Q$ ,  $Q(\mathbf{x}) = e^{-nH(Q)}$ , i.e. its occurrence is given by the entropy of  $Q$ .

It can be shown that the size of a given type class (that is the number of sequences with a given empirical distribution) is asymptotically given by  $|T(P)| \sim e^{nH(P)}$ . This allows us to write the probability of a given type class as

$$Q(T(P)) = |T(P)| e^{-n \left( D(P_{\mathbf{x}}=P \| Q) + H(P) \right)} \sim e^{-nD(P \| Q)}. \quad (2.206)$$

This shows that type class corresponding to the typical distribution is exponentially most probable than any other type. This puts on solid grounds that if one samples a distribution many times the frequency of symbols will be essentially equal to the probabilities.  $\frac{N(a)}{n} = Q(a)$  with high probability.

For example, considering again the tossing-coin example, with a heads probability of  $q$ . For  $n$  samples the types  $\mathcal{P}_n$  can be fully characterized by the number of heads, i.e.  $\mathcal{P}_n = \left\{ \binom{n}{k}, \frac{n-k}{n} \right\}_{k=1}^n$ , and the random variables concentrate around the type  $k$  given by  $\frac{k}{n} = q$  as  $n \rightarrow \infty$ .

Let us return to the likelihood ratio, which given a random sequence  $\mathbf{x}$  is obtained as  $\Lambda(\mathbf{x}) = \frac{p_1(\mathbf{x})}{p_0(\mathbf{x})}$ . In the *i.i.d.* setting we can link the logarithm of the likelihood ratio, which we call the *log-likelihood ratio* and denote by  $\ell$ , to the relative

<sup>36</sup>To see this, we expand the total probability of  $\mathbf{x}$  and exploit the *i.i.d.* property:

$$\begin{aligned} Q(\mathbf{x}) &= \prod_k Q(x_k) = \prod_{a \in \mathcal{A}} Q(a)^{N_{\mathbf{x}}(a)} = \prod_{a \in \mathcal{A}} Q(a)^{nP_{\mathbf{x}}(a)} \\ &= e^{-n \left( \sum_{a \in \mathcal{A}} P_{\mathbf{x}}(a) \log \frac{P_{\mathbf{x}}(a)}{Q(a)} - P_{\mathbf{x}}(a) \log P_{\mathbf{x}}(a) \right)}, \end{aligned}$$

where the last terms in the exponent can be the relative entropy and Shannon entropy respectively.

entropies as

$$\ell(\mathbf{x}) = \log \frac{p_1(\mathbf{x})}{p_0(\mathbf{x})} = \sum_{i=1}^n \log \frac{p_1(x_i)}{p_0(x_i)} \quad (2.207)$$

$$= \sum_{a \in \mathcal{A}} nP_{\mathbf{x}}(a) \log \frac{p_1(a)}{p_0(a)} \quad (2.208)$$

$$\begin{aligned} &= \sum_{a \in \mathcal{A}} nP_{\mathbf{x}}(a) \log \frac{p_{\mathbf{x}}(a)}{p_0(x_i)} - \sum_{a \in \mathcal{A}} nP_{\mathbf{x}}(a) \log \frac{p_{\mathbf{x}}(a)}{p_1(x_i)} \\ &= n \left[ D(P_{\mathbf{x}} \| p_0) - D(P_{\mathbf{x}} \| p_1) \right]. \end{aligned}$$

From here it is immediate to see that the expectation value of log-likelihood under hypothesis  $H_1$  is given by the relative entropy,  $\langle \ell \rangle_1 / n = D(P_1 \| P_0)$ , and under  $H_0$  by  $\langle \ell \rangle_0 / n = -D(P_0 \| P_1)$ . Moreover, from our previous discussion on the theory of types, it also follows that  $\ell/n$  concentrates around these values:

$$\frac{\ell(\mathbf{x})}{n} \xrightarrow[n \rightarrow \infty]{} D(P_1 \| P_0). \quad (2.209)$$

under  $H_1$  (and similarly with  $H_0$ ).

Neyman-Pearson's Lemma in conjunction with the theory of types also allows us to easily assess the asymptotic error rates for *i.i.d.* sampling. It is clear that the decision region depends only on the type class of the observed sequence (in the case of the coin, the total number of head and tails).

Then, for symmetric hypothesis testing, the optimal decision region corresponds to guessing for the most probable hypothesis (given the observed empirical distribution) or equivalently guessing hypothesis  $H_1$  ( $H_0$ ) if  $\ell > 0$  ( $\ell \leq 0$ ). It is easy to check that the error rate for  $P_{err}^{(n)} = \pi_0 \alpha_n + \pi_1 \beta_n$  is dominated by the events where  $\ell = 0$ , i.e.  $D(P_{\mathbf{x}} \| p_0) - D(P_{\mathbf{x}} \| p_1) = 0$ . That is, we obtain the *Chernoff coefficient*:

$$C^* = \lim_{n \rightarrow \infty} -\frac{1}{n} \log \min_{A_n} P_{err}^{(n)}, \quad (2.210)$$

where  $C^* = D(P_{\lambda^*} \| P_1) = D(P_{\lambda^*} \| P_0)$ , meaning that  $\lambda^*$  is determined by the probability distribution  $P_{\lambda}$  for which the relative entropy to  $P_0$  equals the relative entropy to  $P_1$ <sup>37</sup>.

---

<sup>37</sup>Such probability distribution has the structure  $P_{\lambda}(x) = \frac{P_0^{\lambda}(x)P_1^{1-\lambda}(x)}{\sum_{a \in \mathcal{A}} P_0^{\lambda}(a)P_1^{1-\lambda}(a)}$ , which leads to an alternative (and more common) expression for  $C^* = -\min_{0 \leq \lambda \leq 1} \log \sum_{a \in \mathcal{A}} P_0(a)^{\lambda} P_1^{1-\lambda}(a)$

Similarly, in the asymmetric setting one can get the largest error exponent (fastest decay of the error), while keeping a finite probability for the other type error (recall that we wish to diagnose correctly at least a fraction of the healthy patients). This is formalized in the following lemma.

*Stein-lemma.* Let  $\mathbf{x}$  be a sequence of *i.i.d.* samples obtained from  $Q$ . We consider an hypothesis testing scenario where the models are given by  $Q = P_1$  and  $Q = P_0$ . Let  $A_n$  be an acceptance region for hypothesis  $H_1$ , and the error probabilities be  $\alpha_n = P(\bar{A}_n|H_0)$  and  $\beta_n = P(A_n|H_1)$ . Let  $0 < \epsilon < \frac{1}{2}$ , and define  $\beta_n^\epsilon = \min_{A_n} \beta_n$  such that  $\alpha_n < \epsilon$ . Then the optimal error exponent is

$$-\lim_{n \rightarrow \infty} \frac{\log \beta_n^\epsilon}{n} = D(P_0||P_1). \quad (2.211)$$

Note that this rate is independent of the particular value of  $\epsilon$ . Similarly, the  $\alpha$  error is minimized instead (while keeping  $\beta < 1/2$ , then the optimal rate is given by  $D(P_1||P_0)$ ). The relative entropies appearing in the Stein lemma, are strictly larger than the Chernoff coefficient, as it should since in the symmetric scenario both (optimal) errors must decay exponentially, whereas in the asymmetric case only one them remains constant.

Finally, note that the error rates provide an operational interpretation to the relative entropy and the Chernoff coefficient as measures of distinguishability between the two probability distributions.

### Quantum hypothesis testing: a brief comment

While we provided an introduction to the classical hypothesis testing problem, and this suffices for the scope of this thesis, we will here outline how this situation generalizes to the quantum realm. Contrary to the single-shot quantum-state discrimination scenario, we are here given  $N$  copies of the same quantum state, and it is desired to identify its nature among two (or more) alternatives. In this case, since more information is available, our performance is expected to be better on average, similarly to the results discussed in Sec. 2.5.3. The scope of possible strategies is, however, non-trivially larger. In turn, we can choose to perform either a separable quantum measurement (individually measure each copy), an adaptive quantum measurement (where the result of the previous copy conditions the next quantum measurement to be performed), or a joint measurement acting on all the  $N$  copies at once, or even any combination of the above strategies. In this context, a quantum version of the Neyman-Pearson lemma discussed above can be formulated:

*Quantum Neyman-Pearson Lemma* Given a binary hypothesis testing problem with the two hypotheses:  $\rho(H_0)$  vs.  $\sigma(H_1)$ . For  $T > 0$  define a two outcome POVM  $M = \{M_0, M_1 = \mathbb{I} - M_0\}$  with

$$M_0(T) = \mathbb{I}_{(\rho - T\sigma) > 0} \tag{2.212}$$

where  $\mathbb{I}_{\Delta > 0}$  is the projector on the positive part of  $\Delta$ , and the corresponding error probabilities  $\alpha^* = P(M_0|H_1) = \text{Tr}[M_0(T)\rho]$  and  $\beta^* = P(M_0|H_0) = \text{Tr}[M_0(T)\sigma]$ . Given any measurement  $F = \{F_0, F_1\}$  and its associated error probabilities  $\alpha_F$  and  $\beta_F$ . If  $\alpha_F \leq \alpha^*$  then  $\beta_F \geq \beta^*$  (and vice versa).

The projective measurement  $M_0(T)$  is the quantum analog of the likelihood test which defined the acceptance region (for hypothesis  $H_0$ )

$$A = \left\{ \mathbf{x} : \frac{P_0(\mathbf{x})}{P_1(\mathbf{x})} > T \right\} = \{ \mathbf{x} : P_0(\mathbf{x}) - TP_1(\mathbf{x}) > 0 \}$$

Note also from (2.212) that such class of measurements  $M_0(T)$  is optimal when one wishes to minimize the linear combination of errors  $\alpha^* + T\beta^*$ , as for instance in symmetric hypothesis testing taking the parameter  $T$  to be the ratio of priors:  $T = \pi_1/\pi_0$ . Indeed for this case we recover the famous Helstrom bound from last section.

Generally, the optimal error probability exponentially decreases with the number of copies, and deriving rates at which it does so is the matter of *quantum hypothesis testing*. For example, in the *i.i.d.* scenario, generalizations of the classical testing framework are known: while the symmetric error decreases with a rate known as the *quantum Chernoff coefficient* [ACT+07], the asymmetric errors have quantum relative entropies as rates [HP91; ON]. Strikingly, such rates are often obtained by replacing the classical expression of the entropic functions with the quantum one. The optimal measurement attaining such rates might be a global one (*i.e.* acting jointly on all the available copies). We remark, however, that for the case of binary pure state discrimination, it has been shown that one can be asymptotically optimal when measuring locally and updating the priors in a Bayesian manner [ABB+05]. Moreover, the setting of quantum sequential hypothesis testing has recently been formalized and studied in Ref. [MVHS+21], and we will now turn to study the classical version of it.

### 2.5.4 Sequential hypothesis testing

So far we have discussed the hypothesis testing setting in which  $n$  samples of data were presented to us, and a decision needs to be made for the underlying

hypothesis. We introduced different figures of merits, taking into account the importance given to each error type. Following, we introduced the log-likelihood ratio, which via the Neymann-Pearson lemma turns out to be the optimal statistic. Then, we have discussed the asymptotic behaviour of the errors, which in the *i.i.d.* are exponentially decaying, with a rate given by either the relative entropy (asymmetric scenario) or the Chernoff coefficient (symmetric scenario).

Strikingly, by slightly relaxing the setting, more efficient strategies can be brought to the stage. This consists in relaxing the condition that the underlying hypothesis needs to be determined after observing the random sequence  $\mathbf{x}$  (of length  $n$ ). Instead, we could think of allowing for an extra label that deems  $\mathbf{x}$  as inconclusive. In this case, further samples will be required, and the test proceeds in a *sequential* fashion until a given criterium is met.

Thus, a sequential strategy is characterized by (i) a *stopping rule*, that tells whether to stop the sampling process or to demand an additional sample, and (ii) a *decision rule* which selects one hypothesis or the other in case of stopping the test. Moreover, we will consider an scenario in which one can guarantee that *for each realization of the test*, the conditional probability of correctly identifying each of the hypothesis is above some pre-defined threshold. This is known as a *strong errors guarantee*, i.e. given  $\epsilon_k$  ( $k = 0, 1$ ), the test assures that

$$p(H_k|\mathbf{x}_n) \geq 1 - \epsilon_k. \quad (2.213)$$

These conditions cannot always be achieved by a test with a fixed-horizon — since there is a chance that the random sequence is not informative enough to assert the correctness of Eq. 2.213 for *both* hypothesis —. On the contrary, if new samples are required until conditions in Eq. 2.213 are fulfilled, then we can readily devise a test — known as the Sequential Probability Ratio Test (SPRT) [WW48] — that consists on the following. Starting at  $n = 1$ , at each step  $n$  check if

1.  $p(H_1|\mathbf{x}_n) \geq 1 - \epsilon_1$ . If this happens, *stop* and accept  $H_1$ , with a success probability guaranteed to be  $P_{s_1} = 1 - \epsilon_1$ .
2.  $p(H_0|\mathbf{x}_n) \geq 1 - \epsilon_0$ . If this happens, *stop* and accept  $H_0$ , with a success probability guaranteed to be  $P_{s_1} = 1 - \epsilon_0$ .
3. If neither of 1. nor 2. is accomplished, then continue sampling and move to the next step.

This procedure can be casted in terms of the log-likelihood ratio  $\ell_n = \ell(\mathbf{x}_n)$  at step  $n$ . By using Bayes theorem, we can readily construct an *undecision region*

$\Omega = [a_0, a_1]$  such that

$$\begin{aligned} a_1 &= \log \left( \frac{1 - \epsilon_1 p_0}{\epsilon_1 p_1} \right) \\ a_0 &= \log \left( \frac{\epsilon_0 p_0}{1 - \epsilon_0 p_1} \right) \end{aligned} \quad (2.214)$$

and then the test proceeds at step  $n$  as follows:

- Compute the log-likelihood ratio  $\ell_n$
- If  $\ell_n \geq a_1$  *stop* the test, and accept  $H_1$ , with a success probability guaranteed to be  $P_{s_1} = 1 - \epsilon_1$
- Alternatively, if  $\ell_n \leq a_0$ , *stop* the test, and accept  $H_0$ , with a success probability guaranteed to be  $P_{s_0} = 1 - \epsilon_0$
- On the contrary, if  $\ell_n \in \Omega$ , demand an extra sample and repeat the test for step  $n + 1$ .

As explicated by Eq. 2.207, in the *i.i.d.* scenario the log-likelihood ratio can be casted as a sum of contributions  $\ell_k = \ell(x_k)$  that are sequentially-acquired:

$$\ell(\mathbf{x}_n) = \log \frac{p(\mathbf{x}|H_1)}{p(\mathbf{x}|H_0)} = \sum_{k=1}^n \ell_k, \quad \ell_k = \log \frac{p(x_k|H_1)}{p(x_k|H_0)}. \quad (2.215)$$

The beauty of this relies on *the random walk interpretation*: with each new sample, a step of length  $\ell_k$  is taken with probability  $p(\mathbf{x}_k|Q)$ , where  $Q$  is the underlying probability distribution, *i.e.* either  $P_0$  or  $P_1$ . From here, we can readily see that the walker will have a positive (negative) drift when  $H_1(H_0)$  is the underlying hypothesis: by denoting the log-likelihood ratio obtained under hypothesis  $i$  as  $\ell_i$ , we have

$$\begin{aligned} \langle \ell_{|1}(\mathbf{x}_n) \rangle &= nD(P_1||P_0) \\ \langle \ell_{|0}(\mathbf{x}_n) \rangle &= -nD(P_0||P_1). \end{aligned} \quad (2.216)$$

These drifts values indicate that the random walk will likely hit the decision boundary that corresponds to the underlying hypothesis, since it moves with an average speed given by the relative entropy. However, it might well be that the stochastic nature of the process leads the walker to hit the complementary decision boundary. Either the case, Wald proved that the walker eventually exits the undecision region

$\Omega$ , and thus the process stops [WW48]. The time—or sample number— at which the process stops is known as *the stopping time*, and we denote it by  $\tau$ :

$$\tau := \inf_n \{n : \ell(\mathbf{x}_n) \notin \Omega\}. \quad (2.217)$$

Thus, the stopping time refers to the first instance at which the process leaves the  $\Omega$  region, and we have that the probability of the walker to be inside the region  $\Omega$  goes to zero as the number of samples goes to infinity. In the following, we will interchangeably refer to  $\tau$  as the number of samples  $n$  at which the process stops, since this Section will serve as a reference for the continuous-time processes analyzed in Chapter 5.

The average position of the walker at step  $\tau$ , e.g. the value of  $\langle \ell(\mathbf{x}_\tau) \rangle$  can be linked to the average value of the stopping-time  $\tau$  via the Wald's identity [MU05]:

$$\langle \ell(\mathbf{x}_\tau) \rangle_i = \left\langle \sum_{k=1}^{\tau} \ell_k \right\rangle_i = \langle \ell_k \rangle_i, \quad (2.218)$$

resulting in

$$\langle \tau \rangle_1 = \frac{\langle \ell \rangle_1}{D(P_1 || P_0)}, \quad (2.219)$$

$$\langle \tau \rangle_0 = -\frac{\langle \ell \rangle_0}{D(P_1 || P_0)}. \quad (2.220)$$

Thus, it becomes clear that the value of the log-likelihood at the stopping time (the moment when it exits the undecision region) will be essentially given its value at the boundary (essentially  $a_0$  or  $a_1$  for hypothesis  $H_1$  or  $H_0$ , respectively). Hence, as we will show more rigorously below,

$$\langle \tau \rangle_1 \sim \frac{a_1}{D(P_1 || P_0)}, \quad (2.221)$$

$$\langle \tau \rangle_0 \sim -\frac{a_0}{D(P_1 || P_0)}. \quad (2.222)$$

The last equation clearly shows that the average value of the likelihood ratio is proportional to the average-time required to reach the corresponding decision boundary, and the slope is given by the relative entropy between the probability distributions under consideration. Since the relative entropy is generally not symmetric, i.e.  $D(P_0 || P_1) \neq D(P_1 || P_0)$ , we expect the average stopping-times to differ when swapping the underlying model. Thus, it is interesting to ask how many

samples would be required, on average, to stop the process; this will in turn depend on the quality of the sequential test, as given by the strong error conditions of Eq. 2.213.

Let us see now how does SPRT perform regarding the type-I and type-II errors, in particular in the regime where  $|a_i|$  asymptotically large (small errors), where  $\Pr[\ell(\mathbf{x}_n) \in \Omega] \xrightarrow{n \rightarrow \infty} 0$  holds. For simplicity, we define

$$A = e^{a_1} = \frac{1 - \epsilon_1 p_0}{\epsilon_1 p_1}, \quad B = e^{a_0} = \frac{\epsilon_0 p_0}{1 - \epsilon_0 p_1}.$$

Then, the weak errors can be linked to the decision thresholds as follows

$$\alpha = p(\hat{H}_1|H_0) = \sum_{\mathbf{x} \in \mathcal{X}_0} p(\mathbf{x}|H_0) \leq A^{-1} \sum_{\mathbf{x} \in \mathcal{X}_0} p(\mathbf{x}|H_0) = A^{-1}(1 - \beta) \quad (2.223)$$

$$\beta = p(\hat{H}_0|H_1) = \sum_{\mathbf{x} \in \mathcal{X}_1} p(\mathbf{x}|H_1) \leq B \sum_{\mathbf{x} \in \mathcal{X}_1} p(\mathbf{x}|H_0) = B(1 - \alpha), \quad (2.224)$$

where used that, when stopping the test, the log-likelihood ratio is either higher (smaller) than  $a_1(a_0)$ , and defined the decision regions  $\mathcal{X}_k$  under which the test decides for the  $k^{\text{th}}$  hypothesis.

Moreover, the inequalities above saturate if there is no *overshooting*, that is if the process stops exactly at the boundary. This happens either approximately when the step-sizes  $\ell_k$  are small relative to  $a_i$ , as mentioned above, or if the log-likelihood evolves continuously – e.g. if we are sampling from a Wiener-like process, as discussed in Sec. 2.4.4, and analyzed in Chapter 5. In that case the above relations can be inverted to get

$$\alpha = \frac{1 - B}{A - B} = \frac{1 - e^{a_0}}{e^{a_1} - e^{a_0}} = \frac{\epsilon_1(p_1 - \epsilon_0)}{p_0(1 - \epsilon_0 - \epsilon_1)} \quad (2.225)$$

$$\beta = \frac{B(A - 1)}{A - B} = e^{a_0} \frac{e^{a_1} - 1}{e^{a_1} - e^{a_0}} = \frac{\epsilon_0(p_0 - \epsilon_1)}{(1 - \epsilon_0 - \epsilon_1)p_1}. \quad (2.226)$$

Thus, imposing strong-error conditions fixes the value of the weak errors in the SPRT. Note that in the case that the boundaries are very large in absolute value, e.g.  $a_0 \rightarrow -\infty$  and  $a_1 \rightarrow \infty$ , corresponding to very small values of  $\epsilon_0$  and  $\epsilon_1$  respectively, then

$$\alpha \sim e^{a_0} \sim \epsilon_1, \quad \beta \sim e^{-a_1} \sim \epsilon_0. \quad (2.227)$$

$\ell(\mathbf{x}_\tau)$  as *binary random-variable*. In the non-overshooting case, the log-likelihood ratio  $\ell(\mathbf{x}_\tau)$  can take two values once the SPRT stops, which are determined by the region  $\Omega$ , e.g.  $\ell(\mathbf{x}_\tau) \in \{a_0, a_1\}$ . For instance, if the underlying probability distribution



is  $P_1$ , then the walker will arrive to position  $a_1$  with probability  $p(\hat{H}_1|H_1) = 1 - \alpha$ , and otherwise arrive to position  $a_0$  with probability  $p(\hat{H}_0|H_1) = \beta$  (a similar reasoning applies when the underlying model is  $P_0$ ), and thus we have that the mean-value of the  $\ell(\mathbf{x}_\tau)$  reads

$$\langle \ell_{|_1}(\mathbf{x}_\tau) \rangle = a_1(1 - \beta) + a_0\beta, \quad \langle \ell_{|_0}(\mathbf{x}_\tau) \rangle = a_0\alpha + a_1(1 - \alpha). \quad (2.228)$$

From here, we can infer the number of samples required to stop the test, given pre-defined strong error thresholds. In turn, from Eq. 2.228, the Wald identity in Eq. 2.218, and the expressions of the weak errors in terms of the strong ones, one can readily identify the average stopping-time, whose expression is asymptotically given by

$$\langle \tau \rangle_1 \sim -\frac{\log \epsilon_1}{D(P_0||P_1)}, \quad \langle \tau \rangle_0 \sim -\frac{\log \epsilon_0}{D(P_1||P_0)} \quad (2.229)$$

Remarkably, the SPRT is *optimal* regarding such resource [WW48]. This means that, for a pre-defined value of the weak errors  $\alpha$  and  $\beta$ <sup>38</sup>, the number of samples  $-\langle \tau_{|_1} \rangle$  and  $\langle \tau_{|_0} \rangle$ — that a test would require to gather cannot be less than that of the SPRT, when accomplishing to such errors.

In this regard, the SPRT comes with a *bonus*: not we one can attain the desired error rates faster, but also it is possible to certify the error committed for any given trajectory. This is in contrast to the approach presented in Sec. 2.5.3, which contrasts the hypothesis in a *static* way. As expected, allowing the test to stop when enough evidence is gathered, results in a lower average number of samples; and when compared with its *deterministic* counterpart, one can attain better error rates using the same resources (on average). Similarly, these results hold for the symmetric error probability in Eq. 2.184, e.g. when the value  $P_e = p_0\alpha + p_1\beta$  is fixed in advance, and a sequential test is carried out in order to optimize the average stopping time required to reach such error threshold [Sim76].

Thus, we will deem the test presented in Sec. 2.5.3 — which does not process the data sequentially but rather makes a decision once a fixed and pre-defined time-horizon has been reached— as the *deterministic* test. In Chapter 5 we will thoroughly study the performance of sequential tests in continuously-monitored systems and compare their performance with to respect to that of deterministic ones (those that gather data during a pre-established and fixed time).

---

<sup>38</sup>To avoid confusion, recall that such errors are, in the SPRT, are fixed by the values of  $\epsilon_0$  and  $\epsilon_1$ .

### 2.5.5 Parameter estimation

In this Section we discuss parameter estimation, and in particular the limits that (classical) information theory imposes to the accuracy that can be attained when estimating a parameter encoded in a probability distribution function out of a finite number of samples.

Let  $f(x, \theta)$  be a probability distribution function, parametrized by  $\theta$  (as an example, a normal distribution  $\mathcal{N}(\mu, \sigma)$  has parameters  $\mu$  and  $\sigma$ ). We interpret such object as being the probability of a sample to take the value  $x$ , given a known value of  $\theta$ . For instance, if  $f$  has a strong dependence *w.r.t.*  $\theta$ , we expect it to be easier to infer information about  $\theta$ , and fewer samples will be sufficient to estimate the value of the parameter with a given precision. On the contrary, if the landscape of  $f$  looks flatter when varying the value of  $\theta$ , then we will require more samples to be sure that value of  $\theta$  is the right one. This notion is captured by the *score* of  $f$ :

$$s(x, \theta) := \partial_\theta \log f(x, \theta), \quad (2.230)$$

whose mean value vanishes:

$$\mathbb{E}[s(x, \theta)] = \int_{\mathbb{R}} dx \partial_\theta \log f(x, \theta) f(x, \theta) = \partial_\theta \int_{\mathbb{R}} dx f(x, \theta) = 0.$$

where the last equality follows from the normalization  $\int_{\mathbb{R}} dx f(x, \theta) = 1$ . However, its variance contains information about the parameter-landscape, and reads

$$\begin{aligned} I(\theta) := \text{Var}[s] &= \mathbb{E}_x[(\partial_\theta \log f(x, \theta))^2] = \int_{\mathbb{R}} dx f(x, \theta) (\partial_\theta \log f(x, \theta))^2 \\ &= \int_{\mathbb{R}} dx f(x, \theta) \left(\frac{\partial_\theta f}{f}\right)^2. \end{aligned}$$

The variance of the score is known as the *Fisher information*, and we can explicitly see that is linked to the curvature of  $f$  as per

$$\langle \partial_\theta^2 \log f \rangle = \left\langle -\frac{(\partial_\theta f)^2}{f^2} \right\rangle + \left\langle \frac{\partial_\theta^2 f}{f} \right\rangle.$$

Overall, the Fisher information describes how the landscape of  $f$  looks when varying the parameters  $\theta$ , and the following relation holds

$$I(\theta) = \mathbb{E}[(\partial_\theta \log f)^2] = -\mathbb{E}[\partial_\theta^2 \log f]. \quad (2.231)$$

However, we are interested in providing an estimation of the parameter, given that we acquired a random (*i.i.d*) sequence sampled from  $f$ . To this end we define an *estimator*  $\hat{\theta}$  to be an statistic that provides such value  $\hat{\theta}$ . For instance, the sample average  $\langle \mathbf{x} \rangle = \frac{1}{n} \sum_k x_k$  is an (unbiased) estimator for the mean of  $f$ .

Among all possible estimators, we will find it useful to consider the *maximum-likelihood* one, defined as

$$\hat{\theta} = \underset{\theta}{\text{ArgMax}} \log f(\mathbf{x}, \theta), \quad (2.232)$$

where we note that since  $\mathbf{x}$  consists on *i.i.d.* samples, we can factorize the logarithm of the likelihood function as a sum of contributions, similarly to Eq. 2.215. While the maximum-likelihood estimator is *consistent*, in the sense that for a very large number of samples such estimator converges to the correct value almost surely, it might be *biased* (in the sense that its expected value might differ from underlying true parameter we want to estimate)<sup>39</sup>. This is in line with the fact that such estimator retrieves the parameter value which is most consistent with the observed data, but not with the average.

To end this discussion on parameter estimation, let us comment on a fundamental result that relates the ultimate performance of an estimator with the Fisher information. In turn, the *Cramer-Rao bound* provides a lower bound to the variance of any (unbiased) estimator. This means that our confidence in the estimation cannot be higher than the information carried out by the underlying probability distribution function  $f$ , as given by the Fisher Information  $I(\theta)$  introduced above. To see this, we consider an (unbiased) estimator of  $\theta$ , *i.e.*  $\hat{\theta}$ . If it is unbiased, then the following chain of equalities holds:

$$0 = \int dx f(x) (\hat{\theta} - \theta) = \partial_{\theta} \left[ \int dx f(x) (\hat{\theta} - \theta) \right] = -1 + \int dx (\partial_{\theta} f(x)) (\hat{\theta} - \theta),$$

and thus we conclude that  $\int dx f(x) (\partial_{\theta} \log f) (\hat{\theta} - \theta) = 1$  If we now we use the Cauchy-Schwartz inequality, *i.e.*  $(u, v)^2 \leq (u, u)(v, v)$ , we get

$$\begin{aligned} 1^2 &= \left[ \int dx \left( (\hat{\theta} - \theta) \sqrt{f} \right) \left( \sqrt{f} \partial_{\theta} \log f(x) \right) \right]^2 \\ &\leq \int dx \left( (\hat{\theta} - \theta) \sqrt{f} \right)^2 \left[ \int dx \left( \sqrt{f} \partial_{\theta} \log f(x) \right)^2 \right] \\ &= \text{Var}[\hat{\theta}] I(\theta). \end{aligned}$$

---

<sup>39</sup>Note however that any consistent estimator is asymptotically unbiased.

Thus, the Fisher information provides an ultimate (classical) bound for the variance of *any* estimator. Note that in the case of  $n$  trials the Fisher information, in the *i.i.d.* case, is additive:  $I_n(\theta) = nI(\theta)$ , which can be seen from  $\mathbb{E}\partial_\theta^2 \log p(x)^n$ . Moreover, it can also be shown that in this limit the bound can be saturated (by a proper estimator, for instance, maximum-likelihood), and hence  $\text{Var}[\hat{\theta}] \sim \frac{1}{nI(\theta)}$ , a situation known as *shot-noise limit*.

To conclude, we note that similar to hypothesis testing, the problem of parameter estimation gets a twist in the quantum realm, where again a non-trivial optimization over measurements (and probes) is involved. Moreover, a new object known as the *quantum Fisher information* appears in the stage, accompanied by a *quantum Cramer-Rao bound*. We will not discuss this topic, but refer the interested reader to Refs. [Mey21; Hel76; GLM06].

## 2.6 Reinforcement learning

The framework of Reinforcement Learning (RL) is based on the sequential interaction between an agent and an environment during several episodes, and — as many ideas in artificial intelligence — is strongly inspired in biological models of learning [SG18; Sze10a].

At each time-step  $\ell = 0, \dots, L$  of each episode  $t = 1, \dots, T$ , the agent observes its environment in a *state*  $s_\ell^{(t)} \in \mathcal{S}$  and chooses an *action*  $a_\ell^{(t)} \in \mathcal{A}$ ; as a consequence, the agent enjoys a *reward*  $r_{\ell+1}^{(t)} \in \mathcal{R}$  and observes a new state of the environment,  $s_{\ell+1}^{(t)} \in \mathcal{S}$ ; here  $\mathcal{S}$ ,  $\mathcal{A}$  and  $\mathcal{R}$  stand for the sets of states, actions and rewards the agent may experience. Such reward serves as a classical conditioning to the agent, who needs to train its behaviour in order to acquire as much rewards as possible, in a similar way we reward our dogs with soome cheese if they sit after saying “*seu*”<sup>40</sup>. In this sense, one might steer the agent towards optimal desired behaviours which are actually unkown to the designer of the reinforcement learning scenario: while we have the free will to define states, actions, rewards and envionment dynamics, it is highly unlikely that optimal behaviours that lead to high-rewards are known beforehand. This is the spirit of tackling open problems in science with reinforcement learning algorithms, where unknown solutions can readily be found by letting the sequential interaction to evolve over several episodes.

The environment is usually modeled to be Markovian: its dynamics is completely determined by the last time-step via the transition function  $\tau(s', r|s, a)$ , *i.e.*, the conditional probability of ending up in a state  $s'$  and conferring a reward  $r$ , given

---

<sup>40</sup>*Seu* is *sit* in catalan.

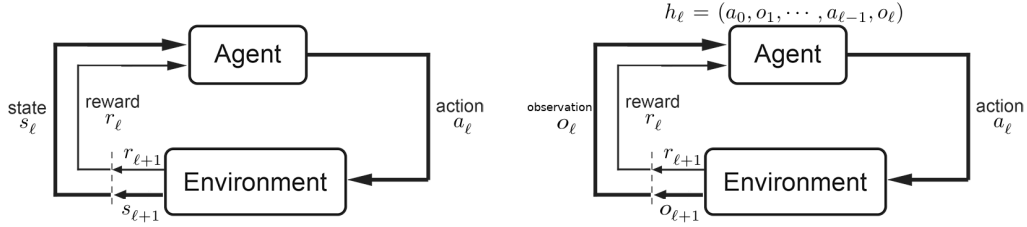


Figure 2.9: We depict the interaction with agent and environment in fully (left) and partially observable (right) cases.

that the previous state was  $s$  and the agent took an action  $a$ ; the next future states accessible from  $s$  are thus restricted to  $\mathcal{S}(s) = \{s' : \tau(s'|s) \neq 0\} \subseteq \mathcal{S}$ .

In general, the agent does not have control of nor access to the transition function, but it will influence the dynamics of the environment by choosing actions according to an interaction *policy*  $\pi(a|s)$ , i.e., the conditional probability of performing an action  $a$  when the observed environment's state is  $s$ ; hence the available actions at a given state may be restricted to a subset  $\mathcal{A}(s) \subseteq \mathcal{A}$ . This setting is usually known as a Markov decision process (MDP).

The agent's objective is to interact with the environment through an *optimal policy*  $\pi^*$ , for which the total reward acquired during an episode is as high as possible. In the dog-training case, this means getting as much cheese as possible, implying correctly execution of all the tricks. To achieve such goal, a *value function* is assigned to each state and optimized over all possible policies, as explained in Sec. 2.6.2.

The Markov assumption is justified whenever the agent's observations provide a complete description of the state of the environment  $s_\ell$ . However, in general this is not the case, and the agent has only access to *partial observations*  $o_\ell \in \mathcal{O}$  at each time-step. This leads the agent to define its own state, e.g. agent's state. While in the fully observable case agent's and environment's states coincide, in partially-observable environments the agent can only rely in such partial observations in order to learn. These observations would not allow to determine the dynamics even if  $\tau$  was known, and they are generated from the current (environment) state and the previous interaction history (e.g. actions performed, observations acquired and environments' states). In RL literature this is known as a partially-observable MDP (POMDP) and developing methods to solve it efficiently constitutes an active area of research [SJJ94; MKS+13; SPK13; Ego15; ZLP+18]; the problem is usually tackled by first reducing it to an effective MDP. The most straightforward approach is to define an effective state that contains all the past history of observations and

actions up to a given time-step, *i.e.*,  $h_\ell = (a_0, o_1, \dots, a_{\ell-1}, o_\ell)$ . In this way, the dynamics observed by the agent can always be described by an effective MDP with transition function  $\tau(h', r|h, a)$ , which is unknown to the agent and determined by the underlying environmental transition function. We will choose such approach when framing quantum-state discrimination as reinforcement learning problem in Chapter 3.

In order to gain some intuition on both the reinforcement learning problem and its difficulty, it is instructive to revisit bandit problem. Such problem is a simplified version of the reinforcement learning setting, that yet captures many of the challenges underlying model-free learning frameworks. After discussing about bandits, we move to define value functions and study their associated Bellman equations in Sec. 2.6.2, whereas in Sec. 2.6.3 we introduce the Q-learning algorithm, used to learn optimal policies in model-free settings. We finally comment on model-aware approaches such as dynamic programming in Sec. 2.6.4.

### 2.6.1 Bandit problems

The multi-armed bandit problem constitutes a simplified version of the RL setting. Here, an agent is faced towards a set of  $N$  slot-machines, and each action  $a \in \mathcal{A} = \{1, \dots, N\}$  stands for the pulling arm of the  $a$ -th machine, which retrieves a reward  $r \in \mathcal{R}$  with an unknown probability  $\tau(r|a)$ . Each episode starts from a single default state, the bandit can only pull one machine at the episode, which ends right after obtaining the reward; thus the MDP is reduced to a Markov Reward Process (MRP). Though it is not clear who the actual *bandit* is, the situation models a gambler trying to maximize its earnings in a casino. Note also that we restrict to the case in which arms' distributions are stationary, *e.g.* do not change in time.

The bandit problem is ideal to discuss our figures of merit when it comes to model-free learning optimal policies, and in particular to characterize learning curves behaviour. On the one hand, we would like the agent to identify the arm which – on average – leads to the highest reward. If the agent knew beforehand all reward distributions  $\tau(r|a)$ , then it would readily know which action to take. Nevertheless, since no access is granted to those distributions (for otherwise that would be a very generous casino), we then monitor how reward acquisition evolves *during* the learning process. In this regard, a reasonable figure of merit is the cumulative reward

$$\mathbf{R}_t(\pi) = \frac{1}{t} \sum_{\nu=1}^t r_\nu \tag{2.233}$$

where  $r_\nu$  is the reward enjoyed by the agent at episode  $\nu$ . As such, the cumulative reward is a stochastic quantity, and depends on agent's policy  $\pi$  (*i.e.* the way it decides for which arm to sample at a given episode). Whenever it is clear from the context, we will drop the dependence on  $\pi$  and simply denote the cumulative reward as  $\mathbf{R}_t$ . Note that the cumulative reward is upper-bounded (on average) by that one associated to the optimal policy, which always samples from  $a^* = \operatorname{argmax}_{a \in \mathcal{A}} Q(a)$ , with  $Q(a) = \mathbb{E}_r[\tau(r|a)]$ . Such expected values  $Q(a)$  quantify how *valuable* each action  $a$  is, and are the analogue of state-action value functions in MRP — a slightly more complex quantity that we will define in Sec. 2.6.2 —. While the agent is not aware of those values, successful learning hinges on how accurately can the agent discriminate between such quantities, and in particular find the optimal one. To this end she updates and estimate of such quantity as new samples are obtained, which we denote by  $\hat{Q}(a)$ . In particular, bandit theory defines the so-called *expected cumulative regret*:

$$\mathcal{L}_t = \mathbb{E} \left[ \sum_{k=1}^t \left( Q(a^*) - Q(a^{(k)}) \right) \right] = t \left( Q(a^*) - \mathbb{E}[\mathbf{R}_t] \right), \quad (2.234)$$

where  $\mathbb{E}$  indicates the expected value with respect to different agents (that experience different realizations of the sampling processes) following the same policy  $\pi$ , and  $a^{(t)}$  is the action actually taken by the agent at episode  $t$ . The cumulative regret is closely related to the (expected) cumulative reward per episode  $\mathbf{R}_t$ , and quantifies the price to pay, or loss, for taking actions different from the optimal one  $a^*$ . In other words, it quantifies the difference in earnings of the agent with respect to those of a model-aware super-agent, which owns the casino and hence has access to  $a^*$ .

One of the most fundamental results in bandit theory is the Lai-Robbins bound [LR85] for the asymptotic behaviour of the expected cumulative regret:

$$\mathcal{L}_t \underset{t > 1}{\gtrsim} \log t \left( \sum_{a \in \mathcal{A} \setminus \{a^*\}} \frac{\Delta_a}{\text{KL}(a||*)} + o(1) \right) := C_{\text{LR}} \log t, \quad (2.235)$$

with  $\Delta_a = Q(a^*) - Q(a)$  and  $\text{KL}(a||*)$  the Kullback-Leibler divergence between the reward distributions  $\tau(r|a)$  and  $\tau(r|a^*)$ .

Recalling that the agent is unaware of the underlying distribution  $\tau(r|a)$ , we note that  $\mathbf{R}_t$  is a figure of merit that detects whether the learning behaviour has improved or not, and is readily available to the agent. In particular, it captures the entire learning process, and measures how well the agent has balanced between

exploring potentially optimal (yet undersampled) arms, or exploiting arms that she consider the best, according to the statistics gathered up to episode  $t$ . Such tradeoff is known as the *exploration-exploitation tradeoff*, and is one of the core concepts in RL theory. It also captures the relevance of bandit problems in real-life applications, where the final success probability of the protocol is not the only figure of merit, but the whole learning process counts. For example, in clinical trials [Tho33] one needs to find the right compromise between advancing in the search of the best treatment (*exploring*) while effectively treating current patients (*exploiting*). Moreover, if one as a scientist is aware of the setting, and aims to monitor the *average* learning behaviour of the agent, the expected regret is readily accessible (provided that enough realisations of the learning process could be simulated), and we know from Lai-Robbins bound that such quantity is asymptotically bounded. In this regard, the general traits of the cumulative reward per episode  $\mathbf{R}_t$  could inspire several ways of quantifying the performance of the learning agent, e.g., (a) the onset episode at which  $\mathbf{R}_t$  starts exceeding a completely random policy (which samples a random arm unconditioned on the history of past rewards acquired); (b) the transient episode at which  $\mathbf{R}_t$  reaches a given fraction of its upper bound  $\mathbf{R}_t(\pi^*)$ ; (c) the learning speed as quantified by the slope of  $\mathbf{R}_t$  after the onset episode; and so on. While little is known on how such traits behave, the expected regret constitutes a route to design policies in bandit problems, which are considered *good* ones if they asymptotically saturate the Lai-Robbins bound. Thus, bandit theory provides us with a framework where some of these notions can be rigorously studied, though extending such formal approach to the general MDP is a challenging task, and constitutes an active area of research [LS18; AOM17; TP21; Liu22; DWC+19].

We will now turn to review some well-known policies that are used in bandit problems, and that will find use in this thesis when dealing with reinforcement learning scenarios (see Sec. 3.3).

Recall that at each episode, the agent keeps an estimate  $\hat{Q}(a)$  of how valuable taking each action is, by estimating the mean reward it provides. Here, the non-trivial question is which arm to try, given the experience gathered so far. The most straightforward policy to use is the  $\epsilon$ -greedy, which is outlined in Algorithm 1, and consists on going greedy (that is, choosing action  $\operatorname{argmax}_{a \in \mathcal{A}} \hat{Q}(a)$ ) with probability  $\epsilon$ , or to randomly choose an action with probability  $1 - \epsilon$ . After enjoying the reward, a Monte-Carlo like update is performed on  $\hat{Q}(a)$ , which sequentially updates such average value according to some learning-rate  $\lambda_t(a)$ , and which might depend on both the arm label and the episode number. While small values of  $\epsilon$  will favour potentially sub-optimal actions that were discovered by the agent in early episodes, large values of  $\epsilon$  imply a random behaviour, and thus a low reward acquisition. In



general, the value of  $\epsilon$  is modified ad-hoc with the episode number.

---

**Algorithm 1:** We detail the  $\epsilon$ -greedy policy for bandit problems.

---

```

input :  $\hat{Q}(a)$  arbitrarily initialized and learning rates  $\lambda_t(a) \in (0, 1]$ 
          $\forall a \in \mathcal{A}, \epsilon \in (0, 1]$ 
1 for  $t$  in  $1 \dots T$  do
2   generate a random number  $j$ 
3   if  $j \leq \epsilon$  then
4     choose  $a$  at random
5   else
6     choose  $a = \operatorname{argmax}_{a \in \mathcal{A}} \hat{Q}(a)$ 
7   observe  $r$ 
8   update  $\hat{Q}$ :
9      $\hat{Q}(a) \leftarrow \hat{Q}(a) + \lambda_t(a)[r - \hat{Q}(a)]$ 

```

---

Note that by choosing the learning rates  $\lambda_t(a)$  to be the inverse of the number of times action  $a$  was visited up to time  $t$ , then

$$\hat{Q}(a) \xrightarrow{t \rightarrow \infty} \sum_{r \in \mathcal{R}} r \tau(r|a) = Q(a) \quad \forall a \in \mathcal{A}. \quad (2.236)$$

Moreover, an  $\epsilon$ -greedy strategy can not attain the logarithmic behaviour for the  $\mathcal{L}_t$  since  $\mathbb{E}[\mathbf{R}_t] < Q(a^*)$  for all  $t$ , because at every episode there is a finite probability  $\epsilon$  that the agent performs a suboptimal action, and thus the expected cumulative regret grows linear with  $t$ . It is then clear that there is room for improvement before saturating the bound in Eq. (2.235).

In what follows we will present two strategies, one based on Upper Confidence Bounds (UCB) [LR85; Agr95; ACBF02] and the other based on Thompson sampling (TS) [Tho33; Tho35; Sco10; RVK+18], which substantially improve the performance of  $\epsilon$ -greedy and even attains the asymptotic logarithmic behaviour for  $\mathcal{L}_t$  [KKM12].

In UCB, the agent keeps a record of the number of times each action  $a$  was selected up to episode  $t$ , which we denote as  $N_t(a)$ . Hoeffding's inequality bounds the probability that the  $\hat{Q}(a)$  underestimates the true value of  $Q(a)$  by more than  $\epsilon(t) > 0$ , as

$$\Pr[\hat{Q}(a) < Q(a) - \epsilon(t)] \leq e^{-2N_t(a)\epsilon(t)^2} =: \mathcal{P}(t). \quad (2.237)$$

Then, for  $N_t(a) > 0$ , the upper confidence bound, defined as

$$\text{ucb}_t(a) := \hat{Q}(a) + \epsilon(t) = \hat{Q}(a) + \sqrt{\frac{-\log \mathcal{P}(t)}{2N_t(a)}}, \quad (2.238)$$

represents an upper bound to the true value  $Q(a)$  with a high probability  $1 - \mathcal{P}(t)$ . This value is used to compare and choose among the different actions, *i.e.*  $a = \operatorname{argmax}_{a \in \mathcal{A}} \operatorname{ucb}_t(a)$ , and responds to the motto *optimism under the face of uncertainty*: actions that have not been visited enough are assigned an *optimistic* estimate value and hence more chances of being picked; in addition, actions whose Q-estimate is accurate but sub-optimal will have little chances to be picked again. The functional form of  $\mathcal{P}(t)$  can be tuned to balance exploration and exploitation. In particular, for the standard choice  $\mathcal{P}(t) = t^{-4}$  it can be proven that the expected cumulative regret  $\mathcal{L}_t$  follows the logarithmic scaling [ACBF02; LS18].

---

**Algorithm 2:** UCB for bandit problems.

---

```

input :  $\mathcal{P}(t)$ , initialize  $\hat{Q}(a), N(a)$  to zero  $\forall a \in \mathcal{A}$  .
1 for  $t$  in  $1, \dots, T$  do
2   |   if  $t \leq |\mathcal{A}|$  then
3   |   |   (choose each action once)  $a = t$ 
4   |   else
5   |   |   choose  $a = \operatorname{argmax}_{a \in \mathcal{A}} \operatorname{ucb}_t(a)$ , (Eq. (2.238))
6   |   observe reward  $r$ 
7   |   record visit:
8   |   |    $N_t(a) \leftarrow N_t(a) + 1$ 
9   |   update Q-value:
10  |   |    $\hat{Q}(a) \leftarrow \hat{Q}(a) + \frac{[r - \hat{Q}(a)]}{N_t(a)}$ 

```

---

Instead of updating an estimate  $\hat{Q}(a)$  for each action, in Thompson sampling (TS) a Bayesian approach is followed, where at every episode a full prior distribution (and not just an expectation value) is assigned to the *expected reward*  $\bar{r}$  of every arm  $a$ ,  $f_t(\bar{r}|a) \forall a \in \mathcal{A}$ . This distribution characterizes the knowledge the bandit has about the expected earnings of each arm,  $Q(a)$ , and at the first episode can be taken to be flat over the whole interval  $[0, 1]$ . The policy then consists in sampling an expected reward  $\bar{r} \sim f_{t-1}(\bar{r}|a)$  for each possible action  $a$  and choosing the action with the largest sample  $\bar{r}$ :  $a = \operatorname{argmax}_{a \in \mathcal{A}} \{\bar{r} \sim f_t(\bar{r}|a)\}$ ; such sampling procedure constitutes an overhead for the bandit, since at each episode  $N$  extra samples are required, which are nonetheless unrelated to the arms (*i.e.* at each episode, the bandit still samples only a single arm). Finally, the distribution for the chosen action is updated according to the true reward  $r$  obtained, using Bayes' theorem.

In order to avoid computationally-expensive Bayesian updates, families of

distributions that are closed under the update rule are used. In the case of Bernoulli bandits, beta-distributions are employed since those are precisely their conjugate priors. That is, given

$$f_t(\bar{r}|a) = \text{Beta}(\mu_t(a), \nu_t(a)) \propto \bar{r}^{\mu_t(a)-1} (1 - \bar{r})^{1-\nu_t(a)}, \quad (2.239)$$

upon obtaining a reward  $r$  the prior is updated to a beta distribution with parameters  $\mu_{t+1}(a) = \mu_t(a) + r$ ,  $\nu_{t+1}(a) = \nu_t(a) + 1 - r$ , where at the first episode it is  $\mu_1(a) = \nu_1(a) = 1 \quad \forall a \in \mathcal{A}$  (flat prior). By mimicking the underlying distributions, TS gauges exploration according to the information acquired so far: if a certain action has not been sampled enough at episode  $t$ , its reward distribution will still be broad and, when sampled, can easily return a higher value of  $\bar{r}$  than that obtained from other (more peaked) distributions; thereby TS will favour to explore such action. At the same time, if a sub-optimal action has been sampled enough episodes, it will be very unlikely that it is sampled again, since the corresponding prior will be highly peaked at low values.

TS algorithm has been introduced surprisingly long ago [Tho35], and it is known to asymptotically attain the Lai-Robbins bound [AG11]. Moreover, its performance has been studied in non-asymptotic regimes [GMS19], where it was shown that sub-leading constants and terms of order  $\log(\log t)$  might be important. Finally, a review about TS and its applicability can be found in Ref. [RVK+18].

Let us conclude this overview of bandit theory by introducing the *simple regret*, another widely used figure of merit that quantifies how well has the agent learned

---

**Algorithm 3:** TS for Bernoulli bandit problems.

---

**input** :  $\mu_1(a), \nu_1(a)$  initialized to one  $\forall a \in \mathcal{A}$

```

1
2 for  $t$  in  $1, \dots, T$  do
3
4     for  $a$  in  $\mathcal{A}$  do
5         |         draw  $\bar{r}_a$  according to  $\text{Beta}(\mu_t(a), \nu_t(a))$ 
6
7     choose  $a = \underset{a}{\text{argmax}} \bar{r}_a$ 
8     observe reward  $r$ 
9     update Beta distribution:
10          $\mu_{t+1}(a) = \mu_t(a) + r$ 
11          $\nu_{t+1}(a) = \nu_t(a) + 1 - r$ 

```

---

to identify the optimal action at episode  $t$ , regardless of her actual performance:

$$\Lambda_t = \mathbb{E}(Q(a^*) - Q(a^{(t)*})), \quad (2.240)$$

where  $a^{(t)*}$  is the agent's *recommendation* of which the optimal action is at episode  $t$ . For example, in an  $\epsilon$ -greedy strategy, the recommendation is given by  $a^{(t)*} = \underset{a}{\operatorname{argmax}} \hat{Q}(a)$ . Thus, strategies designed to minimize  $\Lambda_t$  will prioritize to learn what the optimal arm to pull is, regardless of the rewards acquired during the entire learning process. Recent results [BMS11] show that the exploitation-exploration trade-off manifests itself in the asymptotic scaling of the simple and cumulative regret in the sense that one imposes lower and upper-bounds on the other, and therefore optimizing one usually affects the performance of the other.

Having gained some intuition on why model-free learning schemes are complex, we will not return to study the general MDP setting.

## 2.6.2 Value functions and the Bellman equation

The complexity added in MDPs, as compared to bandit problems (MRPs), is that agent and environment sequentially interact during (a possibly infinite number of) time-steps. Similar to the bandit case, agent's objective is to acquire as much reward as possible during an episode, and as a matter of fact this strongly depends on the policy that the agent follows. At the end of episode  $t$ , in which a sequence of tuples  $\{(s_\ell, a_\ell, r_{\ell+1})\}_{\ell=0}^L$  has been experienced (with  $s_{L+1}$  a *terminal state*, and  $L$  generally varying among different episodes), the agent's performance after each time-step  $\ell$  is evaluated using the so-called *return*,

$$G_\ell^{(t)} = \sum_{i=0}^{L-\ell} \gamma^i r_{i+\ell+1}^{(t)}, \quad (2.241)$$

which is the weighted sum of rewards obtained at all future time-steps, with a *discount factor*  $\gamma \in (0, 1]$ , weighting more the rewards that are closer in the future. Note that for infinite-horizon MDPs, i.e.,  $L \rightarrow \infty$ , it must hold  $\gamma < 1$  to ensure that  $G_\ell$  remains finite.

By introducing the return, it is straightforward to assign a value to a state  $s$  for a given interaction policy  $\pi$ , via the so-called *state value function*:

$$v_\pi(s) = \mathbb{E}\pi[G_\ell | s_\ell = s], \quad (2.242)$$

which is the expected return over all possible trajectories that start from state  $s$ , take actions according to policy  $\pi$  and whose dynamics is governed by  $\tau$ . In other

words, the value function measures how convenient it is to visit state  $s$  when policy  $\pi$  is being followed, and thus provides a route to qualify the goodness of such policy. Note that this quantity is completely determined by the future trajectories accessible from  $s$  and hence its dependence on the time-step  $\ell$  can have at most the effect of restricting the set of states on which  $v_\pi(s)$  is supported at that time; we keep this dependence implicit unless otherwise stated. By writing explicitly the expected value for the first future time-step in Eq. (2.242) and then applying the definition of  $v$  recursively, it is easy to show that the state-value function satisfies, for any policy, the *Bellman equation* [Bel03]:

$$v_\pi(s) = \sum_{\substack{s' \in \mathcal{S}, r \in \mathcal{R} \\ a \in \mathcal{A}}} \tau(s', r | s, a) \pi(a | s) (r + \gamma v_\pi(s')). \quad (2.243)$$

This equation relates the value of a state  $s$  with that of its nearest neighbours  $s'$ , which can be reached with a single action from  $s$ , and with the corresponding reward obtained by performing such action.

Since value functions assign a score to the policy  $\pi$  being followed, a possible route to solve the reinforcement learning problem is that of maximizing state-value functions. Specifically, the optimal policy  $\pi^*$ , maximizes the state-value function for each  $s$  and thus represents the *optimal value function*, which satisfies a particular Bellman equation, which is known as the *optimal Bellman equation*:

$$v^*(s) := v_{\pi^*}(s) = \max_{\pi} v_\pi(s) \quad (2.244)$$

$$= \max_{a \in \mathcal{A}} \sum_{s' \in \mathcal{S}, r \in \mathcal{R}} \tau(s', r | s, a) (r + \gamma v^*(s')). \quad (2.245)$$

In an hypothetical situation where optimal value functions are available, an optimal policy can readily be constructed by selecting an action that, from each state, takes the environment towards the next state whose state-value function is the highest. Nevertheless, to do so one needs to have a precise mapping between actions and next-states, an information which is missing if environment dynamics  $\tau(s' | s, a)$  is unavailable to the agent.

For this reason, we define the state-action value function (or Q-function, or Q-value) as the expected return when starting from state  $s$  and performing action  $a$ :

$$Q_\pi(s, a) = \mathbb{E}_\pi [G_\ell | s_\ell = s, a_\ell = a], \quad (2.246)$$

which is related to the state-value function by  $v_\pi(s) = \sum_{a \in \mathcal{A}} \pi(a | s) Q_\pi(s, a)$ . Whenever it is clear from the context we will drop the dependence on the policy  $\pi$

for  $Q(s, a)$ . Similar to the state value function, by writing explicitly the expected value for the first future time-step, the Bellman equation for state-action value function reads:

$$Q_\pi(s, a) = \sum_{\substack{s' \in \mathcal{S}, r \in \mathcal{R} \\ a \in \mathcal{A}}} \tau(s', r | s, a) (r + \gamma \pi(a' | s') Q_\pi(s', a')).$$

Importantly, the optimal policy  $\pi^*$  can also be obtained by maximizing the Q-function, with a corresponding optimal Bellman equation

$$\begin{aligned} Q^*(s, a) &:= Q_{\pi^*}(s, a) = \max_{\pi} Q_\pi(s, a) \\ &= \sum_{s' \in \mathcal{S}, r \in \mathcal{R}} \tau(s', r | s, a) (r + \gamma \max_{a' \in \mathcal{A}} Q^*(s', a')). \end{aligned} \quad (2.247)$$

Note that the state-action value function  $Q(s, a)$  provides a natural generalization of the expected reward  $Q(a)$  to the case where agent and environment interact sequentially during an episode, where the necessity of defining environment states arises. In the latter case, the Bellman equation trivially reduces to only the first term in the right-hand side of Eq. 2.243.

We are now in position to understand the complexity of the reinforcement learning problem. While agent and environment begin to interact through some policy  $\pi$ , state-action value functions are not available to the agent. In turn, several repetitions of such interaction are required in order to assess which is the relevant subspace of states and actions (recall that environment dynamics  $\tau(s', r | s, a)$  might be stochastic). Since the agent can not do better than randomly sampling actions at the beginning of the learning process, she needs to wait a transient time — which is related to the probability of randomly arriving to a high-reward region of the state-action space — until some reward signal that might guide the search is experienced. Once obtained a reward signal, a balance between exploring new regions of the state-action space, and to better estimate current state-action value functions should be made. Here, even if an action deterministically maps a state  $s$  to a state  $s'$ , the agent needs several episodes in order to estimate value functions  $Q_\pi(s, a)$ , since those depend on the accessible region of the state-action space under policy  $\pi$ , once  $s'$  has been reached, as explicited in the Bellman equation of Eq. 2.247. Moreover, such policy should be modified in order to maximize the expected return. Overall, reaching a reasonable balance between exploration and exploitation in such scenarios is certainly challenging, particularly if the agent has resource constraints such as a finite number of episodes available in order to learn a reasonably good policy.

In this regard, while the model-free assumption is strong, it often happens that such agents discover unknown protocols solving the problem at hand. While the solutions found by RL might not be optimal, they can be better than any previously known one. Moreover, the versatility of the reinforcement learning setting (in particular when defining states, actions and rewards) constitutes an extremely appealing tool the discovery of ansatz to problems that are challenging to tackle (or even to model), as we do in Chapter 3.

In what follows we will detail one out of the – very many – algorithms that can be used in order to learn optimal policies in model-free, reinforcement learning scenarios.

### 2.6.3 Q-learning

In the model-free setting, the agent not only has to find an optimal policy by exploiting valuable actions, but also needs to characterize the environment in the first place by exploring possibly advantageous configurations. In such a case, the Q-value is quite helpful since it associates a value to the transitions determined by taking action  $a$  from state  $s$  and following policy  $\pi$  thereafter.

Q-learning is an algorithm that can be used to learn optimal policies in a model-free way, and it was first proposed by Watkins [Wat89]. This algorithm is often used as a basis for more advanced RL algorithms [MKS+13; SLH+14]. It is based on the observation that any Bellman operator, i.e., the operator describing the evolution of a value function as in Eqs. (2.243,2.244,2.247), is contractive [Sze10b]. This implies that, under repeated applications of a Bellman operator, any value function converges to a fixed point, which by construction satisfies the corresponding Bellman equation. Thus, in order to find  $Q^*(s, a)$ , Q-learning turns the optimal Bellman equation for Q, Eq. (2.247), into an update rule for  $\hat{Q}(s_\ell, a_\ell)$ , i.e., the Q-function's estimate available to the agent at a given time-step  $\ell$  of any episode  $t = 1, \dots, L$ .

After an interaction step  $s_\ell \rightarrow a_\ell \rightarrow r_{\ell+1} \rightarrow s_{\ell+1}$  is experienced, the update rule for the Q-estimate is

$$\hat{Q}(s_\ell, a_\ell) \leftarrow (1 - \lambda_t(s_\ell, a_\ell))\hat{Q}(s_\ell, a_\ell) \tag{2.248}$$

$$+ \lambda_t(s_\ell, a_\ell) \left( r_{\ell+1} + \gamma \max_{a' \in \mathcal{A}(s_{\ell+1})} \hat{Q}(s_{\ell+1}, a') \right), \tag{2.249}$$

where  $\lambda_t(s, a)$  is the learning rate, which depends on the number of times the state-action pair  $(s_\ell, a_\ell)$  has been visited. Note that in order to do the update at each time-step  $\ell$ , it is only necessary to enjoy the next immediate reward  $r_{\ell+1}$  and

observe the next state  $s_{\ell+1}$ ; this method thereby allows an on-line learning of the MDP. A pseudo-code of the algorithm can be found below.

---

**Algorithm 4:** Q-learning pseudo-code.

---

```

input :  $\hat{Q}(s, a)$  arbitrarily initialized  $\forall s \in \mathcal{S}, \forall a \in \mathcal{A}(s)$ ;
         learning rates  $\lambda_t(s_\ell, a_\ell) \in (0, 1], \epsilon > 0$ 
output:  $\hat{Q}(s, a) \sim Q^*(s, a)$ 
1
2 for  $t$  in  $1 \dots T$  do
3   initialize  $s_0$ 
4   for  $step \ell$  in  $episode t$  do
5     take action  $a_\ell$  according to  $\pi$  (e.g.  $\epsilon$ -greedy)
6     observe reward  $r_{\ell+1}$  and next state  $s_{\ell+1}$ 
7     update  $\hat{Q}(s_\ell, a_\ell)$  according to:
8        $\hat{Q}(s_\ell, a_\ell) \leftarrow \hat{Q}(s_\ell, a_\ell) + \lambda(s_\ell, a_\ell)[r_{\ell+1} +$ 
9          $\gamma \max_{a'} \hat{Q}(s_{\ell+1}, a') - \hat{Q}(s_\ell, a_\ell)]$ 
10      if  $s_{\ell+1}$  is terminal state then
11        break
12      else
13         $s_\ell \leftarrow s_{\ell+1}$ 

```

---

After a large number  $n$  of iterations of the update rule Eq. (2.248) for all state-action couples, the convergence of the Q-estimate to the optimal Q-function is guaranteed by two general conditions on the learning rate (also known as Robinson conditions) [Wat89; SG18]:

$$\hat{Q}(s, a) \xrightarrow[k \rightarrow \infty]{} Q^*(s, a) \quad \forall s \in \mathcal{S}, a \in \mathcal{A}(s) \quad (2.250)$$

$$\text{iff } \sum_{t(s,a)} \lambda_t(s, a) = \infty, \quad \sum_{t(s,a)} \lambda_t(s, a)^2 < \infty, \quad (2.251)$$

where the sums are taken over all interactions at which a given state-action couple is visited. Once the optimal Q-function is obtained, an optimal (deterministic) policy can be constructed by “going greedy” with respect to it, i.e.,  $\pi^*(a|s) = \delta(a, \underset{a \in \mathcal{A}}{\operatorname{argmax}} Q^*(s, a))$  for all  $s \in \mathcal{S}$ , where  $\delta(x, y)$  is a Kronecker delta.

In RL literature, Q-learning is classified as an off-policy method [SG18], meaning that it learns the state-action values of a *target policy* - in this case the optimal



policy - by taking actions according to an *interaction policy*, generally differing from the first one. The standard Q-learning method commits to an  $\epsilon$ -greedy interaction policy, where with probability  $\epsilon$  the agent chooses a random action and otherwise it chooses the greedy action that maximizes the current Q-estimate. However, as we will see in Sec. 3.3.4, more general strategies can be considered, including a clever search as inspired by UCB or TS algorithms.

Moreover, Q-learning is also classified as a tabular method [SG18], meaning that it does not approximate or infer the value function of state-actions that have not been visited. In the latter case, it is customary to use an artificial neural network in order to infer the value of the Q-function, and much of the recent success of reinforcement learning algorithms in discovering new protocols for a wide variety of problems hinges on such networks.

There is a plethora of reinforcement learning algorithms that has been proposed according to specific constraints imposed in the setting, and there is generally no best algorithm; on the contrary – as it happens in real-life – the success of an idea generally depends on the context it is carried out. In general, a combination of them needs to be done in order to attain successful learning and to prove an advantage over state-of-the-art techniques, when tackling a problem from the reinforcement learning perspective. Before closing up this section in reinforcement learning, let us comment on the case where a model of environment dynamics is available to the agent.

#### 2.6.4 Model aware methods

In the model-aware scenario, the agent has access to the environment dynamics  $\tau(s'r|s, a)$ . Nonetheless, the question of how to design an optimal policy still holds: while estimating value functions out of samples is no longer required, the agent still needs to solve the Bellman equation in Eq. 2.243 in order to compute  $v_\pi(s)$  for  $s \in \mathcal{S}$ . Moreover, the policy optimization challenge remains, and solving it can be highly non-trivial. Such a problem is known as planning [SG18], and can be solved for finite-horizon MDPs via dynamic programming methods.

Here we distinguish between two approaches. On the one hand, there is the sequential approach, which exploits the structure of Bellman equation and solves the policy optimization by splitting the global policy optimization into a sequence of smaller optimization problems [Bel03]. On the other hand, iterative approaches can be formulated for the policy optimization problem, such as value and policy iteration algorithms. The latter approach exploits the contractive property of state value function, and are slightly more general than the methods we consider here,

since they can readily be applied in scenarios where  $L$  (the number of time-steps in each episode) is not fixed [SG18]). In particular, the state value function can be seen as a fixed point of a contractive operator, called the Bellman operator [Sze10a]. While policy iteration algorithms consist on a loop where value functions are estimated through such contractiveness property and a policy is modified in such a way to optimize such values, value iteration algorithms straightforwardly apply the optimal Bellman operator in order to compute  $v^*(s) \forall s \in \mathcal{S}$ ; more insight can be gained in [this online demonstration](#).

In this thesis we restrict to a sequential optimization of the state value, since we will apply RL algorithms to a problem with a well-defined, sequential structure, namely the calibration of Dolinar receivers. We thus follow the method introduced by Bellman [Bel03], which makes use of the recursive relation of Eq. (2.244) to find the optimal policy step by step, and for this we assume that every episode deterministically ends at a fixed time-step  $L$ , and denote by  $v_\ell^*(s)$  the optimal value function of state  $s$  at time-step  $\ell$ . Here, the so-called *principle of optimality* holds, which is nothing than the optimal value equation in Eq. 2.247, and states that an optimal policy can be obtained by going greedy sequentially, *e.g.* step by step. Hence, since the optimal policy consists in taking the best possible action from any given state, it can be constructed by concatenating optimal (and local) policies at each time-step: we start by solving Eq. (2.244) at the last time-step,

$$v_L^*(s) = \max_{a \in \mathcal{A}(s_L)} \sum_{r \in \mathcal{R}} \tau(r|s, a)r, \quad (2.252)$$

where we used the fact that  $v_{L+1}(s) = 0$ . The solution to Eq. (2.252) provides the optimal action at step  $L - 1$  for each  $s$  and the optimal value function  $v_L^*(s)$ . Then we plug the latter into the optimal Bellman equation for the previous time-step, which in turn can be solved to obtain the optimal action and value function  $v_{L-1}^*(s)$  for every  $s$ . By repeating this procedure iteratively for each time-step  $\ell = L, \dots, 0$ , we can obtain the optimal sequence of actions and value functions for any state at any time-step. This an approach has the nice interpretation of constructing the solution of a large problem by solving smaller problems, each stage essentially adding some complexity into the problem at hand.

The difficulty in this method is that the optimization over actions at a given state *i.e.*  $\max_{a \in \mathcal{A}(s_\ell)}$  might be find problems. As a matter of fact, we will find difficulties in such optimization landscape when calibrating coherent-state receivers in the presence of noisy channels (see Sec. 3.3.5.)

### 2.6.5 Discussion

In this section we have introduced some basic concepts in reinforcement learning, which can readily be used by an agnostic agent in order to learn *good* behaviours when acting in unknown environments. The goodness of such behaviours is formally quantified by the amount of reward that the agent gets, on average, when departing from its initial state (or the average over possible states, if there is a non-vanishing probability of the interaction being initialized in more than one state). For this, we have introduced value functions, which are the expected return (*i.e.* a weighted sum of stochastic rewards obtained during an episode) that the agent enjoys when following policy  $\pi$ , and by doing so we have defined an order relation among policies, *i.e.* a policy  $\pi'$  is better than a policy  $\pi$  if  $v_\pi(s) < v_{\pi'}(s) \forall s \in \mathcal{S}$ . In particular, the agent should seek to perform optimally, and that guarantees her to attain as much rewards as possible, on average, during an episode. We have discussed some methods that the agent can use in order to find such optimal policies, through the optimal state value functions, in Sec. 2.6.4. Nonetheless, such methods can only work if a perfect model of the environment dynamics is available to the agent, that is, if the agent knows exactly which is the probability of enjoying a reward  $r$ , and observe its environment transitioning to state  $s'$ , when departing from state  $s$  and performing action  $a$ . To gain some intuition on the complexity that model-free scenarios carry on, we have stepped back and discussed a simplified setting in Sec. 2.6.1, known as the multi-armed bandit problem, in which we introduced several strategies that balance between exploring how good untaken actions are, and exploiting potentially sub-optimal yet high-reward-retrieving actions; in particular we studied  $\varepsilon$ -greedy, UCB and TS strategies. We then moved to the general MDP case in Sec. 2.6.2, where the usefulness of the state-action value functions  $Q(s, a)$  was highlighted. Finally, we introduced an algorithm that allows the agent to learn optimal policies, called Q-learning, in Sec. 2.6.3.

These concepts and methods will find use in Ch. 3, where a reinforcement learning agent is asked to calibrate a quantum receiver out of several repetitions of the experiment. The model-free features of the algorithms we have discussed can readily be used to deal with scenarios where quantum information is transmitted over unknown quantum channels, a result we will present in Sec. 3.3.5. Let us stress that here, even finding sub-optimal strategies might already be a step forward towards better understanding the structure of quantum communication protocols, since little is known about such systems. In this regard, there is plenty of room for the type of machine learning heuristics that are presented along this thesis, in order to help with the development of quantum information protocols.

*Yo te pido un favor, no me dejes caer  
en las tumbas de la gloria.*

– Fito Paez

# 3

---

## Learning in the darkness

---

Our journey through the learning in the quantum begins, literally, by trial and error. We focus on quantum state discrimination (see Sec. 2.5.2), where we want to tell which is the physical state of our system out of measurement outcomes. In this Chapter, learning occurs from the darkness, and we ask how the discrimination error can be minimized by using only binary reward signals, accounting for the correctness of the guess.

To this end, we will focus in optical systems, and ask a model-free reinforcement-learning agent (see Sec. 2.6) to control an optical table; by departing from complete ignorance about the setting, we require the agent to achieve near-optimal discrimination performance. Our approach is crucially focused in experimental scenarios, where each repetition of the experiment counts, and thus exploiting every measurement outcome is required.

Studying this setting is motivated particularly by long-distance classical communication over quantum channels. Here, classical information is encoded into a quantum state, which is sent by a quantum channel to a receiver. Once arrived, the original information needs be decoded as accurately as possible, and thus a non-trivial optimization over quantum measurements arises. For instance in ground-to-satellite communication, optical signals are sent through the atmosphere, which can degrade signals' intensity to the point that quantum distinguishability effects become relevant. A quantum receiver is thereby used, which decodes the

---

original information sent from free space.

We consider one of the most standard sources of light found in optical laboratories, which are lasers. The quantum-mechanical behaviour of such systems is described by coherent states [Gla63]  $|\alpha\rangle$ , that belong to the set of Gaussian states (see 2.3.1). Specifically, we cast the discrimination of two electromagnetic signals with opposite phases, described by two coherent states of the field,  $|\alpha_k\rangle$ , with  $\alpha^{(k)} = (-1)^k\alpha$ , whose energy is proportional to  $|\alpha|^2$ . When the energy of the signals approaches zero, *i.e.*,  $|\alpha|^2 \ll 1$  (or when losses are present in the transmission channel), quantum effects become evident and it becomes impossible to discriminate between them perfectly. In particular, the ultimate bound given by quantum mechanics imposes bounds to the distinguishability of states, as discussed in Sec. 2.5.2. For the discrimination between two coherent states, the Helstrom bound (e.g. Eq. (2.192)) reads:

$$P_s^{(hel)}(\alpha) = \max_{\mathcal{M}} P_s(\alpha, \mathcal{M}) = \frac{1}{2} \left( 1 + \sqrt{1 - e^{-4|\alpha|^2}} \right), \quad (3.1)$$

where we recall that the overlap between the two states is  $|\langle -\alpha|\alpha\rangle|^2 = e^{-4|\alpha|^2}$ . Thus, as the intensity of the original signals to zero, the success probability gets closer to that of randomly guessing for the phase of the coherent state.

As discussed in Sec. 2.5.2, any binary discrimination protocol is described compactly by a POVM,  $\mathcal{M} = \{M_0, M_1\}$  with  $M_{1,2} \geq 0$  and  $M_1 + M_2 = \mathbb{I}$ . The probability of obtaining measurement outcome  $\hat{k}$  given that hypothesis  $k$  was true is given by  $p(\hat{k}|\alpha_k) = \text{Tr}[M_{\hat{k}}|\alpha^{(k)}\rangle\langle\alpha^{(k)}|]$ . Nevertheless, once outcome  $\hat{k}$  has been obtained, a guess for  $k$  needs to be done, and the best one is – by definition – related to the most likely hypothesis  $k \in \{0, 1\}$ , given outcome  $\hat{k}$ . Thus, the success probability of this setting reads

$$P_s(\alpha, \mathcal{M}) = \sum_{\hat{k}=0,1} \max_{k=0,1} p(\alpha^{(k)}, \hat{k}) \quad (3.2)$$

$$= \sum_{\hat{k}=0,1} \max_{k=0,1} p(\hat{k}|\alpha^{(k)}) p_k. \quad (3.3)$$

In particular, the measurement  $\mathcal{M}$  that achieves the Helstrom success probability is a superposition on the positive and negative part of an operator  $\Lambda = \frac{1}{2}(|\alpha\rangle\langle\alpha| + |-\alpha\rangle\langle-\alpha|)$ . As discussed in the example of Sec. 2.5.2, such projection is obtained by a superposition of the original states, which in the coherent-state discrimination problem leads to a projection over cat-like states of the form  $|\alpha_0\rangle + |\alpha_1\rangle$  [OBH96].

While preparation of such states currently constitutes an experimental challenge [TBW+20; SEL+22], an implementation of this projection can be carried out

using lineal (*i.e.* Gaussian) optics, photon-detectors and feedback operations. This measurement scheme is known as *the Dolinar receiver*, works in a sequential logic and it is proven to be asymptotically optimal [Dol; TSV+05]. Because it only requires lineal optics, photon-detection and feedback operations, this receiver is experimentally appealing, and several proofs of concepts have already been carried out [CMG+07; Ger04; DGD13; DB18]. It constitutes, nevertheless, a non-Gaussian measurement (see Sec. 2.3.4), and as such it presents several implementation challenges; for example, the feedback operation is assumed to act instantaneously, with arbitrarily many operations happening during the measurement process, whereas in practice we are constrained to a limited amount of such operations. Moreover, the presence of noise might alter the performance of the receiver.

However, a strong assumption is needed to implement the Helstrom measurement, or approximate versions of it. Namely, it is (obviously) required that the experimenter knows the quantum states she wants to distinguish. This motivates the current Chapter, where we will focus on the model-free discrimination of coherent states, by means of a Dolinar-like receiver. Our motivation emerges from situations where the experimental setting is not well characterized, and where we need to take advantage of *each* device usage (*i.e.* each measurement).

In this regard, achieving optimality in the implementation of quantum protocols (such as state discrimination) is certainly a goal, but we shall not dismiss sub-optimal strategies. Several examples supporting this claim can be found in our thesis: we find several scenarios in which sub-optimal strategies become state-of-the-art techniques that can readily be used to tackle a problem at hand. For instance, while the Dolinar receiver is asymptotically optimal when enough *measurement layers* are performed, we will see that already a single layer is sufficient to surpass the best Gaussian receiver (that is, a receiver composed only out of Gaussian elements).

Nevertheless, while the performance of such receivers can prove successful even in the case of limited resources (for example, a limited number of such measurement layers available), there are certainly more shortcomes that we need to address in order to model realistic scenarios. Among them, there is always the presence of some noise that underlies every experiment. Moreover, in communication scenarios, an unknown source of noise (such as an uncharacterized quantum channel) may alter the transmitted signals in a non-trivial manner. For example, in the case of long-distance ground-to-satellite communications, such channel is the atmosphere, whose action on the coherent-states can be understood as a lossy channel (see Sec. 2.3.3) whose attenuation suffers non-trivial variations depending on height or temperature, a situation which turns particularly difficult to model [DVR+19; AP05; UHP+12; Pir21a; Pir21b; VSV11; VSV+17].

In this context, our model-free approach to the calibration of resource-limited Dolinar-like receivers finds a solid motivation. Here, an agent needs to find a near-optimal configuration for its initially uncalibrated receiver by trial and error repetitions of the discrimination experiment. At each episode, one of the two possible coherent-states  $|\alpha_k\rangle$  is randomly picked and sent to the agent, who needs to make use of its apparatus in order to guess for the signal's label. Here, we stress that no knowledge about the physics of the setting is assumed, but only a *reward* is given to the agent if the bit of information is correctly decoded. As we will show, this approach is robust under the presence of unknown sources of noise, since the agent is not biased towards the noise-free setting and can readily adapt its calibration strategy to the noisy channel at hand.

This Chapter is structured in the following way. We begin in Sec. 3.1 by revisiting some quantum receivers that are frequently used to discriminate between two coherent-states; in particular we explain the structure of Dolinar receiver in Sec. 3.1.3, and its calibration is casted as a reinforcement-learning problem in Sec. 3.3.1. Before studying the performance of model-free agents, we turn to optimize Dolinar-like receivers in the model-aware case via dynamic programming in Sec. 3.2. We then turn to Sec. 3.3.2 where we present the main results of this Chapter, following by a sequence of experimental limitations considered in Sec. 3.3.5. Finally, we finish with a discussion and prospects for future extensions of this project in Sec. 3.4.

## 3.1 Coherent-state quantum receivers

This section is a tour through Gaussian and *almost-Gaussian* quantum receivers which are used to discriminate between two coherent states  $|\pm\alpha\rangle$  of equal energy and opposite phase. Note that such encoding is known in the literature as Binary Phase-Shifted Key (BPSK) coherent state discrimination, since the (classical) information is encoded into the phase of the coherent states. For simplicity, we assume that the sender and receiver have a shared reference frame, so that we can take the states to be real, e.g.  $\alpha \in \mathbb{R}$ , without loss of generality, as shown in Fig. 3.1. The term quantum receiver stands for a quantum measurement that decodes the information carried out on a quantum state that was sent through a quantum channel. In this Section we will restrict to the case where the signals are not altered by neither the presence of a quantum channel (e.g. we consider the noiseless channel) nor by noise due to malfunctioning devices: we will deal with such scenarios in Sec. 3.3.5.

As discussed in Sec. 2.3.4, operations in continuous-variable systems can be classified as either Gaussian or non-Gaussian. Generally, the Gaussian ones are

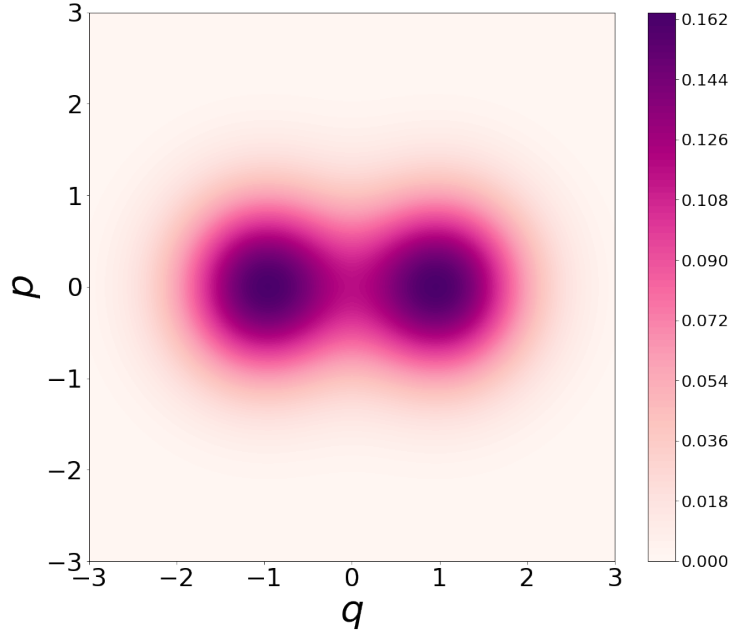


Figure 3.1: We show the Wigner function for the two candidate states  $|\alpha_k\rangle$ ,  $k = 0, 1$ , which are Gaussian distribution functions. We restrict to real amplitudes, since in the binary case one can always rotate the frame (assuming we know the direction to do so).

considered *feasible* in quantum optics laboratories. As such, they include linear operations (*i.e.* displacing, phase-shifting, squeezing and mixing signals by beam-splitters), Gaussian POVMs (*e.g.* homodyne and heterodyne measurements) and laser-beams. In particular, we observed in Sec. 2.3.2 that coherent states are Gaussian states (*e.g.* their Wigner function is a Gaussian distribution in phase-space), are obtained by displacing the vacuum state  $|\alpha\rangle = \hat{D}_\alpha |0\rangle$  and describe laser states.

This Section is structured as follows. We will comment on BPSK Gaussian receivers in Sec. 3.1.1. We then introduce some non-gaussianity through photon-detection measurements in Sec. 3.1.2, where we study the performance of the so-called *Kennedy* receiver. Then, we will introduce the Dolinar receiver in Sec. 3.1.3, and explain its sequential logic throughout the section.



### 3.1.1 Gaussian receivers

Let us briefly retrace our attention to Gaussian measurements for the BPSK coherent-state discrimination problem. As explained in Sec. 2.3.4, such measurements are defined as those that preserve Gaussianity upon post-measurement conditioning or, if the entire state is measured, give rise to Gaussian outcome probability distributions.

A wide class of Gaussian measurements are known as *general-dyne* measurements, which are POVMs that correspond to projections on generic Gaussian pure states, and discussed in the Preliminaries section. Here, note that such projections might not be orthogonal to each other, as happens in the heterodyne case, e.g. projecting over a coherent state, since  $\langle \alpha | \alpha' \rangle \neq 0$ . Let us also recall that any pure Gaussian state can be obtained through a symplectic unitary operation acting on the vacuum state, and Weyl displacements.

When it comes to discriminate between two Gaussian states, it is known that homodyne POVM (given by projections over quadrature states) is optimal among the set of Gaussian operations [RIR+21; TS08; SW21]. For this reason, homodyne success probability (in the discrimination setting) is known as *the homodyne limit*. Note, nevertheless, that this is subset of all POVMs, and as such it might not saturate the Helstrom bound (see Sec. 2.5.2). In turn, as mentioned above for the BPSK case, the Helstrom measurement can be attained by projecting over cat-like states, which is a *non-Gaussian* measurement.

To gain some more insight on the problem, let us derive the success probability attained by homodyne measurement in the BPSK. We consider a quadrature measurement of the form  $\mathcal{M} = \{\Pi_-, \Pi_+\}$ , with  $\Pi_- = \int_{-\infty}^0 dq |q\rangle \langle q|$ , and  $\Pi_+ = \int_0^{\infty} dq |q\rangle \langle q|$ . Given an outcome of the homodyne measurement  $q$ , the guess for the underlying hypothesis  $k$  is chosen according to the sign of  $q$ : if the value is negative, the signal is associated to  $|\alpha^{(1)}\rangle = |-\alpha\rangle$ , whereas if positive to  $|\alpha^{(0)}\rangle = |\alpha\rangle$ . Such a decision rule is denoted by  $\hat{k}(k)$ . With this, the success probability reads:

$$P_s^{\text{hom}} = \sum_k p_k p(\hat{k}|k) \quad (3.4)$$

$$= \frac{1}{2} \left( \int_{-\infty}^0 dq p(q - \alpha) + \int_0^{\infty} dq p(q + \alpha) \right) \quad (3.5)$$

Recalling that  $p(q|\alpha) = |\langle q|\alpha\rangle|^2 = \sqrt{\frac{2}{\pi}} e^{-2(q-\alpha)^2}$ , a straightforward calculation leads to

$$P_s^{\text{hom}} = \frac{1}{2} (1 + \text{Erf}[\sqrt{2}\alpha]). \quad (3.6)$$

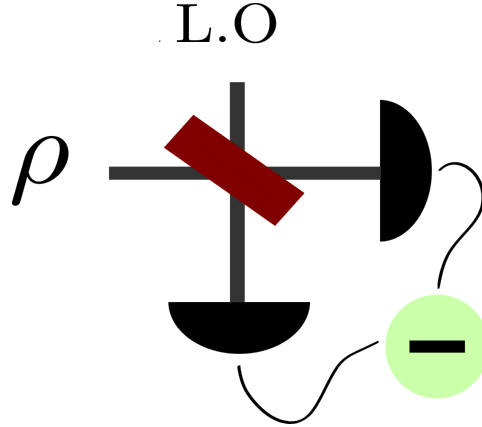


Figure 3.2: An implementation of homodyne measurement is shown, which consists on mixing the incoming signal with a coherent state of strong intensity, known as Local Oscillator, by a balanced beam-splitter. The difference in signals' intensities between the two exiting ports constitutes the homodyne measurement outcome [Ser17a].

As shown in Fig. 3.2, the homodyne receiver can be realized by measuring the intensity of the resulting signals after mixing the state  $\rho$  with a local oscillator (a high-intensity coherent state). Quite surprisingly, the building blocks of such standard Gaussian measurement are photon-detectors, which are actually non-Gaussian.

Experimentally, Gaussian operations can be complemented with on/off photon-detectors, constituted by a POVM with elements  $\{\Pi_0 = |0\rangle\langle 0|, \Pi_1 = \mathbb{I} - |0\rangle\langle 0|\}$ , where the  $\Pi_1$  is a non-Gaussian projection. Using on-off photon-detections is experimentally feasible, and there has been a trend in the recent years to combine Gaussian transformations with photo-detections carried out by the end of the circuit. While in general one can surpass the homodyne limit, it has been shown that optimality might be elusive in this setting [GG21]. Before considering more sophisticated receivers, we will now study an example of the aforementioned Gaussian+on/off receiver, named after Kennedy.

### 3.1.2 Kennedy receivers

When allowing on/off photodetectors, an intuitive approach to the discrimination of binary coherent states can be constructed [Ken73]. Assuming we know the amplitude  $|\alpha|$  of the incoming states, we can readily displace the signal by  $\hat{D}_\beta$  with

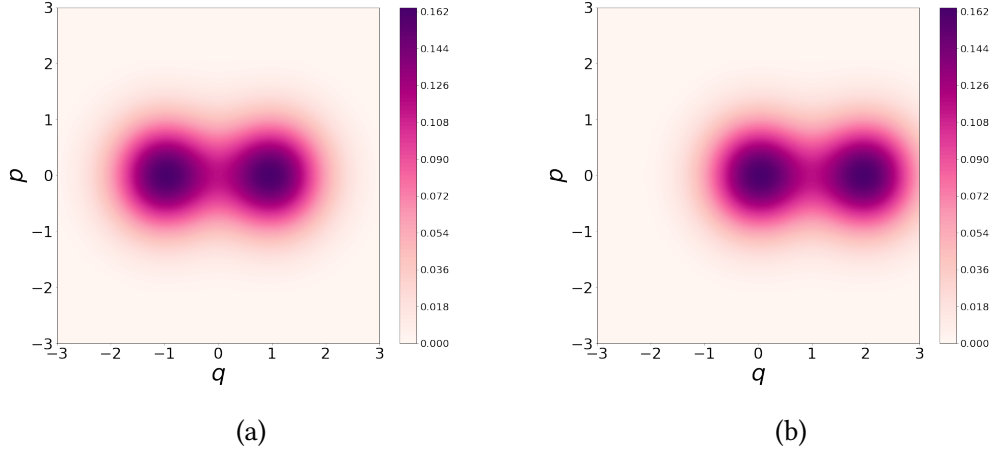


Figure 3.3: We show the Wigner functions for (panel *a*) the coherent states  $|\pm\alpha\rangle$  and (panel *b*) the displaced coherent states  $|0\rangle$  and  $|2\alpha\rangle$

$\beta = \alpha$ . In this way, if  $|-\alpha\rangle$  was received, the resulting state would be  $|0\rangle$ , and if  $|\alpha\rangle$  was received instead, the resulting state would be  $|2\alpha\rangle$ . This is illustrated in Fig. 3.3, where the Wigner function of the incoming states (either  $|\alpha\rangle$  or  $|-\alpha\rangle$ , with equal priors) is shown, along with the displaced signals. After displacing, an on/off photodetector is used, which projects the probe into the vacuum state  $|0\rangle$  or its complement, *i.e.*  $\mathcal{M}_{ph} = \{|0\rangle\langle 0|, \mathbb{I} - |0\rangle\langle 0|\}$ . Such a measurement is a non-Gaussian one, yet is a standard component in modern quantum optics laboratories. The whole apparatus of measuring the displaced signal is known as the *Kenedy receiver*, and is shown in Fig. 3.4.

Thus, once the measurement outcome is obtained, a guess for the most-likely hypothesis needs to be performed. By denoting the measurement outcome by  $o \in \{0, 1\}$  (standing for zero and one or more photons detected respectively), the success probability reads

$$P_s^{ken} = \sum_{o=0,1} \max_k p_k p(o|\alpha^{(k)}), \quad (3.7)$$

where  $p(n|\alpha^{(k)})$  can easily be obtained by noticing that  $p(0|\alpha^{(k)}) = |\langle 0|\hat{D}_\beta|\alpha\rangle|^2 = |(-\beta|\alpha)|^2$ , and  $p(1|\alpha^{(k)}) = 1 - p(0|\alpha^{(k)})$ .

Let us understand the logic behind this receiver. Once displaced, the signal can either be  $|0\rangle$  or  $|2\alpha\rangle$ . In the former case, when measuring with an on/off photodetection — and in the absence of errors such as dark-counts — the only

possible result is  $n = 0$  photons. Thus, if outcome  $n = 1$  was obtained, we can be sure that the signal is  $|\alpha\rangle$ . However, outcome  $n = 0$  does not imply that the signal is  $|\alpha\rangle$  with absolute certainty, since  $|\langle 0|2\alpha\rangle|^2 = e^{-4|\alpha|^2} \neq 0$ . This leads us to evaluate the success probability of Kennedy receiver as:

$$P_s^{\text{ken}}(\alpha) = 1 - \frac{e^{-4|\alpha|^2}}{2}. \quad (3.8)$$

When compared with the homodyne measurement (*i.e.* Eq. 3.6), we observe that an advantage in favour of Kennedy receiver is only attained when  $|\alpha| \gtrsim 0.4$ .

As it happens in cases where little resources are available, each component of a protocol should be optimized if possible. Here, we can readily observe that the displacement value  $\beta = \alpha$  might not be the best choice.

### Optimized Kennedy receiver

By identifying the displacement as a degree of freedom to be optimized, we have *parametrized* the Kennedy receiver, and its performance is now measured by the success probability associated to the specific choice of  $\beta$ . For a fixed  $\beta$ , we will denote such receiver as the  $\beta$ -Kennedy receiver (the original Kennedy receiver is recovered for  $\beta = \alpha$ ).

We note that different values of  $\beta$  will lead to different POVMs  $\mathcal{M}_\beta$ , each belonging to the same family of Kennedy-like receivers. This is analogous to the case of parametrized quantum circuits studied in Sec. 2.2, in which a unitary transformation was constructed from local qubit rotations and entangling CNOT gates, and optimized in such a way to minimize a given cost function. From this quantum machine learning point of view, the parametrized quantum circuit analogous is given by  $\mathcal{M}_\beta$  and the success probability (or, equivalently, the error probability)  $P_e^{\beta\text{-ken}}(\alpha, \beta)$  plays the role of the cost function.

Depending on the context, this problem also falls into the category of *optimal control theory* [Bel64; Ber05], since one is interested in optimizing the cost function.

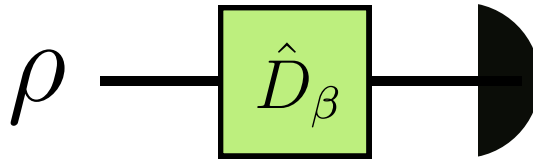


Figure 3.4: We show the Kennedy receiver, which in the case of coherent states' intensity being  $\alpha$ , it reduces to  $\beta = -\alpha$ .

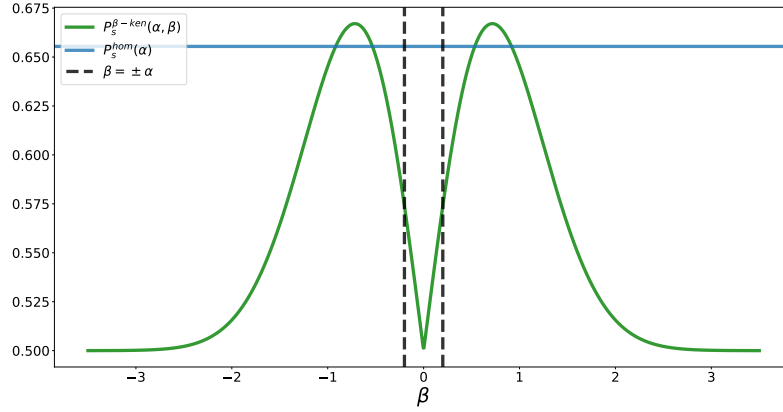


Figure 3.5: We show the optimization landscape for the  $\beta$ -Kennedy receiver, at a fixed intensity  $\alpha = 0.2$ . which in the case of coherent states' intensity being  $\alpha$ , it reduces to  $\beta = -\alpha$ .

This field tries to find answers on how to steer a system towards a target state by performing external controls or *actions* on it. For instance, such system can be thought as our coherent state  $|\alpha^{(k)}\rangle$ , whose phase needs to be deciphered by the  $\beta$ -Kennedy receiver, and different actions correspond to varying the value of  $\beta$ . In this regard, navigating through the optimization landscape is important in order to find which actions are the best, though doing so is generally hard. For instance, one needs to estimate the cost value for each possible value of the control: in our discrimination example, this means estimating the success probability associated to each possible  $\beta$ -Kennedy receiver out of several repetitions of the discrimination experiment. While estimating the entire landscape is generally too costly, we can get some insight into the problem at hand by studying simplified models. In particular, certain scenarios can readily be deemed *hopeless*, in the sense that parameter optimization might only succeed with an exponentially vanishing success probability, as it happens with barren plateaus appearing in certain parametrized quantum circuits (see Sec. 2.2)

In this chapter, the simplified model that allows us to get some insight is precisely the  $\beta$ -Kennedy receiver, whose optimization landscape is shown in Fig. 3.5 for a fixed value of  $\alpha = 0.2$ . In this a region, the homodyne receiver outperforms the Kennedy one, as noticed after inspecting Eq. 3.8. However, we observe that there is an entire region of  $\beta$  values for which the  $\beta$ -Kennedy receiver can actually surpass the homodyne limit. Note also that the optimal solution is degenerate; this is a

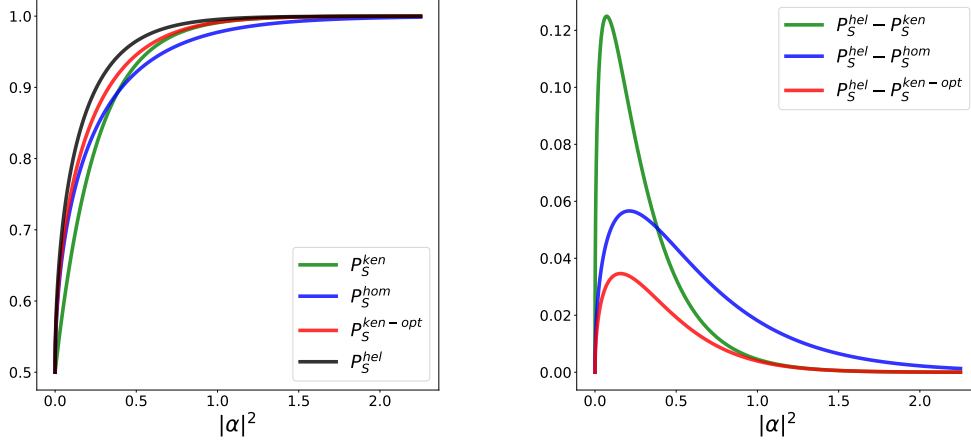


Figure 3.6: We compare the success probabilities for homodyne receiver, Kennedy receiver, and optimized-Kennedy receiver with the Helstrom bound. Data in both panels is the same, however in right one we show the different with Helstrom bound, in order to help visualization.

consequence of the symmetry in testing either  $|\alpha\rangle$  (negative  $\beta$ ) or  $|\alpha\rangle$  (positive  $\beta$ ) using the photon-detector logic explained above.

In this way, the optimization over  $\beta$  can be carried out for each value of  $\alpha$ . The resulting receiver is known as the *optimized Kennedy receiver*, and surpasses the homodyne limit for all intensity values, as shown in Fig. 3.6, where we compare its success probability, denoted by  $P_S^{opt-ken}$ , with that of homodyning and with the original Kennedy receiver  $P_S^{ken}$  (e.g.  $\beta = -\alpha$ ). To aid the comparison, we have depicted the difference with respect to the Helstrom bound: how to approach such a limit is the matter of the next section.

### 3.1.3 Dolinar receivers

Sequential strategies lie in the heart of this thesis, and this Chapter is no exception. The Dolinar receiver [Dol73] is a concatenation of  $\beta$ -Kennedy receivers, each acting on an tiny portion of the original signal.

In particular, this receiver consists on splitting the incoming coherent state by using beam-splitters (BMs), resulting into  $L$  lower-intensity copies, e.g.  $|\alpha_k^{(\ell)}\rangle = |\frac{\alpha_k}{\sqrt{L}}\rangle$  for  $\ell = 1, \dots, L$ . From here, the phase of each state is tested by using a  $\beta$ -Kennedy receiver. Crucially, a sequential logic is applied, in which the displacement

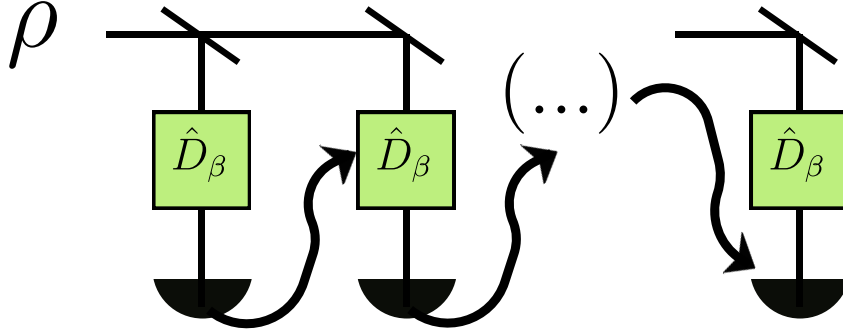


Figure 3.7: We show a Dolinar-like receiver. For an infinite sequence measurements and feedforward operations, such a scheme attains the Helstrom bound for binary coherent-state discrimination.

value at  $\ell$ -th stage depends on the measurement outcomes previously obtained, as shown in Fig. 3.7. This conditioning operation is known as a classical feed-forward, and should in principle be optimized over all possible measurement outcomes. We note that for  $L = 1$  we recover the optimized-Kennedy receiver. However, because of its structure consisting on  $L$  processing *layers*  $\ell = 0, \dots, L$ , Dolinar receiver is considerably more complex than the Kennedy one. Specifically, for each layer  $\ell < L$ , the following operations are applied:

1. The input signal  $|\alpha\rangle$  (or equivalently  $|\alpha\rangle$ ) is split on a BS of transmissivity  $\theta$ , effectively extracting a fraction  $1 - \theta$  of the energy for detection. The BS transforms the input signal and vacuum states as

$$|\alpha\rangle |0\rangle \mapsto |\alpha\sqrt{\theta}\rangle_{\text{tr}} |\alpha\sqrt{1-\theta}\rangle_{\text{ref}}, \quad (3.9)$$

Note that the BS adds a phase to the second mode, which we assumed corrected via a proper phase-shifter not shown in the figure.

2. The reflected part of the signal undergoes a displacement operation  $\hat{D}_\beta$ . Such operation is realizable via interference with a strong coherent signal on a small-reflectivity BS, not shown in the figure. The resulting state is  $|\tilde{\alpha}(\beta, \theta)\rangle_{\text{ref}} = \hat{D}_\beta |\alpha\sqrt{1-\theta}\rangle_{\text{ref}}$ .
3. The displaced signal is measured via a on/off photodetector, which detects no photon, i.e., outcome  $o_{\ell+1} = 0$ , with conditional probability

$$p(o_{\ell+1} = 0 | \alpha, (\beta, \theta)) = |\langle 0 | \tilde{\alpha}(\beta, \theta) \rangle_{\text{ref}}|^2 = e^{-|\tilde{\alpha}(\beta, \theta)|^2}, \quad (3.10)$$

and detects one or more photons, i.e.,  $o_{\ell+1} = 1$ , with probability  $1 - p(o_{\ell+1} = 0 | \alpha, (\beta, \theta))$ .

4. The transmitted part of the signal enters layer  $\ell + 1$ .

Finally, the last processing layer  $\ell = L$  consists in elaborating a guess  $\hat{k}$  of the true hypothesis  $k$ , based on previous measurement outcomes and parameter choices.

Similar to  $\beta$ -Kennedy receivers, Dolinar-like receivers are parametrized by the displacements  $\hat{D}_\beta$ . Nevertheless, the complexity is considerably higher: while  $L$  displacement operations should be performed per signal detection, the feed-forward conditioning non-trivially enlarges the optimization landscape to  $2^L - 1$  possible values.

In addition, we shall now adopt the control-theory mindset introduced in Sec. 2.6, and think of the displacements  $\hat{D}_{\beta_\ell}$  and attenuations  $\theta_\ell$  as *actions*  $a_\ell$  performed on the signal.

Under these definitions, let us now study the success probability associated to a Dolinar-like receiver. For an initial coherent state  $|\alpha\rangle$ , the input state at the  $\ell$ -th layer is  $|\alpha_\ell\rangle = |\alpha\sqrt{\theta_0 \cdots \theta_{\ell-1}}\rangle$ . In principle, we can use all the past history defined as

$$h_\ell = (a_0, o_1, \dots, a_{\ell-1}), \quad (3.11)$$

with  $h_0 = \emptyset$ , to decide the next value of  $(\beta, \theta)$  and the final guess. We label them compactly as  $a_\ell(h_\ell) = (\beta_{h_\ell}, \theta_{h_\ell})$  and  $a_L(h_L) = \hat{k}$ , omitting the label  $\ell$  or the dependence on  $h_\ell$  when it is clear from the context.

Hence, the average success probability of the  $L$ -Dolinar receiver, *calibrated* under actions  $\{a_\ell\}$ , and averaged out over all possible outcomes' sequences  $o_{1:L} = (o_1, \dots, o_L)$  can be written as

$$P_s(\alpha, \{a_\ell\}) = \sum_{o_{1:L}} \prod_{\ell=1}^L p(o_\ell | \alpha^{(k)}, a(h_\ell)) p_k \Big|_{k=a(h_L)}, \quad (3.12)$$

Here, the configuration  $\{a_\ell\}$  is the total set of actions over all histories, and we have written the conditional probability of the sequence of outcomes  $o_{1:L}$  as a product of single-layer conditional probabilities, e.g. Eq.(3.10). In particular, we aim to optimize this success probability over the available actions of the receiver, and this value will denote as

$$P_*^{(L)}(\alpha) = \max_{\{a_\ell\}} P_s(\alpha, \{a_\ell\}). \quad (3.13)$$

We will omit the dependence on  $L$  whenever it is clear from the context.



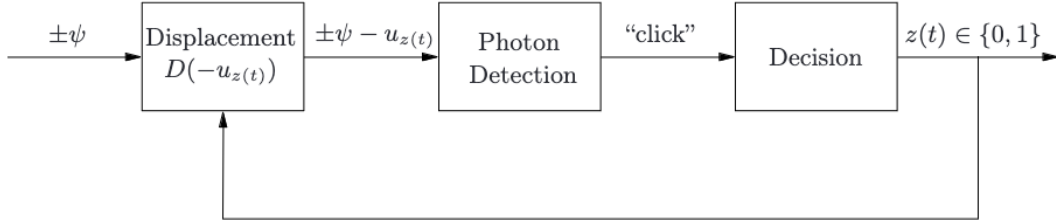


Figure 3.8: We show the original proposal of the Dolinar receiver, which works in continuous-time. This image is taken from Ref. [ADPP11]. Instead of splitting the coherent-state in space by a beam-splitter array, we here have the signal spread in time, as explained in the main-body. The  $\ell$ -dependence in the conditional displacements is now translated into a time-dependence for  $u(t)$ , which in turn depends on the parity of the measurement outcomes as captured by the function  $z(t)$ . After the pulse-duration  $T$ , a decision for the phase of the coherent state is done via the function  $z(T)$ .

The importance of the Dolinar receiver is that *it is known to attain the Helstrom bound* in the limit of many layers, *i.e.*  $L \rightarrow \infty$ , provided that the optimal actions are performed. However, before discussing such optimality and the structure of the receiver further, we will comment on an alternative formulation of this receiver.

### 3.1.4 Time-domain Dolinar receiver & optimality

While we have introduced Dolinar receiver in the spatial-domain, its original formulation was done in the continuous-time domain [Dol]. The two formulations are related to each other by the equivalence between continuous-time photon-counting processes and a (sufficiently-large) sequence of lossless beam-splitters and photon-detectors, as studied in Ref. [Ban94].

In the time-domain picture, the coherent state  $|\pm\alpha\rangle$  is understood as a travelling-mode, and given by the field  $\psi(t) = \pm\psi e^{i\omega t}$ , where  $\omega$  is the optical frequency and  $|\alpha|^2 = \int_0^T |\psi(t)|^2 dt = \psi^2 T$ , with  $T$  the total duration of the pulse. At each time  $t$ , the receiver shifts the incoming-field by a value  $u_{z(t)}(t)$ , and measures it by a photon-detector, as depicted in Fig. 3.8. Such measurement is assumed to be carried out very fast, leading only to binary outcomes (click or no click), which defines a so-called *compound Poisson process*. As shown in Dolinar’s PhD thesis, the most-likely hypothesis  $k$  is given by the parity of the sum of the outcomes obtained up to time  $t$ , and this determines the binary value  $z(t)$  that controls the

nature of the displacements to be done at consecutive times. The optimal shape of the pulses  $u_{z(t)}(t)$  and of the success-probability of this scheme was originally derived in Ref. [Dol] by performing a dynamic programming optimization of the likelihood-ratio; in turn, he was able to show that the receiver attains the Helstrom bound.

In this line, the optimality of Dolinar receiver can also be proved as follows. With sufficiently many measurement layers, say  $L$ , the coherent state will be splitted into  $L$  copies, each having a very low intensity. As such, the support of the attenuated coherent state will be determined by (the Fock states) of  $|0\rangle$  and  $|1\rangle$ . From here, we can recall the results of Ref. [ABB+05], namely that the optimal discrimination between multiple copies of two pure quantum states can be achieved by a sequential strategy, in the asymptotic case of many copies. This strategy consists on a local projective measurement that is optimized at each step according to the most likely hypothesis; most notably the global measurement optimization problem is solved by an adaptive-local measurement, which hinges on bayesian updating the prior probabilities of each hypothesis after each measurement result. Such strategy is shown to achieve the Helstrom bound, and it can readily be applied to the  $L$  copies of the attenuated coherent-state. In Ref. [ADPP11] it was shown that, following this approach, the optimal adaptive measurement coincides with the Dolinar receiver, in the binary coherent-state discrimination problem. Also, let us mention that an alternative proof of Dolinar receiver's optimality is given in Ref. [TSV+05]

However, in practical experimental scenarios it is not possible to perform a large number of measurement layers. Such constraint is justified by (i) the experimental feasibility of constructing the receiver, and (ii) the noise accumulation at every layer.

Moreover, noise might even accumulate with the number of layers, as it happens with one of the most common sources of photon-detectors noise, the dark counts. Here, a click appears in the detector even when measuring the vacuum. Moreover, optimality of Dolinar receiver also requires the feed-forward operation to act before the signal reaches the next measurement layer. However, measuring and classical post-processing do take some time, and this delay is in practice modelled by including losses in between each measurement layer [SL14].

From our previous discussion, it might seem that we are heading towards a paradigm where a set of suggestions is given to the experimentalist, with the optimizing actions in Eq. 3.13 essentially constituting the recipe to follow. This paradigm is actually followed in Chapter 4, where several cost-minimizing quantum circuits are discovered using our VAns algorithm, which is shown to work under

several noise-models. While our experimental simulations should be realistic as possible, the usefulness of the recipe we provide will ultimately depend on how accurate can reality be modeled.

On the contrary, we will here tackle the situation from a different angle, and adapt to reality on the fly. This is the radical and experimentally-centered approach taken in this Chapter. To this end, a reinforcement-learning agent is faced towards the discrimination experiment, and asked to optimize its Dolinar-like receiver by having only access to an answer: *yes/no*, a binary value standing for the correctness of its guess.

### 3.1.5 The sequential structure of Dolinar receiver

Before jumping to the optimization of Dolinar-like receivers (either model-aware or model-free), it is important to further discuss its structure, as captured by Eq 3.12. In turn, we can take profit of the sequential structure, in order to compute Eq. 3.13 in a model-aware setting, a quantity that will help us to benchmark the performance of agnostic agents.

We will refer to a finite Dolinar receiver, *i.e.* one consisting on  $L$  processing layers as  $L$ -Dolinar, and refer to the prior probability of having the state  $|+\alpha\rangle$  as  $\eta_0 = \eta$ ; as a consequence the prior probability of  $|-\alpha\rangle$  is  $\eta_1 = 1 - \eta$ . Whenever it is clear from the context, we will drop the subscripts and simply refer to  $\eta$ .

To this end, we will study the performance of the  $L$ -receiver layer by layer, and first consider a 0-Dolinar receiver. Without any measurement, we only rely on the prior probabilities to asses the label of the states (*e.g.* to guess for the value of  $k$ ). The success probability of such a trivial receiver thus corresponds to the probability of the most likely hypothesis:

$$P_s^{L=0}(\alpha, \eta) = \max \{ \eta, 1 - \eta \}. \quad (3.14)$$

Let us now consider an 1-Dolinar receiver (*e.g.* a Kennedy-like receiver). In such a case, the success probability — when considering the optimal, *i.e.* maximum-likelihood guess— reads

$$P_s^{L=1}(\alpha, \{a\}, \eta) = \sum_{o_1=0,1} \max_k p(\alpha_k, o_1), \quad (3.15)$$

where  $p(\alpha_k, o_1)$  denotes the joint probability of having the state  $|\alpha_k\rangle$  and observing outcome  $o_1$ , and  $\{a\}$  reduces to the displacement with value  $\beta$ , as shown in Fig. 3.4.

The joint probability  $p(\alpha_k, o_1)$  can be written in terms of the prior probability  $\eta_k$  of having  $|\alpha_k\rangle$  times the probability of observing outcome  $o_1$ , given we actually

have such state:

$$P_s^{L=1}(\alpha, \{a\}, \eta) = \sum_{o_1=0,1} \max_k p(o_1|\alpha_k)\eta_k. \quad (3.16)$$

On the other hand, we can also write the joint probability in terms of the total outcome probability, *i.e.*  $p(o_1) = \sum_k p(o_1, \alpha_k) = \sum_k p(o_1|\alpha_k)\eta_k$ , and the *posterior* probability of having  $|\alpha_k\rangle$  once  $o_1$  was observed:

$$\eta_{k|o_1} = p(\alpha_k|o_1) = \frac{p(o_1|\alpha_k)\eta_k}{p(o_1)}. \quad (3.17)$$

With this, the success probability of Kennedy-like receivers reads

$$P_s^{L=1}(\alpha, \{a\}, \eta) = \sum_{o_1=0,1} p(o_1) \max_k p(\alpha_k|o_1) \quad (3.18)$$

$$= \sum_{o_1=0,1} p(o_1) P_s^{L=0}(\eta_{k|o_1}), \quad (3.19)$$

where the dependence on  $\{a\}$  is left implicit. We note that once once the outcome is observed, 1-Dolinar receivers trivially recurs to 0-Dolinar receivers, with a bayesian-updated prior values.

Let us now consider a 2-Dolinar receiver. Here, the success probability reads

$$P_s^{L=2}(\alpha, \{a\}, \eta) = \sum_{o_1=0,1} \sum_{o_2=0,1} \max_k p(\alpha_k, o_1, o_2) \quad (3.20)$$

$$= \sum_{o_1=0,1} \sum_{o_2=0,1} p(o_1, o_2) \max_k p(\alpha_k|o_1, o_2) \quad (3.21)$$

$$= \sum_{o_1=0,1} p(o_1) \sum_{o_2=0,1} p(o_2|o_1) \max_k p(\alpha_k|o_1, o_2) \quad (3.22)$$

$$= \sum_{o_1=0,1} p(o_1) \sum_{o_2=0,1} p(o_2|o_1) \max_k \frac{p(o_2|\alpha_k o_1) p(\alpha_k|o_1)}{p(o_2|o_1)} \quad (3.23)$$

$$= \sum_{o_1=0,1} p(o_1) P_s^{L=1}(\eta_{k|o_1}), \quad (3.24)$$

where we expressed conditional outcomes probabilities as  $p(o_1, o_2) = p(o_2|o_1)p(o_1)$ , and used

$$p(o_2|o_1) = \sum_k p(o_2\alpha_k|o_1) = \sum_k p(o_2|\alpha_k o_1)p(\alpha_k|o_1) = \sum_k p(o_2|\alpha_k o_1)\eta_{k|o_1}. \quad (3.25)$$

Here, we can readily highlight the conditional dependence of the actions. In the 2-Dolinar receiver, they are given by  $\{a_1, a_2^{o_1=0}, a_2^{o_1=1}\}$ , where  $a_1$  denotes the first displacement (which is unconditional) and  $a_2^{o_1}$  stands for the second displacement (conditioned on the outcome  $o_1$ ). While this can be extended to consider the attenuation values as well, we will not reinforcement-learn them, since their contribution to the success probability is very small, as observed in the next Section.

Thus, the conditional dependence of the actions is reflected in the outcome probabilities. For the first outcome  $o_1$ , the dependence is only on  $a_1$  as

$$p(o_1) = \sum_k p(o_1|\alpha_k) = \sum_k p(o_1|\alpha_k)\eta_k = \sum_k (o_1 + (-1)^{o_1} e^{-|\alpha_k + a_1|^2})\eta_k, \quad (3.26)$$

where we have wrote compactly the outcome probability of an on/off photodetector with coherent state  $|\alpha_k + a_1\rangle$  as input. Furthermore, as Eq. 3.25 indicates, the dependence of  $o_2$  with  $o_1$  has a contribution on  $a_2^{o_1}$  (through  $p(o_2|\alpha_k o_1)$ ), but also through the updated prior  $\eta_k|_{o_1}$ . In turn, the sequential interpretation of the Dolinar receiver crucially depends on this bayesian update; as observed in Eq. 3.20, a 2-Dolinar receiver is an extension of a 1-Dolinar receiver, in which the prior probability of each hypothesis is modified according to the first measurement outcome.

Similarly, the sequential structure can be made explicit for the  $L$ -Dolinar receiver. Let us denote with  $o_{\ell_1:\ell'}$  the string of outcomes  $(o_{\ell_1}, o_{\ell_1+1}, \dots, o_{\ell'})$  (with  $\ell < \ell'$ ). Then, the success probability of  $L$ -Dolinar receiver reads

$$P_s^L(\alpha, \eta) = \sum_{o_{1:L}} \max_k p(\alpha_k, o_{1:L}) \quad (3.27)$$

$$= \sum_{o_1} p(o_1) \sum_{o_{2:L}} p(o_{2:L}|o_1) \max_k p(\alpha_k|o_{1:L}) \quad (3.28)$$

$$= \sum_{o_1} p(o_1) \sum_{o_{2:L}} p(o_{2:L}|o_1) \max_k \frac{p(\alpha_k|o_{2:L}) p(\alpha_k|o_1)}{p(o_{2:L}|o_1)} \quad (3.29)$$

$$= \sum_{o_1} p(o_1) \sum_{o_{2:L}} \max_k p(\alpha_k|o_{2:L}) p(\alpha_k|o_1) \quad (3.30)$$

$$= \sum_{o_1} p(o_1) P_s^{L-1}(\eta_k|_{o_1}). \quad (3.31)$$

Thereby, enlarging Dolinar's receiver with an extra layer translates into bayesian updating the prior probability of each hypothesis.

This resembles to the dynamic programming procedure we discussed in Sec. 2.6.4, where the principle of optimality (*i.e.* the recurrence relation given by

the optimal Bellman equation) allowed us to sequentially attain the optimal policy. Here, we observe that the best success probability — *i.e.* that optimized over all actions  $\{a\}$  — can be obtained by a sequential optimization of local problems, in this case given by Kennedy-like receivers with varying priors.

Before jumping to this model-aware optimization, let us highlight some further symmetries that will find use in the following section. In particular, note we that the prior probability of having  $|\alpha\rangle$ , *e.g.*  $\eta_0 = \eta$ , is sufficient to encode all the information required for the binary discrimination problem, since  $\eta_1$  is trivially be obtained from  $\sum_k \eta_k = 1$ . This simple symmetry should also be translated into the success probability. In turn, from Eq. 3.16 it is trivial to see that

$$P_S^{L=1}(\eta, \{a\}) = P_S^{L=1}(1 - \eta, \{-a\}), \quad (3.32)$$

where by  $-a$  we mean that the displacement take now the opposite values. In practice, since  $\eta_k \in [0, 1]$ , this means that it is sufficient to consider half of such an interval, the complementary optimal actions can be obtained from Eq. 3.32.

We will now turn to explain how the sequential structure studied above can be exploited in order to compute the optimal values of the displacements and success probabilities in the fully-aware scenario.

## 3.2 The model-aware approach

We will now discuss the model-aware optimization of Dolinar-like receivers, which is based in the observations outlined previously regarding the sequential structure of such receivers.

As introduced in Sec. 2.6.4, dynamic programming exploits the so-called principle of optimality, which links optimal actions to optimal state value functions through the optimal Bellman equation, see Eq. 2.244. In there, we introduced Markov Decision Problems (MDP) consisting on an agent that sequentially interacts with its environment in order to maximize a figure of merit known as the return. This quantity captures the amount of rewards the agent collects during an episode, and it here translates to the correctness of agent's guess, which is obtained after measuring the state by a Dolinar-like receiver. In the MDP section, we have also made the distinction between agent and environment states, which arises due to the impossibility of the agent to have fully access to the environment's state at a given step, and thus Partially Observable Markov Decision Processes were introduced. Here, we will define the environment states as the underlying quantum states  $|\alpha_k\rangle$  to be discriminated.

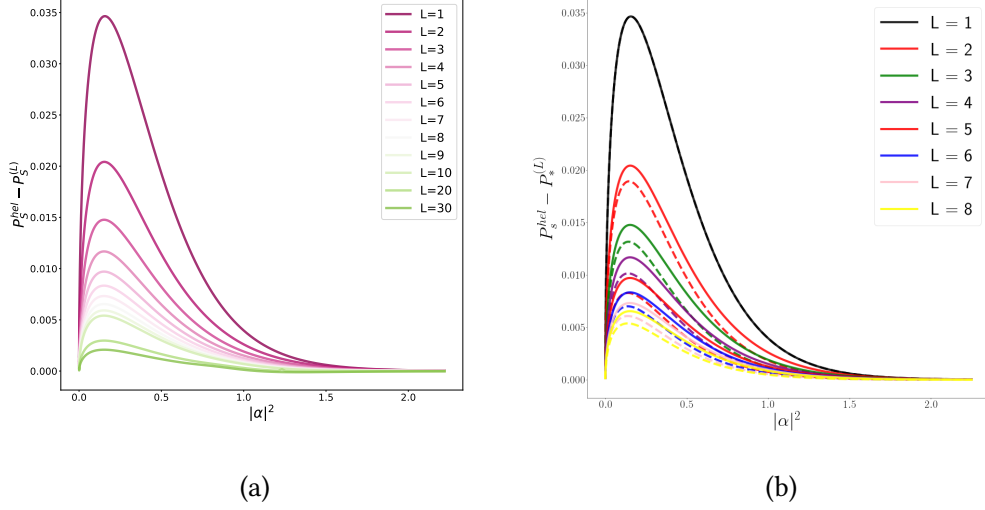


Figure 3.9: We show the difference between the best probability of success attainable for a fix  $L$  and the optimal probability of success in discriminating BPSK coherent states. The results were obtained by dynamic programming. *Left panel:* the attenuations at the beam-splitters are fixed in such a way that each partial measurement deals with a state having the same intensity  $\alpha_k^\ell = \frac{\alpha_k}{\sqrt{L-1}}$  for all  $\ell$ . *Right panel:* we compare the advantage of adapting beam-splitter transmissivities (dashed lines) as compared to the receivers considered in left panel (solid lines).

Since in the model-aware scenario the agent has full access to the outcomes probabilities, it is sensible to define agent's state as the prior probability of  $|\alpha_k\rangle$  at each layer of the receiver. Intuitively, such a quantity stands for the belief distribution over the states  $|\alpha_k\rangle$ , and it is defined as:

$$b_\ell(k) = p(\alpha_\ell^{(k)} | o_\ell, a_{\ell-1}, b_{\ell-1}). \quad (3.33)$$

As new information about the hidden state is collected, the belief evolves. The evolution law of this quantity is obtained by a bayesian update, which reads:

$$b_\ell(k) = \frac{p(\alpha_\ell^{(k)} | o_\ell, a_{\ell-1}) b_{\ell-1}(k)}{\sum_k p(\alpha_\ell^{(k)} | o_\ell, a_{\ell-1}) b_{\ell-1}(k)}. \quad (3.34)$$

We can readily define the state value function  $v(s)$ , recalling it quantifies how *valuable* state  $s$  is in terms of the expected return. In our case, the return reduces

to the final reward obtained after action  $a_L$ , *e.g.* guessing, and the expected return becomes the success probability associated to agent's policy. Specifically, agent's policy consists on specifying the actions of its  $L$ -Dolinar receiver, and we are thus interested in finding the optimal policy, which on the one hand is linked to Eq. 3.13, and on the other hand its state value function should satisfy the optimal Bellman equation.

To understand this last point, let us consider final last step of an  $L$ -Dolinar receiver, which consists on performing a guess. The optimal action to take is the maximum-likelihood guess, and its associated value function reads

$$v_L^*(b_L) = \max_k b_L(k). \quad (3.35)$$

Observe that since  $\sum_k b_\ell(k) = 1$ , it is sufficient to track  $b_\ell(k = 0)$ . The optimal Bellman equation at step  $\ell < L$  instead reads

$$v_\ell^*(b_\ell) = \max_{a \in \mathcal{A}(s_\ell)} \sum_{o_{\ell+1} \in \mathcal{O}} \sum_k p(o_{\ell+1}; b_\ell(k), a_\ell) v_{\ell+1}^*(b_{\ell+1}). \quad (3.36)$$

The last equation shows that the value-functions can be computed *back-wards*. In this sense, we can first compute the optimal state-value functions for the final layer, then move the previous layer (*i.e.*  $\ell = L - 1$ ), and compute optimal ones using Eq. (3.36) for each belief value in a (discretized) set of possible beliefs  $\mathcal{B}$ . Because of the symmetry highlighted in Eq. 3.32, we can consider points in between the interval  $[0, \frac{1}{2}]$ .

However, we note that for a given  $b_\ell \in \mathcal{B}$ , the posterior probability  $b_{\ell+1}$  appearing in optimal Bellman equation might lie outside the set  $\mathcal{B}$ . This constitutes a problem, since its state value function is unavailable from the previous step. To circumvent this issue, we interpolate  $v_{\ell+1}^*(b)$  to the entire interval  $[0, 1]$  using Eq. 3.32, and to this end we require many points in  $\mathcal{B}$  (in our numerics we considered 100).

In our numerics, shown in Fig. 3.9 we were able to optimize up to  $L = 30$ ; the interpolation method is based on Radial Basis Functions (see *e.g.* introduction of Ref. [Rbf]) and the minimization method used was dual annealing [VGO+20]. Moreover, following a similar procedure than outlined above, we can optimize over the attenuations as well. The latter proves advantageous, since from  $L = 4$  and on, adaptive beam-splitters outperform non-adaptive ones, even in the case where an extra layer is included for the latter. To see this, consider Fig. 3.9: we observe that dashed lines (adaptive beam-splitters) for  $L = 5$  outperform solid ones (non-adaptive beam-splitters) even for  $L = 6$ .



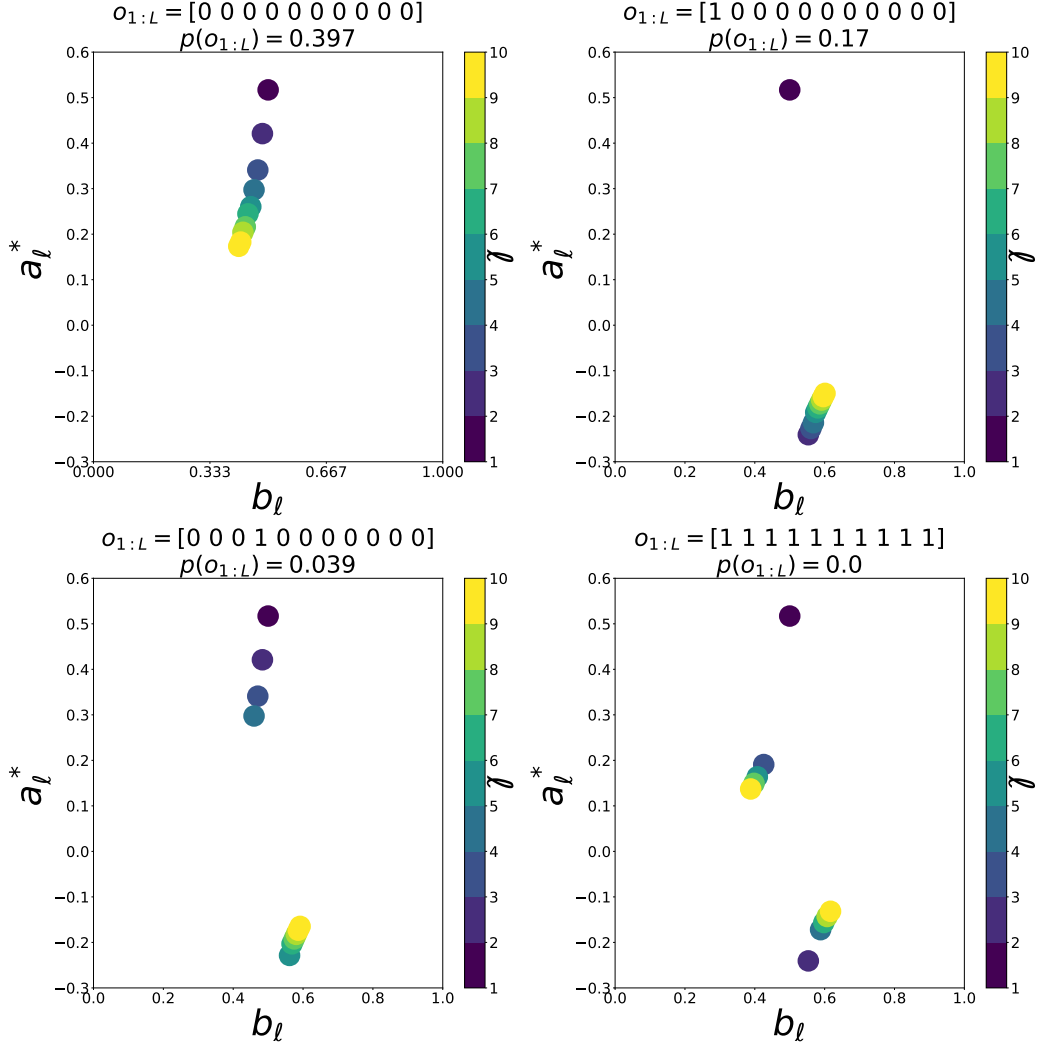


Figure 3.10: We show the structure of the optimal policy for a 10-Dolinar receiver, where we consider  $\alpha = 0.4$ . Each panel corresponds to a different string of observations obtained through a possible experiment realisation. At each layer, the optimal policy obtained through dynamic programming instructs the agent to perform an action (a displacement) according to the current belief  $b_\ell$  on state  $|\alpha\rangle$ ; such belief is updated as more information about the environment state is acquired. Thus, in order to obtain this plot, we propagate the initial prior (set to  $\frac{1}{2}$ ) as per Eq. 3.34, using the corresponding measurement outcome. Then, we find the interpolate the optimal action that corresponding to that belief, and repeat this along the 10 layers. Here, we do also compute the total probability of finding each sequence, as per  $p(o_{1:L}) = \sum_k p(o_{1:L}|\alpha_k)\eta_k$ . Such probability is extremely low for the rightmost sequence (all ones), here rounded to zero, and considerably high for the sequence of all zero outcomes (see discussion in the main body).

### Unveiling Dolinar's strategy from the numerics

It is worth discussing the optimal strategy, *i.e.* which are the displacements obtained of each layer after the dynamic programming optimization for a fixed  $L$ . In Fig. 3.10 we illustrate this optimal strategy, for the 10-Dolinar receiver. Here, we do only consider a subset of the  $2^{10}$  possible combinations of measurement outcomes that can be observed in a single discrimination experiment.

As with Kennedy receiver, Dolinar one is equipped with on/off photodetectors, which projects the incoming state either into the vacuum  $|0\rangle\langle 0|$  (no photons), or into its complement  $\mathbb{I} - |0\rangle\langle 0|$  (one or more photons). Displacing the signal by  $\beta$  right before such measurement acts as a projection  $|\beta\rangle\langle\beta|$  (or  $\mathbb{I} - |\beta\rangle\langle\beta|$ ), and as outlined in Sec. 3.1.2 there is a symbiosis between displacements and photon-detection that surpasses the so-called *homodyne limit*, when optimizing over  $\beta$ . While for Kennedy receivers  $\beta = \alpha$ , leading to measure either  $|0\rangle$  or  $|2\alpha\rangle$ , optimized Kennedy receivers find a better balance between the two contributions by displcing the signal with differnet value (see Fig. 3.6). However, the bottom line is similar: after displacing by a value  $\beta > 0$  and measuring, if outcome 0 was obtained, a bet for  $|-\alpha\rangle$  is done, and if outcome 1 is obtained,  $|\alpha\rangle$  is then the most likely hypothesis for which the agent bets.

Dolinar receiver exploits this idea, and carries out partial tests in a sequential way. By splitting the signal into  $L$  parts, the receiver essentially has now  $L$  copies of the state, each attenuated by a factor  $\sqrt{L}$ . Thus, until a click is detected (measurement outcome 1), it proceeds similarly to the Kennedy receiver, *i.e.* by testing if the state was  $|-\alpha\rangle$ . In this sense, the belief of having such state increases according to the number of 0's, as can be seen in the Fig. 3.10. We also observe that there is a dependence on the value of  $\beta$  with the layer number  $\ell$ , whereas its sign remains positive. However, if the first measurement result happens to be 1, the preferred hypothesis flips to  $|\alpha\rangle$ , since by having performed a positive displacement, that is the most-likely state. From there, Dolinar receiver verifies if such claim holds, by testing on the remaining layers whether the state was  $|\alpha\rangle$ : this is done by changing the sign of the displacement. In turn, if the state  $|\alpha\rangle$  was displaced by a value  $\beta < 0$ , then it is more likely to detect outcome 0. As we can observe in the Figure, if further outcomes happen to be 0, then the receiver keeps testing  $|\alpha\rangle$  through negative displacements. On the contrary, if outcome 1 is obtained, the strategy flips again (bottom-left panel). The success of Dolinar receiver hinges on the fact that having sequences containing so many consecutive 1's is vanishingly small. In turn, for  $\alpha = 0.4$ , we observe that the probability of encountering such a string of outcomes is as small as  $10^{-16}$ .

We remark that a similar reasoning for the structure of the optimal controls

applies in the time-domain Dolinar receiver (see Sec. 3.1.4), as discussed in Refs. [Dol; TSV+05; ADPP11].

We are now in position to turn agnostic, and introduce our reinforcement-learning agent that adapts the receiver to the measurement outcomes sampled from an unknown probability distribution. In such situation, the agent has no other choice than interacting with the receiver through several repetitions of the discrimination experiment, and thus the methods outlined in this Section cannot be applied.

### 3.3 The model-free approach

This Section represents the main contribution of the Chapter, namely the real-time model-free calibration of coherent-state receivers.

Here, the reinforcement-learning agent is asked to optimally-calibrate a  $L$ -Dolinar receiver, but is provided with essentially no information about the setting: she can only select actions according to the experience gathered via sequential repetitions of the discrimination experiment.

By the end of each episode (each discrimination experiment), the agent is only given *one bit of information*, *i.e.* the reward that accounts for the correctness of her guess. Thus, if the agent is able to calibrate the receiver such that it performs near-optimal actions  $\{a_\ell\}$ , then the rewards enjoyed will be higher on average. This stresses the difficulty of the setting proposed: *even if the receiver was calibrated to perform optimally, there is a chance of enjoying no reward*. In turn, the reward is a Bernoulli-distributed random variable with success probability  $P_s^L$ , and even the optimal configuration in Eq. 3.13 will experience a zero reward with probability  $1 - P_*^L$ .

#### 3.3.1 Reinforcement learning the Dolinar receiver

Reinforcement learning was discussed in Sec. 2.6, where the versatility of such framework was stressed. In particular, once a problem of interest is translated into the RL formalism, algorithms such as Q-learning (introduced in Sec. 2.6.3) can readily be applied.

For this purpose, we now need to specify the corresponding definitions of states, actions, rewards and episodes specific in our problem of calibrating the  $L$ -Dolinar receiver. Hence, we now state:

- Each episode  $t$  corresponds to an independent discrimination experiment, with a new default state  $s_0 = \alpha^{(k)}$  sampled from  $p_k$ ,  $k \in \{0, 1\}$

- Each episode consists of  $L + 1$  time-step  $\ell = 0, \dots, L$ , corresponding to the  $L$  detection layers followed by the final guessing stage;
- The possible states of the environment at time-steps  $\ell$  are  $s_\ell = \alpha_\ell^{(k)}$ , i.e., the transmitted part of  $s_0$  at that layer;
- The agent is not aware of the state  $s_\ell$ , in particular it does not know which hypothesis is true, but it can observe the measurement outcome  $o_\ell, 0 < \ell \leq L$ ;
- The actions  $a_\ell$  available at time-step  $0 \leq \ell < L$  are the displacements  $\beta_\ell$  available at that layer, conditioned on the history of observations and actions  $h_\ell = (a_0, o_1, \dots, a_{\ell-1}, o_\ell)$ , while at the last step they constitute the guess,  $a(h_L) = \hat{k} \in \{0, 1\}$ . For a given state  $s_\ell$ , the set of available actions is denoted as  $\mathcal{A}(s_\ell)$ .
- The reward  $r \in \{0, 1\}$  is non-zero only at the end of the episode and provided that the guess is correct, hence the transition function for the environment  $\tau(r, s'|s, a)$  is

$$\tau(\alpha_{\ell+1}^{(k')} | \alpha_\ell^{(k)}, a_\ell) = \delta(k', k) \quad \forall \ell \leq L, \quad (3.37)$$

$$\tau(r_{L+1} | \alpha_L^{(k)}, a_L) = \delta(r_{L+1}, 1) \delta(a_L, k), \quad (3.38)$$

were we omitted the trivial reward for  $\ell \leq L$ .

- Measurement outcomes obtained from the  $\ell$ -th photodetector will be denoted as  $o_\ell$ . The set of possible observations is denoted as  $\mathcal{O}$ .
- A *policy* will correspond to the entire set of conditional actions and guess  $\{a_\ell\}$ , which can be thought as a pre-defined receiver configuration. In particular, we target to find the *optimal* configuration, i.e. that leading to the highest expected reward.

Additionally we set  $\gamma = 1$  since the process has a finite horizon ( $L$ ); recall that such parameter acts as a regularizer for the *return* function in infinite-horizon problems (see, e.g. Eq. 2.241).

Similarly to the model-aware situation studied in Sec. 3.2, we need to specify agent's reconstruction of environment state (recall we are dealing with a Partially Observable Markov Decision Process, and as such the state of the environment is not fully observable by the agent at each time-step). While when applying dynamic programming techniques we defined agent's state to be the belief of

$b_\ell$  of having  $|\alpha\rangle$ , the model-free agent is not able to perform the bayesian update, since it ignores outcomes probabilities distributions.

Thus, we define agent's state as the history  $h_\ell$  of observations and actions made up to the  $\ell$ -th stage. An important consequence of such definitions is that the expected return, departing from the initial state (that is, the state value function of  $s_0$ , independent of agent awareness) is the success probability associated to agent's policy.

In contrast to the dynamic programming approach, the agent needs now to estimate the state-value functions out of several repetitions of the experiment, in order to asses how valuable a given configuraton is.

In particular, in Sec. 3.3.3 we will show that the optimal state-action value functions are in close correspondence to the ultimate attainable success probability  $P_*^L$ . Thus, if the agent is able to find such state-action value functions, then it can readily construct the optimal policy, as described in Sec. 2.6.2. To do so, we now bring the Q-learning algorithm into the play.

### 3.3.2 Q-learning the Dolinar receiver

We will now apply the Q-learning algorithm to  $L$ -Dolinar-receiver calibration problem.

Here, we restrict to  $L = 2$  processing layers and fix the attenuation coefficients to give equal amplitude at each layer, since the advantage obtained in the success probability is small when optimizing over them as well, see Fig. 3.9.

The Q-learning algorithm was introduced in Sec. 2.6.3, and we recall that is consists in exploiting the contractive property of the optimal state-action value function. In our setting, such quantity  $Q^*(h_\ell, a_\ell)$  is the average reward that will be enjoyed by the agent, when departing from state  $h_\ell$ , performing action  $a_\ell$ , and following the optimal policy afterward. As expected, this quantity reduces to the success probability, as explicetely shown in Sec. 3.3.3.

Since the  $Q$  value is unavailable to the agent, she relies on its estimate  $\hat{Q}$ , and a tradeoff between exploring and exploiting *greedy* policies appears, which in Q-learning is tackled by following an  $\epsilon$ -greedy policy. This consists on choosing, for a given state  $h_\ell$ , a greedy action with probability  $\epsilon$  (e.g. the action that maximizes  $\hat{Q}(h_\ell, a_\ell)$ ), or to act randomly with probability  $1 - \epsilon$ .

In contrast to the model-aware case, where the guessing rule was straightforwardly obtained from the Bellman equation at the last time-step, the optimization of state-action value function includes a non-trivial search for the optimal guessing rule, determined by the most likely hypothesis. For a given state  $h_\ell$ , the agent needs

to select a displacement value out of a set  $\mathcal{A}(h_\ell)$ , which here consists on 21 points for each displacement, each one ranging from  $-1$  to  $1$  with step  $0.1$ , leading to a fairly large state-action space: the agent has  $21 \times 2 \times 21 \times 2 \times 2 = 3528$  possible  $Q$ -values to learn from (including the last guess). We note that each discretized displacement is an independent action or “button” in the eyes of the agent —the agent is dispossessed of any notion of closeness between buttons corresponding to similar values.

In particular, we have studied different schedules for  $\epsilon$ . While an  $\epsilon$  close to  $1$  will not demonstrate an empirical success rate better than random guesses, an  $\epsilon$  value that drops to zero too fast will get stuck in sub-optimal actions (since the agent would not dispose of enough episodes in order to explore the entire state-action space). The update rule for the agent’s estimate  $\hat{Q}$  is given by Eq. (2.248), with  $s_\ell \rightarrow h_\ell$  and learning rates  $\lambda_t(h, a) = N_t(h, a)^{-1}$  being the inverse of the number of times a state-action pair has been visited. This choice guarantees convergence as per Eq. (2.250).

As the behaviour of the RL agent strongly depends on the actions chosen at early episodes, we averaged the learning curves over 48 agents. Our results are compared with: (i) the maximum success probability attainable with this number of layers and discretization of displacements, Eq. (3.27), and (ii) the success probability attainable via a standard homodyne measurement, which is optimal among Gaussian receivers.

Following our discussion about the figures of merit in Sec. 2.6.1, we have evaluated the performance of our model-free agents at episode  $t$  using two figures of merit: (i) the cumulative return per episode,

$$\mathbf{R}_t = \frac{1}{t} \sum_{i=1}^t G_0^{(i)} = \frac{1}{t} \sum_{i=1}^t r_{L+1}^{(i)}, \quad (3.39)$$

where  $r_{L+1}^{(i)} = \{1, 0\}$  stands for the correctness of the guess made at episode  $i$ , and (ii) the success probability of the best actions according to the agent, at the current episode,

$$\mathbf{P}_t = P_s(\alpha, \{a_\ell^{(t)*}\}), \quad (3.40)$$

where the best actions  $\{a_\ell^{(t)*}\}$  at episode  $t$  are obtained by going greedy with respect

to the current  $Q$ -estimate, i.e.,

$$\begin{aligned}
 a_0^{(t)*}(h_0) &= \operatorname{argmax}_{a \in \mathcal{A}(h_0)} \hat{Q}(h_0, a) \rightarrow h_1^* = (o_1, a_0^{(t)*}) \\
 a_1^{(t)*}(h_1^*) &= \operatorname{argmax}_{a \in \mathcal{A}(h_1^*)} \hat{Q}(h_1^*, a) \rightarrow h_2^* = (a_0^{(t)*}, o_1, a_1^{(t)*}(h_1^*), o_2) \\
 &\dots \\
 a_L^{(t)*}(h_L^*) &= \operatorname{argmax}_{a \in \mathcal{A}(h_L^*)} \hat{Q}(h_L^*, a).
 \end{aligned} \tag{3.41}$$

The first figure of merit,  $\mathbf{R}_t$ , is usually employed to describe the learning process in reinforcement learning, and it evaluates the success rate obtained by the agent so far. On the other hand, the second figure of merit,  $\mathbf{P}_t$ , is standard in quantum state discrimination, and in our context it evaluates the best strategy discovered by the agent so far. Note that such quantity is unavailable to the agent, but is here used to monitor its learning process.

As  $t \rightarrow \infty$ , for a *good* learner it is expected that  $\mathbf{R}_t \rightarrow \mathbf{P}_t$ , i.e. with enough agent-environment interactions the average reward should tend to the success probability for the best actions found by the agent, which in turn should converge to the optimal success probability  $P_*^{(L)}$ . Therefore, the learner is not only expected to find a good discrimination strategy, but to also follow it: the interaction policy should tend to the optimal policy. This feature is captured by the evolution of  $\mathbf{R}_t$  over different episodes: a good learner is asked to obtain as much reward as possible *during* the learning process.

In Fig. 3.11 we show the evolution of these two figures of merit for  $Q$ -learning agents with three different  $\epsilon$ -greedy interaction policies: (i) a completely random one, i.e.,  $\epsilon = 1$ , (ii) a 0.3-greedy one, i.e.,  $\epsilon = 0.3$ , and (iii) a dynamic one (exp-greedy) that becomes exponentially greedier as time passes, i.e.,  $\epsilon(t) = \max\{e^{-\frac{t}{\tau}}, \epsilon_0\}$ , with  $\epsilon_0 = 10^{-2}$ <sup>1</sup>.

In the first place we note in Fig. 3.11 that, as we mentioned above, a fully random search over the action space (1-greedy policy) leads to the extremely poor cumulative reward per episode of  $\mathbf{R}_t \approx 1/2$ , even for long times, which is expected because a random guess (last action) leads to  $P_s(\alpha, \{a_\ell\}) = 1/2$ . Instead, since all the actions will be sampled enough times for the agent to learn the optimal policy,  $\mathbf{P}_t$  will converge to optimal value at long enough episode number. Nevertheless, if the action space is large, the fully random strategy will require a large number of episodes to explore each action a significant number of times, and for moderate

<sup>1</sup>This choice assures that at initial episodes the agent favours exploration, whereas at  $t = \tau \log \frac{1}{\epsilon_0}$  the agent's behaviour collapses to an  $\epsilon_0$ -greedy policy.

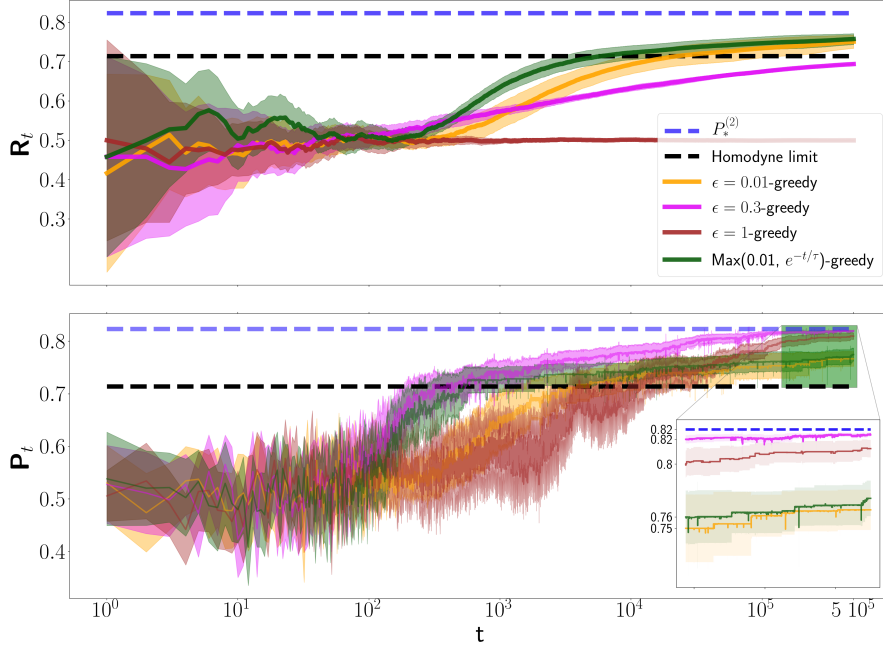


Figure 3.11: We benchmark traditional Q-learning with different *schedules* on  $\epsilon$  as the episode number increases. The figures of merit are averaged over  $A = 48$  agents and show the corresponding uncertainty region.

times a  $\epsilon$ -greedy strategy might reach a better strategy. Indeed, Fig. 3.11 shows that the 0.3-greedy policy has at all episodes a higher  $\mathbf{P}_t$  than the 1-greedy one, being 99% the optimal success probability  $P_*^{(L=2)}$  at episode  $t = 10^5$ . Of course, for 0.3-greedy policy the agent collects many more rewards (actual correct guesses) than for the 1-greedy but it is still limited to  $\mathbf{R}_t \approx 0.7P_*^{(L=2)}$ . In order to reach a better exploration-exploitation trade-off, it is customary to consider an episode-dependent  $\epsilon$ , e.g.  $\epsilon(t) = \max\{e^{-\frac{t}{\tau}}, \epsilon_0\}$ . Fig. 3.11 shows the results for this tunable interaction policy with  $\tau = 2 \cdot 10^2$  and  $\epsilon_0 = 0.01$ . This allows the agent's  $\mathbf{R}_t$  to surpass the homodyne limit at about episode  $\sim 5 \cdot 10^3$  (which is comparable with the size of the action space), while at later times the performance converges to that of the 0.01-greedy policy. Note also that 0.3-greedy discovers a strategy whose  $\mathbf{P}_t$  surpasses the homodyne limit at episode  $\sim 3 \cdot 10^2$ .

In Fig. 3.12 we study the guessing rule discovered by the 0.3-greedy agent at



episode  $t = 5 \cdot 10^5$ . For each sequence of outcomes  $o_1, o_2$ , we plot the difference between the  $Q$ -values of guessing for  $|\!-\alpha\rangle$ , i.e.,  $a_L = 1$ , and  $|\!+\alpha\rangle$ , i.e.,  $a_L = 0$ , as a function of the past actions:

$$\hat{Q}((a_0, o_1, a_1(h_1), o_2), 1) - \hat{Q}((a_0, o_1, a_1(h_1), o_2), 0). \quad (3.42)$$

Note that the sign of Eq. (3.42) corresponds to the agent's best guess for the true hypothesis, since the latter is obtained by going greedy towards  $\hat{Q}(h_L, a_L)$ , as explained below. We compare these results with the optimal guessing rule in the model-aware setting, plotting a shaded region when the maximum-likelihood guess is  $|\pm\alpha\rangle$ . The plot shows the agent perfectly learn the guessing rule at the given resolution. Moreover, the difference between the two  $Q$ -values is more pronounced in the surroundings of the optimal  $\beta$  values, meaning that the agents are more confident about their guess in these regions.

Our numerical results indicate that standard Q-learning successfully trains agents that surpass the homodyne limit of optical detection and discover strategies whose error rate is comparable with that of the optimal receiver. This is remarkable, especially taking into account that the agents are not initially trained for this task, and run in a model-free setting entirely based on the feedback they get (correct/incorrect) on their guess. While many reinforcement learning schemes focus on extracting the optimal policy from the agent (as measured *e.g.* by  $\mathbf{P}_t$ ), our central figure of merit,  $\mathbf{R}_t$ , captures the real performance of the agent, and can actually be assessed by the agent itself. It is hence important to design strategies that not only aim at finding the optimal policy within an episode, but also maximize the cumulative reward per episode, reaching  $\mathbf{R}_t \rightarrow P_*^{(L)}$  as fast as possible.

As we discussed in Sec. 2.6.1, bandit theory precisely captures such tradeoff by means of the expected cumulative regret. In next sections, we will actually borrow ideas from such a field, in order to enhance the Q-learning agents. Nevertheless, before doing so, let us show how the sequential structure of the receiver is exploited by the Q-learning algorithm.

### 3.3.3 Optimal state-action values & convergence

In this section we verify that, by construction, the optimal policy leads to the maximum success probability  $P_*^{(L)}(\alpha)$ . It is assumed that  $Q$  is always associated with the optimal policy  $\pi^*$ ; we simplify notation by  $Q = Q_{\pi^*}$ .

At step  $L$ , given any history  $h_L$ , the actions available to the agent are  $\hat{k} = a_L$ , i.e. guessing for one of the possible phases of the coherent state. The  $Q$ -values at

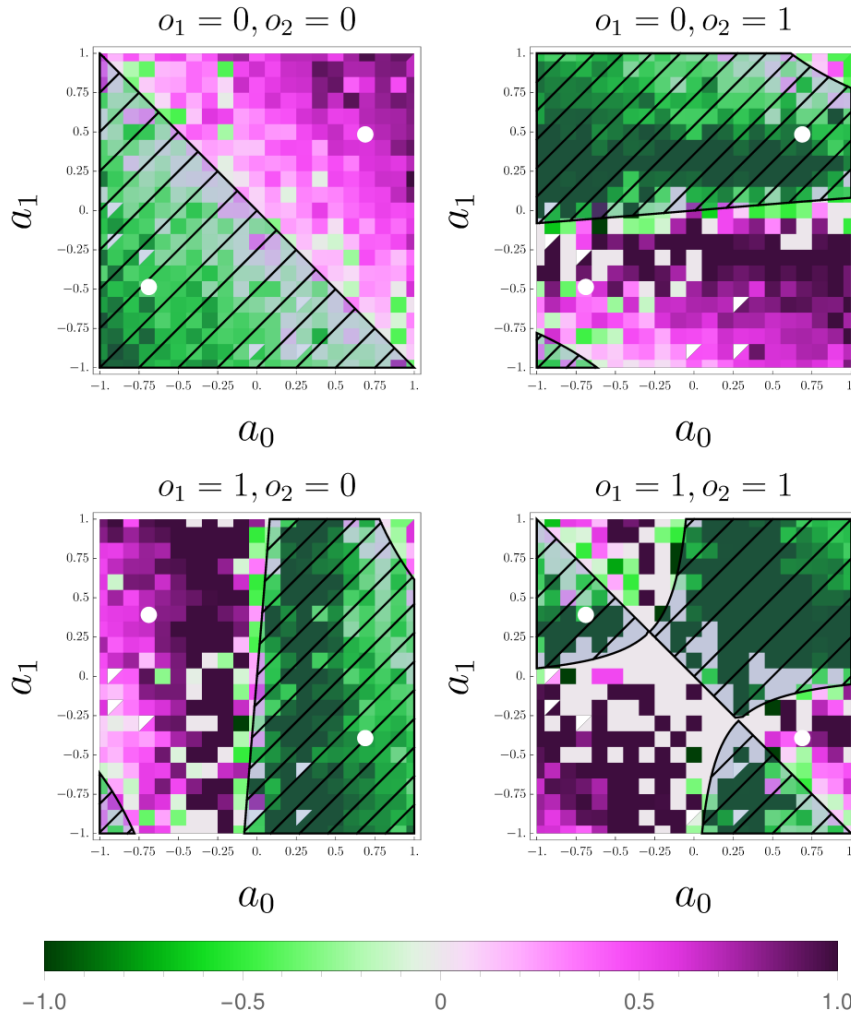


Figure 3.12: Density plot of the difference between the estimated  $Q$ -values for guessing “plus” and “minus” as a function of the displacements at the first and second layer, for each possible sequence of outcomes, with  $\alpha = 0.4$ . The shaded areas correspond to the regions where the optimal guess, taken according to maximum-likelihood, is “plus”. The white dots corresponds to the optimal values of the displacements for the proper discretization).

this time-step read as

$$\begin{aligned} Q(h_L, a_L) &= \mathbb{E}[G_L | h_L, a_L] = \sum_{r_{L+1}} r_{L+1} p(r_{L+1} | h_L; a_L) \\ &= p(\alpha^{(a_L)} | o_{1:L}, a_{0:(L-1)}), \end{aligned} \quad (3.43)$$

with  $o_{1:\ell} = \{o_1, o_2, \dots, o_\ell\}$  the observations obtained up to the  $\ell^{\text{th}}$  photodetector, and  $a_{0:\ell} = \{a_0, a_1, \dots, a_\ell\}$  the actions done up to step  $\ell$ . Such actions are deterministic functions of  $h_\ell$ , and thus we aim to optimize over all possible function choices (and for this reason we will here use the semi-colon notation in the probabilities).

Recalling that the optimal action, given  $h_\ell$ , is obtained from  $Q$  as  $a^*(h_\ell) = \operatorname{argmax}_{a_\ell} Q(h_\ell, a_\ell)$ , the optimal guess  $a_L^*$  is the one of maximum-likelihood:

$$a_L^* = \operatorname{argmax}_{a_L} p(\alpha^{(k)} | o_{1:L}; a_{0:(L-1)}) \Big|_{k=a_L}.$$

By definition of the optimal policy and because the optimal Bellman equation Eq. (2.247) holds, namely

$$Q^*(s, a) := Q_{\pi^*}(s, a) = \max_{\pi} Q_{\pi}(s, a) \quad (3.44)$$

$$= \sum_{s' \in \mathcal{S}, r \in \mathcal{R}} \tau(s', r | s, a) (r + \gamma \max_{a' \in \mathcal{A}} Q^*(s', a')), \quad (3.45)$$

then the optimal action to take given history  $h_{L-1}$  at step  $L - 1$  is

$$\begin{aligned} a_{L-1}^* &= \operatorname{argmax}_{a_{L-1}} Q(h_{L-1}, a_{L-1}) \\ &= \operatorname{argmax}_{a_{L-1}} \sum_{o_L} p(o_L | o_{1:(L-1)}; a_{0:(L-1)}) \max_{a_L} Q(h_L, a_L) \\ &= \operatorname{argmax}_{a_{L-1}} \sum_{o_L} p(o_L | o_{1:(L-1)}; a_{0:(L-1)}) \max_k p(\alpha^{(k)} | o_{1:L}; a_{0:(L-1)}) \\ &= \operatorname{argmax}_{a_{L-1}} \sum_{o_L} \frac{\max_k p(o_{1:L} | \alpha^{(k)}; a_{0:(L-1)}) p_k}{p(o_{1:(L-1)}; a_{0:(L-1)}), \end{aligned} \quad (3.46)$$

where in the last line we have used Bayes theorem. Following this line of reasoning, we can obtain the optimal actions  $a_\ell^*$  at any time-step. In particular, for  $\ell = 0$ , by recursively applying the optimal Bellman equation (Eq. (2.247)) we have

$$\begin{aligned}
 Q(h_0, a_0) &= \sum_{o_1} p(o_1; a_0) \max_{a_1} Q(h_1, a_1) & (3.47) \\
 &= \sum_{o_1} p(o_1; a_0) \max_{a_1} \sum_{o_2} p(o_2|o_1; a_1) \max_{a_2} Q(h_2, a_2) \\
 &= \sum_{o_1} p(o_1; a_0) \max_{a_1} \sum_{o_2} p(o_2|o_1; a_1) \max_{a_2} \sum_{o_3} \left( \right. \\
 &(\dots) \left. \sum_{o_L} p(o_L|o_{1:(L-1)}; a_{1:(L-1)}) \max_{a_L} Q(h_L, a_L) \right) \\
 &= \sum_{o_1} \max_{a_1} \sum_{o_2} \max_{a_2} \sum_{o_3} \left( \right. \\
 &(\dots) \left. \sum_{o_L} \max_k p(o_{1:L}|\alpha^{(k)}; a_{0:(L-1)}) p_k \right).
 \end{aligned}$$

Therefore, by taking the optimal action  $a_0^* = \operatorname{argmax}_{a_0} Q(h_0, a_0)$ , we obtain

$$\max_{a_0} Q(h_0, a_0) = p_*^{(L)}, \quad (3.48)$$

which was defined in Eq. 3.13, though here we dropped the dependence on  $\alpha$ . As pointed out in the main text, the value and action-value functions are related via  $v_\pi(s) = \sum_a \pi(a|s)Q_\pi(s, a)$ . Therefore, the optimal value function for the initial state is the optimal success probability:

$$v_*(h_0) = \sum_a \delta(a, \operatorname{argmax}_a Q(h_0, a)) = Q(h_0, a_0^*) = P_*^L(\alpha). \quad (3.49)$$

In Fig. 3.13 we show several sections of the estimates  $\hat{Q}$ , using 1-greedy as the interaction policy and each update made according  $Q$ -learning, at episode  $t = 10^8$ ,

The success of  $Q$ -learning agents in finding the optimal configurations of Dolinar-like receivers crucially depends on accurately estimating the success probability of the relevant configurations. Such configurations correspond to the *interesting* region of the state-action space (*i.e.* where high rewards are allocated), and as we have seen, they are related to optimal state-action value functions  $Q^*$ .

From such optimal values, the agent can straightforwardly follow an optimal policy  $\pi^*$  by going greedy with respect to them. Nevertheless, as we stressed in Sec. 2.6.2, learning begins without the knowledge of how to access those high-rewarding configurations, and hence the agent needs to balance between exploring

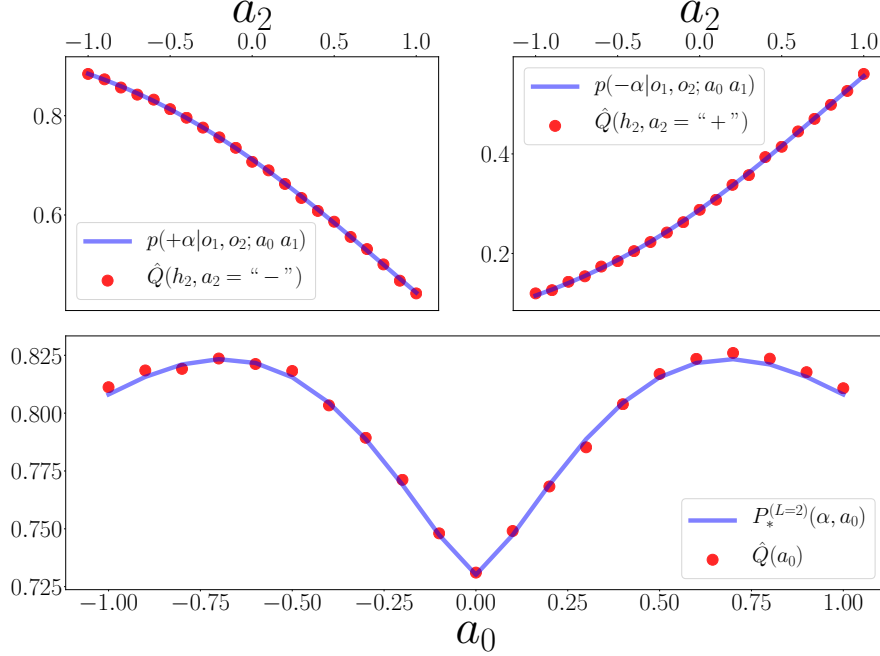


Figure 3.13: We plot different values of the Q-estimates, after  $10^8$  episodes of random exploration ( $\epsilon = 1$ ), updating the Q-estimates according to Q-learning (see Algorithm 1). The random exploration is used in order to ensure that, at *finite* number of episodes, all state-action pairs were equally visited on average. The lower plot corresponds to the estimates  $\hat{Q}(a_0)$ , and it is compared with the optimal success probability as a function of  $a_0$ , i.e.  $P_*^{(L=2)}(\alpha, a_0) = \sum_{o_1} p(o_1; a_0) \max_{a_1} \sum_{o_2} p(o_2 | h_1, a_1) \max_{a_2 = \pm} p(\pm \alpha | h_2) pr(\pm \alpha)$ .

new ones (through displacements that has not been performed previously) and performing displacements that have previously led to a reward of 1 with a high probability. In the latter case, statistical fluctuations when estimating success probabilities out of measurement outcomes can potentially hide the fact that some displacements are actually sub-optimal. To balance such trade-off, Q-learning selects a displacement according to an  $\epsilon$ -greedy strategy.

As introduced in Sec. 2.6.1, bandit theory defines the so-called regret, that precisely captures the *exploration-exploitation* tradeoff in Markov Reward Processes. In that Section, we introduced strategies based on Upper Confidence Bounds (UCB) or Thompson Sampling (TS) which outperforms  $\epsilon$ -greedy ones in terms of the

expected regret. It is thus sensible to ask whether the  $Q$ -learning agent can be enhanced with such bandit-based strategies, and hence we now turn to study how these enhanced agents perform in our BPSK coherent-state discrimination problem.

### 3.3.4 A little help from my bandit friends

As introduced in Sec. 2.6.1, there exists other strategies than  $\epsilon$ -greedy ones, that balance between exploration and exploitation in Markov Reward Processes (MRP). Among them, we studied UCB and TS ones. In this section we will equip the  $Q$ -learning agent introduced in Sec. 3.3.1 to enhance the model-free calibration.

*Calibration of Kennedy-like receivers as a bandit-problem.* We will first consider the model-free calibration of a Kennedy-like receiver as a bandit-problem. To this end, we have numerically compared the performance of the three bandit policies discussed in Sec. 2.6.1. The bandit problem that we consider deals with three different Bernoulli distributions, each associated to a different value of a Kennedy receiver. At each episode the agent observes a reward of value either 0 or 1, obtained by guessing (max-likelihood) for the underlying phase of the coherent state by using the selected receiver. In Fig. 3.14 we show the expected regret  $\mathcal{L}_t$  estimated out of several realizations of such 3-armed bandit problem. The figure shows that the cumulative regret scales linearly with time for the  $\epsilon$ -greedy strategy, while it has a logarithmic scaling for the UCB and TS strategies. The inset shows the cumulative regret as a function of  $\log t$  together with the ultimate bound given by Lai-Robbins bound; we observe that sub-leading constants seem to still be relevant in such a bound, and that whereas asymptotic regime seems not to have been reached, the results are yet consistent with Eq. 2.235.

We will now turn to use this strategies to improve the performance of the  $Q$ -learning agent considered previously. In particular, we will present two alternative strategies for the action exploration to be applied in the  $Q$ -learning algorithm.

The first strategy employs the standard  $Q$ -learning update rule for the estimate  $\hat{Q}$ , as described in Eq. (2.248), but it uses UCB to determine the interaction policy at each time-step of each episode. Here, we recall that UCB chooses the action according to an upper confidence bound of the form of Eq. 2.238:

$$\text{ucb}_t(a) := \hat{Q}(a) + \varepsilon(t) = \hat{Q}(a) + \sqrt{\frac{-\log \mathcal{P}(t)}{2N_t(a)}}, \quad (3.50)$$

which represents an upper bound to the true value  $Q(a)$  with a high probability  $1 - \mathcal{P}(t)$ .

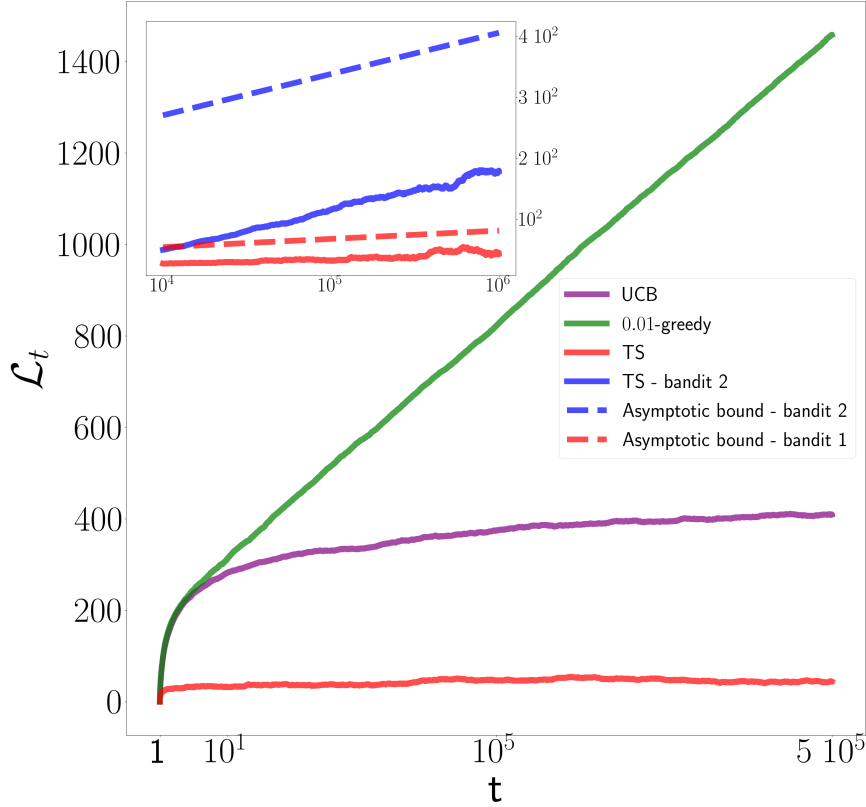


Figure 3.14: We show the evolution of the cumulative regret for three different policies:  $\varepsilon$ -greedy, UCB and TS in a bandit setting. The mean values considered are associated to success probabilities of Kennedy-like receivers, as introduced in Sec. 3.1.2. All curves are averaged over  $10^3$  agents. Specifically, they are Bernoulli distributions with different mean values, each associated to different configurations of a quantum receiver parametrized by a value  $\beta$ . For *bandit problem 1*, we considered  $\beta \in \{0, -\alpha, \beta^*\}$ , with  $\alpha = 0.4$  and  $\beta^* = -0.74$ . Furthermore, we compare the asymptotic behaviour of TS, studying *bandit problem 2*, where  $\beta \in \{-\alpha, \beta^*, -1.5\alpha\}$ , shown in the inset plot.

We have here chosen  $\mathcal{P}(t) = t^{-4}$ . This is implemented by keeping a count of the number of visits of each history-action couple up to the current episode  $t$ , i.e.,  $N_t(h_\ell, a_\ell)$ , which is then used to compute an upper confidence bound,  $\text{ucb}_t(h_\ell, a_\ell)$

as in Eq. (2.238), for each action  $a_\ell$  and history  $h_\ell$ . Finally, at time-step  $\ell$  the agent chooses the greedy action with respect to the UCB, i.e.,  $a_\ell^{(t)} = \underset{a}{\operatorname{argmax}} \operatorname{ucb}_t(h_\ell, a)$ .

The second strategy is instead based entirely on TS, considering each action conditioned on the past history as a bandit problem and rewarding each sequence of actions that led to a successful experiment. In detail, the agent keeps a beta-distribution, Eq. (2.239), of the mean reward obtainable at each time-step  $\ell$  from each action  $a_\ell$  given each possible history  $h_\ell$ , i.e.,  $f_t(\bar{r}|h_\ell, a_\ell)$ . In order to choose a new action at time-step  $\ell$  given history  $h_\ell$ , the agent samples an expected reward  $\bar{r} \sim f_t(\bar{r}|h_\ell, a_\ell)$  for each  $a_\ell$  and selects the action with the largest sample  $\bar{r}$ . At the end of the episode a reward is obtained as usual, and  $f_t(\bar{r}|h_\ell, a_\ell)$  is updated in a Bayesian way for all the history-action couples visited at the episode. In this case, when computing  $\mathbf{P}_t$ , the best actions are chosen by going greedy with respect to their mean reward distribution  $f_t(\bar{r}|h_\ell, a_\ell)$  [RVK+18].

In Fig. 3.15 we show the evolution of the two figures of merit  $\mathbf{R}_t$ ,  $\mathbf{P}_t$  for several agents trained using these two enhanced strategies, as well as for those based on the exp-greedy and 0.3-greedy strategies, considered in Sec. 3.3.2, which had respectively the largest final  $\mathbf{R}_t$  and  $\mathbf{P}_t$  out of all the analyzed strategies. We observe that UCB performs a thorough exploration of the action space and indeed it is able to attain a value of  $\mathbf{P}_t$  close to that of 0.3-greedy. This result comes at the price of a small  $\mathbf{R}_t$  value, which nevertheless shows that UCB has better exploitation properties than 0.3-greedy; in particular it has a strikingly larger slope than the latter at long times. As for TS, we observe that this strategy attains the best  $\mathbf{R}_t$  values, surpassing exp-greedy at intermediate times. Moreover, TS also radically improves the values of  $\mathbf{P}_t$  with respect to exp-greedy and it is even able to attain the performance of the other two strategies that favour exploration. Overall, it appears that for this particular problem setting, and the hyperparameters chosen for all three algorithms, TS provides the most profitable balance of exploration and exploitation.

Overall, we observe that model-free Q-learning agents can readily learn to calibrate Dolinar-like receivers by beginning the learning task from the darkness, meaning that they completely ignore the setting at hand.

While such learning process is interesting on its own, we can now turn to non-ideal scenarios, which arguably is *the* motivation of our setting. The reason for this is, on the one hand, that the presence of unknown quantum channels might non-trivially act on the signals, and thus the optimal configuration will differ to that one of the idealistic, noise-less case. On the other hand, we have assumed so far that there are no imperfections during the measurement process, an issue that is certainly present in experiments. For this reason, we will now turn to benchmark



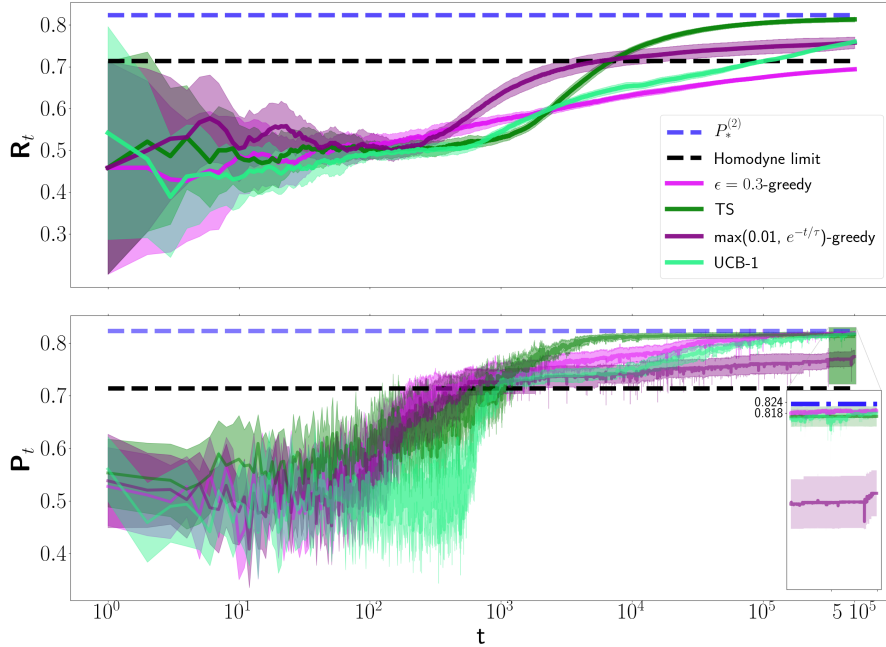


Figure 3.15: We show the learning curves for the enhanced Q-learning agents via bandit methods. On the upper plot we depict  $R_t$ , the agent’s success rate per episode, whereas on the bottom plot we depict  $P_t$ , the success probability of the agent’s recommended actions at episode  $t$ ,  $\{a_\ell^{(t)*}\}$ . Each of the learning curves is averaged over 48 agents; the amplitude was fixed to  $\alpha = 0.4$ .

the performance of our model-free agents in a wide range of noisy scenarios.

### 3.3.5 Noise robustness

We have already seen that our RL agents are able to learn near-optimal discrimination strategies and – most importantly – exploit them in real time, by employing exclusively the detectors’ outcomes and the rewards enjoyed by the end of each episode. In this section we show that such results do not sensibly change in the presence of noise, *i.e.*, that the same Q-learning agents are able to attain near-optimal performances even when unknown errors affect the experiment and hence the learning process. Here, we will consider dark counts, phase flip errors and the

presence of a compound lossy channel acting between sender and receiver.

*Dark counts.* First, we consider a common experimental imperfection known as dark counts: due to the presence of background noise, each photodetector of the receiver has a non-zero probability  $p_{\text{dc}}$  of detecting a photon even when it receives a vacuum signal. Accordingly, the conditional probability of obtaining an outcome 0 given an input state  $|\alpha\rangle$ , Eq. (3.10), is modified by a multiplicative factor  $(1 - p_{\text{dc}})$ . In Fig. 3.16 we plot  $\mathbf{R}_t$  and  $\mathbf{P}_t$  at time  $t = 5 \cdot 10^5$  for several RL strategies as a function of  $p_{\text{dc}} \in [0, 1]$ , along with the maximum success probability attainable by the corresponding receiver. We see that the final values of  $\mathbf{P}_t$  are near-optimal for all values of  $p_{\text{dc}}$ , while  $\mathbf{R}_t$  seems to be slightly affected in an intermediate region of values of  $p_{\text{dc}}$ . Since the agents operate on a completely model-free basis and the reward system has been chosen to ensure convergence of the value function to the true success probability, it can be expected that they are still be able to learn in the long term, as shown by the high values of  $\mathbf{P}_t$  attained. However, since a dark count effectively increases the chance of (not) obtaining a reward for a (correct) wrong action, the time it takes to learn a near-optimal strategy and to start exploiting it might increase, as shown by the behaviour of  $\mathbf{R}_t$ . Note that for  $p_{\text{dc}} \sim 1$  the best guess is the random one and thus easier to learn.

*Phase flips.* Next, we consider the case where the phase of the incoming signal is flipped before arriving to the receiver, with probability  $p_f$ . In this scenario, if the agent guesses for the correct received phase, the corresponding reward will be zero since the phase initially sent was opposite than the received one. In particular, the probability of observing a string of outcomes  $p(o_{1:L}|\alpha, \{a(h_{L-1})\})$  in Eq. (3.10) is modified such that

$$p(o_{1:L}|\alpha, \{a(h_{L-1})\}) \rightarrow (1 - p_f)p(o_{1:L}|\alpha, \{a(h_{L-1})\}) \quad (3.51)$$

$$+ p_f p(o_{1:L} | -\alpha, \{a(h_{L-1})\}) \quad (3.52)$$

In Fig. 3.17 we depict the values of  $\mathbf{R}_t$  and  $\mathbf{P}_t$  attained by several agents at episode  $t = 5 \cdot 10^5$ , as a function of  $p_f \in [0.5, 1]$ , along with the maximum success probability attainable by the corresponding receiver. As in the dark counts case, we see that for all values of  $p_f$ , the agents are able to converge to near-optimal  $\mathbf{P}_t$  values and they exhibit very small variations in  $\mathbf{R}_t$  as  $p_f$  increases.

### Compound lossy channels

Next, we consider the presence of lossy-channels, which are based in Ref. [BRC21]. These channels were introduced in Sec. 2.3.3, and consist on mixing the incoming state with the vacuum state  $|0\rangle$  by a beam-splitter. As such, lossy channels constitute a common model for long-distance optical-fiber and free/deep-space optical

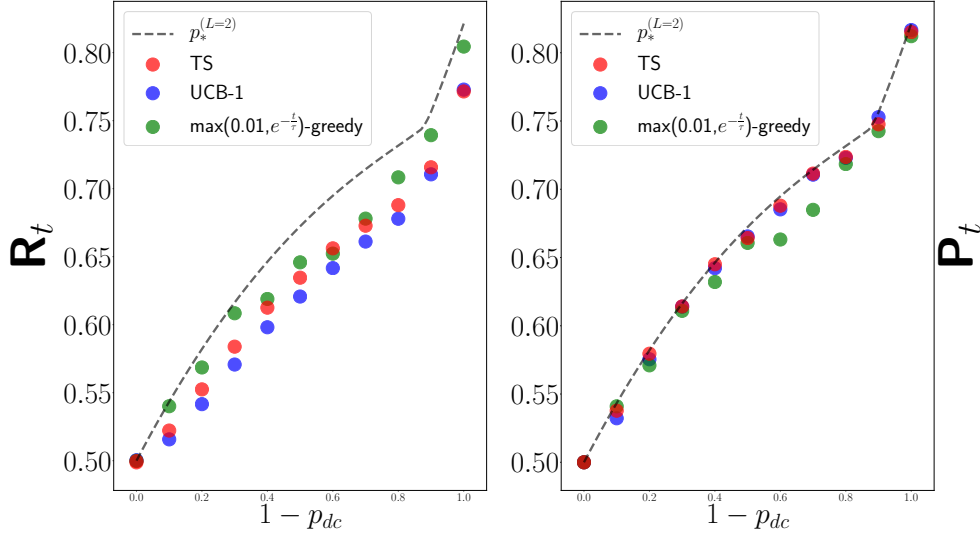


Figure 3.16: We compare the performance of the three RL agents (the same considered in Sec. 3.3.4) at episode  $t = 5 \cdot 10^5$ , for several photo-detection noise values (dark count rates). The amplitude of the coherent states is fixed to  $\alpha = 0.4$ ; all data points are averaged over 48 agents.

communication [DVR+19; AP05; UHP+12; Pir21a; Pir21b; VSV11; VSV+17], since they model attenuations of the incoming signal; their action on coherent states reads

$$\mathcal{L}_\eta : |\alpha\rangle \mapsto |\sqrt{\eta}\alpha\rangle, \quad (3.53)$$

where  $\eta \in [0, 1]$  is the transmissivity or attenuation coefficient of the lossy channel.

While we have seen that RL is effective in calibrating the receiver at different values of the noise-parameters, a considerably more challenging situation arises when practical communication links are affected by noise-parameter variations. In turn, these situation occurs in practice, and can be difficult to estimate and counteract in real-time. In particular, in the case of a lossy channel  $\mathcal{L}_\eta$ , the transmissivity  $\eta$  can be altered over time, a phenomenon known as *fading*, and which is caused by a plethora of different effects that alter the optical signal during its transmission through the atmosphere to/from a satellite [DVR+19; AP05; UHP+12; Pir21a; Pir21b; VSV11; VSV+17]. In this setting, the performance of all the receivers studied in this Chapter can be expected to degrade.

Here, we model such variable-loss channels by restricting to the case of two

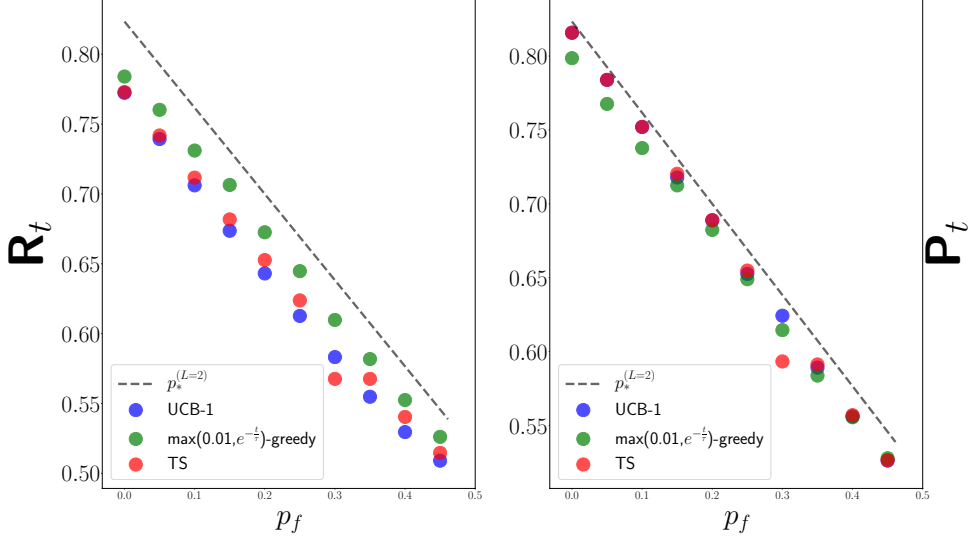


Figure 3.17: We compare the performance of the three RL agents (the same considered in Sec. 3.3.4) at episode  $t = 5 \cdot 10^5$ , as a probability of the signal being phase flipped before arriving to the receiver. The amplitude of the coherent states is fixed to  $\alpha = 0.4$ ; all data points are averaged over 24 agents.

lossy channels  $\{\mathcal{L}_{\eta_0}, \mathcal{L}_{\eta_1}\}$ , happening with probabilities  $\{\pi_0 = \pi, \pi_1 = 1 - \pi\}$ , and hence the success probability is now averaged over the possible channel realizations,

$$\bar{P}_s(\mathcal{M}) := \sum_{i=0,1} \pi_i P_s(\mathcal{M}; \eta_i). \quad (3.54)$$

As a consequence, the Helstrom bound will also be affected, and we will now discuss how to compute it in this case. Each hypothesis state  $|\alpha_k\rangle$  can take two different values, depending on the channel transmissivity. This effectively amounts to discriminate between the two mixed states

$$\rho_x := \sum_{i=0,1} \pi_i |\sqrt{\eta_i} \alpha_k\rangle \langle \sqrt{\eta_i} \alpha_k|, \quad k \in \{0, 1\}. \quad (3.55)$$

The average success probability of a generic measurement then reads

$$\bar{P}_s(\mathcal{M}) = \sum_{k=0,1} p_k \text{Tr}[M_k \rho_k] \quad (3.56)$$

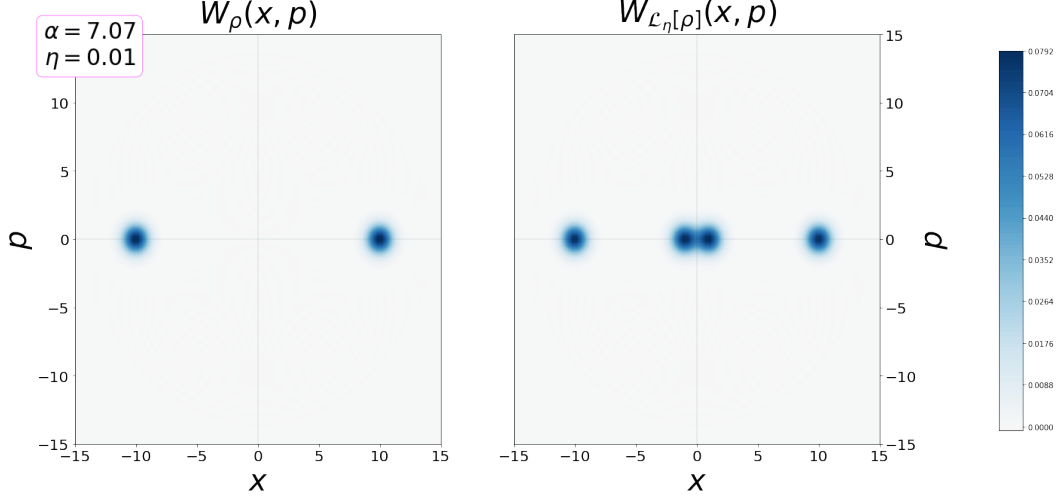


Figure 3.18: We show Wigner function of inputs and outputs states of the compound channel. We observe that each input state is splitted according to the possible attenuation value.

and in order to compute Eq. 2.191, we need to compute the trace-norm of the matrix  $q_0\rho_0 - q_1\rho_1$  in the four-dimensional subspace where it is supported,  $\mathcal{S} := \text{span}\{|\psi_{k+2i}\rangle := |\sqrt{\eta_i}\alpha_k\rangle\}_{i,k=0,1}$ . Since the trace-norm is basis-independent, one can readily find a four-dimensional representation of the involved states by computing the square-root of their Gram matrix,  $B = \sqrt{G}$  with  $G_{ij} = \langle\psi_i|\psi_j\rangle$ . By construction it holds that  $B^\dagger B = G$ , hence the column vectors of  $B$  constitute a representation of the original states in a 4-dimensional subspace of the Hilbert space. By expressing all the operators via this representation, we are then able to compute the Helstrom bound in Eq. (2.191) numerically.

Furthermore, we can readily compute how the homodyne success probability in Eq. 3.6 is affected by this channel,

$$P_s^{\text{hom}} = \frac{1}{2} \left( 1 + \sum_{i=0,1} \pi_i \text{Erf} \left[ \sqrt{2\eta_i} a \right] \right). \quad (3.57)$$

Finally, in the case of Dolinar-like receivers, we need to take into account that the outcome probability should now be averaged out considering the two different transmissivities:

$$p(j|\rho_k^{(\ell)}) := \sum_{i=0,1} \pi_i p(j|\sqrt{\eta_i}\alpha_k^{(\ell)}). \quad (3.58)$$

Thus, we are now in position to compare the success probabilities attained by

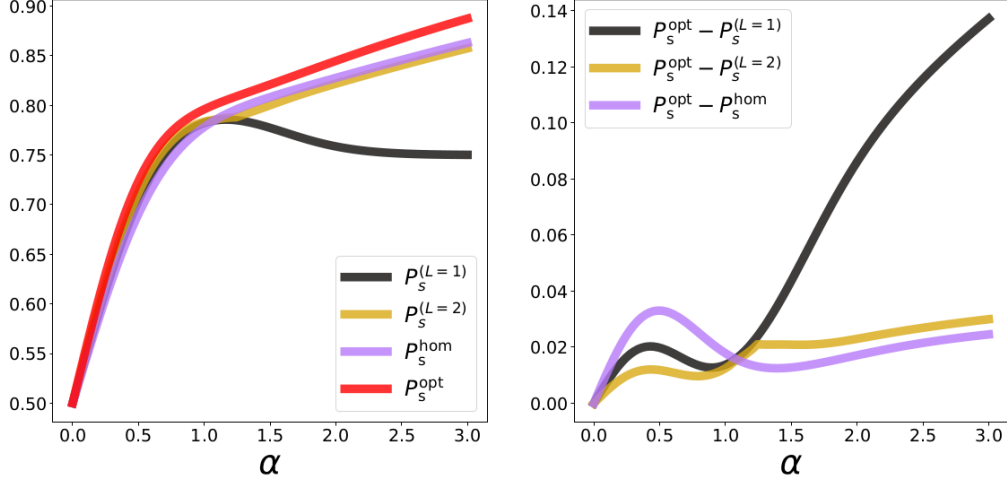


Figure 3.19: Top: We compare the success probabilities attainable by one ( $L = 1$ ) and two ( $L = 2$ ) adaptive receivers with homodyne measurement and Helstrom bound, for different values of signal intensity. Bottom: difference between the success probabilities of the receivers under considerations and the Helstrom bound are shown in order to ease the visualization of different features. In this case, a lossy channel ( $\eta = 10^{-2}$ ) acts on the transmitted state with probability of one half, whereas the signal is transmitted through a noise-less channel otherwise.

the different receivers we have studied. We note that in the case of variable-loss channels, there is no guarantee that Dolinar-like receivers will be able to attain the Helstrom bound, and whether that holds or not is an open question.

In particular, we consider the parameter regime where the two transmissivities are relatively distant from each other, which approximates the physical cases where the channel transmissivity changes abruptly due to weather conditions [VSV+17]. In Fig. 3.19 we compare the behaviour of optimized Kennedy, 2-Dolinar and homodyne receivers across different values of input signal amplitude, together with the Helstrom bound. Displacements in the adaptive receivers have been numerically optimized using simulated annealing [XSF+97; VGO+20]. This optimization proves demanding for higher number of adaptive measurements, and thus the dynamic programming method presented in Sec. 3.2 fails, unless a sufficiently high amount of computational resources is allocated to perform the continuous optimization (for instance, several initializations per optimization seed, and several seeds), though we have not investigated this optimization further.

We observe that, while the homodyne receiver maintains a satisfying perfor-

mance in the whole amplitude range, the behaviour of the Kennedy receiver exhibits a clear transition and its performance degrades at sufficiently large amplitudes. Fortunately, the addition of a second adaptive detection layer appears to be able to correct this trend. Hence the two-layer receiver can get close to the performance of the homodyne receiver in the large-amplitude regime, while showing a clear advantage with respect to the latter in the low-amplitude regime.

The degrade of 1-Dolinar receivers in the high-amplitude regime can be understood as follows. The action of the compound channel, for high-amplitude signals, is that of increasing the possible candidates in the discrimination problem to four; each channel can act with probability  $\frac{1}{2}$ , and while the success probability in Eq. 3.56 weights only the final binary guess, the action of the compound channel is that of averaging out two binary discrimination problems: the original one (with states  $|\pm\alpha\rangle$ ), and the new one (with states  $|\pm\eta\alpha\rangle$ ). In this sense, if initial amplitudes are small, then the attenuated amplitude will be *very* small, and the resulting state will effectively be the vacuum state.

One can understand this results by looking at the phase space in Fig 3.18. The Kennedy receiver effectively projects over  $|\beta\rangle\langle\beta|$ . Therefore it checks whether the hypothesis occupies a small region around  $\beta$  or not. In the case of composite hypothesis the coherent state projection can only cover one of the states in the ensemble. This effect is reduced when the amplitude is very low, and the *blobs* for both coherent states can be partially covered by the coherent state projection. In the limit of very large amplitudes, where the signals are almost orthogonal, a Kennedy receiver will need to specialize on one of the members of the ensemble, which will lead to a success probability of  $\frac{3}{4}$ , instead of 1.

On the contrary, a  $L$ -Dolinar receiver performs a projection on a coherent state  $|\beta_\ell\rangle$  (v. the rest), at each  $\ell$ . Thus, such receiver is expected to attain a probability 1 for large amplitudes, if the number of hypothesis —composite or not— can be covered by  $L$  coherent states.

Finally, in Fig. 3.20 we show that our Q-learning agent is also capable of calibrating a 2-Dolinar receiver under the composite noisy channel, without prior information about the original amplitudes nor the attenuations.

To sum up, our numerical simulations indicate that model-free Q-learning agents can readily adapt to many situations that are present in experimental scenarios. As a matter of fact, we have considered dark counts, phase flips and lossy channels with varying transmissivity.

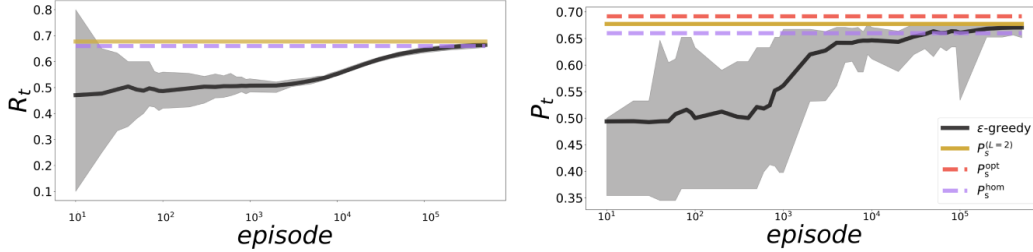


Figure 3.20: We show learning curve behaviour for  $\epsilon$ -greedy Q-learning, with an  $\epsilon$  exponentially decaying as a function of episode number. The figures of merit are averaged over 48 agents and corresponding uncertainty region are shown.

### 3.4 Discussion & Future work

While the reinforcement learning of quantum systems is a relatively young research direction, it has arguably positioned as a state of the art technique in many branches of science. In turn, approaches based in deep reinforcement learning have recently lead to breakthroughs in protein-folding problem [JEP+21], quantum sensing [SFB20], quantum communications [WMD+20], quantum computing [CAB+21b; FNM+21] and more [CT16; TM17; VLH17; CM17; TMC+18; MPK+18; FTW+18; WMD+19; BDS+18; NBS+19].

Whether such a human-computer collaboration will remain as a state-of-the-art paradigm for the discovery of novel protocols in science and technology, or is a stepping stone towards a deeper understanding of knowledge is an open and existing question.

This chapter humbly stands as one of the first steps towards the reinforcement learning of quantum systems, in an experimentally-centered paradigm (this project was initiated back in 2018). While many issues shall be address before proving successful for real-life quantum communications, the approach we have presented can potentially be implemented as a proof of principle. While conditional displacements can be implemented relatively fast using a FPGA, it is a question whether this experimental arrangement is the most suitable for the discrimination of coherent-state signals. In this regard, the original proposal of the experiment, *i.e.* by resolving the states in the time-domain, seems slightly more experimentally friendly, though theoretically equivalent to the spatial-domain studied in this Chapter, as previously discussed. Here, instead of regarding the coherent-state as a train of pulses, whose time-envelope gives rise to a coherent state of intensity  $\alpha$ , we have performed such splitting in space, by using beam-splitters. On the original approach, the feedback



has a time-dependence and acts on the pulses arriving to the receiver. However, we remark that the application of reinforcement-learning algorithms in time-domain is still in the first steps [Doy00; YHL21].

Regarding the machine learning framework we have posed, there is also room for improvement. For example, the actions were considered to be discrete, whereas the displacement actually has a continuous nature. In this regard, this project has hosted a Degree Thesis [JEJ20], in which we have studied the usage of deep neural networks aiming to generalize the experience gathered at episode  $t$  to state-action value functions that has not been visited until that episode. Much of these ideas were inspired by the Deep Deterministic Policy Gradient algorithm [LHP+15], where policy and state-action value functions are now predicted by deep neural networks, and special care is taken about training stability.

While during José's thesis we obtained successful results for the calibration of Kennedy-like receivers, training the networks under the conditioning operations ( $L = 2$ ) proved hard, due to the fact that a huge training set was required in order to successfully train it. A possible circumvent to that situation would be the usage of pre-trained generative models, to have a clever ansatz for the state-action value function profile. Thus, many doors remain open to deeper understand the reachability of reinforcement-learning algorithms to model-free discriminating between quantum states in an agnostic way.

### 3.4.1 Code

The code and simulations supporting this Chapter can be found in the repos: [Bil02a; BC20] (the latter contains a collection of unsuccessful approaches to the deep learning of dolinar-like receivers). We remark that the dynamic programming approach has been carried out mainly by Raúl Morral, an undergraduate student by the time this project was carried out, and the code can be found in Ref. [Bil02b].

# 4

---

## Learning in the twilight

---

The learning-in-the-quantum journey continues: after interacting in a model-free way with quantum systems of light, we will here pose the learning task differently. The agnostic drama will be diminished: we will now profit from our knowledge of quantum theory to aid the discover of novel protocols in NISQ quantum computing. Instead of treating quantum devices as black-boxes, we will inject some knowledge by introducing a semi-agnostic algorithm that discovers useful NISQ circuits. While the agent will have perfect control of the task at hand, the nature of the problem she is faced towards is so hard, that the solutions are unknown to us. Thus, the agent learns in the twilight: while knowledge of NISQ quantum-computing is available, the nature of the solutions she is seeking for is completely unknown.

Here, we will focus on how currently available quantum computers can be used a wide range of problems, ranging from ground-state preparation and training quantum autoencoders to compiling unitary transformations to native hardware. As discussed in Sec. 2.2, it seems natural to adjust the learning algorithm to the task at hand, doing it is often hard, because the learning scenarios are far from ideal.

While in the introductory Section we have stressed these difficulties for the NISQ scenario, let us remark that big efforts have been carried out by the machine-learning community in the recent years in order to injecting prior knowledge into deep-learning architectures [MLE21; SFE+20]. Overall, the design of the machine-learning-algorithm, when taking into account the structure that a potential solution

---

shall obey, is subtle and (as expected) problem-dependent. An example of this is given by Convolutional Neural Networks [LHB04; KSH12; ON15], designed in such a way to sequentially process an image in a task-oriented manner. Here, the network is able to identify *features* that are invariant under translation-like transformations, such as eyes or mouths in human pictures, and thus exploits the correlations hidden in the image.

However, by such prior-knowledge injection we unavoidable induce some bias to model, and we can readily elucidate a tradeoff. In this chapter, rather than specifically-tailoring an algorithm to come up with a task-oriented solution, we will study the performance of a general-purposed algorithm, which adapts its structure to the specific cost-function at hand, by means of machine-learning inspired techniques.

We will focus on NISQ computing, which were first introduced in Sec. 2.2. Here, an intermediate-scale (of about  $\sim 100$  qubits) noisy quantum device needs to be configured as a special-purpose machine. A promising route to do so is given by Variational Quantum Algorithms (VQAs), as discussed in Sec. 2.2.1. In VQAs, a parametrized quantum circuit (PQC) is used to estimate cost-function values, whose minima encodes solutions to the problem at hand. In order to reach such minima, a classical algorithm is used to optimize available degrees of freedom such as qubit-rotation values, or even circuit's structure. The latter plays a crucial role in NISQ computing: while the presence of noise essentially forbids the implementation of large-depth circuits, special care must be taken even when dealing with shallow circuits due to the appearance of Barren Plateaus (BPs). As discussed in Sec. 2.2.4, under the presence of a BP, estimating cost-minimizing directions is very hard, since the gradients exponentially concentrate around zero, which in turn implies that any optimizer will effectively get stucked. Addressing the BP issue currently constitutes one of the major research directions in the field of NISQ computing. Here, the ansatz used for the quantum circuit (*e.g.* the structure under which quantum gates act on the available qubits) plays a fundamental role, and developing tools to discover useful quantum-circuit structures that can be trained under the VQA framework is of utmost importance.

To this end, we combine several features of recently proposed methods and introduce the Variable Ansatz (VAns) algorithm [BCV+21] to generate variable structure ansatzes for generic VQA applications. As shown in Fig. 4.1, VAns iteratively grows the parameterized quantum circuit by adding blocks of gates initialized to the identity, but also prevents the circuit from over-growing by removing gates and compressing the circuit at each iteration. In this sense, VAns produces shallow circuits that are more resilient to noise, and that have less trainable parameters

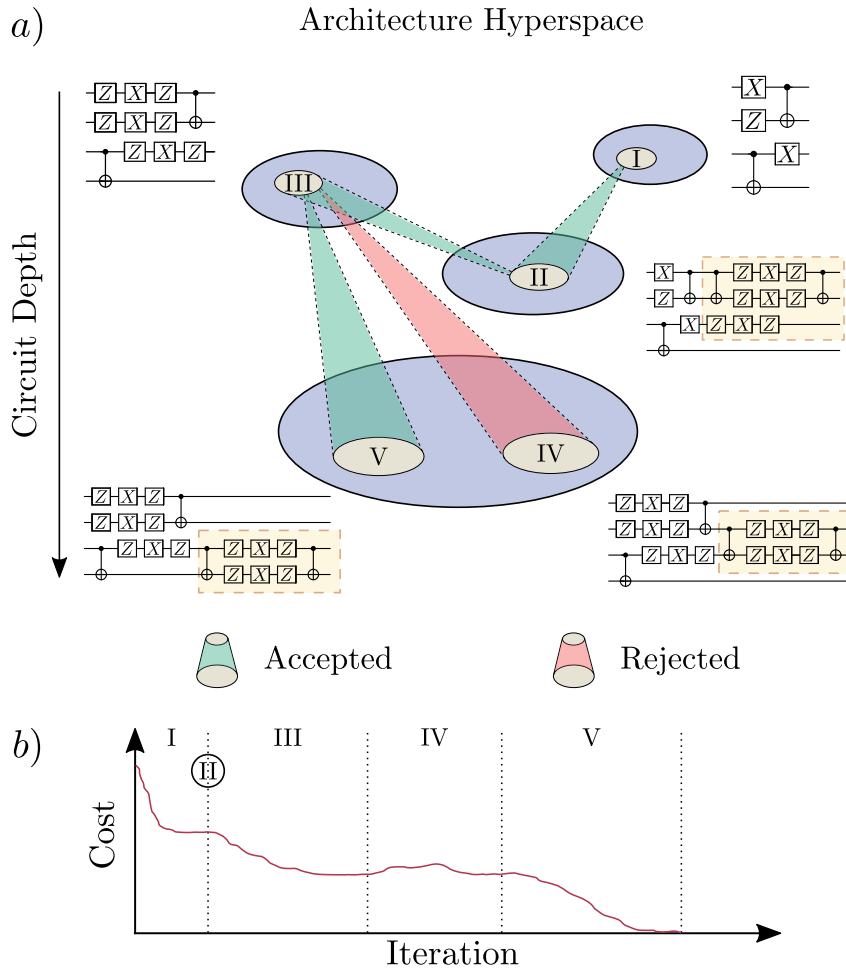


Figure 4.1: We show a schematic diagram of the VAns algorithm. (a) VAns explores the hyperspace of architectures of parametrized quantum circuits to create short depth ansatzes for VQA applications. VAns takes a (potentially non-trivial) initial circuit (step I) and optimizes its continuous parameters  $\theta$  until convergence. At each step, VAns inserts blocks of gates into the circuit which are initialized to the identity (indicated in a box in the figure), so that the ansatzes at contiguous steps belong to an equivalence class of circuits leading to the same cost value (step II). VAns then employs a classical algorithm to simplify the circuit by eliminating gates and finding the shortest circuit (step II to III). The ovals represent subspaces of the architecture hyperspace connected through VAns. While some regions may be smoothly connected by placing identity resolutions, VAns can also explore regions that are not smoothly connected via a gate-simplification process. VAns can either reject (step IV) or accept (step V) modifications in the circuit structure. Here  $Z$  ( $X$ ) indicates a rotation about the  $z$  ( $x$ ) axis. (b) Schematic representation of the cost function value versus the number of iterations for a typical VAns implementation which follows the steps in (a).

to avoid trainability issues. Our approach provides a simple yet effective way to address the ansatz design problem, without resorting to resource-expensive computations similar to recently evolutionary-oriented proposals.

Our algorithm is inspired by the strong presence of noise in NISQ devices. As such, it potentially hinders any successful application of NISQ computers. In this context, we should take advantage of every component of the quantum device that could be optimized over, and this generally includes the circuit’s layout. In this regard, some approaches –as EVQE [RHP+19] or MoG-VQE [CSU+20] –have been introduced in the past, as discussed in Sec. 2.2.2. VAns algorithm differs from them in at least two important aspects: (i) it considers general cost functions, and thus can be considered a general-purposed quantum-machine learning algorithm, and (ii) the method incorporates knowledge on quantum computing via specific circuit compression rules.

Thus, the learning problem we are dealing with consists in finding cost-minimizing quantum circuits, where not only the parameters can be optimized, but also the structure of the circuit itself. While some light is shed into the learning problem through the aforementioned compression rules, an important *learning-in-the-darkness* component remains. In particular, no structure is imposed on how our algorithm grows the quantum circuit, and at each time-step quantum gates are randomly placed across the circuit, acting on qubits that are also randomly selected. Thus, the learning paradigm departs from the one considered in Chapter 3, and rather than assuming complete ignorance of the setting, the agent (VAns algorithm) is now equipped with partial information about the setting, henceforth it learns in the twilight through the VAns algorithm, which we now turn to explain in detail.

## 4.1 The Variable Ansatz (VAns) Algorithm

The goal of the VAns algorithm is to adaptively construct trainable ansatzes for generic problems encoded into a cost function, using parametrized quantum circuits. We will here consider general cost-functions of the form<sup>1</sup>

$$C(\mathbf{k}, \boldsymbol{\theta}) = \sum_i^M f_i (\text{Tr}[O_i U(\mathbf{k}, \boldsymbol{\theta}) \rho_i U^\dagger(\mathbf{k}, \boldsymbol{\theta})]), \quad (4.1)$$

where  $\{\rho_i\}_{i=1}^M$  are  $n$ -qubit states forming a training-set, and  $U(\mathbf{k}, \boldsymbol{\theta})$  is our NISQ circuit parametrized by continuous parameters  $\boldsymbol{\theta}$  (e.g., rotation angles) and by

---

<sup>1</sup>More details can be found in Sec. 2.2.1.

discrete parameters  $\mathbf{k}$  (e.g., gate placements); moreover  $O_i$  are observables and  $f_i$  are functions that encode the optimization task at hand.

We define  $\mathcal{C}_l$  as the architecture hyperspace of quantum circuits having  $l$  gates in the layout. VAns takes as input:

- A cost function  $C(\mathbf{k}, \boldsymbol{\theta})$  to minimize.
- A dictionary  $\mathcal{D}$  of parametrized gates that compile to the identity. That is, for  $V(\vec{\gamma}) \in \mathcal{D}$  there exists a set of parameters  $\vec{\gamma}^*$  such that  $V(\vec{\gamma}^*) = \mathbb{1}$ .
- An initial circuit configuration  $U^{(0)}(\mathbf{k}, \boldsymbol{\theta}) \in \mathcal{C}_{l_0}$  of depth  $l_0$ .
- **Circuit Insertion** rules which stochastically take an element  $V(\vec{\gamma}^*) \in \mathcal{D}$  and randomly insert it in the circuit. The insertion is a map  $\mathcal{I} : \mathcal{C}_l \rightarrow \mathcal{C}_{l'}$  with  $l' \geq l$ .
- **Circuit Simplification** rules to eliminate unnecessary gates, redundant gates, or gates that do not have a large effect on the cost. The simplification is a map  $\mathcal{S} : \mathcal{C}_l \rightarrow \mathcal{C}_{l'}$  with  $l' \leq l$ .
- An optimization algorithm for the continuous parameters  $\boldsymbol{\theta}$ , and an optimization algorithm for the discrete parameters  $\mathbf{k}$ .

Given these inputs, VAns outputs a circuit architecture and set of parameters that approximately minimize the cost function in Eq. 4.1.

In what follows we describe the overall structure of VAns (whose pseudocode is provided later in this Section), and in the next sections we provide additional details for the **Insertion** and **Simplification** modules. We remark that the steps presented here are aimed at giving a general overview of the method and are intended to be used as building blocks for more advanced versions of VAns (which we discuss in Sec. 4.3).

The first ingredient of VAns (besides the cost function, which is defined by the problem at hand) is a dictionary  $\mathcal{D}$  of parametrized gates that can compile to identity and which VAns employs to build the ansatz. A key aspect here is that  $\mathcal{D}$  can be composed of any set of gates, so that one can build a dictionary specifically tailored for a given application. In addition, it is usually convenient to have the unitaries in  $\mathcal{D}$  expressed in terms of gates native to the specific quantum hardware employed, as this avoids compilation depth overheads (e.g. a larger number of gates would otherwise be required). In our numerics, we will consider a dictionary composed of one and two qubit identity block of gates that can resolve to the identity; the alphabet is given by single-qubit rotations and CNOT acting on any

two qubits present in the circuit (see Sec. 4.1.1 for further details). As stressed in Sec. 2.2.1, we will not consider issues arising from connectivity constraints.

Once the gate dictionary is set, the ansatz is initialized to a given configuration  $U^{(0)}(\mathbf{k}, \boldsymbol{\theta})$ . As discussed in Sec. 2.2.3, we then need to employ an optimizer to control the parameters  $\boldsymbol{\theta}$  of the initial ansatz until the convergence is reached. We recall that in Sec. 2.2.2 we discussed several strategies that define the ansatz structure. In particular, the two non-trivial ansatzes shown in Fig. 2.2 are employed in our numerical simulations as initialization strategies: (i) the *separable* product ansatz which generates no entanglement, and (ii) the alternating *Hardware Efficient Ansatz* which entangles neighboring qubits; note that only shallow circuits (e.g. a low number of HEA-layers) are considered for initialization purposes. While the choice of an appropriate initial ansatz can lead to faster convergence, VAns can in principle transform a simple initial ansatz into a more complex one as part of its architecture search.

From this point, VAns enters a nested optimization loop. In the outer loop, VAns explores the architecture hyperspace to optimize the ansatz's discrete parameters  $\mathbf{k}$  that characterize the circuit structure. Then, in the inner loop, the ansatz structure is fixed and the continuous parameters  $\boldsymbol{\theta}$  are optimized.

At the start of the outer loop, VAns employs its `Insertion` rules to stochastically grow the circuit. The fact that these rules are stochastic guarantees that different runs of VAns explore distinct regions of the architecture hyperspace. As previously mentioned, the gates added to the circuit compile to the identity so that circuits that differ by gate insertions belong to an equivalence class of circuits leading to the same cost function value (we remark that this strategy has been also introduced in [RHP+19]). As discussed below, the `Insertion` rules can be such that they depend on the current circuit they act upon. For instance, VAns can potentially add entangling gates to qubits that were not previously connected.

To prevent the circuit from constantly growing each time gates are inserted, VAns follows the `Insertion` step by a `Simplification` step. Here, the goal is to determine if the circuit can be compressed without significantly modifying the cost function value in a systematic way, as proposed in [MDM+08]. This is a fundamental step of VAns as it allows the algorithm to explore and jump between different regions of the architecture hyperspace which might not be trivially connected. Moreover, unlike other variable-ansatz strategies that continuously grow the number of gates present in the circuit, or which randomly remove gates, the `Simplification` step allows VAns to find shorter ansatzes by deleting gates in an informed manner.

Taken together, `Insertion` and `Simplification` provide a set of dis-

crete parameters  $\mathbf{k}$ . However, to determine if this new circuit structure can improve the cost function value it is necessary to enter the inner optimization loop and train the continuous parameters  $\theta$ . Here, we remark that a global optimization of all parameters shall be performed, since the landscape might contain a newer minima. In practice, when entering the continuous optimization, the parameters are initialized such that the circuit does not exactly compile to the same unitary than that of the previous step, but rather small perturbations are injected in order to help exiting potential local minima. When convergence in the optimization is reached, the final cost-function value is compared to the cost in the previous iteration. Updates that lead to equi-cost values or to smaller costs are accepted, while updates leading to higher cost functions are accepted with exponentially decaying probability in a manner similar to a Metropolis-Hastings step [Has70]. Here one accepts an update which increases the cost value with a probability given by  $\exp(-\beta \frac{\Delta C}{C_0})$ , with  $\frac{\Delta C}{C_0}$  being increment in the cost function with respect to the initial value, and  $\beta > 0$  a *temperature* factor. The previous optimizations in inner and outer loops are repeated until a termination condition  $f_{\text{Term}}$  is reached, e.g. distance to a target cost function value (if a lower bound is available), maximum VAns iteration number, or an user-specified function that might depend on variables such as circuit structure and cost value reached.

In the following we will give more details about VAns specific modules.

### 4.1.1 Insertion method

The `Insertion` step stochastically grows the circuit by inserting a parametrized block of gates from the dictionary  $\mathcal{D}$  which compiles to the identity. In order to facilitate the exploration of a larger architecture-hyperspace region, in practice we allow some deviation from the identity by slightly modifying the continuous parameters so that the new gate deviates slightly from the identity. In Fig. 4.2 we show examples of two parametrized quantum circuits that can compile to the identity.

There are many choices for how VAns determines which gates are chosen from  $\mathcal{D}$  at each iteration, and where they should be placed. When selecting gates, we have here taken a uniform sampling approach, where every sequence of gates in  $\mathcal{D}$  has an equal probability to be selected.

### 4.1.2 Simplification method

The `Simplification` steps in VAns are aimed at eliminating unnecessary gates, redundant gates, or gates that do not have a large effect on the cost. For this



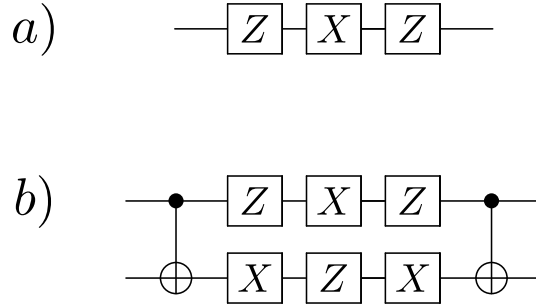


Figure 4.2: We show the two words from the dictionary  $\mathcal{D}$  used during the *Insertion* steps. Here we show two types of the parametrized gate sequences composed of CNOTs and rotations about the  $z$  and  $x$  axis. Specifically, one obtains the identity if the rotation angles are set to zero. Using the circuit in (a), one inserts a general unitary acting on a given qubit, while the circuit in (b) correlates the two qubits it acts upon.

purpose, *Simplification* moves gates in the circuit using the commutation rules shown in Fig. 4.3(a) to group single qubits rotations and CNOTs together. Once there are no further commutations possible, the circuit is scanned and a sequence of simplification rules are successively applied. For instance, assuming that the input state is initialized to  $|0\rangle^{\otimes n}$ , we can define the following set of simplification rules.

1. Scan the circuit for possible commutations: if possible place CNOTs at the left, and rotations at the right side.
2. CNOT gates acting at the beginning of the circuit are removed.
3. Rotations around the  $z$ -axis acting at the beginning of the circuit are removed.
4. Consecutive CNOT sharing the same control and target qubits are removed.
5. Two or more consecutive rotations around the same axis and acting on the same qubit are compiled into a single rotation (whose value is the sum of the previous values).
6. If three or more single-qubit rotations are sequentially acting on the same qubit, they are simplified into a general single-qubit rotation of the form  $R_z(\theta_1)R_x(\theta_2)R_z(\theta_3)$  or  $R_x(\theta_1)R_z(\theta_2)R_x(\theta_3)$  which has the same action as the previous product of rotations.

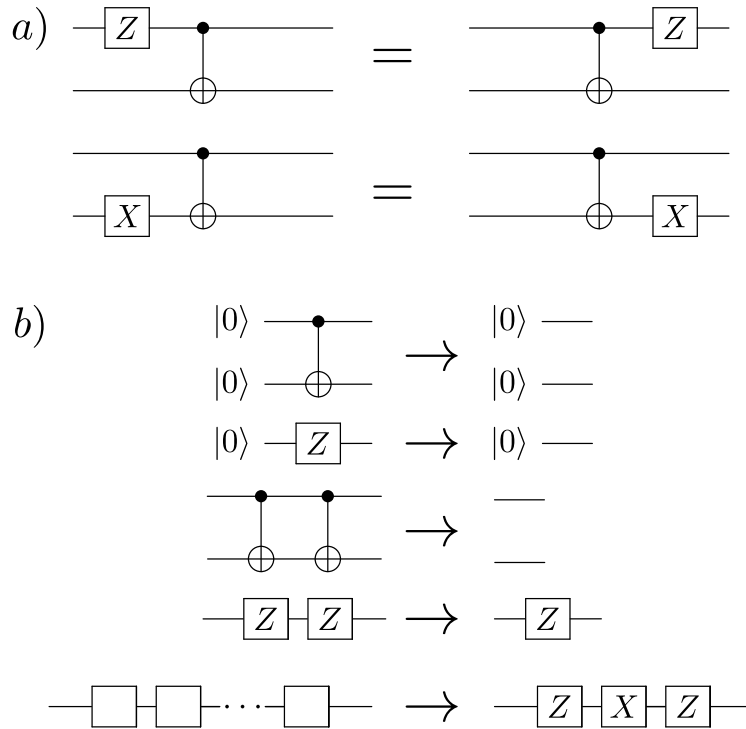


Figure 4.3: We depict the rules for the **Simplification** steps. (a) Commutation rules used by VAns to move gates in the circuit. As shown, we can commute a CNOT with a rotation  $Z$  ( $X$ ) about the  $z$  ( $x$ ) axis acting on the control (target) qubit; when possible, CNOTs are moved to the left-side of the circuit, and rotations to the right one. (b) Example of simplification rules used by VAns to reduce the circuit depth. Here we assume that the circuit is initialized to  $|0\rangle^{\otimes n}$ .

7. Gates whose presence in the circuit does not considerably reduce the cost are removed.

Rules (2) – (6) are schematically shown in Fig. 4.3(b). We remark that a crucial feature of these **Simplification** rules is that they can be performed using a classical computer that analyzes the circuit structure and hence do not lead to additional quantum-computation resources.

As indicated by step (6), the **Simplification** steps can also delete gates whose presence in the circuit does not considerably reduce the cost. Here, given a parametrized gate, one can remove it from the circuit and compute the ensuing cost function value. If the resulting cost is not increased by more than some threshold value, the gate under consideration is removed and the simplification rules (1) – (5)

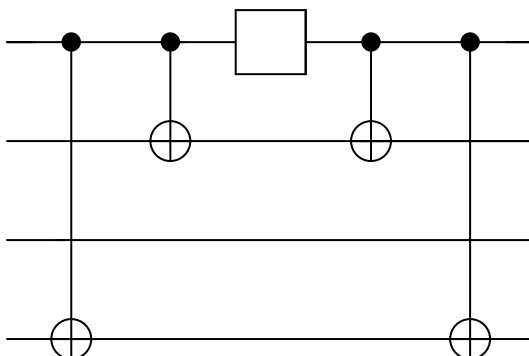


Figure 4.4: We show a non-trivial circuit structure that can be obtained by VAns using the Insertion and Simplification steps and the gate dictionary in Fig. 4.2.

are again implemented. Here, we can use information from the inner optimization loop to find candidate gates for removal. For instance, when employing a gradient descent optimizer, we may attempt to remove gates whose parameters lead to small gradients. Note that, unlike the simplification steps (1) – (5) in Fig. 4.3(b), when using the deletion process in (6) one needs to call a quantum computer to estimate the cost function, and hence these come at an additional quantum-resource overhead which scales linearly with the number of gates one is attempting to remove.

An interesting aspect of the Simplification method is that it allows VAns to obtain circuit structures that are not contained in the initial circuit  $U^{(0)}(\mathbf{k}, \boldsymbol{\theta})$  or in the gate dictionary  $\mathcal{D}$ , and hence to explore new regions of the architecture hyperspace. For instance, using the gate dictionary in Fig. 4.2, VAns can obtain a gate structure as the one shown in Fig. 4.4.

---

**Algorithm 5:** Pseudo-code for VAns

---

**Input:** Cost function  $C(\mathbf{k}, \boldsymbol{\theta})$ ; initial circuit  $U^{(0)}(\mathbf{k}, \boldsymbol{\theta})$ ; dictionary of gates  $\mathcal{D}$ ; **Insertion** rules which take gates from  $\mathcal{D}$  and appends them to a circuit; **Simplification** rules; optimization algorithm  $\text{Opt}_C$  for continuous parameters  $\boldsymbol{\theta}$ ; optimization algorithm  $\text{Opt}_D$  for discrete parameters which accepts or rejects an ansatz update given changes in the cost function value; termination condition function  $f_{\text{Term}}(n, C(\mathbf{k}, \boldsymbol{\theta}), U(\mathbf{k}, \boldsymbol{\theta}))$ ,  $n \in \mathbb{N}$

**Output:** Optimized ansatz  $U^{(f)}$ .

**Init:** Randomly initialize the parameters  $\boldsymbol{\theta}$ ; initialize the ansatz  $U^{(f)} \leftarrow U^{(0)}(\mathbf{k}, \boldsymbol{\theta})$ ;  $C^{(f)} \leftarrow 0$ ;  $\mathbf{k}^{(f)} \leftarrow \mathbf{k}$ ;  $\boldsymbol{\theta}^{(f)} \leftarrow \boldsymbol{\theta}$ ;  $\mathcal{U}(\mathbf{k}, \boldsymbol{\theta}) \leftarrow \mathbb{1}$ ;  $\text{Term} \leftarrow \text{false}$ ;  $n \leftarrow 0$ .

- 1 Optimize  $\boldsymbol{\theta}$  with  $\text{Opt}_C$  and store result in  $\boldsymbol{\theta}^{(f)}$ ;  $C^{(f)} \leftarrow C(\mathbf{k}, \boldsymbol{\theta})$ .
- 2 **while**  $\text{Term}$  is false **do**
- 3      $n \leftarrow n + 1$
- 4      $\text{Accept} \leftarrow \text{false}$
- 5     **while**  $\text{Accept}$  is false **do**
- 6         Use **Insertion** in  $U^{(f)}$  and store new sets of discrete parameters, continuous parameters and ansatz in  $\boldsymbol{\theta}$ ,  $\mathbf{k}$ , and  $\mathcal{U}(\mathbf{k}, \boldsymbol{\theta})$ , respectively.
- 7         Use rules [1-6] of **Simplification** on  $\mathcal{U}(\mathbf{k}, \boldsymbol{\theta})$  and store new sets of discrete parameters, continuous parameters and ansatz in  $\boldsymbol{\theta}$ ,  $\mathbf{k}$ , and  $\mathcal{U}(\mathbf{k}, \boldsymbol{\theta})$ , respectively.
- 8         Optimize continuous parameters in  $\mathcal{U}(\mathbf{k}, \boldsymbol{\theta})$  with  $\text{Opt}_C$  and store result in  $\boldsymbol{\theta}$ ;  $C^{(f)} \leftarrow C(\mathbf{k}, \boldsymbol{\theta})$ .
- 9         Repeat step 7, with rules [1-7] of **Simplification**.
- 10         Given  $C(\mathbf{k}, \boldsymbol{\theta})$  and  $C^{(f)}$ , optimize discrete parameters in  $\mathcal{U}(\mathbf{k}, \boldsymbol{\theta})$  with  $\text{Opt}_D$  and store result in  $\text{Accept}$ .
- 11      $\mathbf{k}^{(f)} \leftarrow \mathbf{k}$ .
- 12      $\boldsymbol{\theta}^{(f)} \leftarrow \boldsymbol{\theta}$ .
- 13      $U^{(f)} \leftarrow \mathcal{U}(\mathbf{k}, \boldsymbol{\theta})$ .
- 14      $C^{(f)} \leftarrow C(\mathbf{k}, \boldsymbol{\theta})$ .
- 15      $\text{Term} \leftarrow f_{\text{Term}}(n, C^{(f)}, U^{(f)})$

---

### 4.1.3 Scaling of VAns

With the previous overview of the VAns method in mind, let us now discuss the computational complexity arising from using VAns versus that of using a standard fix architecture scheme. In the following discussion we will not include any computational cost or complexity of the continuous-parameter optimizer as we assume that the same tools could be used for fixed or variable ansatzes.

Firstly, we note that any additional computational cost comes due to circuit manipulations, meaning that we should study the scaling of the `Insertion` and `Simplification` methods. On the one hand, adding gates via `Insertion` is stochastic, and independent of the number of qubits or the current circuit depth, that is: its complexity is always in  $\mathcal{O}(1)$ . Then, removing gates via `Simplification` has a cost which increases with the number of gates in the circuit. If we have  $M$  gates, then running the `Simplification` scheme has a cost  $\mathcal{O}(M)$ . We note that such computational complexity is similar to that of using gradient-free versus gradient-based methods, as the computational cost of the latter also scale as  $\mathcal{O}(M)$ . Notably, since the goal of VAns is to produce short-depth circuits, the algorithm itself tries to reduce its own computational cost during training. As we will see below in our numerical examples, VAns is always able to find short-depth circuits whose solutions are better than those arising from fixed structure ansatzes, meaning that the extra complexity of VAns can be justified in terms of its performance.

With an understanding of VAns mechanism, we will now turn to use it in a plethora of different scenarios.

## 4.2 Using VAns

In this Section, we present our results obtained from simulating VAns to solve paradigmatic problems in condensed matter, quantum chemistry, quantum autoencoding and quantum compiling problems in ideal conditions, e.g. without considering any noise-model of the quantum computer.

We first use VAns in the Variational Quantum Eigensolver (VQE) algorithm [PMS+14] to obtain the ground state of the Transverse Field Ising model (TFIM), the XXZ Heisenberg spin model, and the  $H_2$  and  $H_4$  molecules. We then apply VAns to a quantum autoencoder [ROAG17] task for data compression. We then move to use VAns to compile a Quantum Fourier Transform unitary in systems up to 10 qubits. In all these cases, we have performed the simulations under the unrealistic scenario in which neither shot-noise nor hardware noise were considered; this will be the matter of further sections.

The simulations presented here were performed using Tensorflow Quantum [BVM+20]. Adam [KB15] and qFactor [YC] were employed to optimize the continuous parameters  $\theta$ , e.g. `Opt_C` in the pseudo-code previously presented. While VAns is a stochastic algorithm, and the number of iterations required until meeting a convergence criteria should in principle be a random variable, we remark that all the results shown in this Section were obtained from *a single instance* of the algorithm (for each of the problems considered). While a statistical analysis should in principle be carried out in order to analyze VAns' average performance, the latter fact intuitively highlights the way in which VAns is able to exploit computational resources.

The dictionary  $\mathcal{D}$  of gates used consisted on block of single single-qubit and two-qubit gates that can resolve to identity. Each is composed by single-qubit rotations around  $x$  and  $z$  axis, and CNOTs gates between all qubits present in the circuit. As already mentioned, we have assumed no connectivity constraints under the quantum circuits under consideration. In the following examples, VAns was initialized to either a separable ansatz or an  $L$ -HEA, with  $L < 3$  (recall that in Fig. 2.2 we showed the separable and 2-HEA circuits).

### 4.2.1 Transverse Field Ising model (TFIM)

We now consider a cyclic TFIM chain. The Hamiltonian of the system reads

$$\hat{H} = -J \sum_{j=1}^n \sigma_j^x \sigma_{j+1}^x - g \sum_{j=1}^n \sigma_j^z, \quad (4.2)$$

where  $\sigma_j^{x(z)}$  is the Pauli  $x$  ( $z$ ) operator acting on qubit  $j$ , and where  $n + 1 \equiv 1$  to indicate periodic boundary conditions. Here,  $J$  indicates the interaction strength, while  $g$  is the magnitude of the transverse magnetic field. As mentioned in Section 2.2.1, when using the VQE algorithm the goal is to optimize a parametrized quantum circuit  $U(\mathbf{k}, \boldsymbol{\theta})$  to prepare the ground state of  $H$  so that the cost function becomes

$$C_{\text{VQE}}(\mathbf{k}, \boldsymbol{\theta}) = \text{Tr}[\hat{H}U(\mathbf{k}, \boldsymbol{\theta})\rho U^\dagger(\mathbf{k}, \boldsymbol{\theta})], \quad (4.3)$$

where one usually employs  $\rho = |0\rangle\langle 0|$  with  $|0\rangle = |0\rangle^{\otimes n}$ .

In Fig. 4.5 we show results obtained from employing VAns to find the ground state of a TFIM model of Eq. (4.2) with  $n = 4$  qubits (a) and with  $n = 8$  qubits (b), field  $g = 1$ , and different interactions values. To quantify the performance of the algorithm, we additionally show the relative error  $|\Delta E/E_0|$ , where  $E_0$  is the exact ground state energy  $E_0$ ,  $\Delta E = E_{\text{VAns}} - E_0$ , and  $E_{\text{VAns}}$  the best energy obtained

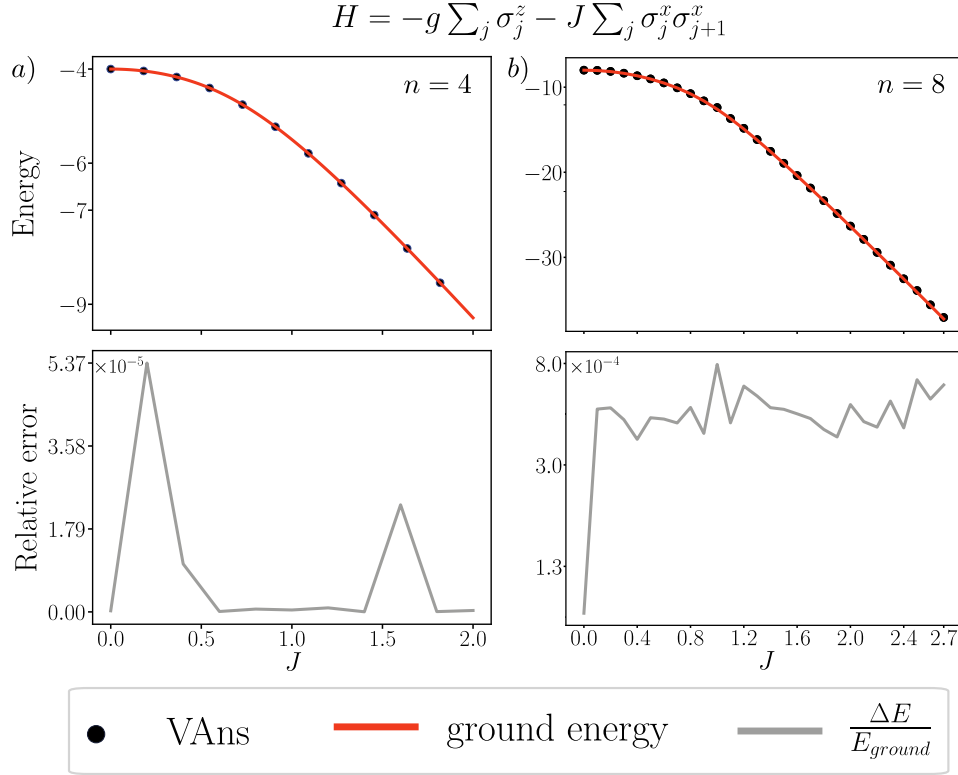


Figure 4.5: Results of using VAns to obtain the ground state of a Transverse Field Ising model: we use VAns in the VQE algorithm for the Hamiltonian in Eq. 4.2 with (a)  $n = 4$  qubits and (b)  $n = 8$  qubits, field  $g = 1$ , for different values of the interaction  $J$ . Top panels: solid lines indicate the exact ground state energy, and the markers are the energies obtained using VAns. Bottom panels: Relative error in the energy for the same interaction values.

through VAns. For 4 qubits, we see from Fig. 4.5 that the relative error is always smaller than  $6 \times 10^{-5}$ , showing that the ground-state energy was obtained for all coupling values  $J$ . Then, for  $n = 8$  qubits, VAns obtains the ground state of the TFIM with relative error smaller than  $8 \times 10^{-4}$ .

To gain some insight into the learning process, in Fig. 4.6 we show the cost function value, number of CNOTs, and the number of trainable parameters in the circuit discovered by VAns as different modifications of the ansatz are accepted to minimize the cost in an  $n = 8$  TFIM VQE implementation. Specifically, in Fig. 4.6(top) we see that as VAns explores the architecture hyperspace, the cost function value continually decreases until one can determine the ground state of the

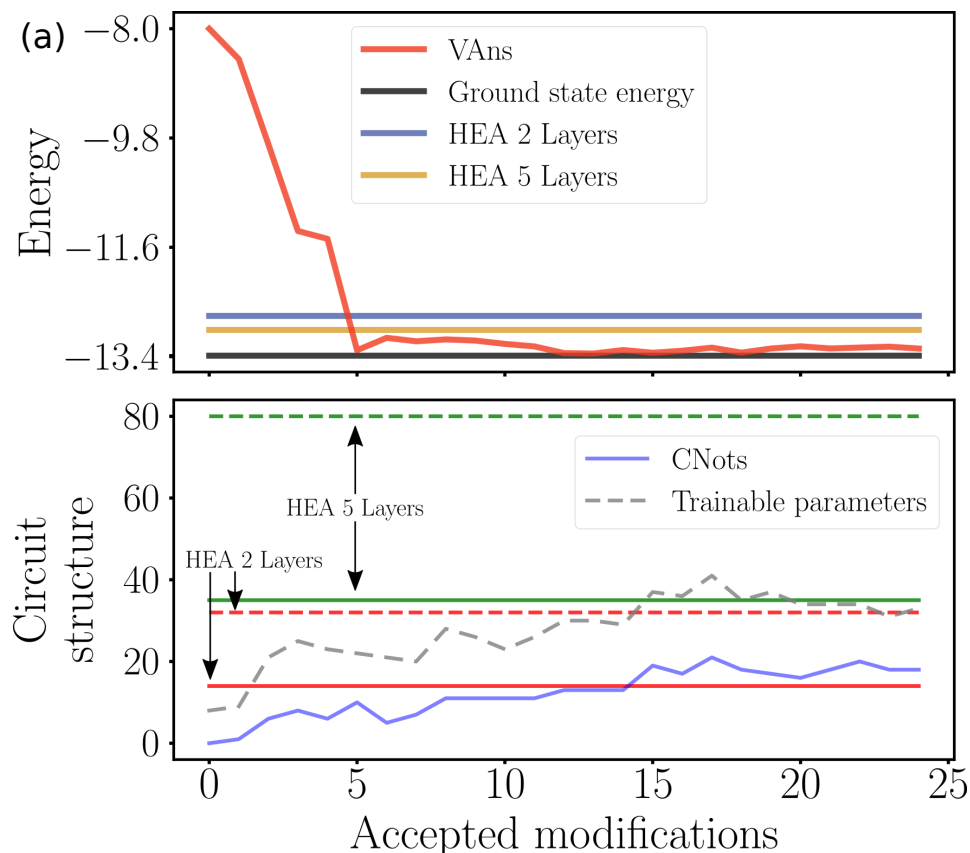


Figure 4.6: We show VAns learning process: (a) we show an instance of running the algorithm for the Hamiltonian in Eq. (4.2) with  $n = 8$  qubits, field  $g = 1$ , and interaction  $J = 1.5$ . The top panel shows the cost function value and the bottom panel depicts the number of CNOTs, and the number of trainable parameters versus the number of modifications of the ansatz accepted in the VAns algorithm. Top: As the number of iterations increases, VAns minimizes the energy until one finds the ground state of the TFIM. Here we also show the best results obtained by training a fixed structure layered Hardware Efficient Ansatz (HEA) with 2 and 5 layers, and in both cases, VAns outperforms the HEA. Bottom: While initially the number of CNOTs and number of trainable parameters increases, the Simplification method in VAns prevents the circuit from constantly growing, and can even lead to shorter depth circuits that achieve better solutions. Here we also show the number of CNOTs (solid line) and parameters (dashed line) in the HEA ansatzes considered, and we see that VAns can obtain circuits with less entangling and trainable gates.



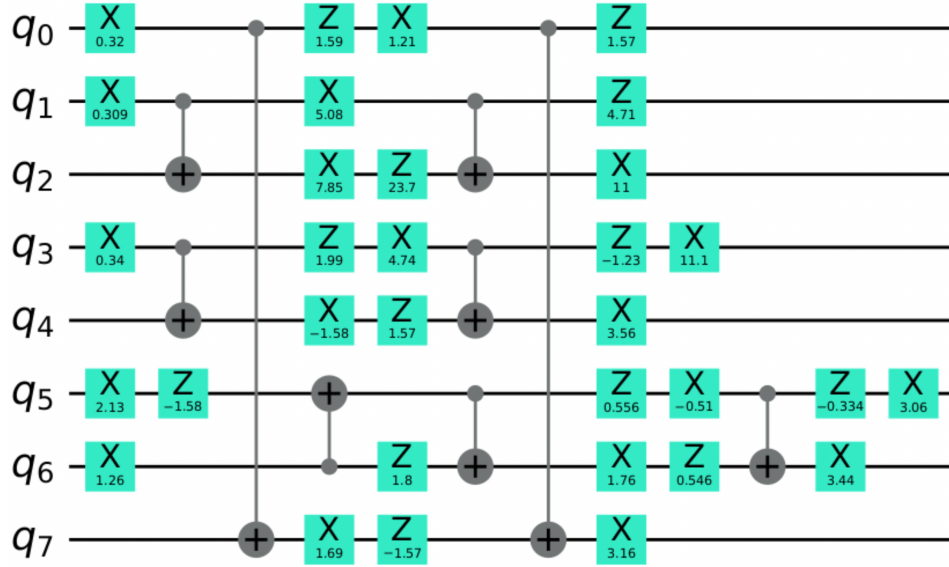


Figure 4.7: We show a low-depth, ground-state preparing circuits found by VAns during the learning process; here  $Z$  ( $X$ ) indicates a rotation about the  $z$  ( $x$ ) axis, about the corresponding value appearing below.

TFIM. Fig. 4.6(bottom) shows that initially VAns increases the number of trainable parameters and CNOTs in the circuit via the `Insertion` step. However, as the circuit size increases, the action of the `Simplification` module becomes more relevant as we see that the number of trainable parameters and CNOTs can decrease throughout the computation. Moreover, here we additionally see that reducing the number of CNOTs and trainable parameters can lead to improvements in the cost function value. The latter indicates that VAns can indeed lead to short depth ansatzes which can efficiently solve the task at hand, even without the presence of noise.

Finally, in Fig. 4.7 we also compare the performance of VAns with that of 2-HEA and 5-HEA. We specifically compare against those two fixed structure ansatzes as the first (latter) has a number of trainable parameters (CNOTs) comparable to those obtained in the VAns circuit. In all cases, we see that VAns can produce better results than those obtained with the Hardware Efficient Ansatz.

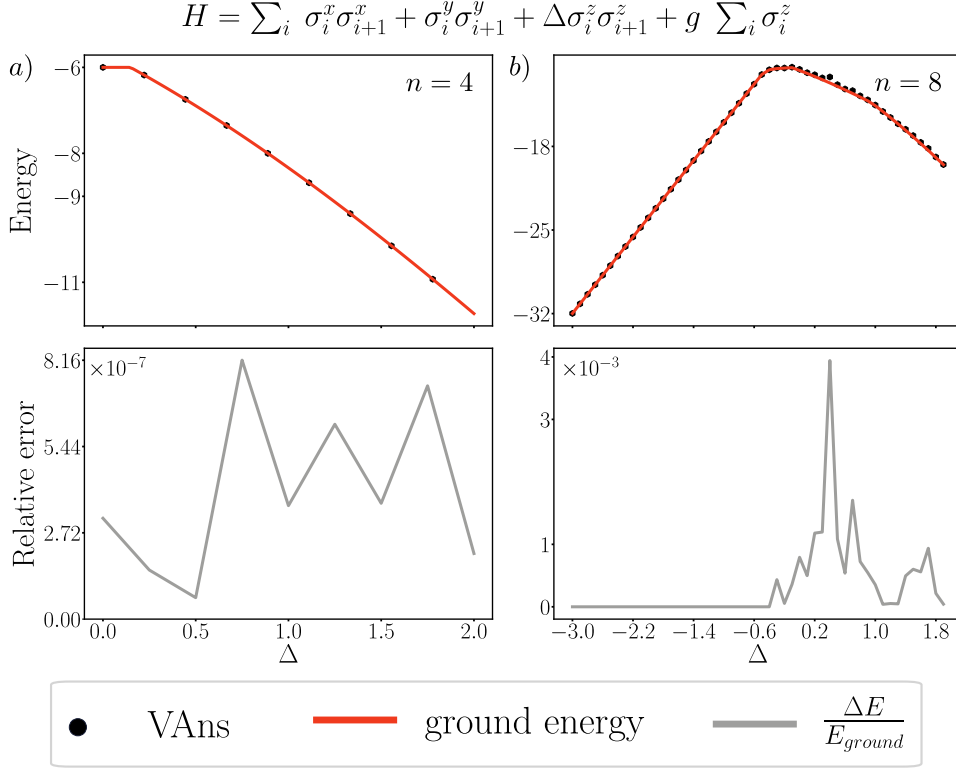


Figure 4.8: We show results of using VAns to obtain the ground state of a Heisenberg  $XXZ$  model. Here, we consider the VQE algorithm for the Hamiltonian in Eq. 4.4 with (a)  $n = 4$  and (b)  $n = 8$  qubits, field  $g = 1$ , and indicated anisotropies  $\Delta$ . Top panels: The solid line indicates the exact ground state energy, and the markers are the energies obtained using VAns. Bottom panels: Relative error in the energy for the same anisotropy values.

## 4.2.2 $XXZ$ Heisenberg Model

Here we use VAns in a VQE implementation to obtain the ground state of a periodic  $XXZ$  Heisenberg spin chain in a transverse field. The Hamiltonian of the system is

$$H = \sum_{j=1}^n \sigma_j^x \sigma_{j+1}^x + \sigma_j^y \sigma_{j+1}^y + \Delta \sigma_j^z \sigma_{j+1}^z + g \sum_{j=1}^n \sigma_j^z, \quad (4.4)$$

where again  $\sigma_j^\mu$  are the Pauli operators (with  $\mu = x, y, z$ ) acting on qubit  $j$ ,  $n+1 \equiv 1$  to indicate periodic boundary conditions, and where  $\Delta$  is the anisotropy. We recall that  $H$  commutes with the total spin component  $S_z = \sum_j \sigma_j^z$ , meaning that its

eigenvectors have definite magnetization  $M_Z$  along  $z$  [CRC+17].

In Fig. 4.8 we show numerical results for finding the ground state of (4.4) with  $n = 4$  and  $n = 8$  qubits, field  $g = 1$ , and for different anisotropy values. For 4 qubits, we see that VANS can obtain the ground-state energy with relative errors which are always smaller than  $9 \times 10^{-7}$ . In the  $n = 8$  qubits case, the relative error is of the order  $10^{-3}$ , with error increasing in the region  $0 < \Delta < 1$ ; in such case, as we discuss next, the algorithm finds the first-excited energy state.

We remark that a similar phenomenon is observed in [CLKAG20], where errors in preparing the ground state of the  $XXZ$  chain increase in the same region. The reason behind this phenomenon is that the optimizer can get stuck in a local minimum where it prepares excited states instead of the ground state. Moreover, it can be verified that while the ground state and the three first excited states all belong to the same magnetization sub-space of state with magnetization  $M_Z = 0$ , they have in fact different local symmetries and structure. Several of the low-lying excited states have a Néel-type structure of spins with non-zero local magnetization along  $z$  of the form  $|\uparrow\downarrow\uparrow\downarrow \cdots\rangle$ . On the other hand, the state that becomes the ground-state for  $\Delta > 1$  is a state where all spins have zero local magnetization along  $z$ , meaning that the local states are in the  $xy$  plane of the Bloch sphere. Since there is a larger number of excited states with a Néel-type structure (and with different translation symmetry) variational algorithms tend to prepare such states when minimizing the energy. Moreover, since mapping a state with non-zero local magnetization along  $z$  to a state with zero local magnetization requires a transformation acting on all qubits, any algorithm performing local updates will have a difficult time finding such mapping.

### 4.2.3 Molecular Hamiltonians

We will now focus on quantum chemistry problems, which consist in preparing the ground-state of a molecular hamiltonian. We will first outline how such systems can be modelled under the NISQ framework, and present our VQE-results afterwards. While our introduction to quantum chemistry suffices for the purposes of presenting the problem we benchmark VANS with, the curious reader can find more information on quantum chemistry in Refs. [MEAG+20a; MSK+19; MEAG+20b].

The molecular Hamiltonian, under the Born-Oppenheimer approximation, can

be written both in the phase-space and in the (fermionic) Fock space as per

$$H = \frac{1}{2} \sum_{i \neq j} \frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|} - \frac{1}{2} \sum_i \nabla_{r_i}^2 - \sum_{i,j} \frac{Z_j}{|\mathbf{R}_j - r_i|} + \sum_{i < j} \frac{1}{|r_i - r_j|}, \quad (4.5)$$

$$= h_{nuc} + \sum_{pq} h_{pq}(\mathbf{R}) a_p^\dagger a_q + \frac{1}{2} \sum_{pqrs} h_{pqrs}(\mathbf{R}) a_p^\dagger a_q^\dagger a_r a_s, \quad (4.6)$$

where we have used atomic units, and where  $Z_i$  represents the nuclear charge of  $i^{\text{th}}$  nuclei whose position is  $\mathbf{R}_i$ . Moreover,  $a$  and  $a^\dagger$  represent the fermionic annihilation and creation operators obeying the anticommutation relations  $\{a_i, a_j^\dagger\} = \delta_{ij}$ , and we have also included the *one and two-electron integrals*, which relate the phase-space representation of the Hamiltonian to the Fock-space one. To this end, we use single-electron wave-functions (known as *electron spin-orbitals*, or *atomic orbitals*)  $\psi(\sigma)$  – generally obtained via a Hartree-Fock approach – with  $\sigma = (\mathbf{r}_i, s_i)$  standing for the electron’s position and spin respectively, whose expression is given by

$$h_{pq}(\mathbf{R}) = - \int d\sigma \phi_p^*(\sigma) \left( \frac{\nabla_{\mathbf{r}}^2}{2} + \sum_i \frac{Z_i}{|\mathbf{R}_i - \mathbf{r}|} \right) \phi_q(\sigma) \quad (4.7)$$

$$h_{pqrs}(\mathbf{R}) = \int d\sigma_1 d\sigma_2 \frac{\phi_p^*(\sigma_1) \phi_q^*(\sigma_2) \phi_p(\sigma_1) \phi_q(\sigma_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}, \quad (4.8)$$

and remark that such quantities depend on the geometry of the molecule under consideration, as given by the position of all nuclei  $\mathbf{R} = \{\mathbf{R}_j\}$ . Computing such integrals is a non-trivial task, and we here rely on specific libraries to do so [MSK+19].

From here, a basis-set of states known as *molecular orbitals* is used to describe the molecule’s ground-state, and many methods have been developed in the past decades to classically solve this problem [MEAG+20a]. The fermionic state is generally given by a superposition of Slater determinants, generated by such basis-set.

If we aimed to obtain the exact ground state of the molecule, then we could variationally minimize its energy, and this constitutes the Full Configuration Interaction (FCI) approach; we note however that the number of determinants buildable from a given basis-set is exponentially large, and this approach is not practical – although we will use it for benchmarking purposes in our numerics –.

On the contrary, the Hartree-Fock (HF) method keeps a single determinant, where the molecular orbitals are obtained as lineal combinations of atomic orbitals; the combination coefficients are optimized by following a mean-field approach where each electron is assumed to suffer from an effective potential generated by the remaining ones, and the state  $|HF\rangle$  is thereby obtained.

However, since the HF approach treats the electrons independently, it does not capture correlation effects generally present in the ground-state. To this end, an iterative approach can be considered, which applies *excitation* operators to the  $|HF\rangle$ , in order to bring it closer to the true molecular ground-state. Each excitation swaps an electron from one molecular orbital to another, and if all excitation operators  $T_n$  were to be considered, then any state could be reachable within this approach, *i.e.* we get FCI. However, and in order to remain in the scope of classically-tractable solutions, only one and two excitation operators are often considered. To this end, the Unitary Coupled Cluster with Single and Double excitations (UCCSD) ansatz reads  $|\psi_{UCC}\rangle = e^{T-T^\dagger} |HF\rangle$  with  $T = \sum_{i=1}^2 T_i$ , and

$$T_1 = \sum_{i,\alpha} t_{i,\alpha} a_i^\dagger a_\alpha, \quad T_2 = \sum_{ij,\alpha\beta} t_{ij,\alpha\beta} a_i^\dagger a_j^\dagger a_\alpha a_\beta. \quad (4.9)$$

From here, we variationally minimize the energy associated to  $|\psi_{UCC}\rangle$ , by modifying the coefficients  $t_{i,\alpha}$  and  $t_{ij,\alpha\beta}$ .

As discussed in Sec. 2.2.2, to implement Eq. 4.5 in a digital quantum computer one needs to map the fermionic operators into qubits operators (usually through a Jordan Wigner or Bravyi-Kitaev transformation). Here we employed the OpenFermion package [MSK+19] to map Eq. 4.5 into a Hamiltonian expressed as a linear combination of  $n$ -qubit Pauli strings of the form

$$H = \sum_{\vec{z}} c_{\vec{z}} P_{\vec{z}}, \quad (4.10)$$

with  $P_{\vec{z}} \in \{\mathbb{1}, \sigma^x, \sigma^y, \sigma^z\}^{\otimes n}$ ,  $c_{\vec{z}}$  real coefficients, and  $\vec{z} \in \{0, x, y, z\}^{\otimes n}$ . Here, we stress that the coefficients  $c_{\vec{z}}$  generally depend on the electronic integrals, which in turn depend on the geometry of the molecule under consideration.

In all cases, the basis set used to approximate atomic orbitals was the *STO-3g* one, a neutral molecule was always considered, and the Jordan-Wigner transformation was used. While for the  $H_2$  the number of qubits required is four ( $n = 4$ ), this number is doubled for the  $H_4$  chain ( $n = 8$ ). Here, we note that the  $H_4$  chain that we consider might not be found in equilibrium conditions (*e.g.* the geometry that we consider may not be stable). This consists on an equally-separated array of Hydrogen atoms, by a distance that we deem *bond-length*, and in our numerics we vary such distance in order to reconstruct the energy curve.

Thus, in Fig. 4.9 we show the results obtained for finding the ground-state energy of the Hydrogen molecule (left) and  $H_4$  chain (right), at different bond lengths. As stressed above, equal bond distance were considered for the  $H_4$  chain. Noticeably, the dictionary of gates  $\mathcal{D}$  chosen here is not a *chemical-inspired* one (*e.g.*,

it does not contain single and double excitation operators nor its hardware-efficient implementations), yet VAns is able to find ground-state preparing circuits within *chemical accuracy*, a term which stands for the ultimate accuracy experimentally reachable in these systems [MEAG+20b]. Moreover, as shown in the bottom panels, VAns usually requires less than 15 iterations until reaching convergence, showing that the algorithm quickly finds a way through the architecture hyperspace towards a solution.

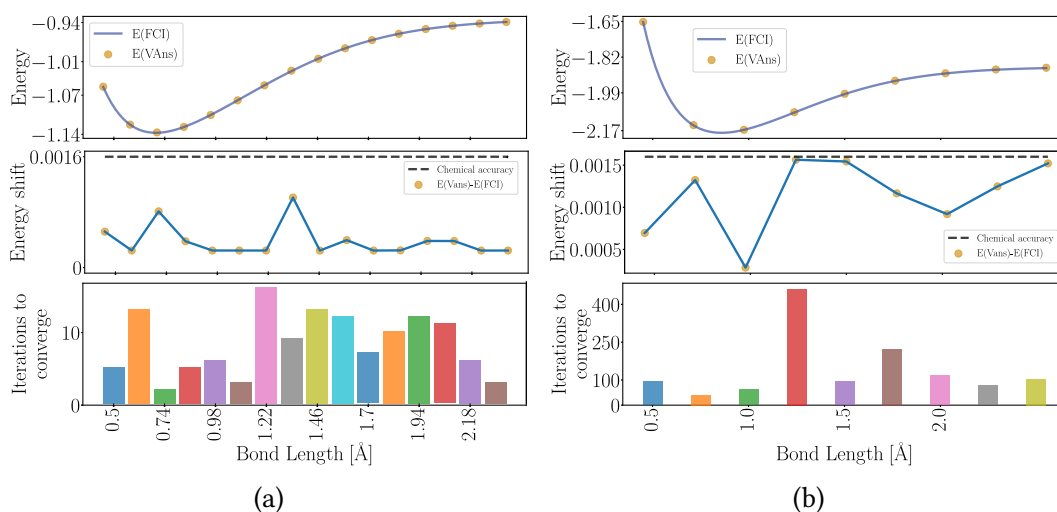


Figure 4.9: We show results of using VAns to obtain the ground state of a Hydrogen (H4) molecule, at different bond lengths (for the H4 molecule, this means the distance between the equally-separated atoms, which are disposed in a lineal array). Here we use VAns in the VQE algorithm for the molecular Hamiltonian obtained after a Jordan-Wigner transformation, leading to a 4(8)-qubit circuit in left(right) panels. Top: Solid lines correspond to ground state energy as computed by the Full Configuration Interaction (FCI) method, whereas points correspond to energies obtained using VAns. Middle: Differences between exact and VAns ground state energies are shown. Dashed line corresponds to chemical accuracy, which stands for the ultimate accuracy experimentally reachable in such systems. Bottom: Number of iterations required by VAns until convergence are shown.

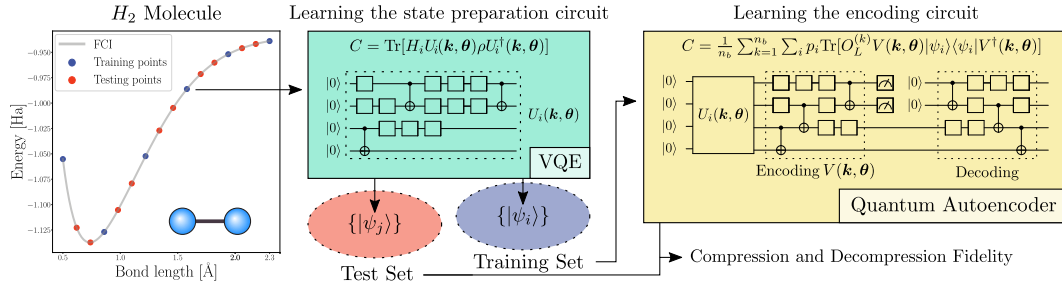


Figure 4.10: Schematic diagram of the quantum autoencoder implementation. We first employ VANS to learn the circuits that prepare the ground states  $\{|\psi_i\rangle\}_{i=1}^M$  of the  $H_2$  molecule for different bond lengths. These ground states are then used to create a training set and test set for the quantum autoencoder implementation. The goal of the autoencoder is to train an encoding parametrized quantum circuit  $V(\mathbf{k}, \boldsymbol{\theta})$  to compress each  $|\psi_i\rangle$  into a subsystem of two qubits so that one can recover  $|\psi_i\rangle$  from the reduced states, and we quantify this using the cost in Eq. 4.13.

#### 4.2.4 Quantum Autoencoder

We will now focus on the quantum autoencoder. This is a paradigmatic quantum-machine learning application which was first introduced in [ROAG17], where the quantum information stored in a multipartite quantum state is aimed to be compressed in terms of qubits' number.

In the following, we will first introduce the quantum-autoencoder and then study how VANS can be applied to train it. In particular we will show results when using VANS to train the autoencoder to compress the ground-states of the Hydrogen molecule, from 4 qubits to 2.

We consider a bipartite quantum system  $AB$  of  $n_A$  and  $n_B$  qubits, respectively. Let  $\{p_i, |\psi_i\rangle\}$  be a training set of pure states on  $AB$ . The goal of the quantum autoencoder is to train an encoding parametrized quantum circuit  $V(\mathbf{k}, \boldsymbol{\theta})$  to compress the states in the training set onto subsystem  $A$ , so that we can discard the qubits in subsystem  $B$  without losing much information. This can be quantified by introducing an ancillary system  $B'$ , and computing how close can the autoencoder transformation get to the initial training set. The performance of this task can be quantified by the average fidelity between the input and output states (as a

shorthand, we will denote  $V(\mathbf{k}, \boldsymbol{\theta})$  as  $V_{AB}$ , when acting on the system (AB)).

$$C_1(\boldsymbol{\theta}) = 1 - \sum_i p_i \mathcal{F}(|\psi_i\rangle, \rho_{i,out}), \quad (4.11)$$

$$\rho_{i,out} = V_{AB'}^\dagger \text{Tr}_B \left[ V_{AB} \left( |\psi_{i,AB}\rangle \otimes [0_{B'}] V_{AB}^\dagger \right) \right] V_{AB'}, \quad (4.12)$$

where we note that  $\rho_{i,out} \in \mathcal{H}_{AB'}$  (for our purposes,  $B'$  is just a copy of  $B$ , and they have the same dimension), and the autoencoder is represented by the unitary  $V$  we aim to train. Moreover, we have used the notation  $[\psi] = |\psi\rangle\langle\psi|$ , and denoted the fidelity between to states by  $\mathcal{F}$ .

Thus, the quantum autoencoder consists on the following steps:

- Encode the quantum state  $|\psi_{i,AB}\rangle$  to obtain  $[\psi'_{AB}] = V_{AB} \left( |\psi_{i,AB}\rangle \right) V_{AB}^\dagger$ .
- Discard the  $B$  system by tracing over it.
- Construct a new version of the global system  $AB'$ , by introducing a reference state  $[0_{B'}]$  in  $B'$ . This results in an state  $\text{Tr}_B[\psi'_{i,AB}] \otimes [0_{B'}]$ <sup>2</sup>.
- Decode the quantum state to obtain an approximate version of  $|\psi_i\rangle$  by revering the unitary transformation (e.g. by applying  $V^\dagger$ ).

In summary,  $V$  is aimed to decouple subsystem  $A$  from subsystem  $B$ , so that the resulting state is completely compressed into subsystem  $A$  if the qubits in  $B$  are found in the fixed target state  $[0_B]$ . In turn, assuming that  $V |\psi_{i,AB}\rangle = |\phi_{i,A}\rangle |0_B\rangle$ , then it follows that  $\rho_{i,out} = [\psi_{i,AB'}]$ , which lead to an average fidelity of 1.

As shown in [ROAG17], optimizing the average fidelity between the input and output states is equivalent to impose that the reduced state of system  $B$ , known as *trash state* e.g.  $\text{Tr}_A[\psi'_{i,AB}]$  after encoding, matches the reference state  $[0_{B'}]$ . This is done by introducing a swap between systems  $B$  and  $B'$  in between the encoding and decoding steps. Thus, the autoencoder task can be accomplished by training only on the trash state. Inspired by this, and the fact that computing the fidelities (for instance, via the SWAP) test requires a considerably complex circuit, whose compilation would lead to potentially many gates in the NISQ-context, Ref. [ROAG17] considers an alternative cost-function of the form

$$C(\mathbf{k}, \boldsymbol{\theta}) = 1 - \sum_i p_i \text{Tr} \left[ \left( |\vec{0}\rangle\langle\vec{0}|_B \otimes \mathbb{1}_A \right) V |\psi_i\rangle\langle\psi_i| V^\dagger \right] \quad (4.13)$$

---

<sup>2</sup>For our purposes, we can think on the previous item and this one as resetting the state of sub-system  $B$  to  $[0_{B'}]$ .



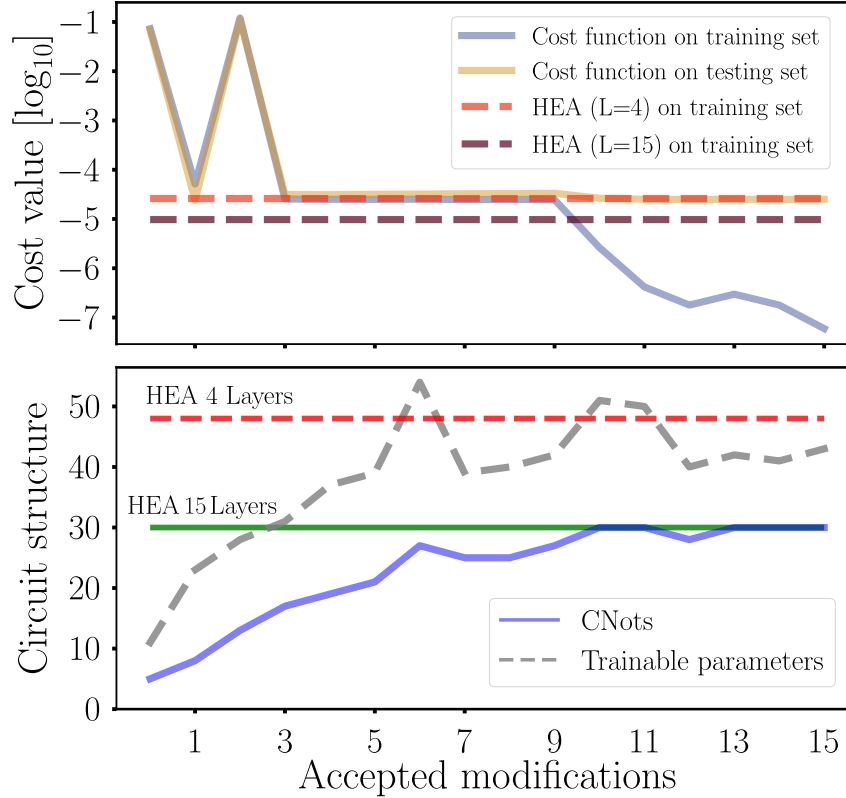


Figure 4.11: Results of using VAns to train a quantum autoencoder. Here we use VAns to train an encoding parametrized quantum circuit by minimizing Eq. (4.13) on a training set comprised of six ground states of the hydrogen molecule. We here also show the lowest cost function obtained for a  $L$ -HEA of  $L = 4$  and  $L = 15$  layers. Top panel: the cost function evaluated at both versus accepted VAns circuit modifications. In addition, we also show results of evaluating the cost on the testing set. Bottom panel: number of CNOTs, and number of trainable parameters versus the number of modifications of the ansatz accepted in the VAns algorithm. Here we additionally show the number of CNOTs (solid line) and parameters (dashed line) in the HEA ansatzes considered. We remark that for 15-HEA has 180 parameters, and hence the curve is not shown as it would be off the scale.

where  $\mathbb{1}_A$  is the identity on subsystem  $A$ . Here we see that if the reduced state in  $B$  is  $|\vec{0}\rangle_B$  for all the states in the training set, then the cost is zero.

As shown in Fig. 4.10, we employ the ground states  $|\psi_i\rangle$  of the  $H_2$  molecule (for  $M$  different bond lengths) to create a training set of six states and a test set of ten states. Here, the circuits obtained through VAns in the previous section serve as (fixed) state-preparation circuits for the  $H_2$ -molecule ground-states. We then use

VAns to learn an encoding circuit  $V(\mathbf{k}, \boldsymbol{\theta})$  which can compress the states  $|\psi_i\rangle$  into a subsystem of two qubits.

Fig. 4.11 presents results obtained by minimizing the cost in Eq. 4.13, for a single run of the VAns algorithm. As seen, within 15 accepted architecture modifications, VAns can decrease the training cost function down to  $10^{-7}$ , by departing from a separable product ansatz (see Fig. 2.2). We here additionally show results obtained by training the Hardware Efficient Ansatz of Fig. 2.2(b) with 4 and 15 layers (as they have a comparable number of trainable parameters and CNOTs, respectively, compared to those obtained with VAns). In all cases, VAns achieves the best performance. In particular, it is worth noting that VAns has much fewer parameters ( $\sim 45$  versus 180) than the 15-layer HEA, while also achieving a cost value that is lower by two orders of magnitude. Hence, VAns obtains better performance with fewer quantum resources.

To end this Section showcasing VAns in ideal scenarios, we will now discuss the case of unitary compilation.

## 4.2.5 Unitary compilation

Unitary compilation is a task in which a target unitary is decomposed into a sequence of quantum gates that can be implemented on a given quantum computer. As discussed in Sec. 2.2, current quantum computers are limited by the depth of quantum circuits that can be executed on them, which makes the compilation task very important in the near term. Indeed, we would like to decompose a given unitary using as few gates as possible to maximally reduce the effect of noise.

We will illustrate our approach by compiling Quantum Fourier Transform (QFT) on systems up to  $n = 10$  qubits. Apart from VAns, we also compile the QFT unitary using standard HEA and compare the performance of both methods.

The cost function for unitary compilation is defined as follows. First, a training set is selected

$$\{(|\psi_j\rangle, U_{\text{QFT}}^{(n)} |\psi_j\rangle)\}_{j=1}^M, \quad (4.14)$$

where  $U_{\text{QFT}}^{(n)}$  is a target QFT unitary on  $n$  qubits and  $|\psi_j\rangle$  are  $M$ , randomly selected input states. We assume that the states  $|\psi_j\rangle$  are pairwise orthogonal to avoid potential optimization problems caused by similarities in the training set. The cost function takes the form

$$C(\mathbf{k}, \boldsymbol{\theta}) = \sum_{j=1}^N \|U_{\text{QFT}}^{(n)} |\psi_j\rangle - V(\mathbf{k}, \boldsymbol{\theta}) |\psi_j\rangle\|^2. \quad (4.15)$$

Note that the cost function introduced in Eq. (4.15) becomes equivalent to a more standard one,  $C'(\mathbf{k}, \boldsymbol{\theta}) = \|U_{\text{QFT}}^{(n)} - V(\mathbf{k}, \boldsymbol{\theta})\|^2$ , when  $M = 2^n$ . While  $C(\mathbf{k}, \boldsymbol{\theta})$  measures the distance between the exact output of QFT and the one returned by  $V(\mathbf{k}, \boldsymbol{\theta})$  only on selected input states,  $C'(\mathbf{k}, \boldsymbol{\theta})$  measures the discrepancy between full unitaries  $U_{\text{QFT}}^{(n)}$  and  $V(\mathbf{k}, \boldsymbol{\theta})$ . It has recently been shown [CHC+22] that  $M \ll 2^n$  is sufficient to accurately decompose  $U_{\text{QFT}}^{(n)}$ . More precisely, a constant number of training states  $M$  (independent of  $n$ ) can be used to ensure small value of  $C'(\mathbf{k}, \boldsymbol{\theta})$ , while minimizing the cost function in Eq. 4.15. This observation provides an exponential speedup in evaluating the cost function for unitary compilation. Indeed, the cost of evaluating  $C(\mathbf{k}, \boldsymbol{\theta})$  in Eq. (4.15) is  $M \cdot 2^n$  (assuming the circuit  $V(\mathbf{k}, \boldsymbol{\theta})$  consists of few body gates and the states  $U_{\text{QFT}}^{(n)} |\psi_j\rangle$  are given in computational basis), while the cost of computing  $C'(\mathbf{k}, \boldsymbol{\theta})$  is  $4^n$ . The number of training states  $M$  which lead to small value of  $C'(\mathbf{k}, \boldsymbol{\theta})$  depends on the number of independent variational parameters in  $V(\mathbf{k}, \boldsymbol{\theta})$ . Suboptimal decompositions  $V(\mathbf{k}, \boldsymbol{\theta})$  (in terms of number of parametrized gates) will require larger  $M$  to achieve good compilation accuracy. We have used  $M = 15$  for  $n = 10$  qubit compilation with VAns, and a much larger training set in an approach that uses HEA; the increase in  $M$  is necessary since the latter approach needed a much deeper circuit, as discussed below. We have used most general two-qubit gates as the building block in VAns and to construct HEA. This is a slight generalization to the `Insertion` and `Simplification` steps in the VAns algorithm discussed above. General two-qubit gates can be decomposed in terms of CNOTs and one-qubit rotations using standard methods.

The method based on HEA requires very deep circuits (at  $n = 10$ ). They consists of so many gates that the regular optimization has very small success probability. We therefore modify the method based on HEA and utilize the recursive structure of  $U_{\text{QFT}}^{(n)}$ . In the modified approach, we use HEA to compile  $U_{\text{QFT}}^{(n-1)}$  and then use it to create an ansatz for  $U_{\text{QFT}}^{(n)}$ . The ansatz for larger system size additionally consists of several layers of HEA. We apply the above growth technique starting from  $n = 3$  to eventually build the ansatz for  $n = 10$ . We stress that VAns does not require such simplification and is capable of finding the decomposition with high success probability directly at  $n = 10$  while initialized randomly.

Figure 4.12 shows VAns results for  $n = 10$  qubit QFT compilation. Panel (a) depicts how the value of the cost function  $C(\mathbf{k}, \boldsymbol{\theta})$  is minimized over the iterations. We also show the corresponding value of  $C'(\mathbf{k}, \boldsymbol{\theta}) = \|U_{\text{QFT}}^{(n)} - V(\mathbf{k}, \boldsymbol{\theta})\|^2$ . We observe strong correlation between both cost functions.  $C'(\mathbf{k}, \boldsymbol{\theta})$  is eventually minimized below  $10^{-9}$  at the end of the optimization. Panel (b) shows how the

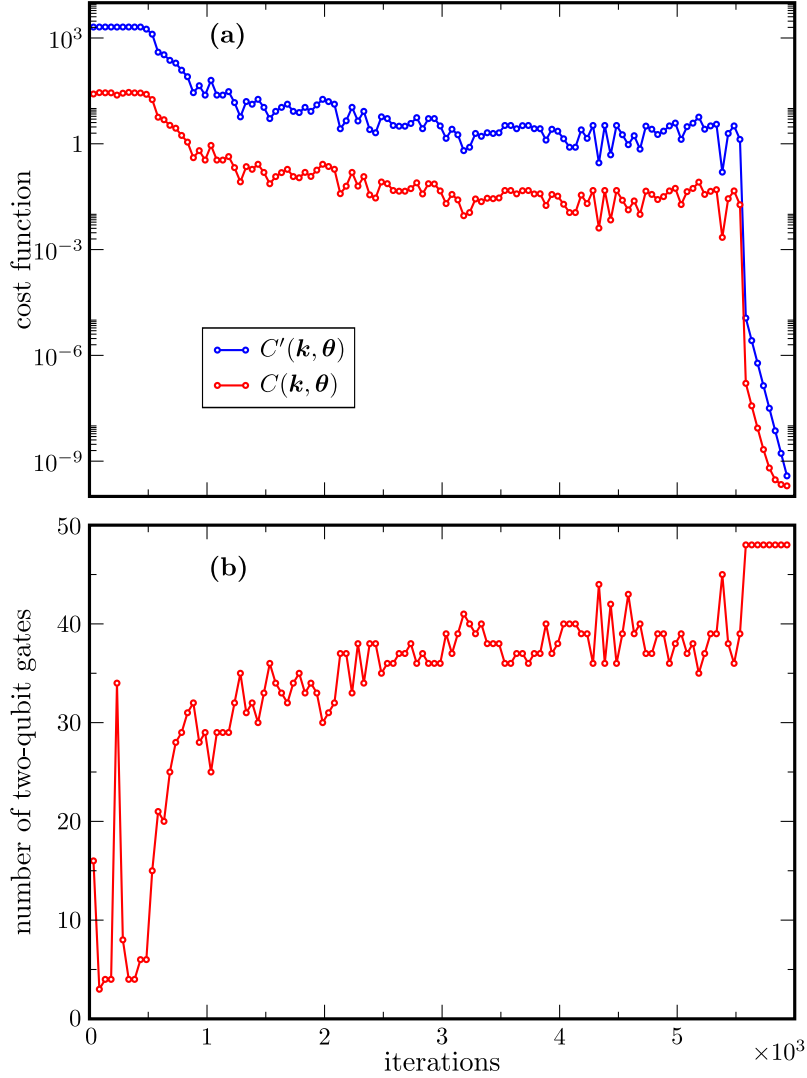


Figure 4.12: Results of using VAns for unitary compilation. Here we use VAns to find a decomposition of QFT unitary defined on  $n = 10$  qubits, by minimizing a cost function  $C(\mathbf{k}, \boldsymbol{\theta})$  (red line in panel (a)) defined in Eq. (4.15). The cost evaluates a difference between exact output of QFT and the one returned by a current circuit, on a small number of input states only ( $M = 15$ ). The blue line shows corresponding difference between full unitaries,  $C'(\mathbf{k}, \boldsymbol{\theta}) = \|U_{\text{QFT}}^{(n)} - V(\mathbf{k}, \boldsymbol{\theta})\|^2$ . We observe a high correlation between those two cost functions. Panel (b) shows how VAns modifies the number of two-qubit gates as it approaches the minimum of  $C(\mathbf{k}, \boldsymbol{\theta})$ . The minimum is found with 48 gates, which is  $\sim 4.5$  times less than the decomposition found with HEA (not shown).

number of two-qubit gates evolves as VAns optimization is performed. Excluding the initial *warm-up* period, during which the cost function has very large (close to maximal possible) value, the number of two-qubit gates is steadily grown reaching 48 at the end of the optimization. That number is only slightly larger than the number of two-qubit gates (45) used in the textbook-QFT circuit for  $n = 10$ .

The approach based on HEA requires 219 general two-qubit gates to decompose  $n = 10$  QFT, which is over 4 times more than the best circuit found by VAns. The HEA approach uses recursive structure of QFT to find accurate decomposition, while VAns does not rely on that property and finds a solution in fewer number of iterations. Finally, VAns takes advantage of the generalization bound [CHC+22], finding solution with near optimal number of variational parameters in  $V(\mathbf{k}, \boldsymbol{\theta})$ ; the training set size  $M$  required for small generalization error is small resulting in fast cost function evaluations.

#### 4.2.6 Noise and the $\lambda$ -model

The results previously shown were obtained without considering hardware noise. We observed that VAns was able to better exploit the quantum resources at hand (*i.e.* attain a lower cost-function value) as compared to its fix-structure counterpart (*e.g.* HEA).

We now consider the case where noisy channels are present in the quantum circuit, an unavoidable situation in current experimental setups, with noise essentially forbidding large-depth quantum circuits to preserve quantum coherence. In the context of QML, the overall effect of noise is that of degrading the cost-function value and, if its strength is sufficiently high, then short-depth circuits turn to be favoured even at the cost of expressibility. For instance, increasing the number of layers in HEA ansatz might not reduce the cost function and even increase it, since noise accumulates due to the presence of gates.

There are several sources of noise in quantum computers. For instance, experimental implementation of a quantum gate takes a finite amount of time, which in turn depends on the physical qubit at hand, the latter subjected to thermal relaxation errors. Relaxation and dephasing errors depend, in general, on each particular qubit (*i.e. the qubit label*). The overall effect of the gate implementation is often modeled by a depolarizing quantum channel, followed by phase flips and amplitude damping channels, whose strength depends on the aforementioned parameters (gate implementation time, qubit label), gate type and environment temperature. For instance, an entangling gate such as a CNOT injects considerably more noise to the circuit than a single-qubit rotation. Moreover, state initialization and readout

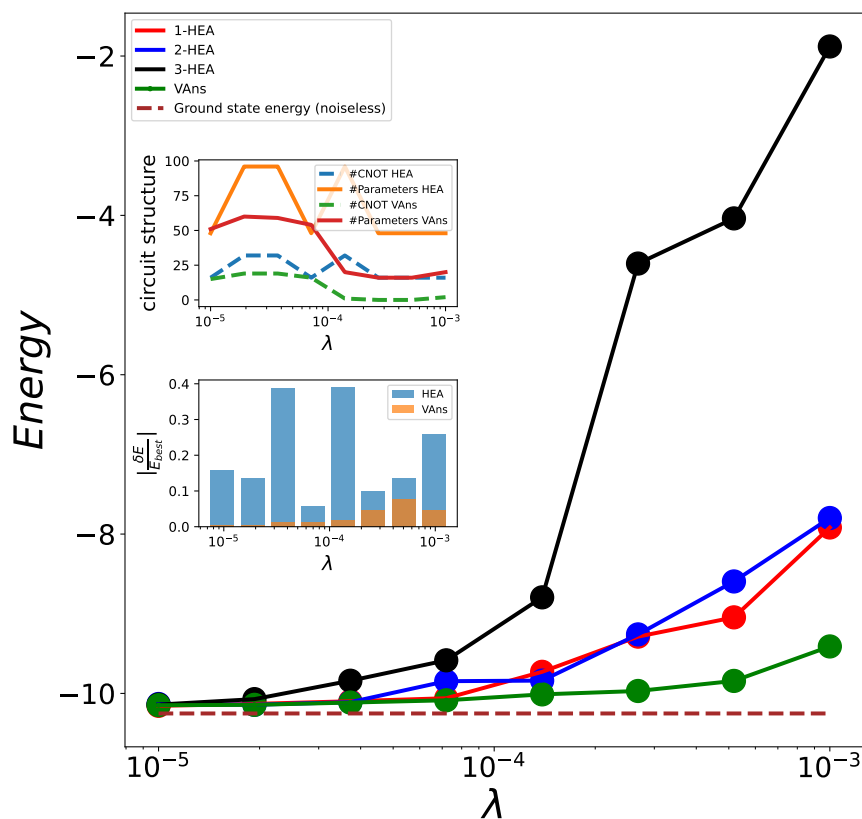


Figure 4.13: Results of using VAns for VQE under the  $\lambda$ -model. Here we consider the TFIM for 8 qubits, with  $g = J = 1$ . The results are obtained after repeating 50 iterations of optimizations with VAns and HEA respectively. We observe that VAns discovers much more efficient quantum circuits as compared to HEA. As shown in the upper inset, VAns automatically adjusts the circuit layout according to the noise strength at hand, a feature that fix-structure ansatzes lack. In the lower inset we show the relative errors (e.g. standard deviation over optimal cost found) for both ansatzes, across the 50 iterations; we observe that VAns is more precise in reaching a minimum as compared to HEA. We note that in this experiment we have initialized VAns to a 1-HEA, which is in turn inconvenient for a sufficiently high value of  $\lambda$ . Yet, VAns learns how to adapt the ansatz (in this case, finding a separable one) so to reach the lowest cost value. In all cases VAns termination criteria was set to a maximum number of 30 iterations.

errors should be taken into account. We refer the reader to find further details on noise modeling in Ref. [GEZ21]. We also note that additional sources of noise should ultimately be considered, such as idle noise and cross-talk effects [LMR+22; MMM+20]. Because the complexity of noise modelling in NISQ devices is particularly high, we here propose a sufficiently simple model that yet captures the essential noise sources.

**The  $\lambda$ -model.** While a noise-model is ultimately linked to the specific quantum hardware at hand and depends on several factors, we propose a unifying and simplified one that depends on a single parameter. The main motivation behind this is that of benchmarking the performance of different ansatzes in the presence of noise. In more complex scenarios, one should consider specific process matrices obtained from *e.g.* process tomography [OPG+04; YZKK+14], which would here obscure the benchmark and also bias it towards specific quantum hardware. In particular, our model is inspired by Refs [BPR+20; GEZ21], which in turn are implemented in the `aer noise simulator` of IBM, and consists of the following models. State preparation and measurement errors are modeled via bit flip channels acting on each qubit, with strength  $\lambda 10^{-2}$ , happening before the circuit  $U(\mathbf{k}, \boldsymbol{\theta})$  and measurements respectively. Noise due to gate implementation is modeled as a depolarizing channel, followed by a phase flip and amplitude damping channels acting on the target qubit right after the gate. In principle, the strength of the channel should depend on the specific qubit, and gate type, but in order to keep the model simple enough we have assumed no noise dependence on the qubit label. Moreover, a two-qubit gate is considerably more noisy than a single-qubit one, which in the  $\lambda$ -model is reflected by the fact that noise strengths are an order of magnitude higher, in the depolarizing channel, than in single-qubit gates, the latter being  $\lambda 10^{-5}$ . Finally, the strengths of the phase flip and amplitude damping channels are set to  $\lambda 10^{-3}$ . We note that, while not considered here, different situations can easily be incorporated such as qubit connectivity constraints, or differences in qubits' quality (some qubits might be noisier than others). In such cases, we expect VAns to find circuits which automatically balance the trade-offs at hand, *i.e.* minimize the number of gates acting on such noisier qubits.

With this model at hand, we have explored the region of  $\lambda$  in which the action of the noise becomes *interesting*. The results of running VAns under the  $\lambda$ -model for ground state preparation (VQE) of TFIM 8-qubit system are shown in Fig. 4.13. Here, the noise strength is sufficiently high so to affect the ground-state energy (which can otherwise be attained by either VAns or a 3-HEA). We thus sweep the value of  $\lambda$  by two orders of magnitude, and compare the results that VAns reaches with those of HEA (varying the number of layers of the latter). We see that for a

sufficiently high noise strength, increasing the number of gates (e.g. the number of layers in HEA) eventually degrades the cost-function value, as opposed to the noise-less scenario. On the contrary, we observe that even if the noise-strength is sufficiently high, VAns considerably outperforms HEA by automatically adjusting the depth of the circuit to the noise-strength at hand. Thus, if the noise is large, shallow circuits are found, whereas if the noise-strength is low, deeper circuits are allowed to be explored, since the penalty of adding new gates is smaller. In general, we observe that VAns is capable of finding the best possible circuit under given conditions, which is something that HEA simply can not accomplish.

### 4.3 Discussion and future perspectives

In this Chapter, we have introduced the VAns algorithm, a semi-agnostic method for building variable-structure parametrized quantum circuits.

At each iteration of the optimization, VAns stochastically grows the circuit to explore the architecture hyperspace. Crucially, VAns also compresses and simplifies the circuit by removing redundant and unimportant gates. This is a key aspect of our method, as it differentiates VAns from other variable ansatz approaches and allows us to produce shallow circuits, which can potentially mitigate the effect of noise.

To showcase the performance of VAns, we simulated our algorithm for several paradigmatic problems in the so-called quantum machine learning framework. Namely, we implemented VAns to find ground states of condensed matter systems and molecular Hamiltonians, for a quantum autoencoder problem and for 10-qubit QFT compilation. In all cases, VAns was able to satisfactorily create circuits that optimize the cost. Moreover, due to VAns' specific circuit-compression rules, these optimal circuits contain a small number of trainable parameters and entangling gates. Here we also compared the result of VAns with those obtained by using a Hardware Efficient Ansatz with either the same number of entangling gates or the same number of parameters: in all cases we found that VAns could achieve the best performance. This point is crucial for the success of VAns in the presence of noisy channels, as it automatically adapts the circuit layout to the situation at hand (e.g. noise strength). For instance, under the  $\lambda$ -model (which is the noise model we have proposed and implemented), VAns notably outperforms HEA under ground-state preparation tasks.

While we provided the basic elements and structure of VAns (*i.e.*, the gate `Insertion` and gate `Simplification` rules), these should be considered as blueprints for variable ansatzes that can be adapted and tailored to more spe-



cific applications. For instance, the gates that VAns inserts can preserve a specific symmetry in the problem. Moreover, one can cast the VAns architecture optimization (*e.g.*, removing unimportant gates) in more advanced learning frameworks. Examples of such frameworks include supervised learning [QBW22] or reinforced learning schemes [FNM+21; MPR+21; HM22], which could potentially be employed to detect which gates are the best candidates for being removed.

We expect VAns to be especially useful for abstract applications, such as linear systems [BPLC+19; HBR19; XSE+21], factoring [AOAG+19], compiling [KLP+19; SKC+20], metrology [BCS+22; KEJ+20], and data science [LTOJ+19; CSA+22; BWP+17; SSP14; ASZ+21; VMN+19], where physically motivated ansatzes are not readily available. In addition, VAns will likely find use even for physical applications such as finding ground states of molecular and condensed matter systems, as it provides an alternative to physically motivated ansatzes for mitigating the impact of noise, as shown in our noisy simulations. This is particularly promising since we have seen that VAns readily adapts the ansatz to the noise situation at hand.

Let us now discuss how VAns is expected to deal with barren plateaus, which currently constitute one of the major barriers for the success of VQA frameworks.

### 4.3.1 Mitigating the effect of barren plateaus

First, consider the type of BPs that are caused by the circuit approaching an approximate 2-design [MBS+18]. Approximating a 2-design requires a circuit that both has a significant number of parameters and also has a significant depth [BHH16; DCE+09; HL09; HM18; Haf22]. Hence, reducing either the number of parameters or the circuit depth can combat the appearance of barren plateaus. VAns attempts to reduce both the number of parameters and the depth and consequently attempts to avoid approximating a  $t$ -design.

Second, consider the BPs that are caused by hardware noise [WFC+21]. For such barren plateaus, it was shown that the circuit depth is the key parameter, as the gradient vanishes exponentially with the depth. As VAns actively attempts to reduce the number of CNOTs, it also reduces the circuit depth. Hence VAns will mitigate the effect of noise-induced barren plateaus by keeping the depth shallow during the optimization. As we have seen in our noisy simulations in Sec. 4.2.6, VAns automatically adjusts the circuit layout in such a way that the cost function reaches a minima, which translates to short-depth circuits in noisy scenarios.

### 4.3.2 Future directions

#### BP-aware implementation

While in the previous subsection we have presented general arguments as to why VAns can improve trainability, here we instead present a practical method that combines VAns with the recent techniques of Ref. [SMM+22] for mitigating barren plateaus using classical shadows.

As discussed in Sec. 2.2.4, it is known that the presence of barren plateaus is intrinsically related to the entanglement generated in the circuit [SCC+22; PNG+21; MKW21]. That is, circuits generating large amounts of entanglement are prone to barren plateaus. With this remark in mind, the authors in [SMM+22] propose to detect the onset of a barren plateau by monitoring, at each iteration, the entanglement of the resulting state. This can be achieved by computing, via classical shadows [HKP20], the second Rényi entanglement entropy  $S_2(\rho_R) = -\log(\text{Tr}[\rho_R^2])$ , where

$$\rho_R = \text{Tr}_{\bar{R}} [U(\mathbf{k}, \boldsymbol{\theta}) \rho_i U^\dagger(\mathbf{k}, \boldsymbol{\theta})] \quad (4.16)$$

denotes a reduced state on a subset of  $R$  qubits. As such, if  $S_2(\rho_R)$  approaches the maximal possible entanglement of the  $S$  qubits, given by the so-called Page value  $S^{\text{page}} \sim k \log(2) - \frac{1}{2^{m-2k+1}}$ , one knows that the optimization is leading to a region of high entanglement, and thus of barren plateaus.

The key proposal in [SMM+22] is to tune the optimizer (*e.g.*, by controlling the gradient step) so that regions of large entropies are avoided. This technique is shown to work well with an identity block initialization [GWO+19], whereby the parameters in the trainable unitary are chosen such that  $U(\mathbf{k}, \boldsymbol{\theta}) = \mathbb{1}$  at the start of the algorithm. Note that, in principle, this is still a fixed-ansatz method, as some circuit structure has to be fixed beforehand, and as no gates are ever removed. Hence, the methods in Ref. [SMM+22] can be readily combined with VAns to variationally explore the architecture hyperspace while keeping track of the reduced state entropy. In practice, this means that one can modify the VAns update rule to allow for steps that do not significantly increase entropy, while favouring steps that keep the entropy constant, or even that reduce it (*e.g.* by removing gates during the Simplification modules).

#### Reinforcement-learning

Recapting the formalism of RL introduced in the previous Chapter, we might be tempted to define state-action value functions as done in Chapter 3, where the problem was to find quantum receiver configurations. Nonetheless, we are here

dealing with a task of higher complexity, as illustrated when casting it in terms of reinforcement learning. For instance, we can define  $\mathcal{C}(\theta)$  to be the state  $s$  of the circuit (*e.g.* a classical description of circuit's layout and parameters value  $\theta$ ), and the actions  $a \in \mathcal{A}$  consisting on placing a parametrized gate  $G_x(\alpha) \in \mathcal{A}$ , at the  $x$ -th position in the circuit, where  $\mathcal{A}$  is a dictionary of available quantum gates that can be used to grow the circuit. Moreover, we can define the reward to be the global minima of the cost-function associated to state  $s$  (assuming the classical optimization algorithm is capable of finding it). From here, we could opt to consider an infinite-horizon scheme, where the environment dynamics  $\tau$  becomes a deterministic function of  $s$  and  $a$ <sup>3</sup>. However, inferring the state-action value function constitutes an extremely hard task, since one should consider any possible circuit layout reachable from  $s'$  on. Moreover, the presence of quantum correlations makes the encoding of circuit's state a highly non-trivial task; here, we would need an encoding that allows to process the circuit by *e.g.* a neural network (or more sophisticated RL agents).

Nevertheless, we could think of an hybrid approach, under which a subset of candidate quantum circuits is found by VAns, and an RL-agent is rewarded by filtering the important features of cost-minimizing circuits, and repeating such approach iteratively, which would fall under the umbrella of batched reinforcement learning.

### CV systems, receivers and beyond

The VAns algorithm can also be casted in discovering useful continuous-variable quantum neural network structures [KBA+19]. In this regard, we shall not restrict our attention to quantum-computing problems, and we note that VAns could also be applied to similar problems than those dealt during Chapter 3. In turn, very little is known about the structure of near-optimal and implementable quantum receivers (particularly if the states form an  $M$ -ary set of coherent-states, with  $M \geq 3$ ). Here, the cost-function to be optimized is the error probability,  $\mathbf{k}$  would be the POVM structure, in terms of symplectic transformations (given by phase-shifts, beam-splitters, etc.) and active transformations (squeezing, displacements) parametrized by  $\theta$ . In this regard, similar problems could be considered, such as channel-discrimination and parameter estimation (where the cost-function will be linked to the fisher information).

---

<sup>3</sup>We could slightly complexify the model, for instance by including circuit compression rules carried out after each gate-insertion

### 4.3.3 Code

The code and simulations supporting this Chapter (but the large-scale one, which has not been carried out by myself) can be found in the repos [Bil02c; Bil02d]. Finally, let us mention that a tutorial used for introducing quantum variational methods to undergraduate quantum physics students can be found in [BC21]

### 4.3. DISCUSSION AND FUTURE PERSPECTIVES

---

*Morelliana. Si el volumen o el tono de la obra pueden  
llevar a creer que el autor intentó una suma,  
apresurarse a señalarle que está ante la tentativa  
contraria, la de una resta implacable*

— Julio Cortazar

# 5

---

## Learning in the daylight

---

### 5.1 Introduction

So far in this thesis we have studied scenarios in which an agent had either none or little information about the encomended task. On the one hand, we studied the performance of an agent that learns *in the darkness*, by introducing a model-free reinforcement-learning approach to the calibration of coherent-state receivers in Chapter 3. On the other hand, we presented a semi-agnostic method that finds new NISQ circuits *in the twilight*. This is done by the help of VAns algorithm, discussed in Chapter 4, and which consists on randomly modifying quantum circuit's structure. While the opsome light is shed to the agent by allowing the usage of specifically-tailored compression rules that help with the navigation over the architecture-hyperspace.

This Chapter departs from the learning-with-sparse-information scenario, and takes the agent to a daytrip. Here, the agent has a full description of its environment, and we will ask her to tackle statistical-inference problems, a framework that has already been introduced in Sec. 2.5.

In the first part of the Chapter, we will introduce sequential hypothesis testing to continuously-monitored quantum systems —discussed in Sec. 2.4— where a model for the dynamics needs to be distinguished between two candidates. The second part of the Chapter is about parameter estimation, where an unique model

of the environment is available, and the agent needs to tell the value of a certain parameter, e.g. the frequency of an harmonic oscillator.

## 5.2 Hypothesis testing in continuously-monitored systems

Opposite to other approaches, in which a *deterministic* approach is followed [KM18]<sup>1</sup>, a sequential strategy processes the data *on the fly*, and potentially deems the data is not sufficiently informative (and thus further samples are demanded). As pointed out in Sec. 2.5.4, the Sequential Probability Ratio Test (SPRT) analyzes the log-likelihood ratio  $\ell$  between the two hypothesis, and provides strong error guarantees, in the sense that the error committed for a given sample sequence can be certified. This comes with a cost, namely that the length of the sequence becomes a random variable itself. The conditions imposed on the SPRT are stronger than those of the *deterministic* test: in a deterministic setting the errors are only accounted for in average, over all possible measurement sequence, so the conditional error for a single measurement outcome could be actually larger (or smaller). Yet, the average number of samples required to reach a certain error is generally smaller, for the SPRT than for the deterministic test.

Our main contribution in this Chapter is to study this advantage in continuously-monitored quantum systems. Note that an introduction to the topic of continuously-monitored systems is provided in Sec. 2.4.

The problem of sequential hypothesis testing has been studied in the quantum realm [MVHS+21; LHT22] in a setting where copies of a quantum system (either in  $\rho = \rho_0$  or in  $\rho = \rho_1$ ) are provided on demand. The ultimate quantum bound on the mean stopping time (or mean number of sampled copies) has been recently shown to follow the “quantized version” (naively exchanging probability distributions  $p_{0/1}(x)$  for quantum states  $\rho_{0/1}$ ):  $\langle \tau \rangle \sim -\frac{\log \epsilon}{D(\rho_1 || \rho_0)}$ .

Here, we move to a completely different setting where instead of performing a sequence of measurements on an increasing number of copies, we perform a (continuous) sequence of measurements on the *very same quantum system*, and the question is to discriminate between two possible internal dynamics of the monitored system. We envision a quantum sensor, in particular an optomechanical sensor, whose dynamics is affected by the presence of an external mass — or some

---

<sup>1</sup>In such deterministic approach, a batch of data is analyzed and a decision on the underlying hypothesis (*i.e.* model) is made afterwards

other force — and our task is to be able to detect this presence by observing a single (possibly long) measurement signal.

In particular, we consider an optomechanical system that consists on a mechanical mode coupled to an optical mode, the latter often called the *cavity* mode because it is stored in a cavity. The mechanical mode describes one of the ending mirrors of the cavity, whose free evolution is given by a quantum harmonic oscillator Hamiltonian, with frequency  $\omega_m$ . The other extreme-mirror is assumed to be imperfect, allowing information (e.g. light) to leak. Such leaked photons couple to a bosonic bath, which is assumed to be in the ground state, and such bath is continuously-monitored via homodyne measurement. While the leaked information comes from the cavity mode, such is coupled to the mechanical mode via an interaction of the form  $V_{int} \sim q_m q_c$ , where  $q_m$  is the position operator of the mechanical mode and  $q_c$  is the cavity mode's quadrature. Overall, by measuring the photons leaked from the cavity, we can infer information about the mechanical mode thanks to such interaction. Moreover, the cavity-mode normally evolves much quickly than the mechanical system, and an approximation called *adiabatic elimination* is carried out, where an effective stochastic master equation for the reduced system (i.e. the mechanical mode alone) under homodyne detection is obtained, as discussed in Sec. 2.4.7 (see also Ref. [DJ99]).

Generally, the evolution of the mechanical mode under continuous-homodyne monitoring of the environment modes  $\mathbf{c} = (c_1, \dots, c_k)^T$  to which is coupled with is given by a Belavkin-Zakai equation of the form

$$d\rho_t = dt\mathcal{L}_\theta[\rho_t]dt + \mathcal{H}[\eta\mathbf{c}]\rho_t \cdot d\mathbf{W}_t \quad (5.1)$$

$$\mathcal{L}[\rho] = -i[H, \rho] + \mathcal{D}[\mathbf{c}]\rho \quad (5.2)$$

$$\mathcal{D}[\mathbf{c}]\rho = \mathbf{c}\rho\mathbf{c}^\dagger - \left\{ \frac{\mathbf{c}^\dagger\mathbf{c}}{2}, \rho \right\} \quad (5.3)$$

$$\mathcal{H}[\eta\mathbf{c}]\rho = \mathcal{H}_{\eta\mathbf{c}}[\rho] - \text{Tr}[\mathcal{H}_{\eta\mathbf{c}}[\rho]]$$

$$\mathcal{H}_{\eta\mathbf{c}}\rho = \sqrt{\eta}\mathbf{c}\rho + \rho\sqrt{\eta}\mathbf{c}^\dagger$$

where a semi-definite complex-valued matrix  $\eta$  generalizes our discussion in Sec. 2.4.3 regarding detection efficiencies to consider several modes [WD05; WM09]. The term  $\mathcal{D}[\mathbf{c}]$  represents a diffusive contribution of the external bath, and the stochastic contribution  $d\mathbf{W}_t$  stand for the *innovations*, i.e. how much the measurement outcome  $dy_t$  deviates from its expected value:

$$d\mathbf{W}_t = dy_t - \text{Tr}[\mathcal{H}_{\eta\mathbf{c}}[\rho_t]]. \quad (5.4)$$

Here, the probability of having measurement outcome  $dy_t$  at time  $t$  is described by



the Gaussian observational model

$$P(d\mathbf{y}_t|\rho_t) = \frac{e^{-\frac{d\mathbf{W}_t \cdot d\mathbf{W}_t}{2dt}}}{\sqrt{2\pi k dt}} \quad (5.5)$$

$$= e^{-\frac{1}{2}[\text{Tr}[\mathcal{H}_{\eta\mathbf{c}}[\rho_t]]]^2 + \text{Tr}[\mathcal{H}_{\eta\mathbf{c}}[\rho_t]] \cdot d\mathbf{y}_t} P_{d\mathbf{W}_t}(d\mathbf{y}_t), \quad (5.6)$$

where  $P_{d\mathbf{W}_t}(d\mathbf{y}_t)$  is the probability distribution of  $k$  independent Wiener processes such that  $\mathbb{E}[dW_i dW_j] = dt\delta_{ij}$ . Accordingly, the measurement record is defined as  $\mathcal{Y}_t = (dy_0, \dots, dy_t)$ , and its probability, conditioned on the evolution of  $\rho_t$ , is given by

$$P(\mathcal{Y}_t|\rho_t) = e^{\lambda(\mathcal{Y}_t|\rho_t)P_W(\mathcal{Y}_t)}, \quad (5.7)$$

where  $P_W(\mathcal{Y}_t) = \prod_{\tau=0}^t P_{d\mathbf{W}}(d\mathbf{y}_\tau)$  is just the joint probability of the independent Wiener noises. Here,  $\lambda(\mathcal{Y}_t|\rho_t)$  is the log-likelihood distribution associated to the sequence of outcomes  $\mathcal{Y}_t$  and whose evolution is inferred from Eq. 5.5 and Eq. 5.7 to be

$$d\lambda(\mathcal{Y}_t|\rho_t) = \text{Tr}[\mathcal{H}_{\eta\mathbf{c}}[\rho_t]] \cdot d\mathbf{y}_t - \frac{1}{2}[\text{Tr}[\mathcal{H}_{\eta\mathbf{c}}[\rho_t]]]^2 dt. \quad (5.8)$$

We remark that the quantum state in turn  $\rho_t$  depends on the specific realization of  $\mathcal{Y}_t$  via Eq. (5.1). Moreover, the realizations of such process — either the quantum trajectories or the measurement record  $\mathcal{Y}_t$  — will also depend on the specific structure of system dynamics. In this regard, we will here focus on two possible models, *each associated to a different set of parameters* that govern the dynamics. Such parameters can, for example, be the cavity's decay rate value, or specific constants appearing in the free evolution of the mechanical mode, such as the mechanical-mode frequency.

By defining as  $\theta_0$  and  $\theta_1$  the two possible candidates for the parameter(s) we aim to distinguish<sup>2</sup>, then we can readily compute the log-likelihood ratio

$$\ell(\mathcal{Y}_t) = \log \frac{P_1(\mathcal{Y}_t)}{P_0(\mathcal{Y}_t)}, \quad (5.9)$$

where  $P_i(\mathcal{Y}_t)$  stands for the probability of the random sequence under parameter  $\theta = \theta_i$ . In such case, the solution of Eq. 5.1 at time  $t$  will be denoted as  $\rho_{t,k} = \rho_{t,\theta_k}$ , with  $k$  being either 0 or 1, defining the underlying model for the dynamics. As in Sec. 2.5.1, we will deem  $k = 1$  as the alternative hypothesis, and  $k = 0$  as the null

---

<sup>2</sup>In case the model comprises more than one parameter,  $\theta$  should be understood as a tuple containing such parameter values. However, in our numerics and case-studies we will restrict to the single-parameter case

one. From Eq. 5.8 we can readily find a dynamical equation for the log-likelihood ratio as per

$$d\ell(\mathcal{Y}_t) = \text{Tr}[\mathcal{H}_{\eta\mathbf{c}}[\rho_{t,1} - \rho_{t,0}]] \cdot d\mathbf{y}_t - \frac{1}{2}(|\mathcal{H}_{\eta\mathbf{c}}[\rho_{t,1}]|^2 - |\mathcal{H}_{\eta\mathbf{c}}[\rho_{t,0}]|^2)dt, \quad (5.10)$$

where we remark that  $\rho_i$  is obtained by solving the Belavkin equation up to time  $t$  under the hypothesis  $\theta = \theta_k$ <sup>3</sup>.

Moreover, if we consider that hypothesis  $k$  is the underlying *true* model for the dynamics, meaning that the measurement outcomes are generated from  $\rho_k$ , according to

$$d\mathbf{y}_t = d\mathbf{y}_{t|k} = \text{Tr}[\mathcal{H}_{\eta\mathbf{c}}[\rho_{t,k}]]dt + d\mathbf{W}_t, \quad (5.11)$$

then the log-likelihood ratio  $\ell(\mathcal{Y}_{t|k})$  can be written as a function of the innovations according to

$$d\ell(\mathcal{Y}_{t|k}) = (-)^{k+1} \frac{1}{2} |\mathcal{H}_{\eta\mathbf{c}}[\rho_1(\mathcal{Y}_{t|k}) - \rho_0(\mathcal{Y}_{t|k})]|^2 dt + \text{Tr}[\mathcal{H}_{\eta\mathbf{c}}[\rho_1(\mathcal{Y}_{t|k}) - \rho_0(\mathcal{Y}_{t|k})]] \cdot d\mathbf{W}_t \quad (5.12)$$

This last equation indicates that the log-likelihood ratio obeys an stochastic equation, whose drift value at time  $t$  happens to be the squared of the diffusion value at such time, multiplied by a factor 2. Moreover, the drift will be positive for  $k = 1$ , and negative for  $k = 0$ . This should be intuitively clear, as  $\ell$  should (on average) tend towards positive values when hypothesis  $k = 1$  is *the true one*, and towards negative values when the null hypothesis  $k = 0$  stands for the underlying model that generates the data, as discussed in Sec. 2.5.1.

### 5.3 Sequential testing & Gaussian systems

We are interested in casting how sequential strategies perform in continuously-monitored systems. In Sec. 2.4.7 we discussed a particular model for an optomechanical system under continuous homodyne detection, whose dynamics for the mechanical mode (after adiabatic elimination of the cavity mode) under the Gaussian assumption obeys an stochastic linear system for the first two moments, *i.e.*  $\bar{\mathbf{r}} = \text{Tr}[\rho \hat{\mathbf{r}}]$  and  $\Sigma = \text{Tr}[\{\hat{\mathbf{r}} - \bar{\mathbf{r}}, \rho\}]$  given by

$$d\bar{\mathbf{r}}_t = (A - \chi(\Sigma_t)C)\bar{\mathbf{r}}_t dt + \chi(\Sigma_t)d\mathbf{y}_t = A\bar{\mathbf{r}}_t dt + \chi(\Sigma_t)d\mathbf{W}_t, \quad (5.13)$$

---

<sup>3</sup>As explained above, the dependence on  $\theta$  in Eq. 5.1 is not made explicit, and is rather encoded in the values of the physical constants appearing in the equation, such as decay rates, frequencies, efficiencies, etc.

where the measurement outcome  $dy_t$  is

$$dy_t = C(\bar{\mathbf{r}}_t dt + C^{-1})d\mathbf{W}_t. \quad (5.14)$$

We remark that this last equation Eq. 5.14 is not a dynamical equation for the measurement signal, but a compact way to provide its (stochastic) value at time  $t$ ; in turn, the measurement record is given by the tuple  $\mathcal{Y}_t = (dy_0, \dots, dy_t)$ . Moreover, the evolution for the quantum system's covariance reads

$$d\Sigma_t = A\Sigma_t + \Sigma_t A^T + D - \chi(\Sigma_t)^T \chi(\Sigma_t), \quad \chi(\Sigma) = \Sigma C^T + \Gamma, ,$$

with the coefficient matrices being

$$A = \begin{pmatrix} -\frac{\gamma}{2} & -\omega \\ \omega & -\frac{\gamma}{2} \end{pmatrix}, \quad D = \left[ \gamma \left( n + \frac{1}{2} \right) + \kappa \right] \mathbb{I}_2, \quad C = \sqrt{4\eta\kappa} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad (5.15)$$

and  $\Gamma = 0$ . We recall that  $\gamma$  is the damping rate of the mechanical mode, whereas  $\omega$  is its frequency,  $\kappa$  the measurement strength and  $\eta$  the efficiency of the measurement. This model also considers an extra thermal bath that is in contact with the cavity, with  $n$  being the average number of thermal excitations. While the latter bath is *not* being measured, we consider the system to be in contact with another bosonic bath, found in the ground-state, and this is being continuously-monitored. Note also that we included  $C^{-1}$ , the pseudo-inverse of the matrix  $C$ , in order to evidence that only one component of the Wiener noise  $d\mathbf{W}_t$  is injected to the system (*i.e.* the quadrature that is being measured). A realization of this process is shown in Fig 5.1.

Furthermore, under the rotating frame that demodulates the signal – using the mechanical-mode frequency, as discussed in Sec. 2.4.7–, the matrices are modified to

$$C \rightarrow 2\sqrt{\eta\kappa} \mathbb{I}_2, \quad A \rightarrow -\frac{\gamma}{2} \mathbb{I}_2, \quad (5.16)$$

with the remaining coefficient matrices being untouched. In such scenario, both (rotating) quadratures do suffer from such noise injection. Finally, a *driving* term can be incorporated in these equations by adding a term in  $d\bar{\mathbf{r}}_t$ ; such term appears when driving the cavity with a laser, a matter that will be discussed in Sec. 5.4 when estimating external signals.

Let us discuss further the structure of Eqs. 5.13 and Eq.5.15. Firstly, we note that the innovations  $d\mathbf{W}_t$  do only appear in the evolution for the first moment  $\bar{\mathbf{r}}_t$ . In fact, the evolution for the covariance matrix given in Eq. 5.15 is deterministic, and known as the continuous Ricatti equation; moreover we note that the final

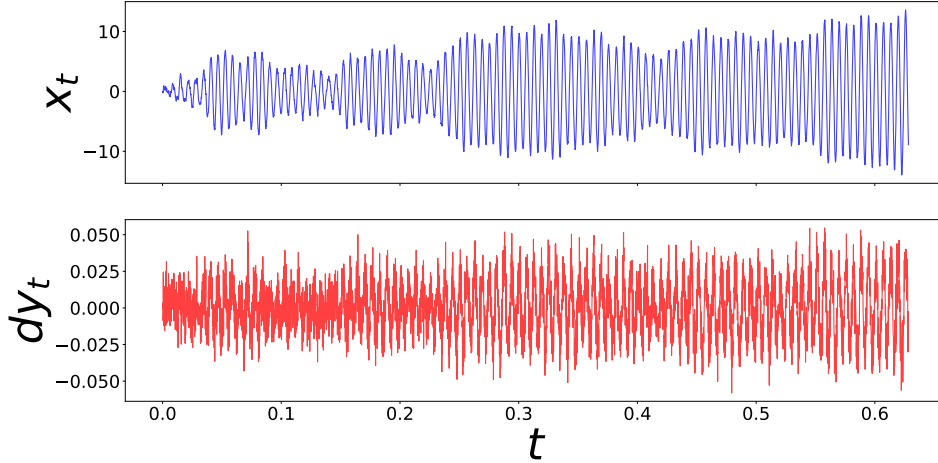


Figure 5.1: We show a realization of a quantum trajectory (top) along with the measurement record (bottom), for the process described by Eq. 5.13

term causes a reduction in the uncertainty about the system state and is related to the measurement via  $\chi(\Sigma)$ . Furthermore, it can be shown that the Riccati equation admits a stationary solution if the eigenvalues of  $A$  have a non-positive real-part part, what is known as the Hurwitz condition<sup>4</sup>, then the covariance will eventually converge to its stationary value (*i.e.*  $d\Sigma_t = 0$ ), also known as stabilizing solution [WD05].

Secondly, we note that Eq. 5.13 is shown in two forms: either with the measurement outcome  $dy_t$  or with the innovation  $dW_t$ . While the latter provides a clear interpretation of the evolution in the Ito form, being  $A$  and  $\chi(\Sigma_t)$  the drift and diffusion terms, the innovations are not experimentally accessible. Rather, we do only have access to the measurement outcomes  $dy_t$  in order to update our knowledge of the quantum-state, and thus the evolution should be understood from this perspective. In turn, this makes explicit the *back-action* phenomena: gathering information about the system unavoidable affects its evolution. In this regard, the system can be understood as a Hidden markov model, where the measurement outcome  $dy_t$  conditions the value of the hidden state  $\bar{r}_{t+dt}$ ; such hidden state (as well as the covariance) is unavailable to us, but can be *tracked* by means of Eqs. 5.13

---

<sup>4</sup>Intuitively, the imaginary part will contribute to the oscillatory behaviour, whereas the real part will either damp (positive) or magnify (negative) the signal, since the solutions are of the form  $\langle \bar{r}(t) \rangle = e^{At} \bar{r}(0)$

and 5.15.

While the bayesian structure of quantum theory via the Born rule and the post-measurement states result in the aforementioned set of stochastic lineal equations for the Gaussian model under consideration, there is a classical counterpart. This analogy with the Kalman filter was firstly highlighted in Ref. [DJ99], and a similar structure is obtained for the equations that update the first two moments of a Hidden Markov (Gaussian) system. In this regard, while the initial states of the system might be unknown, the equations steer an estimate of the first two moments in a contractive way, and via the measurement result  $dy_t$ , towards their (hidden) *true* value. This holds, as long as the right dynamics is used in order to update such moments. On the contrary, if a wrong dynamics was used, a discrepancy between the measurement result (which takes the estimate towards the *true* hidden state) and the update via the *wrong* model will arise. This is captured by the log-likelihood distribution, *i.e.* the probability of observing such measurement results under the corresponding model, and its evolution reduces in the Gaussian to

$$d\lambda_t = -\frac{1}{2} \|C\bar{\mathbf{r}}_t\|^2 dt + C\bar{\mathbf{r}}_t \cdot dy_t, \quad (5.17)$$

where  $\bar{\mathbf{r}}_t$  is the solution of Eq. 5.13 at time  $t$ , and its dependence on  $\mathcal{Y}_t$  is implicit.

Whether we are using the right dynamics in order to update the hidden state value or not, is the matter hypothesis testing. As discussed in Sec. 2.5.1, the discrepancy between the two hypothesis is captured via the log-likelihood ratio, which for the Gaussian system under consideration obeys the dynamical equation

$$d\ell(\mathcal{Y}_{t|k}) = \frac{(-1)^{k+1}}{2} \|C\Delta\bar{\mathbf{r}}_t\|^2 dt + C\Delta\bar{\mathbf{r}}_t \cdot d\mathbf{W}_t \quad (5.18)$$

where  $\Delta\bar{\mathbf{r}}_t = \bar{\mathbf{r}}_1(\mathcal{Y}_{t|k}, t) - \bar{\mathbf{r}}_0(\mathcal{Y}_{t|k}, t)$  is the vector difference between the first moment of the Gaussian state generated through the null and alternative hypothesis dynamics. Here, each of the moments associated to the corresponding hypothesis is obtained by (numerically) solving the set of stochastic lineal equations Eqs. 5.13 and Eq. 5.15 using the respective corresponding value for the coefficient matrices.

We will study two different scenarios: damping and frequency discrimination. The former will be analyzed in detailed, as some analytical results can be derived. Since our goal is to introduce sequential strategies for hypothesis testing in continuously-monitored systems, the very same approach can be used to test different models, and thus using it on frequency discrimination problems serves as an extra numerical *check*.

### Damping discrimination

We will now consider the scenario where a model is captured by a specific value of the mechanical mode's damping rate  $\gamma$ . The null hypothesis is given by  $\gamma = \gamma_0$  and the alternative hypothesis by  $\gamma = \gamma_1$ , with all the remaining coefficients (namely  $\eta$ ,  $n$  and  $\kappa$ ) being shared by the two models. We will study the case in which the system is analyzed in a frame that rotates with the mechanical-mode's frequency, and thus the coefficient matrices  $A$  and  $C$  are given by Eq. 5.16. In this case, the quadratures become uncoupled (since  $A$  is now diagonal), and both quadratures are affected by the same dynamics. Thus, we can expect the stationary value of the covariance matrix to be diagonal. Moreover, in this case the Riccati equation admits a closed-form solution for its stationary value under hypothesis  $k$ , namely  $\Sigma_k = \sigma_k \mathbb{I}_2$  with

$$\sigma_k = \frac{\gamma_k}{8\eta\kappa} \left( \sqrt{1 + \frac{16\eta\kappa\sigma_{k,uc}}{\gamma_k}} - 1 \right) \quad (5.19)$$

$$\sigma_{k,uc} = n + \frac{1}{2} + \frac{\kappa}{\gamma_k}, \quad (5.20)$$

where  $\sigma_{k,uc}$  is the solution for the unconditional evolution of the covariance in Eq. (2.168). Because Hurwitz conditions are satisfied for the system under consideration (and in order to save some computing resources), our simulations already start from the stationary state. In case that a closed-expression for such stable solution is not available, as it happens when studying cavity dynamics with non-trivial Hamiltonians [FRT+22], or without moving to a rotating-frame, we rely on numerical methods to solve the Riccati equation and provided by Refs. [VGO+20; Lau78].

Let us now study the evolution of the hidden-state  $\bar{\mathbf{r}}_{k,t} \equiv \bar{\mathbf{r}}_k(t)$ . Assuming hypothesis  $k$  is the true one, the solution of  $\bar{\mathbf{r}}_k(t)$  reads

$$\bar{\mathbf{r}}_k(t) = e^{A_k t} \bar{\mathbf{r}}_k(0) + \chi(\Sigma_k) \int_0^t e^{A_k(t-\tau)} d\mathbf{W}_\tau, \quad (5.21)$$

where we have expressed Eq. 5.13 as an integral stochastic equation. From our discussion of Ito calculus in Sec. 2.4.4, we can expect that the average value<sup>5</sup>  $\langle \bar{\mathbf{r}}_{k,t} \rangle \xrightarrow[t \rightarrow 0]{} 0$  for any initial condition  $\bar{\mathbf{r}}(0) = (q_0, p_0)$ . Moreover, we can readily

---

<sup>5</sup>Note that this average is over several realizations of the stochastic process, *i.e.* over different quantum trajectories, and as such is a *classical* one.

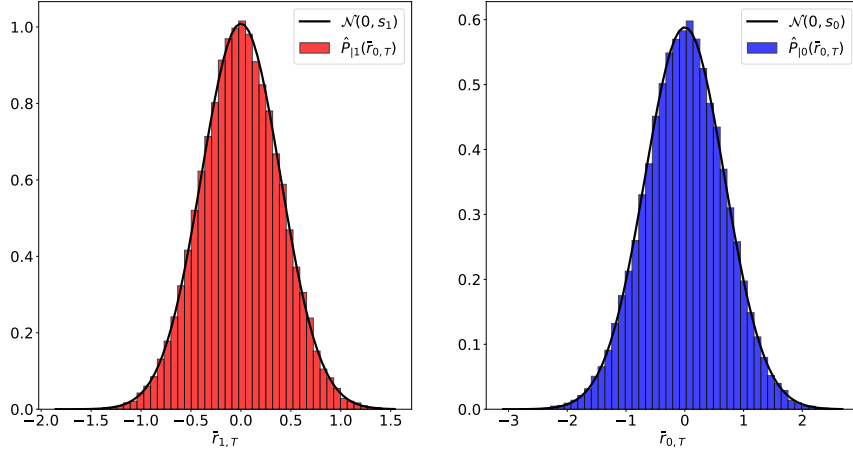


Figure 5.2: The distributions for the hidden states are shown, simulated under the alternative ( $\gamma = 429$  Hz) and null hypothesis ( $\gamma = 100$  Hz) respectively. The histograms are obtained by simulating  $2 \cdot 10^4$  quantum trajectories, and the time for which the distributions are displayed is  $T = 4s$ .

compute the variance in the long-time regime, *i.e.*  $\mathbb{E}[\bar{r}_k^2]$ , which reads

$$\mathbb{E}[\bar{r}_k^2] = \frac{8\eta\kappa\sigma_k^2}{\gamma_k} := s_k^2 \quad (5.22)$$

Thus, the first moment of the hidden state  $\rho_k$ , evolved under hypothesis  $k$ , is Gaussian distributed with zero mean and a variance  $s_k^2$  given by Eq. 5.22. This is illustrated in Fig. 5.2, where the probability distribution of the hidden state is shown at a large time, and compared with numerical simulation of the (Gaussian) quantum trajectories. Here, and in most of the plots shown in this section the simulations are based on the following choice of parameters:  $\gamma_0 = 100$  Hz for  $H_0$  and  $\gamma_1 = 429$  Hz for  $H_1$ , while the rest of the parameters are taken to be same for both hypotheses  $n := n_0 = n_1 = 1, \eta = \eta_0 = \eta_1 = 1, \kappa = \kappa_0 = \kappa_1 = 9$  Hz.

### 5.3.1 The log-likelihood ratio

As discussed above, chances exists that the *state of knowledge* of our quantum system is being updated with the *wrong* model. For simplicity, let us assume that hypothesis 1 generates the data, and thus  $dy_t = C(\bar{r}_{1,t} + C^{-1}d\mathbf{W}_t)$ . While both hypothesis are updated using  $dy_t$ , as per Eq. 5.13, the crucial matter is to reduce

the possibilities of telling that null hypothesis  $k = 0$  is the true one (we will later consider the symmetric error). Recalling that the evolution for  $\bar{\mathbf{r}}_0(t)$  reads

$$d\bar{\mathbf{r}}_{0,t} = [A_0 - \chi(\boldsymbol{\Sigma}_0)]\bar{\mathbf{r}}_{0,t}dt + \chi(\boldsymbol{\Sigma}_0)C\bar{\mathbf{r}}_{1,t}dt + \chi(\boldsymbol{\Sigma}_0)d\mathbf{W}_t, \quad (5.23)$$

where we have expanded the  $d\mathbf{y}_t$  term in Eq. 5.13, then the integral expression reads<sup>6</sup>

$$\bar{\mathbf{r}}_0(t) = \chi(\boldsymbol{\Sigma}_0)C \int_{-\infty}^t e^{\left[A_0 - \chi(\boldsymbol{\Sigma}_0)C\right](t-\tau)} \bar{\mathbf{r}}_1(\tau)d\mathbf{W}_\tau + \chi(\boldsymbol{\Sigma}_0) \int_0^t e^{\left[A_0 - \chi(\boldsymbol{\Sigma}_0)C\right](t-\tau)} d\mathbf{W}_\tau. \quad (5.24)$$

Recalling that  $\bar{\mathbf{r}}_1(t)$  is given by Eq. 5.21 for  $k = 1$ , and using the stochastic integration rule outlined in Sec. 2.4.4, we can readily compute the following expected values in the long-time regime ( $t \rightarrow \infty$ )<sup>7</sup>:

$$\mathbb{E}_1[\bar{\mathbf{r}}_0^2] = \frac{8\sigma_0^2\eta\kappa}{\gamma_0 + 8\sigma_0\eta\kappa} \left( 1 + \frac{16\sigma_1\eta\kappa}{\gamma_0 + \gamma_1 + 8\sigma_0\eta\kappa} + \frac{(8\sigma_1\eta\kappa)^2}{\gamma_1(\gamma_0 + \gamma_1 + 8\sigma_0\eta\kappa)} \right), \quad (5.25)$$

$$\mathbb{E}_1[\bar{\mathbf{r}}_0\bar{\mathbf{r}}_1] = \frac{32\sigma_0\sigma_1\eta\kappa}{\gamma_0 + \gamma_1 + 8\sigma_0\eta\kappa} \left( 1 + \frac{4\sigma_1\eta\kappa}{\gamma_1} \right). \quad (5.26)$$

We note that in the case that hypothesis 0 was the underlying one (*i.e.* generating  $d\mathbf{y}_t$ ) then the above expressions are identical, but swapping the values of  $\gamma_1$  and  $\gamma_0$ .

Combining these expressions with Eq. 5.18, we can readily find the expected value of  $\ell$  under hypothesis  $k$ . To see this, we take the expectation value, to get

$$\mathbb{E}_k[d\ell_t] = \frac{(-1)^{k+1}}{2} \mathbb{E} \left[ \|C\Delta\mathbf{r}_t\|^2 \right] dt = (-1)^{(k+1)} \mu_k dt \quad (5.27)$$

where in the last equality we have used the stationary value (long time regime) of the effective drift coefficient. Inserting the previous results we get an analytical expression for the value of  $\mu_1$  at  $t \rightarrow \infty$ ,

$$\mu_1 = \frac{c^2(\gamma_1^2\chi_0^2 + 2c\gamma_0(\chi_0 - \chi_1)^2 + \gamma_0^2\chi_1^2 + \gamma_0\gamma_1(\chi_0^2 - 4\chi_0\chi_1 + \chi_1^2))}{\gamma_0(\gamma_1 + 2c\chi_1)(\gamma_0 + \gamma_1 + 2c\chi_1)}, \quad (5.28)$$

where  $\chi_k = c\sigma_k$ , with  $\sigma_k$  given by Eq. 5.19 and  $c = \sqrt{4\eta\kappa}$ . The expression for  $\mu_0$  is obtained by swapping  $\gamma_0$  and  $\gamma_1$ . It is important to note that the two drifts do

---

<sup>6</sup>Note that we have dropped the deterministic term proportional to  $e^{(A-\chi(\boldsymbol{\Sigma}_0)t)}\bar{\mathbf{r}}_0(0)$  since we are concerned with the long-time regime.

<sup>7</sup>Exact expressions for all times can be also computed, but we refrain from giving them explicitly here.



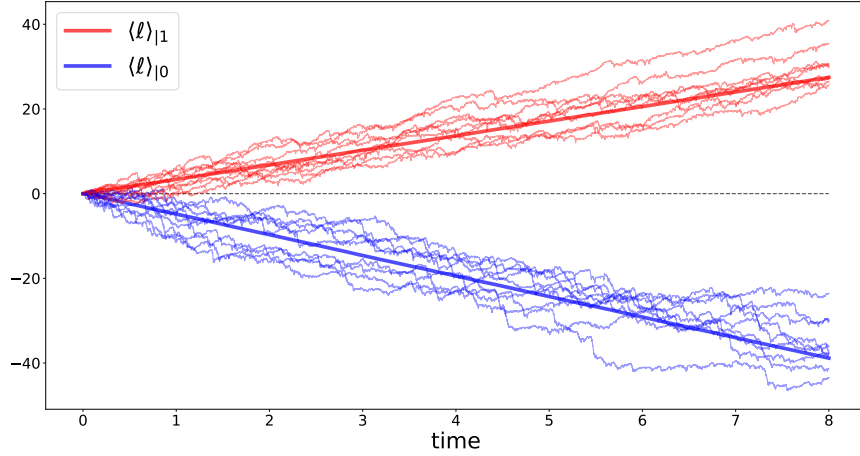


Figure 5.3: We show the stochastic evolution of the log-likelihood ratio under both models, together with its empirical average computed over  $2 \cdot 10^4$  trajectories, for the damping discrimination problem considered in this Section.

always differ. Similarly, we can compute the variance of  $\ell$ , which we anticipate scales linearly with time, and its full analytical expression is, at least, scaring. We will discuss an alternative approach to compute these two moments in (for general system parameters) later in the Sec. 5.3.3, where we consider the evolution of an extended system, comprising the dynamics of both models,  $\bar{r}_0$  and  $\bar{r}_1$ , effectively coupled by the common measurement record, as an Orstein-Uhlenbeck process.

From (5.27) and (5.28) we readily find that the expected value of the log-likelihood function at long times is given by

$$\mathbb{E}_k[\ell_t] = \int_0^t \mathbb{E}_k[d\ell_t] = (-1)^{k+1} \mu_k t + \mathcal{O}(1) \quad (5.29)$$

where  $\mathcal{O}(1)$  accounts for the finite contribution to the integral of the time it takes for  $\mu_1(t)$  to reach its stationary value.

We have conducted numerical simulations of a possible measurement records  $\mathcal{Y}_t$  obtained for both models under consideration and computed the the log-likelihood statistic  $\ell_t(\mathcal{Y}_t)$  according to the above sequential prescription. This is depicted in Fig. 5.3, where the average value of  $\ell_t$  is shown for both models, accompanied by some realizations of the process. As expected, we observe a positive drift when the alternative hypothesis is the underlying one, and a negative drift in the opposite

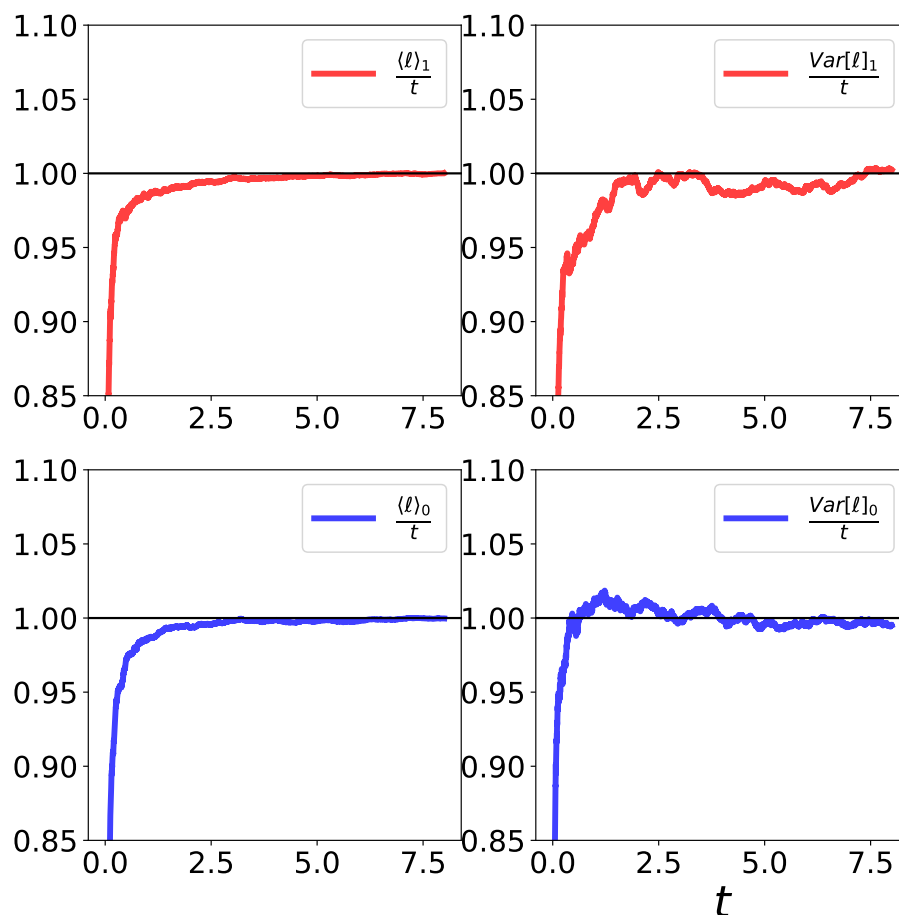


Figure 5.4: We compare the first two moments estimated over  $2 \cdot 10^4$  trajectories with their predicted theoretical value, for both hypothesis, in the damping discrimination scenario under consideration.

case. Moreover, the mean and variance of the log-likelihood (averaged out over different quantum trajectories) are compared to their theoretical values in Fig. 5.4.

Having validated the stochastic equation for the log-likelihood we can compute the average stopping time of a sequential test. Recall that the Sequential Probability Ratio Test (SPRT) stops as soon as one can guarantee a large probability of success

(what we called strong error conditions):

$$p(H_1|\mathcal{Y}_t) \geq 1 - \epsilon_0, \quad (5.30)$$

$$p(H_0|\mathcal{Y}_t) \geq 1 - \epsilon_1. \quad (5.31)$$

If none of these two conditions are satisfied, the test continues, *i.e.* further samples are requested. Following the same reasoning than standard SPRT for the *i.i.d.* case discussed in Sec. 2.5.4, these conditions can be equivalently expressed in terms of the log-likelihood ratio  $\ell$ , by fixing an *undecision region*  $\Omega = [-a_0, a_1]$ . For simplicity of the presentation we will treat the hypothesis on an equal footing, and set  $\epsilon = \epsilon_0 = \epsilon_1$  and equal prior  $\pi_0 = \pi_1 = 1/2$  (the results are easily extended to more general settings), resulting in

$$a = a_0 = -a_1 = \log \frac{1-\epsilon}{\epsilon} \quad (5.32)$$

. The test stops as soon  $\ell_t$  first hits a boundary of  $\Omega$ <sup>8</sup> and guesses for  $\hat{H}_1$  ( $\hat{H}_0$ ) when it hits the upper (lower) boundary.

Recall that the values of type I and II errors for such sequential test can be computed from the consistency equation (*i.e.* Eq. 2.225),

$$\alpha = p(\hat{H}_1|H_0) = e^{-a}(1 - \beta) \quad (5.33)$$

$$\beta = p(\hat{H}_0|H_1) = e^{-a}(1 - \alpha), \quad (5.34)$$

leading to  $\alpha = \beta = \frac{1}{1+e^a} = \epsilon$ , as it should since the error is guaranteed to be  $\epsilon$  for each single trajectory (the inequalities in strong error conditions (5.30) are saturated because the likelihood is a continuous stochastic function).

Next note that the limit of large  $a$  (long times),

$$\mathbb{E}_k[\ell_\tau] = \mathbb{E}_k\left[\int_0^\tau d\ell_t\right] = \mathbb{E}_k\left[\int_0^\infty \mathcal{I}_{t<\tau} d\ell_t\right] = \int_0^\infty \mathbb{E}_k[\mathcal{I}_{t<\tau}] \mathbb{E}_k[d\ell_t] \quad (5.35)$$

$$= \mu_k \int_0^\infty \mathbb{E}_k[\mathcal{I}_{t<\tau}] dt + \mathcal{O}(1) = \mu_k \mathbb{E}_k \left[ \int_0^\infty \mathcal{I}_{t<\tau} dt \right] + \mathcal{O}(1) = \quad (5.36)$$

$$= \mu_k \mathbb{E}_k \left[ \int_0^\tau dt \right] + \mathcal{O}(1) = \mu_k \mathbb{E}_k[\tau] + \mathcal{O}(1) \quad (5.37)$$

where we have defined the indicator function  $\mathcal{I}_C = 1$  if condition  $C$  is fulfilled and  $\mathcal{I}_C = 0$  otherwise, we have used that  $\mathcal{I}_{t<\tau} = \mathcal{I}_{\ell_t \notin \Omega}$  and  $d\ell_t$  are independent stochastic variables, and that  $\mathbb{E}_k[d\ell_t]$  becomes constant  $\mu_k$  in a finite decay time.

---

<sup>8</sup>Note that if the test did not stopped, then chances are that  $\ell$  hit such boundaries also at later times for the same trajectory.

Because of the continuity of  $\ell_t$ , there is no overshooting, and the stopping time  $\ell_\tau$  has the value at the boundary  $\ell_\tau = \pm a$ . Therefore, following the same arguments than in the *i.i.d.* case in Eq. (2.228), we get

$$\langle \ell_\tau \rangle_1 = a(1 - \beta) - a\beta = a = \frac{1-\epsilon}{\epsilon} \quad (5.38)$$

$$\langle \ell_\tau \rangle_0 = -a(1 - \alpha) + a\alpha = a = \log \frac{1-\epsilon}{\epsilon} \quad (5.39)$$

Putting the pieces together, we finally arrive to a closed expression for the mean stopping time

$$\mathbb{E}_k[\ell_\tau] = (-1)^{k+1} \mu_k \mathbb{E}_k[\tau] + \mathcal{O}(1) = (-1)^{k+1} a + \mathcal{O}(1) \quad (5.40)$$

↓

$$\mathbb{E}_k[\tau] = \frac{a}{\mu_k} + \mathcal{O}(1) = \frac{\log \frac{1-\epsilon}{\epsilon}}{\mu_k} + \mathcal{O}(1) \quad (5.41)$$

where, as mentioned above, the  $\mathcal{O}(1)$  term is due to the stabilization time of  $\mu_k(t)$  (see Fig. 5.4). Moreover, in the asymptotic limit of small errors we get

$$\mathbb{E}_k[\tau] = -\frac{\log \epsilon}{\mu_k} + \mathcal{O}(1). \quad (5.42)$$

In Fig. 5.6 we show the mean stopping time as a function of the error threshold corresponding to our numerical “experiments”, for each of the hypothesis, and also the average for both equiprobable hypotheses. We see a perfect consistency w.r.t. our analytical predictions, namely the asymptotic linear scaling of stopping time  $\langle \tau \rangle = \mu_k^{-1} \log \epsilon$ . This results agree very well with the *i.i.d.* case,  $\langle \tau \rangle_k = D(P_k || P_{k+1})^{-1} \log \epsilon$ , were now we replace the relative entropy  $\langle \ell_i \rangle_k = D(P_{k+1} || P_k)$  by its regularized version  $\lim_{t \rightarrow \infty} \langle \ell_t \rangle_k = \mu_k$ .

Note that Eq. (5.41) gives an accurate assessment of the time it takes to reach a decision (withing a guaranteed error bound), though however it only provides its average value. Thus, a real-life experiment could actually end before or after such average value. In order to characterize the size of the fluctuations we need to understand how  $\ell_t$  is distributed (see Fig. 5.6).

Having the full distribution  $P(\ell_t)$  would also allow us to compute the probability of error for deterministic strategies (that is, for experiments with a *fixed* duration  $t = T$ ), since Neyman-Pearson tells us that the log-likelihood ratio statistic is all we need to optimally guess true Hypothesis (see Sec. 2.5.1).

The full probability distribution of  $\ell_t$  can be seen in Fig. 5.7 where we give the histograms for  $\ell_t$  at different time slices under each of the hypothesis. We note

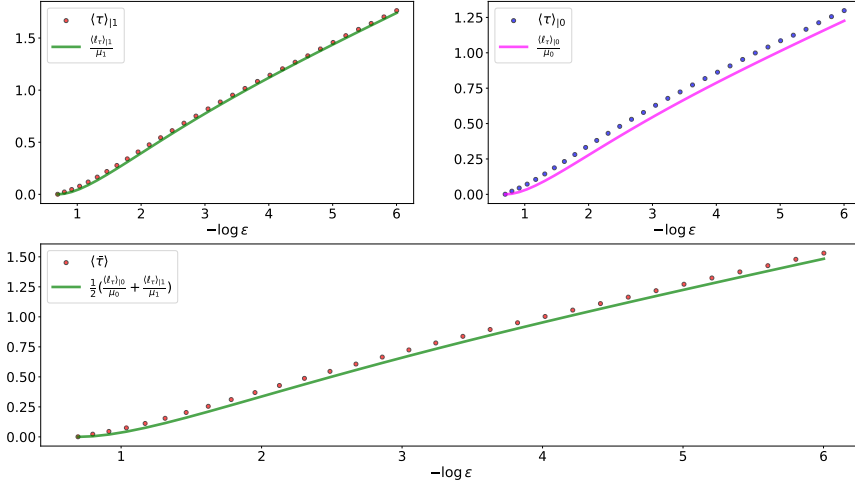


Figure 5.5: We compare the average stopping time (under each hypothesis and also its mean value) with our analytical curve given in Eq. 5.40.

that for initial times, the distributions considerably overlap as expected —with a low number of samples it is hard to distinguish the underlying model. As time progresses the two distributions separate from each other at a linear pace. We also observe that the distributions seem to be well approximated by a Gaussian. Indeed, from Eq. 5.18, and the shape of the histograms in Fig. 5.2 we might be tempted to say that  $\Delta \bar{\mathbf{r}}_t$  concentrates around its mean  $\mu_k$ , and tends to a Gaussian with all cumulants of order 3 or higher scaling sub-linearly with  $t$ . Under such assumption, we could replace  $\Delta \bar{\mathbf{r}}_t$  by  $\mu_k$  in the dynamical evolution, and thus the process would reduce to an Ornstein-Uhlenbeck process, discussed in Sec. 2.4.4. The solution of such process is a Gaussian distribution whose mean is drifted linearly in time and diffuses with a variance also proportional to  $t$ . As we will see in more detail below, Sect. 5.3.2, this intuition is only partially true.

### 5.3.2 To be or not to be Gaussian

The numerics and the heuristic argument given above seems to support that the distribution  $P_k(\ell_t)$  is indeed Gaussian. We begin this Section by providing a simple no-go theorem that shows  $P_k(\ell_t)$  cannot be Gaussian. We will then revisit again the heuristic argument and compute several quantities of interest under this, strictly

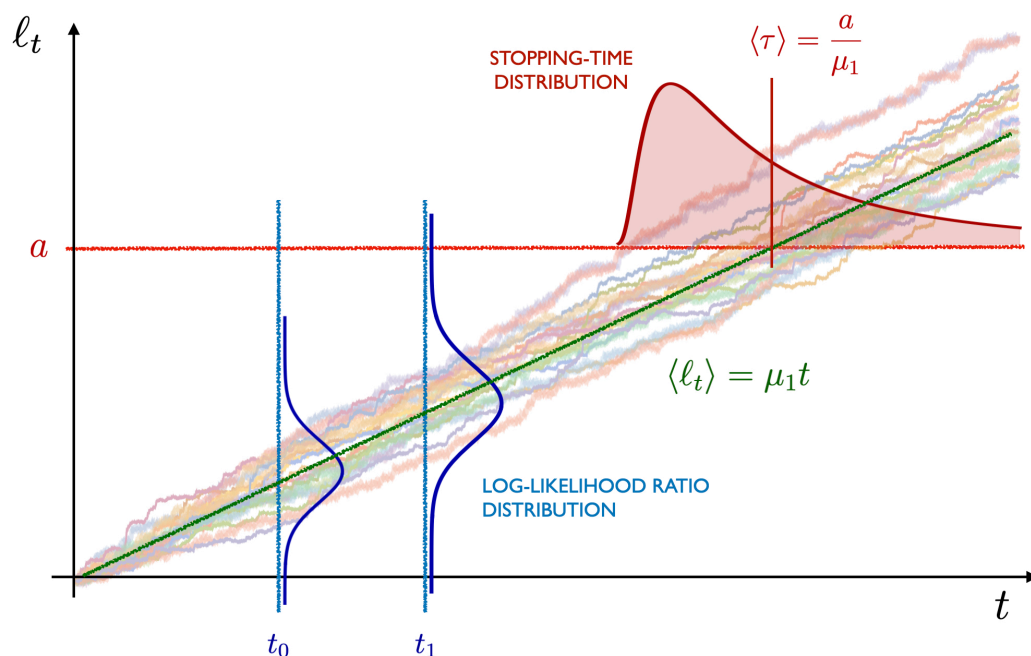


Figure 5.6: We illustrate how the stopping time distribution arises from the random trajectories of  $\ell_t$ ; here we show some realisations of the  $\ell$  process, along with some time-slices at  $t_0$  and  $t_1$  shown in blue together with the corresponding distributions  $P(\ell_t)$ . The horizontal line in red corresponds to the fixed threshold  $\ell = a$ , leading to an arrival time distribution  $P(\tau)$ . Such distribution appears as a consequence of a difference in the arrival times for trajectories of  $\ell_t$ , which are governed by a Gaussian distribution that is drifted and diffuses in time.

speaking, wrong Gaussian assumption. We will show that the bulk of the distribution is correctly described by a Gaussian, but it completely fails to describe the tails (or large deviations).

*No-go theorem for Gaussians:* We will start by showing that assuming Gaussianity in both of the hypotheses implies  $\mathbb{P}_1(\ell_t = x) = \mathbb{P}_0(\ell_t = -x)$ , that is, the distributions are mirrored versions of each-other. To see this, we note that the following relation between the characteristic functions holds true:

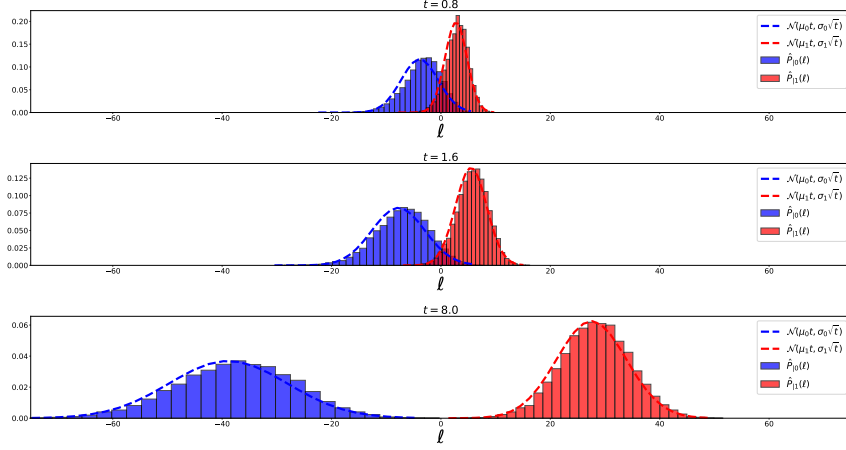


Figure 5.7: We show the distributions for the log-likelihood ratio at different times, for both hypothesis, computed over  $2 \cdot 10^4$  for the damping discrimination problem considered in this Section. Here, we have also plotted the distributions associated to the Gaussian model discussed in Sec.5.3.2.

$$\begin{aligned} \chi_\ell(q-1) &\equiv \mathbb{E}_1[e^{(q-1)\ell_t}] = \int \mathcal{D}\mathcal{Y}_t P_1(\mathcal{Y}_t) e^{(q-1)\ell_t} = \int \mathcal{D}\mathcal{Y}_t P_0(\mathcal{Y}_t) e^{\ell_t} e^{(q-1)\ell_t} = \\ &= \int \mathcal{D}\mathcal{Y}_t P_0(\mathcal{Y}_t) e^{q\ell_t} = \mathbb{E}_0[e^{q\ell_t}] \equiv \chi_\ell(q). \end{aligned} \quad (5.43)$$

where we have used the fact that  $\ell(\mathcal{Y}_t) = \log \frac{P_1(\mathcal{Y}_t)}{P_0(\mathcal{Y}_t)}$ , to write  $e^\ell = \frac{P_1(\mathcal{Y}_t)}{P_0(\mathcal{Y}_t)}$  to perform a change of measure. Now recalling that the characteristic function for a Gaussian distribution to be  $\chi(q) = e^{(q\mu + \frac{q^2}{2}\sigma^2)}$  we obtain

$$(q-1)\mu_1 + \frac{(q-1)^2}{2}\sigma_1^2 = -q\mu_0 + \frac{q^2}{2}\sigma_0^2. \quad (5.44)$$

Choosing now  $q = 0$ ,  $q = 1/2$  and  $q = 1$ , we get the following identities:  $(-)^{k+1}2\mu_k = \sigma_k^2$  and  $2\mu_1 + \sigma_1^2 = -2\mu_0 + \sigma_0^2$ , which combined together give  $\mu_1 = -\mu_0$  and  $\sigma_0^2 = \sigma_1^2 = 2\mu_1$ , showing that  $\mathbb{P}_1(\ell_t = x) = \mathbb{P}_0(\ell_t = -x)$ . The latter condition that cannot be fulfilled by the model we study, since  $\mu_0 \neq -\mu_1$ , as we can explicitly check from Eq. 5.28 and clearly appreciate from Fig.5.7. *End of No-Go Theorem*

In the remaining of this Section we will see how far can we go analyzing the hypothesis testing scenario assuming the Gaussianity of  $\mathbb{P}_0(\ell_t)$ , particularly because this allow us to get expressions for the weak errors under the deterministic setting, and to fully characterize the stopping-time probability distribution for the sequential state.

The central-limit theorem (CLT) guarantees that for *i.i.d* random variables, the sample average will asymptotically normally distributed, centered in the mean value of the distribution and with a standard deviation that decreases as  $\sim 1/\sqrt{N}$ . If we consider the process

$$d\ell = (-1)^{k+1}\mu_k dt + \sigma_k dW_t, \quad (5.45)$$

then we can readily find its associated probability distribution by recalling our discussion on OU-like processes from the previous Section and Sec.2.4.4. Such distribution reads

$$P_k(\ell) = \frac{1}{\sqrt{2\pi\sigma_k^2 t}} e^{-\frac{(\ell + (-1)^{k+1}\mu_k t)^2}{2\sigma_k^2 t}}. \quad (5.46)$$

From Eq. 5.45 and the CLT we expect the normalized variable  $\frac{\ell_G}{t}$  to concentrate around its mean  $(-1)^{k+1}\mu_k$  as  $t \rightarrow \infty$ , and *approximating* its distribution by a Gaussian function around the mean-value should actually be *always valid*, at least in the long-time regime. That is, we do expect the Gaussian approximation to hold under Eq. 5.18 around the mean-value of  $\ell$ : *in the long-time regime, the probability distribution should look Gaussian in the bulk*. However, we do know that the Gaussian model should not hold, and we have already provided evidence of this in Fig. 5.7 and Fig. 5.4: while the distributions *do* look Gaussian around their mean-value, the tails of the distribution are clearly not in correspondance with the Gaussian model.

In Sec. 2.5.1 we have discussed two different approaches to the hypothesis testing scenario, namely the deterministic and the sequential one. We will now analyze such strategies for the system under consideration, when assuming that  $\ell$  is normally-distributed.

*Error probabilities for deterministic strategies.* The deterministic tests consists on computing the value of  $\ell(\mathcal{Y}_t)$ , for a fixed value of  $t$ , and making a decision for the undelying hypothesis based on its value. In this context, we will pay attention to the symmetric error, and thus the decision countour will be set at  $\ell = 0$ . Moreover, we will assume that the hypothesis are equally likely. Based on the log-likelihood



ratio, the weak errors are defined as

$$\begin{aligned}\alpha &= P(\hat{h}_1|h_0) = \int_0^\infty P_0(\ell) d\ell \\ \beta &= P(\hat{h}_0|h_1) = \int_{-\infty}^0 P_1(\ell) d\ell,\end{aligned}\tag{5.47}$$

where for compactness of notation we have defined  $\ell = \ell_t$  at the given end time  $t$  of the experiment. Under the Gaussian model for  $P_k(\ell)$ , the errors can easily be computed and we get

$$\begin{aligned}\beta &= \frac{1}{2} \left( 1 + \operatorname{Erf} \left( \frac{\mu_0 t}{\sqrt{2\sigma_0^2 t}} \right) \right) \xrightarrow{t \rightarrow \infty} \sqrt{\frac{\sigma_0^2}{2\mu_0 t}} e^{-\frac{\mu_0^2 t}{2\sigma_0^2}}, \\ \alpha &= \frac{1}{2} \left( 1 + \operatorname{Erf} \left( \frac{\mu_1 t}{\sqrt{2\sigma_1^2 t}} \right) \right) \xrightarrow{t \rightarrow \infty} \sqrt{\frac{\sigma_1^2}{2\mu_1 t}} e^{-\frac{\mu_1^2 t}{2\sigma_1^2}},\end{aligned}\tag{5.48}$$

In Fig. 5.8 we compare these expressions with the numerical estimation of both type of errors, and for the symmetric one. Note that since the errors are computed from large deviation statistics, many quantum trajectories need to be simulated in order to have enough statistics to compute such quantities. As expected, the agreement between the Gaussian model and our numerics deprecates, particularly for long-times where we can differentiate the tails from the bulk.

*Stopping time distribution for sequential hypothesis testing.* Our result in Eq.(5.41) already gives the average stopping-time. However, we note that if  $P_k(\ell)$  was a Gaussian distribution as per Eq. 5.46 then we would also fall in a new contradiction, since  $\mu_k$  can only coincide with the relative entropy between the two Gaussian distribution only if  $\mu_0 = \mu_1$ <sup>9</sup>.

However, if we ignored all of the issues pointed out above – for instance, assume that the drifts do only differ slightly – then an expression for the full stopping-time

<sup>9</sup>Recall that the relative entropy between two Gaussian distributions  $\mathcal{N}_1(\mu_1, \sigma_1)$  and  $\mathcal{N}_0(\mu_0, \sigma_0)$  is given by

$$D(\mathcal{N}_1||\mathcal{N}_0) = \log \frac{\sigma_0}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_0)^2}{2\sigma_0^2} - \frac{1}{2},\tag{5.49}$$

which for the case under consideration ( $\sigma_k = \sqrt{2\mu_k}$ ) reduces to

$$D(\mathcal{N}_1||\mathcal{N}_0) = \frac{1}{2} \log \frac{\mu_0}{\mu_1} + \frac{2\mu_1 + (\mu_1 - \mu_0)^2}{4\mu_0} - \frac{1}{2}.\tag{5.50}$$

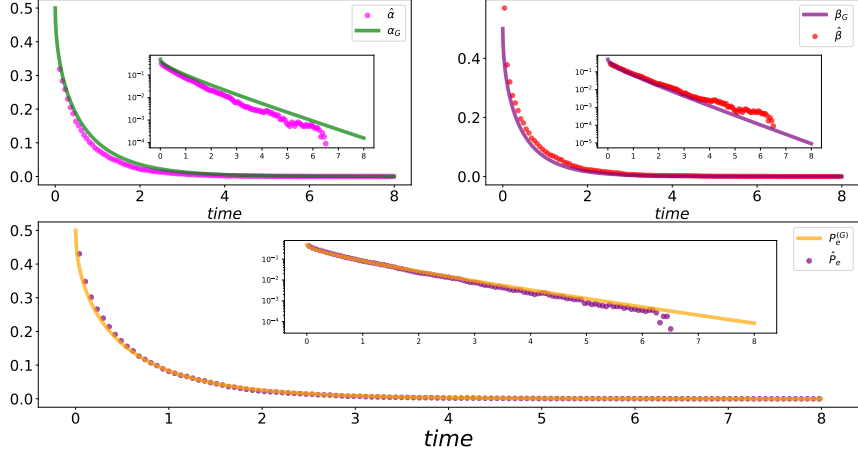


Figure 5.8: We show the weak errors computed for the deterministic test, and compare them with the expressions of Eq 5.48, which are obtained by assuming that the distribution  $P_k(\ell)$  is Gaussian.

probability probability distribution can be obtained. Thus, we will consider an stochastic process

$$d\ell_t = \mu dt + \sigma dW_t, \tag{5.51}$$

whose probability distribution at time  $t$  will be denoted by  $P(\ell_t)$ , and is understood as the probability that the state takes the value  $\ell$  at time  $t$ .

Assuming that the SPRT eventually stops, *i.e.*  $\ell$  eventually exits the undecision region  $\Omega$ , then the stopping-time probability distribution  $P(\tau = t)$ , *i.e.* the probability that the stopping time takes a value of  $t$  can be written in terms of its cumulative distribution as

$$P(\tau = t) = -\frac{d}{dt}P(\tau > t) = -\frac{d}{dt} \int_t^\infty P(\tau) d\tau. \tag{5.52}$$

Now, the probability that the stopping-time is larger than  $t$  is given by the total probability of finding the stochastic process  $\ell$ , at times smaller or equal to  $t$ , inside the undecision region  $\Omega$ :

$$P(\tau > t) = \int_{-a_0}^{a_1} \mathcal{P}(\ell, t) d\ell. \tag{5.53}$$

Note our slight abuse of notation: we understand  $\int d\ell$  as an integral over measurement record realizations  $\mathcal{Y}_t$  that lead to  $\ell(\mathcal{Y}_t) = \ell$ . Moreover, in the last integral we

assume that the values of  $(\ell, t)$  we are integrated over have not reached the decision boundary  $\Omega$  for times smaller than  $t$ , thus  $\mathcal{P}(\ell, t)$  is deemed *survival probability*; due to this fact we will use the  $\mathcal{P}$  notation for this process. This quantity can be obtained as a solution of the Fokker-Planck equation (plus absorbing boundary conditions), as per

$$\partial_t \mathcal{P}(\ell, t) = -\mu \partial_\ell \mathcal{P}(\ell, t) + \frac{\sigma^2}{2} \partial_\ell^2 \mathcal{P}(\ell, t), \quad (5.54)$$

$$\mathcal{P}((\ell, t) \notin \Omega) = 0. \quad (5.55)$$

The initial condition for this problem is given by  $\int_\Omega P(\ell, 0) d\ell = 1$ . Combining the above equations and using the Leibniz integral rule, we obtain

$$P(t) = \frac{\sigma^2}{2} \partial_\ell \mathcal{P}(\ell_t) \Big|_{\ell=a_0}^{a_1} + \frac{\sigma^2}{2} \int_{a_0(t)}^{a_1(t)} \left( \dot{a}_0 \partial_{a_0(t)} + \dot{a}_1 \partial_{a_1(t)} \right) \mathcal{P}(\ell, t) d\ell, \quad (5.56)$$

where we have account for the possibility that the decision boundaries change in time. In turn, if we now change of variables as  $l_t = \ell_t - \mu t$ , then  $\Omega$  becomes time-depentend as  $\Omega(t) = [a_0(t), a_1(t)]$  with  $a_i(t) = a_i + \mu t$ . The Fokker-Planck equation associated with the process  $l$  is symmetric under reflection operations of the form  $l \rightarrow 2\beta - l$  with  $\beta \in \mathbb{R}$ , allowing for the use of the image charge method to find an explicit solution in the form of an infinite series for the survival probability  $P(\ell, t)$ :

$$\begin{aligned} \mathcal{P}(\ell, t) &= \bar{P}(\ell, t) \\ &+ \sum_{n=1}^{\infty} \left\{ \bar{P}(2n(a_1(t) - a_0(t)) + \ell, t) - \bar{P}(2na_0(t) - 2(n-1)a_1(t) - \ell, t) \right. \\ &\left. + \bar{P}(2n(a_0(t) - a_1(t)) + \ell, t) - \bar{P}(2na_1(t) - 2(n-1)a_0(t) - \ell, t) \right\} \end{aligned} \quad (5.57)$$

with

$$\bar{P}(\ell, t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{\ell^2}{2\sigma^2 t}} \quad (5.58)$$

is the solution of the differential problem  $\partial_t \mathcal{P}(l, t) = \frac{\sigma^2}{2} \partial_l^2 \mathcal{P}(l, t)$  with initial condition  $\mathcal{P}(l, 0) = \delta(l)$ . Substituting Eq. (5.57) in Eq. (5.52), and after some (tedious) algebra and keeping only the lowest-order terms associated with the limit  $A = \min(|a_0|, |a_1|) \rightarrow \infty$ , *i.e.* very low errors which is also associated to long-times, we get that the stopping-time probability distribution is an inverse-gaussian

(or Wald distribution):

$$P(t) = \frac{|a_k|}{t^{3/2} \sqrt{2\pi\sigma_k^2}} e^{-\frac{(a_k + (-1)^{k+1} \mu t)^2}{2\sigma_k^2 t}}, \quad (5.59)$$

where we have taken into account that hypothesis  $k$  is the underlying one when taking the limit.

While the solution for the process in Eq. 5.51 corresponds to a drifted and diffused Gaussian distribution whose expression is explicated in Eq. 5.46, we are now interested in the probability that a trajectory drawn from such Gaussian distribution hits the decision boundary  $\Omega = [-a, a]$ . Since the mean of the  $\ell$  increases as  $\mu t$ , and its standard deviation as  $\sigma\sqrt{t}$ , the resulting projection of such probability into the line  $\ell = a$  (for hypothesis  $k = 1$ ) leads to a tilted distribution, as illustrated in Fig. 5.6 which is known as the Wald distribution, and reads

$$P(\tau) = \frac{a}{\sqrt{2\pi\sigma^2 t^3}} e^{-\frac{(a - \mu t)^2}{2\sigma^2 t}}, \quad (5.60)$$

where  $\tau$  stands for the time at which the  $\ell_\tau = a$ , *i.e.* the stopping time. We note that a similar distribution is obtained for  $k = 0$ . Moreover, the mean of this distribution is

$$\langle \tau \rangle = \frac{a}{\mu}, \quad (5.61)$$

and we note that this coincides with the Wald identity in Eq. 5.41 (see also Eq. 2.229) for large values of  $a \sim \log \epsilon$ .

From our discussion regarding the validity of the Gaussian distribution, we do expect the Wald distribution to match our numerics close to the bulk, and this is shown in Fig. 5.9. While the agreement is particularly good for the bulk of the distributions, we note that the tails are quite off, particularly for  $k = 0$ , for which the drift is higher.

### 5.3.3 Joint system evolution as an OU-process

We have already shown that  $\mathbb{E}_k \left[ \|\Delta \bar{\mathbf{r}}_t\|^2 \right] = \mu_k t$ , and we will next outline an alternative method to compute its variance, which scales lineary with time as  $\text{Var}_k \left[ \|\Delta \bar{\mathbf{r}}_t\|^2 \right] = \sigma_k^2 t$ . The fact that that  $\ell$  is not compatible with a Gaussian distribution can be understood from the back-action term due to the measurement record, which couples the dynamical equations for the two hypothesis. Having this in mind, it is convenient to treat the problem in an extended vector space where

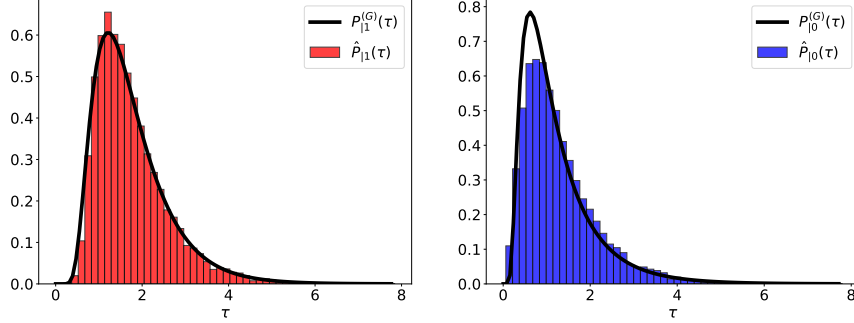


Figure 5.9: We show the stopping time probability distributions, under both hypothesis, obtained numerically by carrying out the SPRT on 20K trajectories for the damping discrimination problem under consideration. Here, we have chosen a value of  $a = 6$ , *i.e.*  $\epsilon \sim 0.0025$ .)

the state of the system is defined as  $\mathbf{X}_t = (\bar{\mathbf{r}}_0(t), \bar{\mathbf{r}}_1(t))^T$  and  $\mathcal{S}t = \Sigma_0(t) \oplus \Sigma_1(t)$  whose evolution is given by

$$\begin{aligned} d\mathbf{X}_t &= (\mathcal{A} - \chi(\mathcal{S}t)\Pi_k\mathcal{C})\mathbf{X}_tdt + \chi(\Sigma)d\mathbf{w}_t \\ \dot{\mathcal{S}}_t &= \mathcal{A}\mathcal{S}_t + \mathcal{S}_t\mathcal{A} + \mathcal{D} - \chi(\mathcal{S}_t)\chi(\mathcal{S}_t)^T \end{aligned} \quad (5.62)$$

with  $\mathcal{A} = A_0 \oplus A_1$ ,  $\mathcal{C} = C \oplus C$  and  $\mathcal{D} = D_0 \oplus D_1$ ,  $\chi(\mathcal{S}_t) := \mathcal{S}_t\mathcal{C}^T + \tilde{\Gamma}$ ,  $\tilde{\Gamma} = \Gamma_0 \oplus \Gamma_1$  and

$$\Pi_0 = \begin{pmatrix} 0 & 0 \\ \mathbf{1} & -\mathbf{1} \end{pmatrix}, \quad \Pi_1 = \begin{pmatrix} \mathbf{1} & -\mathbf{1} \\ 0 & 0 \end{pmatrix}$$

with  $k$  denoting the hypothesis under which the signal  $\mathcal{Y}_t$  is generated (*i.e.*  $k = 0, 1$ ). Under this notation, the evolution for the log-likelihood ratio reads:

$$d\ell(\mathcal{Y}_{t|\theta_k}, t) = \frac{(-)^{k+1}}{2} |\Delta^T \mathcal{C} \mathbf{X}_t|^2 dt + \Delta^T \mathcal{C} \mathbf{X}_t d\mathbf{w}_t \quad (5.63)$$

with  $\Delta^T = (1, -1)^T$ . Under the assumption that the evolution for the covariance admits a stationary solution  $\mathcal{S}_{st} = \sigma_0 \oplus \sigma_1$ , the probability distribution of  $\mathbf{X}_t$  is asymptotically given by:

$$P_{st,k}(\mathbf{X}) = \frac{1}{(2\pi)^{n/2} \det[\boldsymbol{\omega}]^{1/2}} \exp\left(-\frac{1}{2} \mathbf{X}^T \boldsymbol{\omega}^{-1} \mathbf{X}\right) \quad (5.64)$$

where  $\boldsymbol{\omega}$  is the solution of the Lyapunov equation  $(\mathcal{A} - \chi(\Sigma_t)\Pi_k\mathcal{C})\boldsymbol{\omega} + \boldsymbol{\omega}(\mathcal{A} - \chi(\mathcal{S}_{st})\Pi_k\mathcal{C})^T = \chi(\mathcal{S}_{st})\chi(\mathcal{S}_{st})^T$ <sup>10</sup> Using this notation, we can express the asymptotic behaviour for the expected value of  $\ell$  as

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{\mathbb{E}_k[d\ell(\mathcal{Y}_t)]}{t} &= \frac{(-)^{k+1}}{2} \mathbb{E}[|\Delta^T \mathcal{C} \mathbf{X}|^2] \\ &= \frac{(-)^{k+1}}{2} \text{Tr}[\mathcal{C}^T \Delta \Delta^T \mathcal{C} \boldsymbol{\omega}]. \end{aligned} \quad (5.65)$$

From here, we can insert the formal solution of  $\mathbf{X}_t$  to get

$$\Delta^T \mathcal{C} \mathbf{X}_t = \Delta^T \mathcal{C} e^{(\mathcal{A} - \chi(\mathcal{S}_{st})\Pi_k\mathcal{C})(t-t_0)} \mathbf{X}_{t_0} + \chi(\mathcal{S}_{st}) \int_{t_0}^t \Delta^T \mathcal{C} e^{(\mathcal{A} - \chi(\mathcal{S}_{st})\Pi_k\mathcal{C})(t-\tau)} d\mathbf{W}_\tau. \quad (5.66)$$

Finally, by defining

$$\begin{aligned} g(t - t_0) &= \Delta^T \mathcal{C} e^{(\mathcal{A} - \chi(\Sigma_{st})\Pi_k\mathcal{C})(t-t_0)} \mathbf{X}_{t_0} \\ f(t - s) &= \Delta^T \mathcal{C} e^{(\mathcal{A} - \chi(\Sigma_{st})\Pi_k\mathcal{C})(t-s)} \chi(\Sigma_{st}) \mathbb{Q} \end{aligned} \quad (5.67)$$

with  $\mathbb{Q} = (1, 1)^T$ , we arrive to an asymptotic expression for the mean and covariance of  $\ell$ :

$$\begin{aligned} \mathbb{E}_k[\ell_t] &= \lim_{t \rightarrow \infty} \frac{(-)^{k+1}}{2} \left\{ \left( \int_{t_0}^t g(\tau - t_0) d\tau \right)^2 - \int_{t_0}^t d\tau \int_{t_0}^\tau f(\tau - s)^2 ds \right\} \\ \mathbb{E}_k[\ell_t^2] - \mathbb{E}_k[\ell_t]^2 &= \lim_{t_0 \rightarrow -\infty} \int_{t_0}^t d\tau \int_{t_0}^\tau f(\tau - s)^2 ds \\ &\quad - 2 \int_{t_0}^t d\tau_1 \int_{t_0}^{\tau_1} d\tau_2 f(\tau_1 - \tau_2) \int_{t_0}^{\tau_2} ds f(\tau_1 - s) f(\tau_2 - s) \\ &\quad + \int_{t_0}^t d\tau_1 \int_{t_0}^{\tau_1} d\tau_2 \left( \int_{t_0}^{\tau_2} f(\tau_1 - s) f(\tau_2 - s) ds \right)^2. \end{aligned}$$

Overall, this constitutes an alternative approach to compute the first two moments of  $\ell$ , and for the damping discrimination case-study we can readily check that the above expression matches Eq. 5.28. Moreover, in Fig. 5.4 we have compared our numerics with the analytical expressions that we get using these formulae for the damping discrimination case, for both the average and variance of  $\ell$ . As we can see,

<sup>10</sup>This equation arises from a generalization of the OU-process discussed in Sec. 2.4.4 to multi-variable systems, and we refer the interested reader to Ref. [Gar04] for more details.

after a transient time, a good agreement can be observed. We remark that numerical errors associated to the integration routine lead to considerable computational-resources usage. Here, we have used the Python package `sdeint` [Abu17] which implements an order 1 strong-convergence stochastic Runge-Kutta method, introduced in Ref. [Rö9]<sup>11</sup>

### 5.3.4 Deterministic vs Sequential

While we have discussed the validity of the Gaussian model for  $\ell$ , let us note that the expected time required to reach a certain threshold for the error probability is 4 times lower in favour of the SPRT. This can be seen from the asymptotic behaviour of the asymmetric errors in Eq. 5.48, *i.e.* we obtain  $\log \alpha_k \sim -\frac{\mu_k t}{4}$ , with  $\alpha_0 = \alpha$  and  $\alpha_1 = \beta$ .

For the general case, we do expect an advantage in favour of the sequential, but (for the moment) we rely on numerical methods to compute the advantage. In Fig. 5.10 we compare the average time required by the SPRT to reach a given error threshold for the symmetric error probability, as given by Eq. 2.225, with that of deterministic test. For the latter, we compute the empirical probability of error for several times, and use interpolation to find the time at which a given mean error is reached. In both cases, we compare with the theoretical models, namely the rigorous results for the sequential case, and the calculations based on the Gaussian approximation for the deterministic case.

To sum up, we have introduced a sequential hypothesis testing approach to the discrimination between two physical models in continuously-monitored systems, showing a clear advantage of the latter as compared to the deterministic test. While we have focused on the damping-rate discrimination case, since some analytical insight can be gained, let us stress that the sequential approach is ought to be taken not only for a different set of parameters in the (quantum) Gaussian model for the optomechanical system, but also for general stochastic evolutions.

#### Frequency discrimination

To complement our numerical study of the sequential hypothesis test in continuously-monitored systems, we here study the case of frequency discrimination. In Fig. 5.11 we show how the probability distribution of  $\ell$  looks for the two hypothesis, each associated to a different value of the mechanical-mode frequency. We observe that,

---

<sup>11</sup>Whereas strong convergence stands for the average error committed per time-step, the *weak* convergence refers to the error of the mean values obtained by averaging out the integrated trajectories. The order is measured in terms of  $dt$ , for instance Euler algorithm has order  $\frac{1}{2}$ .

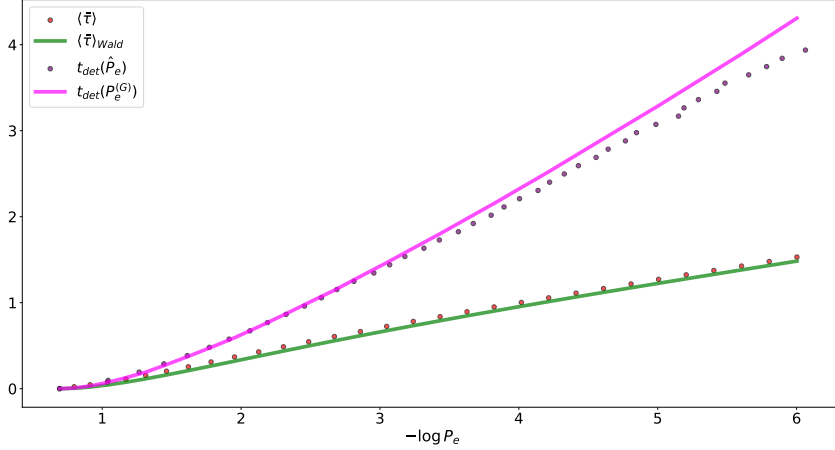


Figure 5.10: We compare the performance of the sequential test against the deterministic one, in terms of the average time that takes the test to reach a certain error threshold. In the deterministic case, we invert the Gaussian model for the symmetric error, to get the desired time. This was done by interpolating the time that the superposition of Lorentzian distributions requires to value the pre-defined error threshold  $P_e$ , as given when fixing  $\epsilon$  in Eq. 2.225 and computing  $\frac{1}{2}(\alpha + \beta)$ . Such a Gaussian model is also compared with our numerics, and in this plot we show the numerical value of the symmetric error probability in the deterministic case ( $x$  axis), at the different times it was computed for ( $y$  axis). For the sequential test, we show the average stopping time as a function of  $P_e$ , and compare this with the quantity that Wald identity predicts.

for the times under consideration, the distributions are clearly not Gaussian. Note that here we are not analyzing the data in the rotating-frame, since the frequency is unknown. One of the consequences of this fact is that a transcendental equation for the stationary value of the covariance matrix arises in Eq. 5.13, and thus little can be done analytically-wise when aiming to obtain the values of the drift and diffusion for  $\ell_t$ .

Following a similar analysis than in the damping discrimination case, in Fig. 5.12 we compare the sequential and deterministic strategies' performance, showing also a clear advantage in favour of the former.



5.3. SEQUENTIAL TESTING & GAUSSIAN SYSTEMS

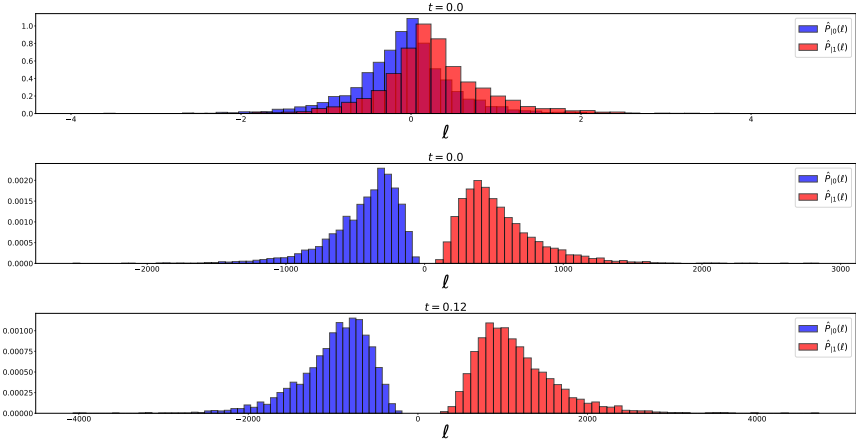


Figure 5.11: We show the distributions of the  $\ell$  for the frequency discrimination problem, where  $\omega_0 = 10^4 Hz$  and  $\omega_1 = 1.1 \cdot 10^4 Hz$ ,  $\eta = n = 1$  and  $\kappa = 1000 Hz$ . The reason we take such a high value of  $\kappa$  is to aid with the computing resources and time simulation. The results are shown for  $10^4$  quantum trajectories.

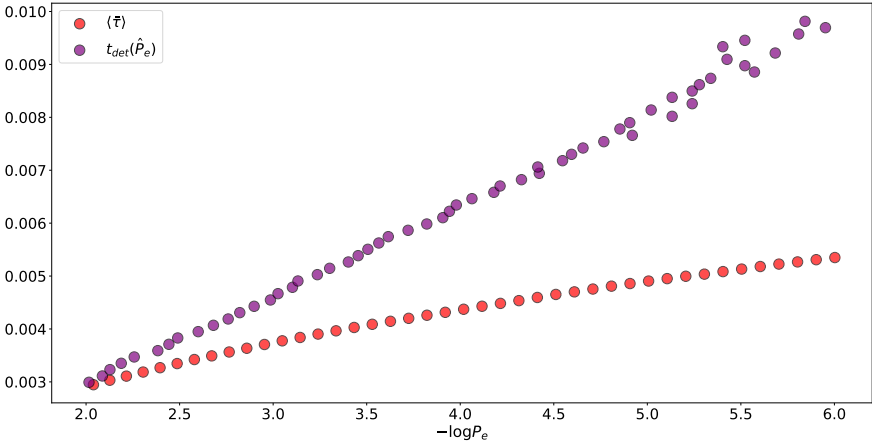


Figure 5.12: Here we show the same comparison between the performance of both discrimination strategies (sequential vs. deterministic), similarly to Fig. 5.10, but for the frequency discrimination case.

## 5.4 Parameter estimation & continuously-monitoring

In the previous Section we have discussed how to discriminate between two different physical models, and provided strong numerical evidence in favour of the sequential strategy. Moreover, in the examples we have considered the damping discrimination case, and the analytical results we obtained hinge on the fact that the mechanical-mode's frequency is known. By the end of the previous Section we discussed the case in which such value is discerned between two possible candidates. However, in general we need to estimate this quantity, what reduces to a different problem in statistical-inference, namely the (classical) parameter estimation discussed in Sec. 2.5.5.

In this Section we will discuss a work in progress, which is about estimating a parameter encoded in a physical system that is being continuously monitored. Our preliminary results consist on implementing the maximum-likelihood with the automatic-differentiation machinery, and the ultimate goal of this (ongoing) project is that of machine-learning the *dynamics* of a continuously-monitored system. By this we understand asking the machine-learning to provide a set of dynamical equations based solely on the observed data, similar to the proposals of Ref. [BPK16] (though taking noise and back-action terms into account). The latter can be useful when aiming to model external signals, and we will here study simple examples that enter in the evolutino as external forces *e.g.* drives.

### 5.4.1 The spectral power and the Lorentizan fit

To infer the value of the frequency, it is customary to fit a Lorentizan function to the spectral power of the signal. The power spectrum of a stationary stochastic process, say a quadrature of a mode,  $x_t$ , is the Fourier transformed of its autocorrelation function, *i.e.*  $S(\omega) = \int_{-\infty}^{\infty} dt' e^{i\omega t'} \langle x_t x_{t+t'} \rangle$ . For a damped oscillator system as the one described in Eq. 5.13, such quantity is given by a Lorentzian peaked at the mechanical-mode frequency as per

$$S(\omega) \sim A \frac{\gamma}{(\omega - \omega_m)^2 + \frac{\gamma}{2}}, \quad (5.68)$$

where  $A$  is a constant factor related to bath's temperature [CDG+10]. Here, we are interested in how to estimate a physical parameter out of a *single* quantum trajectory, whose access is granted via the measurement record. Whence, we consider a single process realization (and not the average), and replace the hidden state  $\mathbf{r}_t$  by the measurement record  $dy_t$ , given by the homodyne measurement

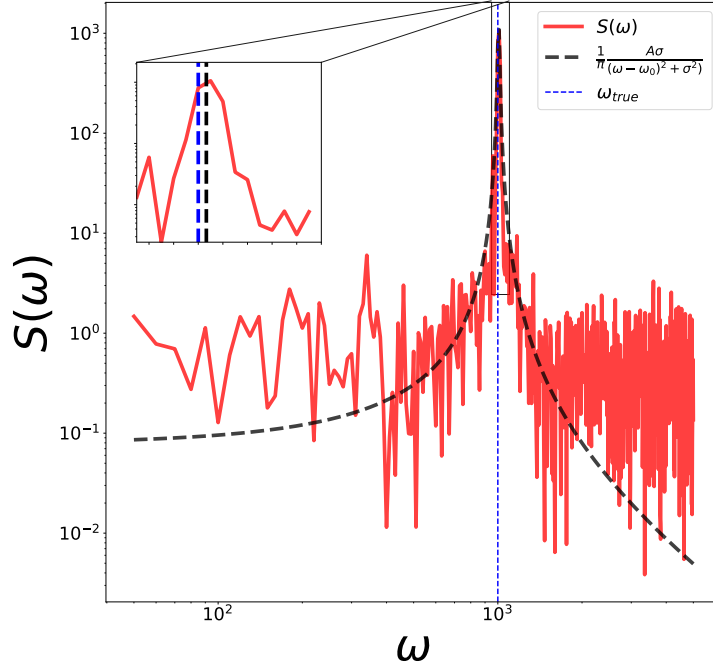


Figure 5.13: We show the power spectra for a realization of the stochastic process, and the corresponding Lorentzian fit.

that projects onto the  $q$  quadrature of the hidden state (*i.e.* we are not considering the rotating-frame discussed in the previously-discussed damping discrimination case). In this context, the spectral power becomes a random-variable itself, and similar to our discussion about the concentration of  $\ell_t$  around its mean, we do expect the former quantity to be peaked at the mechanical-mode frequency in a sufficiently-large time. This is shown in the left-panel of Fig. 5.13, accompanied by a Lorentzian fit which — see the inset — falls into a small error when compared with the underlying true value of  $\omega_m$ . Such error contributes to the estimator’s variance, and we here estimate this quantity by averaging up the squared discrepancies over several quantum trajectories.

### 5.4.2 Evolution of the Fisher information

As discussed in Sec. 2.5.5, when estimating the value of a classical parameter, the variance of an estimator is lower bounded by the Fisher information, via the Cramér-Rao bound. Here, recall that given  $P(\mathcal{Y}_t; \theta)$  the probability of  $\mathcal{Y}_t$  given a known value of  $\theta$ , and thus the log-likelihood is defined as  $\lambda(\mathcal{Y}_t; \theta) = \log P(\mathcal{Y}_t; \theta)$ . Then, the Fisher information is given by

$$I_t(\theta) = \langle (\partial_\theta \lambda(\mathcal{Y}_t; \theta))^2 \rangle = \langle -\partial_\theta^2 \lambda(\mathcal{Y}_t, \theta) \rangle \quad (5.69)$$

For the Gaussian system in Eq. 5.13, we can readily find a dynamical equation for  $\lambda_t \equiv \lambda(\mathcal{Y}_t, \theta)$ . In turn, by defining  $\mathbf{u}_t = C\mathbf{r}_t$ , we can write the evolution of  $\partial_\theta \lambda_t$ , obtained by taking the derivative w.r.t.  $\theta$  in Eq. 5.17, as

$$d\partial_\theta \lambda_t = -\mathbf{u}_t \cdot \partial_\theta \mathbf{u}_t dt + (\partial_\theta) \cdot d\mathbf{y}_t. \quad (5.70)$$

We can readily compute the second derivative as

$$d\partial_\theta^2 \lambda_t = -\|\partial_\theta \mathbf{u}_t\|^2 dt - \mathbf{u}_t \cdot (\partial_\theta^2 \mathbf{u}_t) dt + (\partial_\theta^2 \mathbf{u}_t) \cdot d\mathbf{y}_t \quad (5.71)$$

$$= -\|\partial_\theta \mathbf{u}_t\|^2 dt + (\partial_\theta^2 \mathbf{u}_t) \cdot d\mathbf{W}_t, \quad (5.72)$$

where in the last line we have used the expression for the measurement outcome  $d\mathbf{y}_t$  in terms of the Wiener noise. In turn, the last expression with the Wiener noise turns useful, since we can readily see that the expected value  $\mathbb{E}[(\partial_\theta^2 \mathbf{u}_t) \cdot d\mathbf{W}_t]$  vanishes<sup>12</sup>. Thus, the Fisher info is obtained as

$$I_t(\theta) = -\mathbb{E}[\partial_\theta^2 \lambda_t] = -\int_0^t \mathbb{E}[d\partial_\theta^2 \lambda_t], \quad (5.73)$$

and we only need to integrate the values  $(\bar{\mathbf{r}}_t, \partial_\theta \bar{\mathbf{r}}_t)$  in order to get it.

<sup>12</sup>To see this, consider the solution

$$\begin{aligned} \mathbf{r}_t &= \chi(\Sigma) \int_0^t e^{-A(t-t')} dW_{t'} \\ \partial_\theta \mathbf{r}_t &= \partial_\theta \chi(\Sigma) \int_0^t e^{-A(t-t')} dW_{t'} + \chi(\Sigma) (\partial_\theta A) \int_0^t (t-t') e^{-A(t-t')} dW_{t'} \\ \partial_\theta^2 \mathbf{r}_t &= \partial_\theta^2 \chi(\Sigma) \int_0^t e^{-A(t-t')} dW_{t'} + 2(\partial_\theta \chi(\Sigma)) (\partial_\theta A) \int_0^t (t-t') e^{-A(t-t')} dW_{t'} \\ &\quad + (\partial_\theta \chi(\Sigma)) (\partial_\theta A)^2 \int_0^t (t-t')^2 e^{-A(t-t')} dW_{t'} + \chi(\Sigma) (\partial_\theta^2 A) \int_0^t (t-t') e^{-A(t-t')} dW_{t'}. \end{aligned}$$

As expected  $\partial_\theta^2 \mathbf{r}(t)$  is a non-anticipating function, as defined in Sec. 2.4.4, and thus it is uncorrelated with the Wiener noise  $d\mathbf{W}_t$  at time  $t$ , i.e.  $\mathbb{E}[dW_t dW_{t'}] = \delta(t-t') dt = 0$  since  $0 < t' < t$ .

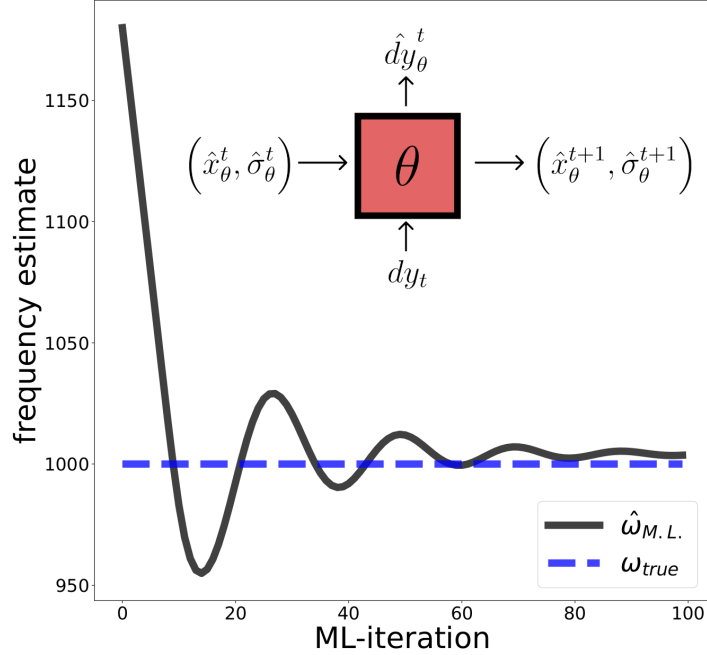


Figure 5.14: We show the evolution for the optimization of the log-likelihood using AD, and an accompanying depiction of the recurrent cell.

### 5.4.3 Maximum-likelihood estimation via automatic-differentiation

While we have provided a dynamical equation for the (derivative of) the log-likelihood  $\lambda(\mathcal{Y}_t, \theta)$ , here we take a step further. With the help of an automatic-differentiation (AD) library [ABC+16]<sup>13</sup> we are able to implement the maximum-likelihood estimator, *i.e.*

$$\hat{\theta} = \underset{\theta}{\text{ArgMax}} \lambda(\mathcal{Y}_t, \theta). \quad (5.74)$$

This optimization is done via a recurrent cell, whose structure is also shown in the Fig. 5.14. The mechanism of the cell is the following. The training consists on episodes  $L = 1, \dots, N$ , and at each episode the cell has an estimate  $\hat{\theta}_L$ . Each episode consists on as many time-steps as the time-trace signal  $\mathcal{Y}_t$  has. At each time-step, the cell is fed with the value of the measurement outcome  $dy_t$ , and evolves its

<sup>13</sup>We refer the reader to our discussion on AD in Sec. 2.2.3

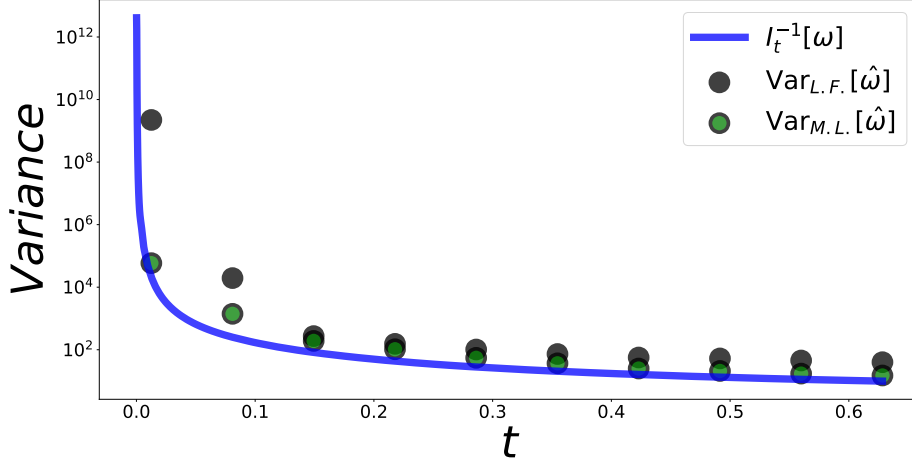


Figure 5.15: We compare the mean square error of the two methods discussed in the main body with the Cramér-Rao bound.

internal state  $\hat{r}_{\hat{\theta}_L}$  as per Eq. 5.13 (using the value  $\hat{\theta}_L$ ). The cell keeps a record of such estimated hidden state, which is forwarded to the next time-step. Moreover, at each time-step, the cell outputs a predicted value for the measurement signal, given by  $C\hat{r}_t dt$ . By the end of the time-trace, the mean squared error between all predicted values and  $\mathcal{Y}_t$  is computed. Note that the latter quantity coincides with the log-likelihood of observing  $\mathcal{Y}_t$  under  $\hat{\theta}_L$ . Finally, the AD mechanism is used to optimize such likelihood over  $\hat{\theta}$ , which is done by re-tracing the (derivative of) the likelihood and using a gradient-based method. Figure 5.14 we also show how the estimate of the frequency improves with the number of optimization iterations.

Similarly to the Lorentzian fit, we can see in Fig. 5.14 that the AD mechanism does also commits a certain error, which we expect it to be (on average) higher than the (inverse of)  $I(\theta)$  — that is, if assuming  $\mathcal{Y}_t$  was an *i.i.d.* sequence—. To this end, in Fig. 5.15 we compare the estimated variances of the Lorentzian fit, and the AD method, with the inverse of the Fisher information, for the frequency estimation case. This is done by carrying out the procedures outlined above for  $10^3$  quantum trajectories. As can be seen, there seems to be a slight advantage in favour of the maximum-likelihood method.

#### 5.4.4 Learning a linear force

We have previously introduced a recurrent cell that, aided with the automatic-differentiation machinery, is able to perform maximum-likelihood estimation. Moreover, we provided an example of frequency estimation, and this constitutes an alternative to the Lorentzian fit that is traditionally used in experimental labs.

However, our approach is not aimed to replace the Lorentzian fit, but rather to complement it in cases where such fit cannot help with estimating a certain parameter. This is the case of an external drive  $f_\theta^t \equiv f(t; \theta)$ , which appears in the dynamics of the first moment  $\bar{\mathbf{r}}_t$  as an additive term [WM09]. Thus, we consider the following evolution:

$$d\mathbf{r}_t = (A - \chi(\Sigma))\bar{\mathbf{r}}_t + \chi(\Sigma)d\mathbf{y}_t + f_\theta^t dt, \quad (5.75)$$

Here, the value of the external signal might be time-dependent and depend on parameter(s)  $\theta$ . Note that we have chosen the covariance matrix to take its stationary value (which does not depend on the force).

In our numerics, we consider two simple scenarios, (i) a constant force and (ii) an exponentially decaying force shown in Fig. 5.16. In the first case, we show how the expected measurement signal value  $\langle \hat{d\mathbf{y}}_{t\hat{\theta}_L} \rangle$  differs from the beginning of the training to the end of it. Note however that even for the untrained model, a *trend* appears to be correctly reproduced. This is due to the innovation term appearing in the evolution of the hidden state, for which the *true* value of  $d\mathbf{y}_t$  is injected at each time. Such fact is also illustrated in the right panel of the figure, for the exponentially-decaying case. In such example, large values of time will lead to no external signal (and thus the model will reproduce the correct dynamics), although for small values the same effect is observed. Behind such *trend-reproducing* effect is the magic of Kalman filtering, which updates the probability distribution for the hidden state by means of the available observations, which are generated under the underlying truth.

While the examples considered in this Section are arguably simple, this method is aimed to be extended to more complex systems. In particular, we are interested in discovering dynamical equations of motion for the external signals, in the same spirit as in Ref. [BPK16], though further investigation needs to be carried out, *i.e.* by the time of writing this thesis, the current is an ongoing project.

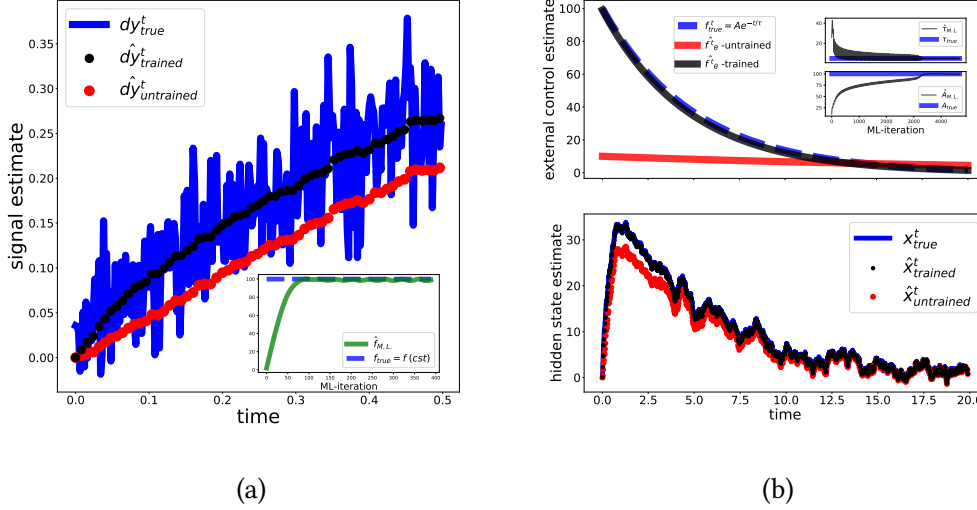


Figure 5.16: We show the results of external force estimation. While the left panel consists on estimating a constant force, the right panel shows that the method learns both the amplitude and decay rate of an exponentially decaying external signal.

## 5.5 Discussion & future perspectives

In this Chapter we have analyzed statistical inference problems for continuously monitored quantum systems.

Our main contribution is that of introducing sequential hypothesis testing strategies in such quantum regime, and elucidated an advantage in terms of a deterministic, *i.e.* fixed-measurement time, strategies. On the one hand, we provided strong analytical insight on the problem, by focusing on the damping discrimination scenario. On the other, we have shown that the sequential approach can be used in different contexts, such as frequency discrimination problems.

Moreover, we have studied parameter estimation problems, and analyzed the performance of maximum-likelihood strategies, which are carried out by automatic differentiation modules. Such strategy is compared with a Lorentzian fit done on the power spectra and the mean squared errors of both strategies are compared with the Fisher information. The latter quantity is obtained by numerical integration. Furthermore, we have investigated how this approach can be used to infer parameters of external signals, and tested this on some simple examples.



Finally, let us remark that it would be very interesting to derive ultimate quantum bounds for the mean stopping times under general (continuous) measurement schemes, analogous to the bound found in [MVHS+21; LHT22]. A natural way to study this problem would be to fix the coupling with the readout system (leaked light from the cavity) and optimize over the measurements done on this ancillary system.

### 5.5.1 Code

We refer the interested reader to Ref. [GG] for the hypothesis testing results (see branches *damping-disc* and *freq-disc*), and for the estimation results (see branch *estimation* of the same reference).

---

# Bibliography

---

- [NC00] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.
- [Man19] Daniel Manzano. “A short introduction to the Lindblad Master Equation”. In: (2019). DOI: [10.1063/1.5115323](https://doi.org/10.1063/1.5115323). eprint: [arXiv:1906.04478](https://arxiv.org/abs/1906.04478).
- [Wat18] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. DOI: [10.1017/9781316848142](https://doi.org/10.1017/9781316848142).
- [Pre98] John Preskill. *Lecture notes for physics 229: Quantum information and computation*. 1998.
- [Lin76] G. Lindblad. “On the generators of quantum dynamical semigroups”. In: *Communications in Mathematical Physics* 48.2 (1976), pp. 119–130. DOI: [cmp/1103899849](https://doi.org/10.1007/BF01645742).
- [WM09] Howard M. Wiseman and Gerard J. Milburn. *Quantum Measurement and Control*. Cambridge University Press, 2009. DOI: [10.1017/CBO9780511813948](https://doi.org/10.1017/CBO9780511813948).
- [Pre18] John Preskill. “Quantum Computing in the NISQ era and beyond”. In: *Quantum* 2 (2018), p. 79. DOI: [10.22331/q-2018-08-06-79](https://doi.org/10.22331/q-2018-08-06-79).
- [FMM+12] Austin G. Fowler et al. “Surface codes: Towards practical large-scale quantum computation”. In: *Phys. Rev. A* 86 (3 2012), p. 032324. DOI: [10.1103/PhysRevA.86.032324](https://doi.org/10.1103/PhysRevA.86.032324).

- 
- [AAB+19] Frank Arute et al. “Quantum supremacy using a programmable superconducting processor”. In: *Nature* 574.7779 (2019), pp. 505–510. ISSN: 1476-4687. DOI: [10.1038/s41586-019-1666-5](https://doi.org/10.1038/s41586-019-1666-5).
- [PCZ22] Feng Pan, Keyang Chen, and Pan Zhang. “Solving the Sampling Problem of the Sycamore Quantum Circuits”. In: *Phys. Rev. Lett.* 129 (9 2022), p. 090502. DOI: [10.1103/PhysRevLett.129.090502](https://doi.org/10.1103/PhysRevLett.129.090502).
- [LLZ+22] Seunghoon Lee et al. *Is there evidence for exponential quantum advantage in quantum chemistry?* 2022. eprint: [arXiv:2208.02199](https://arxiv.org/abs/2208.02199).
- [ABB+21] Yuri Alexeev et al. “Quantum Computer Systems for Scientific Discovery”. In: *PRX Quantum* 2 (1 2021), p. 017001. DOI: [10.1103/PRXQuantum.2.017001](https://doi.org/10.1103/PRXQuantum.2.017001).
- [MBS+18] Jarrod R McClean et al. “Barren plateaus in quantum neural network training landscapes”. In: *Nature Communications* 9.1 (2018), pp. 1–6. DOI: [10.1038/s41467-018-07090-4](https://doi.org/10.1038/s41467-018-07090-4).
- [CSV+21] M. Cerezo et al. “Cost function dependent barren plateaus in shallow parametrized quantum circuits”. In: *Nature Communications* 12.1 (2021), pp. 1–12. DOI: [10.1038/s41467-021-21728-w](https://doi.org/10.1038/s41467-021-21728-w).
- [PCW+21a] Arthur Pesah et al. “Absence of Barren Plateaus in Quantum Convolutional Neural Networks”. In: *Physical Review X* 11.4 (2021), p. 041011. DOI: [10.1103/PhysRevX.11.041011](https://doi.org/10.1103/PhysRevX.11.041011).
- [HAY+21] Zoë Holmes et al. “Barren plateaus preclude learning scramblers”. In: *Physical Review Letters* 126.19 (2021), p. 190501. DOI: [10.1103/PhysRevLett.126.190501](https://doi.org/10.1103/PhysRevLett.126.190501).
- [ZG21] Chen Zhao and Xiao-Shan Gao. “Analyzing the barren plateau phenomenon in training quantum neural networks with the ZX-calculus”. In: *Quantum* 5 (June 2021), p. 466. ISSN: 2521-327X. DOI: [10.22331/q-2021-06-04-466](https://doi.org/10.22331/q-2021-06-04-466).
- [TWN+21] Supanut Thanasilp et al. “Subtleties in the trainability of quantum machine learning models”. In: *arXiv preprint arXiv:2110.14753* (2021).
- [SCC+22] Kunal Sharma et al. “Trainability of dissipative perceptron-based quantum neural networks”. In: *Physical Review Letters* 128.18 (2022), p. 180505. DOI: [10.1103/PhysRevLett.128.180505](https://doi.org/10.1103/PhysRevLett.128.180505).
-

## BIBLIOGRAPHY

---

- [PNG+21] Taylor L Patti et al. “Entanglement devised barren plateau mitigation”. In: *Physical Review Research* 3.3 (2021), p. 033090. DOI: [10.1103/PhysRevResearch.3.033090](https://doi.org/10.1103/PhysRevResearch.3.033090).
- [MKW21] Carlos Ortiz Marrero, Mária Kieferová, and Nathan Wiebe. “Entanglement-induced barren plateaus”. In: *PRX Quantum* 2.4 (2021), p. 040316. DOI: [10.1103/PRXQuantum.2.040316](https://doi.org/10.1103/PRXQuantum.2.040316).
- [CC21] M. Cerezo and Patrick J Coles. “Higher order derivatives of quantum neural networks with barren plateaus”. In: *Quantum Science and Technology* 6.2 (2021), p. 035006. DOI: [10.1088/2058-9565/abf51a](https://doi.org/10.1088/2058-9565/abf51a).
- [ACC+21] Andrew Arrasmith et al. “Effect of barren plateaus on gradient-free optimization”. In: *Quantum* 5 (2021), p. 558. DOI: [10.22331/q-2021-10-05-558](https://doi.org/10.22331/q-2021-10-05-558).
- [BK21] Lennart Bittel and Martin Kliesch. “Training Variational Quantum Algorithms Is NP-Hard”. In: *Phys. Rev. Lett.* 127 (12 2021), p. 120502. DOI: [10.1103/PhysRevLett.127.120502](https://doi.org/10.1103/PhysRevLett.127.120502).
- [BWP+17] Jacob Biamonte et al. “Quantum machine learning”. In: *Nature* 549.7671 (2017), pp. 195–202. DOI: [10.1038/nature23474](https://doi.org/10.1038/nature23474).
- [SK22] Maria Schuld and Nathan Killoran. “Is Quantum Advantage the Right Goal for Quantum Machine Learning?” In: *PRX Quantum* 3 (3 2022), p. 030101. DOI: [10.1103/PRXQuantum.3.030101](https://doi.org/10.1103/PRXQuantum.3.030101).
- [Aar15] Scott Aaronson. “Read the fine print”. In: *Nature Physics* 11.4 (2015), pp. 291–293. ISSN: 1745-2481. DOI: [10.1038/nphys3272](https://doi.org/10.1038/nphys3272).
- [PMS+14] Alberto Peruzzo et al. “A variational eigenvalue solver on a photonic quantum processor”. In: *Nature Communications* 5.1 (2014), pp. 1–7. DOI: [doi.org/10.1038/ncomms5213](https://doi.org/10.1038/ncomms5213).
- [PCW+21b] Arthur Pesah et al. “Absence of Barren Plateaus in Quantum Convolutional Neural Networks”. In: *Phys. Rev. X* 11 (4 2021), p. 041011. DOI: [10.1103/PhysRevX.11.041011](https://doi.org/10.1103/PhysRevX.11.041011).
- [SCH+22] Kunal Sharma et al. “Reformulation of the No-Free-Lunch Theorem for Entangled Datasets”. In: *Phys. Rev. Lett.* 128 (7 2022), p. 070501. DOI: [10.1103/PhysRevLett.128.070501](https://doi.org/10.1103/PhysRevLett.128.070501).
- [GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016.

- 
- [CAB+21a] M. Cerezo et al. “Variational quantum algorithms”. In: *Nature Reviews Physics* 3.9 (2021), pp. 625–644. ISSN: 2522-5820. DOI: [10.1038/s42254-021-00348-9](https://doi.org/10.1038/s42254-021-00348-9).
- [BCLK+22] Kishor Bharti et al. “Noisy intermediate-scale quantum algorithms”. In: *Rev. Mod. Phys.* 94 (1 2022), p. 015004. DOI: [10.1103/RevModPhys.94.015004](https://doi.org/10.1103/RevModPhys.94.015004).
- [LLD22] Weikang Li, Zhide Lu, and Dong-Ling Deng. “Quantum Neural Network Classifiers: A Tutorial”. In: *SciPost Phys. Lect. Notes* (2022), p. 61. DOI: [10.21468/SciPostPhysLectNotes.61](https://doi.org/10.21468/SciPostPhysLectNotes.61).
- [FN18] Edward Farhi and Hartmut Neven. *Classification with Quantum Neural Networks on Near Term Processors*. 2018. eprint: [arXiv:1802.06002](https://arxiv.org/abs/1802.06002).
- [SBS+20] Maria Schuld et al. “Circuit-centric quantum classifiers”. In: *Phys. Rev. A* 101 (3 2020), p. 032308. DOI: [10.1103/PhysRevA.101.032308](https://doi.org/10.1103/PhysRevA.101.032308).
- [BBF+20] Kerstin Beer et al. “Training deep quantum neural networks”. In: *Nature Communications* 11.1 (2020), p. 808. ISSN: 2041-1723. DOI: [10.1038/s41467-020-14454-2](https://doi.org/10.1038/s41467-020-14454-2).
- [CCL19a] Iris Cong, Soonwon Choi, and Mikhail D Lukin. “Quantum convolutional neural networks”. In: *Nature Physics* 15.12 (2019), pp. 1273–1278. DOI: [10.1038/s41567-019-0648-8](https://doi.org/10.1038/s41567-019-0648-8).
- [EFH+22] Andreas Elben et al. “The randomized measurement toolbox”. In: *arXiv preprint arXiv:2203.11374* (2022).
- [YEZ+19] Xiao Yuan et al. “Theory of variational quantum simulation”. In: *Quantum* 3 (Oct. 2019), p. 191. ISSN: 2521-327X. DOI: [10.22331/q-2019-10-07-191](https://doi.org/10.22331/q-2019-10-07-191).
- [VPB18] Guillaume Verdon, Jason Pye, and Michael Broughton. “A universal training algorithm for quantum deep learning”. In: *arXiv preprint arXiv:1806.09729* (2018).
- [KAC+20] Jonas M Kübler et al. “An adaptive optimizer for measurement-frugal variational algorithms”. In: *Quantum* 4 (2020), p. 263. DOI: [10.22331/q-2020-05-11-263](https://doi.org/10.22331/q-2020-05-11-263).
- [ACS+20] Andrew Arrasmith et al. “Operator Sampling for Shot-frugal Optimization in Variational Algorithms”. In: *arXiv preprint arXiv:2004.06252* (2020).

## BIBLIOGRAPHY

---

- [SIK+20] James Stokes et al. “Quantum natural gradient”. In: *Quantum* 4 (2020), p. 269. DOI: [10.22331/q-2020-05-25-269](https://doi.org/10.22331/q-2020-05-25-269).
- [KB19] Bálint Koczor and Simon C Benjamin. “Quantum natural gradient generalised to non-unitary circuits”. In: *arXiv preprint arXiv:1912.08660* (2019).
- [NFT20] Ken M Nakanishi, Keisuke Fujii, and Synge Todo. “Sequential minimal optimization for quantum-classical hybrid algorithms”. In: *Physical Review Research* 2.4 (2020), p. 043158. DOI: [10.1103/PhysRevResearch.2.043158](https://doi.org/10.1103/PhysRevResearch.2.043158).
- [FCA+20] Enrico Fontana et al. “Optimizing parametrized quantum circuits via noise-induced breaking of symmetries”. In: *arXiv preprint arXiv:2011.08763* (2020).
- [HSC+22] Zoë Holmes et al. “Connecting ansatz expressibility to gradient magnitudes and barren plateaus”. In: *PRX Quantum* 3 (1 2022), p. 010313. DOI: [10.1103/PRXQuantum.3.010313](https://doi.org/10.1103/PRXQuantum.3.010313).
- [Gri19] van de Arianne Griend. “Constrained quantum CNOT circuit re-synthesis using deep reinforcement learning”. MA thesis. Faculteit der Sociale Wetenschappen: Rijksuniversiteit Groningen, 2019.
- [Wik] Wikipedia. *Ansatz*.
- [ZHS+98] Karol Zyczkowski et al. “Volume of the set of separable states”. In: *Phys. Rev. A* 58 (2 1998), pp. 883–892. DOI: [10.1103/PhysRevA.58.883](https://doi.org/10.1103/PhysRevA.58.883).
- [KMT+17] Abhinav Kandala et al. “Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets”. In: *Nature* 549.7671 (2017), pp. 242–246. ISSN: 1476-4687. DOI: [10.1038/nature23879](https://doi.org/10.1038/nature23879).
- [FGG14] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. “A quantum approximate optimization algorithm”. In: *arXiv preprint arXiv:1411.4028* (2014).
- [HWO+19] Stuart Hadfield et al. “From the quantum approximate optimization algorithm to a quantum alternating operator ansatz”. In: *Algorithms* 12.2 (2019), p. 34. DOI: [10.3390/a12020034](https://doi.org/10.3390/a12020034).
- [CRO+19] Yudong Cao et al. “Quantum chemistry in the age of quantum computing”. In: *Chemical reviews* 119.19 (2019), pp. 10856–10915.

- 
- [BM07] Rodney J Bartlett and Monika Musiał. “Coupled-cluster theory in quantum chemistry”. In: *Reviews of Modern Physics* 79.1 (2007), p. 291.
- [GEB+19] Harper R Grimsley et al. “An adaptive variational algorithm for exact molecular simulations on a quantum computer”. In: *Nature Communications* 10.1 (2019), pp. 1–9. DOI: [10.1038/s41467-019-10988-2](https://doi.org/10.1038/s41467-019-10988-2).
- [TSB+21] Ho Lun Tang et al. “qubit-adapt-vqe: An adaptive algorithm for constructing hardware-efficient ansätze on a quantum processor”. In: *PRX Quantum* 2.2 (2021), p. 020310. DOI: [10.1103/PRXQuantum.2.020310](https://doi.org/10.1103/PRXQuantum.2.020310).
- [ZKK+21] Zi-Jian Zhang et al. “Mutual information-assisted adaptive variational quantum eigensolver”. In: *Quantum Science and Technology* (2021). DOI: [10.1088/2058-9565/abdca4](https://doi.org/10.1088/2058-9565/abdca4).
- [RHP+19] Arthur G Rattew et al. “A domain-agnostic, noise-resistant, hardware-efficient evolutionary variational quantum eigensolver”. In: *arXiv preprint arXiv:1910.09694* (2019).
- [CSU+20] D Chivilikhin et al. “MoG-VQE: Multiobjective genetic variational quantum eigensolver”. In: *arXiv preprint arXiv:2007.04424* (2020).
- [CRS+21] Lukasz Cincio et al. “Machine Learning of Noise-Resilient Quantum Circuits”. In: *PRX Quantum* 2 (1 2021), p. 010324. DOI: [10.1103/PRXQuantum.2.010324](https://doi.org/10.1103/PRXQuantum.2.010324).
- [CSS+18] Lukasz Cincio et al. “Learning the quantum algorithm for state overlap”. In: *New Journal of Physics* 20.11 (2018), p. 113022. DOI: [10.1088/1367-2630/aae94a](https://doi.org/10.1088/1367-2630/aae94a).
- [DHY+20] Yuxuan Du et al. “Quantum circuit architecture search: error mitigation and trainability enhancement for variational quantum solvers”. In: *arXiv preprint arXiv:2010.10217* (2020).
- [ZHZ+20] Shi-Xin Zhang et al. “Differentiable Quantum Architecture Search”. In: *arXiv preprint arXiv:2010.08561* (2020).
- [GPRS+21] Guillermo García-Pérez et al. “Learning to Measure: Adaptively Informationally Complete Generalized Measurements for Quantum Algorithms”. In: *PRX Quantum* 2 (4 2021), p. 040342. DOI: [10.1103/PRXQuantum.2.040342](https://doi.org/10.1103/PRXQuantum.2.040342).
- [Alg] .

## BIBLIOGRAPHY

---

- [CWM+20] Daniel Claudino et al. “Benchmarking adaptive variational quantum eigensolvers”. In: *Frontiers in Chemistry* 8 (2020), p. 1152. DOI: [10.3389/fchem.2020.606863](https://doi.org/10.3389/fchem.2020.606863).
- [PT20] Mohammad Pirhooshyaran and Tamas Terlaky. “Quantum Circuit Design Search”. In: (2020).
- [BCV+21] M Bilkis et al. “A semi-agnostic ansatz with variable structure for quantum machine learning”. In: *arXiv preprint arXiv:2103.06712* (2021).
- [KB15] Diederik P Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization”. In: *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*. 2015.
- [GLD+21] Andi Gu et al. “Adaptive shot allocation for fast convergence in variational quantum algorithms”. In: (2021). DOI: [10.48550/arXiv.2108.10434](https://doi.org/10.48550/arXiv.2108.10434). eprint: [arXiv:2108.10434](https://arxiv.org/abs/2108.10434).
- [BB17] Jacob Biamonte and Ville Bergholm. “Tensor Networks in a Nutshell”. In: (July 2017). arXiv: [1708.00006](https://arxiv.org/abs/1708.00006) [[quant-ph](https://arxiv.org/archive/quant)].
- [Orú19] Román Orús. “Tensor networks for complex quantum systems”. In: *Nature Reviews Physics* 1.9 (2019), pp. 538–550. ISSN: 2522-5820. DOI: [10.1038/s42254-019-0086-7](https://doi.org/10.1038/s42254-019-0086-7).
- [BVM+20] Michael Broughton et al. “Tensorflow quantum: A software framework for quantum machine learning”. In: *arXiv preprint arXiv:2003.02989* (2020).
- [LLZ+20] Xiu-Zhe Luo et al. “Yao.jl: Extensible, Efficient Framework for Quantum Algorithm Design”. In: *Quantum* 4 (Oct. 2020), p. 341. ISSN: 2521-327X. DOI: [10.22331/q-2020-10-11-341](https://doi.org/10.22331/q-2020-10-11-341).
- [REF+20] Christopher Rackauckas et al. “Generalized Physics-Informed Learning through Language-Wide Differentiable Programming”. In: *AAAI Spring Symposium: MLPS*. 2020.
- [LLW+19] Hai-Jun Liao et al. “Differentiable Programming Tensor Networks”. In: *Phys. Rev. X* 9 (3 2019), p. 031041. DOI: [10.1103/PhysRevX.9.031041](https://doi.org/10.1103/PhysRevX.9.031041).
- [SSK+21] Frank Schäfer et al. “Control of stochastic quantum dynamics by differentiable programming”. In: *Machine Learning: Science and Technology* 2.3 (2021), p. 035004. DOI: [10.1088/2632-2153/abec22](https://doi.org/10.1088/2632-2153/abec22).



- 
- [Wik23] Wikipedia. *Automatic differentiation* — *Wikipedia, The Free Encyclopedia*. <http://en.wikipedia.org/w/index.php?title=Automatic%20differentiation&oldid=1149933866>. [Online; accessed 18-April-2023]. 2023.
- [ABC+16] Martín Abadi et al. “Tensorflow: A system for large-scale machine learning”. In: *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*. 2016, pp. 265–283.
- [MDA15] Dougal Maclaurin, David Duvenaud, and Ryan P Adams. “Autograd: Effortless Gradients in Numpy”. In: *ICML 2015 AutoML Workshop*. 2015.
- [Mac16] Dougal Maclaurin. “Modeling, Inference and Optimization with Composable Differentiable Procedures”. PhD thesis. Harvard University, 2016.
- [BPR+18] Atilim Gunes Baydin et al. “Automatic Differentiation in Machine Learning: a Survey”. In: *Journal of Machine Learning Research* 18.153 (2018), pp. 1–43.
- [MNK+18] K. Mitarai et al. “Quantum circuit learning”. In: *Phys. Rev. A* 98 (3 2018), p. 032309. DOI: [10.1103/PhysRevA.98.032309](https://doi.org/10.1103/PhysRevA.98.032309).
- [SBG+19] Maria Schuld et al. “Evaluating analytic gradients on quantum hardware”. In: *Phys. Rev. A* 99 (3 2019), p. 032331. DOI: [10.1103/PhysRevA.99.032331](https://doi.org/10.1103/PhysRevA.99.032331).
- [WIW+22] David Wierichs et al. “General parameter-shift rules for quantum gradients”. In: *Quantum* 6 (Mar. 2022), p. 677. ISSN: 2521-327X. DOI: [10.22331/q-2022-03-30-677](https://doi.org/10.22331/q-2022-03-30-677).
- [SWM+20] Ryan Sweke et al. “Stochastic gradient descent for hybrid quantum-classical optimization”. In: *Quantum* 4 (Aug. 2020), p. 314. DOI: [10.22331/q-2020-08-31-314](https://doi.org/10.22331/q-2020-08-31-314).
- [HN21] Aram W Harrow and John C Napp. “Low-depth gradient measurements can improve convergence in variational hybrid quantum-classical algorithms”. In: *Physical Review Letters* 126.14 (2021), p. 140502.
- [BIS+18] Ville Bergholm et al. “Pennylane: Automatic differentiation of hybrid quantum-classical computations”. In: *arXiv preprint arXiv:1811.04968* (2018).

## BIBLIOGRAPHY

---

- [LCS+22] Martin Larocca et al. “Diagnosing Barren Plateaus with Tools from Quantum Optimal Control”. In: *Quantum* 6 (Sept. 2022), p. 824. ISSN: 2521-327X. DOI: [10.22331/q-2022-09-29-824](https://doi.org/10.22331/q-2022-09-29-824).
- [SJAG19] Sukin Sim, Peter D. Johnson, and Alán Aspuru-Guzik. “Expressibility and Entangling Capability of Parameterized Quantum Circuits for Hybrid Quantum-Classical Algorithms”. In: *Advanced Quantum Technologies* 2.12 (2019), p. 1900070. DOI: <https://doi.org/10.1002/qute.201900070>.
- [VC21] Tyler Volkoff and Patrick J Coles. “Large gradients via correlation in random parameterized quantum circuits”. In: *Quantum Science and Technology* 6.2 (2021), p. 025008. DOI: [10.1088/2058-9565/abd89](https://doi.org/10.1088/2058-9565/abd89).
- [SMM+21] Andrea Skolik et al. “Layerwise learning for quantum neural networks”. In: *Quantum Machine Intelligence* 3.1 (2021), pp. 1–11. DOI: [10.1007/s42484-020-00036-4](https://doi.org/10.1007/s42484-020-00036-4).
- [GWO+19] Edward Grant et al. “An initialization strategy for addressing barren plateaus in parametrized quantum circuits”. In: *Quantum* 3 (2019), p. 214. DOI: [10.22331/q-2019-12-09-214](https://doi.org/10.22331/q-2019-12-09-214).
- [VBM+19] Guillaume Verdon et al. “Learning to learn with quantum neural networks via classical neural networks”. In: *arXiv preprint arXiv:1907.05415* (2019).
- [CCL19b] Iris Cong, Soonwon Choi, and Mikhail D. Lukin. “Quantum convolutional neural networks”. In: *Nature Physics* 15.12 (2019), pp. 1273–1278. ISSN: 1745-2481. DOI: [10.1038/s41567-019-0648-8](https://doi.org/10.1038/s41567-019-0648-8).
- [WFC+21] Samson Wang et al. “Noise-induced barren plateaus in variational quantum algorithms”. In: *Nature Communications* 12.1 (2021), pp. 1–11. DOI: [10.1038/s41467-021-27045-6](https://doi.org/10.1038/s41467-021-27045-6).
- [SFGP21] Daniel Stilck França and Raul Garcia-Patron. “Limitations of optimization algorithms on noisy quantum devices”. In: *Nature Physics* 17.11 (2021), pp. 1221–1227. DOI: [10.1038/s41567-021-01356-3](https://doi.org/10.1038/s41567-021-01356-3).
- [ZHL+20] Kaining Zhang et al. “Toward Trainability of Quantum Neural Networks”. In: *arXiv preprint arXiv:2011.06258* (2020).

- 
- [BH21] Kishor Bharti and Tobias Haug. “Quantum-assisted simulator”. In: *Physical Review A* 104.4 (2021), p. 042418. DOI: [10.1103/PhysRevA.104.042418](https://doi.org/10.1103/PhysRevA.104.042418).
- [CSA+22] M. Cerezo et al. “Variational Quantum State Eigensolver”. In: *npj Quantum Information* 8.1 (2022), pp. 1–11. DOI: [10.1038/s41534-022-00611-6](https://doi.org/10.1038/s41534-022-00611-6).
- [Ser17a] Alessio Serafini. *Quantum Continuous Variables: A Primer of Theoretical Methods*. CRC Press, 2017.
- [FOP05] Alessandro Ferraro, Stefano Olivares, and Matteo G. A. Paris. *Gaussian states in continuous variable quantum information*. Napoli: Bibliopolis, 2005. ISBN: ISBN 88-7088-483-X.
- [RZB+94] Michael Reck et al. “Experimental realization of any discrete unitary operator”. In: *Phys. Rev. Lett.* 73 (1 1994), pp. 58–61. DOI: [10.1103/PhysRevLett.73.58](https://doi.org/10.1103/PhysRevLett.73.58).
- [WPGP+12] Christian Weedbrook et al. “Gaussian quantum information”. In: *Rev. Mod. Phys.* 84 (2 2012), pp. 621–669. DOI: [10.1103/RevModPhys.84.621](https://doi.org/10.1103/RevModPhys.84.621).
- [Wil36] John Williamson. “On the Algebraic Problem Concerning the Normal Forms of Linear Dynamical Systems”. In: *American Journal of Mathematics* 58.1 (1936), pp. 141–163. ISSN: 00029327, 10806377.
- [JPP21] Leonardo Banchi Jason Pereira and Stefano Pirandola. “Symplectic decomposition from submatrix determinants”. In: *Proceedings A* 477 (2 2021), p. 20210513. DOI: <https://doi.org/10.1098/rspa.2021.0513>.
- [PSL09] Stefano Pirandola, Alessio Serafini, and Seth Lloyd. “Correlation matrices of two-mode bosonic systems”. In: *Phys. Rev. A* 79 (5 2009), p. 052327. DOI: [10.1103/PhysRevA.79.052327](https://doi.org/10.1103/PhysRevA.79.052327).
- [Wal21] Mattia Walschaers. “Non-Gaussian Quantum States and Where to Find Them”. In: *PRX Quantum* 2 (3 2021), p. 030204. DOI: [10.1103/PRXQuantum.2.030204](https://doi.org/10.1103/PRXQuantum.2.030204).
- [Oli12] S. Olivares. “Quantum optics in the phase space”. In: *The European Physical Journal Special Topics* 203.1 (2012), pp. 3–24. ISSN: 1951-6401. DOI: [10.1140/epjst/e2012-01532-4](https://doi.org/10.1140/epjst/e2012-01532-4).

## BIBLIOGRAPHY

---

- [DVR+19] Daniele Dequal et al. “Feasibility of satellite quantum key distribution with continuous variable”. In: *Quantum Inf. Meas. V Quantum Technol.* Vol. Part F165-. Washington, D.C.: OSA, 2019, T5A.89. ISBN: 978-1-943580-56-9. DOI: [10.1364/QIM.2019.T5A.89](https://doi.org/10.1364/QIM.2019.T5A.89). arXiv: [2002.02002](https://arxiv.org/abs/2002.02002).
- [AP05] Larry C. Andrews and Ronald L. Phillips. *Laser Beam Propagation through Random Media*. 1000 20th Street, Bellingham, WA 98227-0010 USA: SPIE, 2005. ISBN: 9780819459480. DOI: [10.1117/3.626196](https://doi.org/10.1117/3.626196).
- [UHP+12] Vladyslav C Usenko et al. “Entanglement of Gaussian states and the applicability to quantum key distribution over fading channels”. In: *New J. Phys.* 14.9 (2012), p. 093048. ISSN: 1367-2630. DOI: [10.1088/1367-2630/14/9/093048](https://doi.org/10.1088/1367-2630/14/9/093048).
- [Pir21a] Stefano Pirandola. “Limits and security of free-space quantum communications”. In: *Phys. Rev. Res.* 3.1 (2021), p. 013279. ISSN: 2643-1564. DOI: [10.1103/PhysRevResearch.3.013279](https://doi.org/10.1103/PhysRevResearch.3.013279).
- [Pir21b] Stefano Pirandola. “Satellite quantum communications: Fundamental bounds and practical security”. In: *Phys. Rev. Res.* 3.2 (2021), p. 023130. ISSN: 2643-1564. DOI: [10.1103/PhysRevResearch.3.023130](https://doi.org/10.1103/PhysRevResearch.3.023130).
- [VSV11] D. Yu. Vasylyev, A. A. Semenov, and W. Vogel. “Toward Global Quantum Communication: Beam Wandering Preserves Nonclassicality”. In: (2011). DOI: [10.1103/PhysRevLett.108.220501](https://doi.org/10.1103/PhysRevLett.108.220501). arXiv: [1110.1440](https://arxiv.org/abs/1110.1440).
- [VSV+17] D. Vasylyev et al. “Free-space quantum links under diverse weather conditions”. In: (2017). DOI: [10.1103/PhysRevA.96.043856](https://doi.org/10.1103/PhysRevA.96.043856). arXiv: [1707.04932](https://arxiv.org/abs/1707.04932).
- [LMN08] Adriana E. Lita, Aaron J. Miller, and Sae Woo Nam. “Counting near-infrared single-photons with 95% efficiency”. In: *Opt. Express* 16.5 (2008), pp. 3032–3040. DOI: [10.1364/OE.16.003032](https://doi.org/10.1364/OE.16.003032).
- [GIC02] Géza Giedke and J. Ignacio Cirac. “Characterization of Gaussian operations and distillation of Gaussian states”. In: *Phys. Rev. A* 66 (3 2002), p. 032316. DOI: [10.1103/PhysRevA.66.032316](https://doi.org/10.1103/PhysRevA.66.032316).
- [Jac14] Kurt Jacobs. *Quantum Measurement Theory and its Applications*. Cambridge University Press, 2014. DOI: [10.1017/CBO9781139179027](https://doi.org/10.1017/CBO9781139179027).

- 
- [GC85] C. W. Gardiner and M. J. Collett. “Input and output in damped quantum systems: Quantum stochastic differential equations and the master equation”. In: *Phys. Rev. A* 31 (6 1985), pp. 3761–3774. DOI: [10.1103/PhysRevA.31.3761](https://doi.org/10.1103/PhysRevA.31.3761).
- [DJ99] A. C. Doherty and K. Jacobs. “Feedback control of quantum systems using continuous state estimation”. In: *Phys. Rev. A* 60 (4 1999), pp. 2700–2711. DOI: [10.1103/PhysRevA.60.2700](https://doi.org/10.1103/PhysRevA.60.2700).
- [WM93a] H. M. Wiseman and G. J. Milburn. “Interpretation of quantum jump and diffusion processes illustrated on the Bloch sphere”. In: *Phys. Rev. A* 47 (3 1993), pp. 1652–1666. DOI: [10.1103/PhysRevA.47.1652](https://doi.org/10.1103/PhysRevA.47.1652).
- [JMnT+18] Ricardo Jiménez-Martínez et al. “Signal Tracking Beyond the Time Resolution of an Atomic Sensor by Kalman Filtering”. In: *Phys. Rev. Lett.* 120 (4 2018), p. 040503. DOI: [10.1103/PhysRevLett.120.040503](https://doi.org/10.1103/PhysRevLett.120.040503).
- [LJP+03] Wei Lu et al. “Real-time detection of electron tunnelling in a quantum dot”. In: *Nature* 423.6938 (2003), pp. 422–425. ISSN: 1476-4687. DOI: [10.1038/nature01642](https://doi.org/10.1038/nature01642).
- [AKM14] Markus Aspelmeyer, Tobias J. Kippenberg, and Florian Marquardt. “Cavity optomechanics”. In: *Rev. Mod. Phys.* 86 (4 2014), pp. 1391–1452. DOI: [10.1103/RevModPhys.86.1391](https://doi.org/10.1103/RevModPhys.86.1391).
- [Dol73] S. J. Dolinar. “Communication and sciences engineering”. In: *Q. Prog. Rep. (Research Lab. Electron.* 111 (1973), p. 115.
- [JS06] Kurt Jacobs and Daniel A. Steck. “A straightforward introduction to continuous quantum measurement”. In: *Contemporary Physics* 47.5 (2006), pp. 279–303. ISSN: 0010-7514. DOI: [10.1080/00107510601101934](https://doi.org/10.1080/00107510601101934).
- [Ser17b] Alessio Serafini. *Quantum continuous variables: a primer of theoretical methods*. CRC press, 2017.
- [GLS16] Marco G. Genoni, Ludovico Lami, and Alessio Serafini. “Conditional and unconditional Gaussian quantum dynamics”. In: *Contemporary Physics* 57.3 (2016), pp. 331–349. ISSN: 0010-7514. DOI: [10.1080/00107514.2015.1125624](https://doi.org/10.1080/00107514.2015.1125624).
- [WM93b] H. M. Wiseman and G. J. Milburn. “Quantum theory of field-quadrature measurements”. In: *Phys. Rev. A* 47 (1 1993), pp. 642–662. DOI: [10.1103/PhysRevA.47.642](https://doi.org/10.1103/PhysRevA.47.642).

## BIBLIOGRAPHY

---

- [Gar04] C. W. Gardiner. *Handbook of stochastic methods for physics, chemistry and the natural sciences*. Third. Vol. 13. Springer Series in Synergetics. Berlin: Springer-Verlag, 2004, pp. xviii+415. ISBN: 3-540-20882-8.
- [WD05] H. M. Wiseman and A. C. Doherty. “Optimal Unravellings for Feedback Control in Linear Quantum Systems”. In: *Phys. Rev. Lett.* 94 (7 2005), p. 070405. DOI: [10.1103/PhysRevLett.94.070405](https://doi.org/10.1103/PhysRevLett.94.070405).
- [SDH+11] A. Szorkovszky et al. “Mechanical Squeezing via Parametric Amplification and Weak Measurement”. In: *Phys. Rev. Lett.* 107 (21 2011), p. 213603. DOI: [10.1103/PhysRevLett.107.213603](https://doi.org/10.1103/PhysRevLett.107.213603).
- [DSH+12] Andrew C. Doherty et al. “The quantum trajectory approach to quantum feedback control of an oscillator revisited”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 370.1979 (2012), pp. 5338–5353. DOI: [10.1098/rsta.2011.0531](https://doi.org/10.1098/rsta.2011.0531). eprint: <https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2011.0531>.
- [SL13] B. Sahakian and J.N. LaBuzetta. *Bad Moves: How decision making goes wrong, and the ethics of smart drugs*. EBSCOhost ebooks online. OUP Oxford, 2013. ISBN: 978-0-19-164552-5.
- [HW94] Paul Hausladen and William K. Wootters. “A ‘Pretty Good’ Measurement for Distinguishing Quantum States”. In: *Journal of Modern Optics* 41.12 (1994), pp. 2385–2390. DOI: [10.1080/09500349414552221](https://doi.org/10.1080/09500349414552221). eprint: <https://doi.org/10.1080/09500349414552221>.
- [BC09] Stephen M. Barnett and Sarah Croke. “Quantum state discrimination”. In: *Adv. Opt. Photon.* 1.2 (2009), pp. 238–278. DOI: [10.1364/AOP.1.000238](https://doi.org/10.1364/AOP.1.000238).
- [BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. ISBN: 0521833787.
- [ACT+07] K. M. R. Audenaert et al. “Discriminating States: The Quantum Chernoff Bound”. In: *Phys. Rev. Lett.* 98 (16 2007), p. 160501. DOI: [10.1103/PhysRevLett.98.160501](https://doi.org/10.1103/PhysRevLett.98.160501).
- [HP91] Fumio Hiai and Dénes Petz. “The proper formula for relative entropy and its asymptotics in quantum probability”. In: *Communications in Mathematical Physics* 143.1 (1991), pp. 99–114. ISSN: 1432-0916. DOI: [10.1007/BF02100287](https://doi.org/10.1007/BF02100287).

- [ON] Tomohiro Ogawa and Hiroshi Nagaoka. “Strong Converse and Stein’s Lemma in Quantum Hypothesis Testing”. In: *Asymptotic Theory of Quantum Statistical Inference*, pp. 28–42. DOI: [10.1142/97898125630710003](https://doi.org/10.1142/97898125630710003). eprint: <https://www.worldscientific.com/doi/pdf/10.1142/97898125630710003>.
- [ABB+05] A. Acín et al. “Multiple-copy two-state discrimination with individual measurements”. In: *Phys. Rev. A* 71 (3 2005), p. 032338. DOI: [10.1103/PhysRevA.71.032338](https://doi.org/10.1103/PhysRevA.71.032338).
- [MVHS+21] Esteban Martínez Vargas et al. “Quantum Sequential Hypothesis Testing”. In: *Phys. Rev. Lett.* 126 (18 2021), p. 180502. DOI: [10.1103/PhysRevLett.126.180502](https://doi.org/10.1103/PhysRevLett.126.180502).
- [WW48] A. Wald and J. Wolfowitz. “Optimum Character of the Sequential Probability Ratio Test”. In: *The Annals of Mathematical Statistics* 19.3 (1948), pp. 326–339. ISSN: 00034851.
- [MU05] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005. DOI: [10.1017/CBO9780511813603](https://doi.org/10.1017/CBO9780511813603).
- [Sim76] Gordon Simons. “An Improved Statement of Optimality for Sequential Probability Ratio Tests”. In: *The Annals of Statistics* 4.6 (1976), pp. 1240 –1243. DOI: [10.1214/aos/1176343655](https://doi.org/10.1214/aos/1176343655).
- [Mey21] Johannes Jakob Meyer. “Fisher Information in Noisy Intermediate-Scale Quantum Applications”. In: *Quantum* 5 (Sept. 2021), p. 539. ISSN: 2521-327X. DOI: [10.22331/q-2021-09-09-539](https://doi.org/10.22331/q-2021-09-09-539).
- [Hel76] Carl W Helstrom. *Quantum Detection and Estimation Theory*. Vol. 84. New York: Academic press, 1976, p. 309. ISBN: 9780123400505.
- [GLM06] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. “Quantum Metrology”. In: *Phys. Rev. Lett.* 96 (1 2006), p. 010401. DOI: [10.1103/PhysRevLett.96.010401](https://doi.org/10.1103/PhysRevLett.96.010401).
- [SG18] Richard Sutton and Andrew G. Barto. *Reinforcement Learning Sutton*. MIT Press, 2018. ISBN: 9780262039246.
- [Sze10a] Csaba Szepesvári. *Algorithms for Reinforcement Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan and Claypool Publishers, 2010.

## BIBLIOGRAPHY

---

- [SJJ94] Satinder P Singh, Tommi Jaakkola, and Michael I Jordan. “Learning Without State-Estimation in Partially Observable Markovian Decision Processes”. In: *Mach. Learn. Proc. 1994*. 1994, pp. 284–292. DOI: [10.1016/b978-1-55860-335-6.50042-8](https://doi.org/10.1016/b978-1-55860-335-6.50042-8).
- [MKS+13] Volodymyr Mnih et al. “Playing Atari with Deep Reinforcement Learning”. In: (2013). arXiv: [1312.5602](https://arxiv.org/abs/1312.5602).
- [SPK13] Guy Shani, Joelle Pineau, and Robert Kaplow. “A survey of point-based POMDP solvers”. In: *Auton. Agent. Multi. Agent. Syst.* 27.1 (2013), pp. 1–51. ISSN: 13872532. DOI: [10.1007/s10458-012-9200-2](https://doi.org/10.1007/s10458-012-9200-2).
- [Ego15] Maxim Egorov. *Deep reinforcement learning with pomdps*. Tech. rep. Technical Report, Stanford University, 2015. arXiv: [1903.07765](https://arxiv.org/abs/1903.07765).
- [ZLP+18] Pengfei Zhu et al. “On Improving Deep Reinforcement Learning for POMDPs”. In: (2018). arXiv: [1804.06309](https://arxiv.org/abs/1804.06309).
- [LR85] T. L. Lai and Herbert Robbins. “Asymptotically efficient adaptive allocation rules”. In: *Adv. Appl. Math.* 6.1 (1985), pp. 4–22. ISSN: 10902074. DOI: [10.1016/0196-8858\(85\)90002-8](https://doi.org/10.1016/0196-8858(85)90002-8).
- [Tho33] William R. Thompson. “On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples”. In: *Biometrika* 25.3/4 (1933), p. 285. ISSN: 00063444. DOI: [10.2307/2332286](https://doi.org/10.2307/2332286).
- [LS18] Tor Lattimore and Csaba Szepesvari. “Bandit Algorithms”. In: *Cambridge Univ. Press* (2018), p. 542.
- [AOM17] Mohammad Gheshlaghi Azar, Ian Osband, and Rémi Munos. *Minimax Regret Bounds for Reinforcement Learning*. 2017. eprint: [arXiv: 1703.05449](https://arxiv.org/abs/1703.05449).
- [TP21] Damianos Tranos and Alexandre Proutiere. *Regret Analysis in Deterministic Reinforcement Learning*. 2021. eprint: [arXiv: 2106.14338](https://arxiv.org/abs/2106.14338).
- [Liu22] S. Liu. “Regret Analysis for Discounted Reinforcement Learning”. PhD thesis. 2022. eprint: <https://escholarship.org/uc/item/3kd0q8g0>.
- [DWC+19] Kefan Dong et al. *Q-learning with UCB Exploration is Sample Efficient for Infinite-Horizon MDP*. 2019. eprint: [arXiv: 1901.09311](https://arxiv.org/abs/1901.09311).



- 
- [Agr95] Rajeev Agrawal. “Sample mean based index policies by  $O(\log n)$  regret for the multi-armed bandit problem”. In: *Adv. Appl. Probab.* 27.4 (1995), pp. 1054–1078. ISSN: 0001-8678. DOI: [10 . 2307 / 1427934](https://doi.org/10.2307/1427934).
- [ACBF02] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. “Finite-time analysis of the multiarmed bandit problem”. In: *Mach. Learn.* 47.2-3 (2002), pp. 235–256. ISSN: 08856125. DOI: [10 . 1023/A : 1013689704352](https://doi.org/10.1023/A:1013689704352).
- [Tho35] William R. Thompson. “On the Theory of Apportionment”. In: *Am. J. Math.* 57.2 (1935), p. 450. ISSN: 00029327. DOI: [10 . 2307 / 2371219](https://doi.org/10.2307/2371219).
- [Sco10] Steven L Scott. “A modern Bayesian look at the multi-armed bandit”. In: *Appl. Stoch. Model. Bus. Ind.* 26.6 (2010), pp. 639–658. ISSN: 15241904. DOI: [10 . 1002/asmb . 874](https://doi.org/10.1002/asmb.874).
- [RVK+18] Daniel J. Russo et al. *A tutorial on Thompson sampling*. 2018. DOI: [10 . 1561/22000000070](https://doi.org/10.1561/22000000070). arXiv: [1707 . 02038](https://arxiv.org/abs/1707.02038).
- [KKM12] Emilie Kaufmann, Nathaniel Korda, and Rémi Munos. “Thompson Sampling: An Asymptotically Optimal Finite-Time Analysis”. In: *Algorithmic Learning Theory*. Ed. by Nader H. Bshouty et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 199–213. ISBN: 978-3-642-34106-9.
- [AG11] Shipra Agrawal and Navin Goyal. *Analysis of Thompson Sampling for the multi-armed bandit problem*. 2011. eprint: [arXiv : 1111 . 1797](https://arxiv.org/abs/1111.1797).
- [GMS19] Aurélien Garivier, Pierre Ménard, and Gilles Stoltz. “Explore first, exploit next: The true shape of regret in bandit problems”. In: *Math. Oper. Res.* 44.2 (2019), pp. 377–399. ISSN: 15265471. DOI: [10 . 1287 / moor . 2017 . 0928](https://doi.org/10.1287/moor.2017.0928). arXiv: [1602 . 07182](https://arxiv.org/abs/1602.07182).
- [BMS11] Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. “Pure exploration in finitely-armed and continuous-armed bandits”. In: *Theor. Comput. Sci.* 412.19 (2011), pp. 1832–1852. ISSN: 03043975. DOI: [10 . 1016 / j . tcs . 2010 . 12 . 059](https://doi.org/10.1016/j.tcs.2010.12.059).
- [Bel03] Richard Bellman. *Dynamic Programming (Reprinted version)*. Dover Publications, 2003, p. 384. ISBN: 0486428095.
- [Wat89] C.J.C.H. Watkins. ““Learning from delayed rewards”. PhD thesis”. PhD thesis. Cambridge, 1989.
-

## BIBLIOGRAPHY

---

- [SLH+14] David Silver et al. “Deterministic Policy Gradient Algorithms”. In: *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*. ICML’14. Beijing, China: JMLR.org, 2014, I–387–I–395.
- [Sze10b] Csaba Szepesvári. “Algorithms for Reinforcement Learning”. In: *Synthesis Lectures on Artificial Intelligence and Machine Learning* (2010).
- [Gla63] Roy J. Glauber. “Coherent and Incoherent States of the Radiation Field”. In: *Phys. Rev.* 131 (6 1963), pp. 2766–2788. DOI: [10.1103/PhysRev.131.2766](https://doi.org/10.1103/PhysRev.131.2766).
- [OBH96] Masao Osaki, Masashi Ban, and Osamu Hirota. “Derivation and physical interpretation of the optimum detection operators for coherent-state signals”. In: *Phys. Rev. A* 54 (2 1996), pp. 1691–1701. DOI: [10.1103/PhysRevA.54.1691](https://doi.org/10.1103/PhysRevA.54.1691).
- [TBW+20] G. S. Thekkadath et al. “Engineering Schrödinger cat states with a photonic even-parity detector”. In: *Quantum* 4 (Mar. 2020), p. 239. ISSN: 2521-327X. DOI: [10.22331/q-2020-03-02-239](https://doi.org/10.22331/q-2020-03-02-239).
- [SEL+22] V. V. Sivak et al. “Model-Free Quantum Control with Reinforcement Learning”. In: *Phys. Rev. X* 12 (1 2022), p. 011059. DOI: [10.1103/PhysRevX.12.011059](https://doi.org/10.1103/PhysRevX.12.011059).
- [Dol] S. J. Dolinar. “02830334-MIT.pdf”. PhD thesis.
- [TSV+05] Masahiro Takeoka et al. “Implementation of projective measurements with linear optics and continuous photon counting”. In: *Phys. Rev. A - At. Mol. Opt. Phys.* 71.2 (2005), pp. 1–10. ISSN: 10502947. DOI: [10.1103/PhysRevA.71.022318](https://doi.org/10.1103/PhysRevA.71.022318). arXiv: [0410133 \[quant-ph\]](https://arxiv.org/abs/0410133).
- [CMG+07] Robert L. Cook et al. “Optical coherent state discrimination using a closed-loop quantum measurement”. In: *Nature* 446.7137 (2007), pp. 774–777. ISSN: 14764687. DOI: [10.1038/nature05655](https://doi.org/10.1038/nature05655).
- [Ger04] JM Geremia. “Distinguishing between optical coherent states with imperfect detection”. In: *Phys. Rev. A* 70.6 (2004), p. 062303. ISSN: 1050-2947. DOI: [10.1103/PhysRevA.70.062303](https://doi.org/10.1103/PhysRevA.70.062303). arXiv: [0407205 \[quant-ph\]](https://arxiv.org/abs/0407205).

- 
- [DGD13] Marcus P. Da Silva, Saikat Guha, and Zachary Dutton. “Achieving minimum-error discrimination of an arbitrary set of laser-light pulses”. In: *Phys. Rev. A - At. Mol. Opt. Phys.* 87.5 (2013). ISSN: 10502947. DOI: [10.1103/PhysRevA.87.052320](https://doi.org/10.1103/PhysRevA.87.052320). arXiv: [arXiv:1201.6625](https://arxiv.org/abs/1201.6625).
- [DB18] M. T. DiMario and F. E. Becerra. “Robust Measurement for the Discrimination of Binary Coherent States”. In: *Phys. Rev. Lett.* 121.2 (2018), p. 023603. ISSN: 0031-9007. DOI: [10.1103/PhysRevLett.121.023603](https://doi.org/10.1103/PhysRevLett.121.023603).
- [RIR+21] David E. Roberson et al. “Limit of Gaussian operations and measurements for Gaussian state discrimination and its application to state comparison”. In: *Phys. Rev. A* 103 (2 2021), p. 022423. DOI: [10.1103/PhysRevA.103.022423](https://doi.org/10.1103/PhysRevA.103.022423).
- [TS08] Masahiro Takeoka and Masahide Sasaki. “Discrimination of the binary coherent signal: Gaussian-operation limit and simple non-Gaussian near-optimal receivers”. In: *Phys. Rev. A* 78.2 (2008), p. 022320. ISSN: 1050-2947. DOI: [10.1103/PhysRevA.78.022320](https://doi.org/10.1103/PhysRevA.78.022320).
- [SW21] Krishna Kumar Sabapathy and Andreas Winter. *Bosonic data hiding: power of linear vs non-linear optics*. 2021. eprint: [arXiv:2102.01622](https://arxiv.org/abs/2102.01622).
- [GG21] Christos N. Gagatsos and Saikat Guha. “Impossibility to produce arbitrary non-Gaussian states using zero-mean Gaussian states and partial photon number resolving detection”. In: *Phys. Rev. Research* 3 (4 2021), p. 043182. DOI: [10.1103/PhysRevResearch.3.043182](https://doi.org/10.1103/PhysRevResearch.3.043182).
- [Ken73] Robert S. Kennedy. “Near-Optimum Receiver for the Binary Coherent State Quantum Channel”. In: *MIT Res. Lab. Electron. Q. Prog. Rep.* 108 (1973), p. 219.
- [Bel64] J. S. Bell. *On the Einstein Podolsky Rosen paradox*. 1964. DOI: [10.1002/prop.19800281202](https://doi.org/10.1002/prop.19800281202). arXiv: [1409.4807](https://arxiv.org/abs/1409.4807).
- [Ber05] Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control*. 3rd. Vol. I. Belmont, MA, USA: Athena Scientific, 2005.
- [ADPP11] Antonio Assalini, Nicola Dalla Pozza, and Gianfranco Pierobon. “Revisiting the Dolinar receiver through multiple-copy state discrimination theory”. In: *Phys. Rev. A* 84 (2 2011), p. 022342. DOI: [10.1103/PhysRevA.84.022342](https://doi.org/10.1103/PhysRevA.84.022342).

## BIBLIOGRAPHY

---

- [Ban94] Masashi Ban. “Quasicontinuous measurements of photon number”. In: *Phys. Rev. A* 49 (6 1994), pp. 5078–5085. DOI: [10.1103/PhysRevA.49.5078](https://doi.org/10.1103/PhysRevA.49.5078).
- [SL14] Denis Sych and Gerd Leuchs. “Optimal receiver for binary coherent signals”. In: *arXiv* (2014), p. 1404.5033. DOI: [10.1103/PhysRevLett.117.200501](https://doi.org/10.1103/PhysRevLett.117.200501). arXiv: [1404.5033](https://arxiv.org/abs/1404.5033).
- [Rbf] “Radial basis interpolation”. In: *University of Stellenbosch (thesis)* (2008).
- [VGO+20] Pauli Virtanen et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17 (2020), pp. 261–272. DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).
- [BRC21] M. Bilkis, Matteo Rosati, and John Calsamiglia. “Reinforcement-learning calibration of coherent-state receivers on variable-loss optical channels”. In: *2021 IEEE Information Theory Workshop (ITW)*. 2021, pp. 1–6. DOI: [10.1109/ITW48936.2021.9611396](https://doi.org/10.1109/ITW48936.2021.9611396).
- [XSF+97] Y Xiang et al. “Generalized simulated annealing algorithm and its application to the Thomson model”. In: *Physics Letters A* 233.3 (1997), pp. 216–220. ISSN: 0375-9601. DOI: [https://doi.org/10.1016/S0375-9601\(97\)00474-X](https://doi.org/10.1016/S0375-9601(97)00474-X).
- [JEP+21] John Jumper et al. “Highly accurate protein structure prediction with AlphaFold”. In: *Nature* 596.7873 (2021), pp. 583–589. ISSN: 1476-4687. DOI: [10.1038/s41586-021-03819-2](https://doi.org/10.1038/s41586-021-03819-2).
- [SFB20] Jonas Schuff, Lukas J Fiderer, and Daniel Braun. “Improving the dynamics of quantum sensors with reinforcement learning”. In: *New Journal of Physics* 22.3 (2020), p. 035001. DOI: [10.1088/1367-2630/ab6f1f](https://doi.org/10.1088/1367-2630/ab6f1f).
- [WMD+20] Julius Wallnöfer et al. “Machine Learning for Long-Distance Quantum Communication”. In: *PRX Quantum* 1 (1 2020), p. 010301. DOI: [10.1103/PRXQuantum.1.010301](https://doi.org/10.1103/PRXQuantum.1.010301).
- [CAB+21b] M. Cerezo et al. “Variational quantum algorithms”. In: *Nature Reviews Physics* 3.9 (2021), pp. 625–644. ISSN: 2522-5820. DOI: [10.1038/s42254-021-00348-9](https://doi.org/10.1038/s42254-021-00348-9).
- [FNM+21] Thomas Fösel et al. *Quantum circuit optimization with deep reinforcement learning*. 2021. eprint: [arXiv:2103.07585](https://arxiv.org/abs/2103.07585).

- 
- [CT16] Giuseppe Carleo and Matthias Troyer. “Solving the Quantum Many-Body Problem with Artificial Neural Networks”. In: *Science* 355.6325 (2016), pp. 602–606. ISSN: 0036-8075. DOI: [10.1126/science.aag2302](https://doi.org/10.1126/science.aag2302). arXiv: [1606.02318](https://arxiv.org/abs/1606.02318).
- [TM17] Giacomo Torlai and Roger G. Melko. “Neural Decoder for Topological Codes”. In: *Phys. Rev. Lett.* 119.3 (2017). ISSN: 10797114. DOI: [10.1103/PhysRevLett.119.030501](https://doi.org/10.1103/PhysRevLett.119.030501). arXiv: [1610.04238](https://arxiv.org/abs/1610.04238).
- [VLH17] Evert P.L. Van Nieuwenburg, Ye Hua Liu, and Sebastian D. Huber. “Learning phase transitions by confusion”. In: *Nat. Phys.* 13.5 (2017), pp. 435–439. ISSN: 17452481. DOI: [10.1038/nphys4037](https://doi.org/10.1038/nphys4037). arXiv: [1610.02048](https://arxiv.org/abs/1610.02048).
- [CM17] Juan Carrasquilla and Roger G. Melko. “Machine learning phases of matter”. In: *Nat. Phys.* 13.5 (2017), pp. 431–434. ISSN: 17452481. DOI: [10.1038/nphys4035](https://doi.org/10.1038/nphys4035). arXiv: [1605.01735](https://arxiv.org/abs/1605.01735).
- [TMC+18] Giacomo Torlai et al. “Neural-network quantum state tomography”. In: *Nat. Phys.* 14.5 (2018), pp. 447–450. ISSN: 17452481. DOI: [10.1038/s41567-018-0048-5](https://doi.org/10.1038/s41567-018-0048-5). arXiv: [1703.05334](https://arxiv.org/abs/1703.05334).
- [MPK+18] Alexey A Melnikov et al. “Active learning machine learns to create new quantum experiments.” In: *Proc. Natl. Acad. Sci. U. S. A.* 115.6 (2018), pp. 1221–1226. ISSN: 1091-6490. DOI: [10.1073/pnas.1714936115](https://doi.org/10.1073/pnas.1714936115). arXiv: [1706.00868](https://arxiv.org/abs/1706.00868).
- [FTW+18] Thomas Fösel et al. “Reinforcement Learning with Neural Networks for Quantum Feedback”. In: *Phys. Rev. X* 8.3 (2018). ISSN: 21603308. DOI: [10.1103/PhysRevX.8.031084](https://doi.org/10.1103/PhysRevX.8.031084). arXiv: [1802.05267](https://arxiv.org/abs/1802.05267).
- [WMD+19] Julius Wallnöfer et al. “Machine learning for long-distance quantum communication”. In: (2019). arXiv: [1904.10797](https://arxiv.org/abs/1904.10797).
- [BDS+18] Marin Bukov et al. “Reinforcement Learning in Different Phases of Quantum Control”. In: *Phys. Rev. X* 8.3 (2018). ISSN: 21603308. DOI: [10.1103/PhysRevX.8.031086](https://doi.org/10.1103/PhysRevX.8.031086). arXiv: [1705.00565](https://arxiv.org/abs/1705.00565).
- [NBS+19] Murphy Yuezhen Niu et al. “Universal quantum control through deep reinforcement learning”. In: *AIAA Scitech 2019 Forum*. Vol. 5. 1. Nature Publishing Group, 2019, p. 33. ISBN: 9781624105784. DOI: [10.2514/6.2019-0954](https://doi.org/10.2514/6.2019-0954). arXiv: [1803.01857](https://arxiv.org/abs/1803.01857).

## BIBLIOGRAPHY

---

- [Doy00] Kenji Doya. “Reinforcement Learning in Continuous Time and Space”. In: *Neural Computation* 12.1 (2000), pp. 219–245. DOI: [10.1162/089976600300015961](https://doi.org/10.1162/089976600300015961).
- [YHL21] Çağatay Yıldız, Markus Heinonen, and Harri Lähdesmäki. *Continuous-Time Model-Based Reinforcement Learning*. 2021. eprint: [arXiv: 2102.04764](https://arxiv.org/abs/2102.04764).
- [JEJ20] (directed by John Calsamiglia) José Esté Jaloveckas. *Model-free approach to coherent state discrimination*. 2020.
- [LHP+15] Timothy P. Lillicrap et al. *Continuous control with deep reinforcement learning*. 2015. eprint: [arXiv:1509.02971](https://arxiv.org/abs/1509.02971).
- [Bil02a] Matías Bilkis. 202.
- [BC20] Matias Bilkis and John Calsamiglia. 2020.
- [Bil02b] Matías Bilkis. 202.
- [MLE21] Vishal Monga, Yuelong Li, and Yonina C. Eldar. “Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing”. In: *IEEE Signal Processing Magazine* 38.2 (2021), pp. 18–44. DOI: [10.1109/MSP.2020.3016905](https://doi.org/10.1109/MSP.2020.3016905).
- [SFE+20] Nir Shlezinger et al. “ViterbiNet: A Deep Learning Based Viterbi Algorithm for Symbol Detection”. eng. In: *IEEE Transactions on Wireless Communications* 19.5 (2020), p. 13. ISSN: 1536-1276.
- [LHB04] Yann LeCun, Fu Jie Huang, and Léon Bottou. “Learning methods for generic object recognition with invariance to pose and lighting”. English (US). In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2* (2004). Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004 ; Conference date: 27-06-2004 Through 02-07-2004, pp. II97–II104. ISSN: 1063-6919.
- [KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Advances in Neural Information Processing Systems*. Ed. by F. Pereira et al. Vol. 25. Curran Associates, Inc., 2012.
- [ON15] Keiron O’Shea and Ryan Nash. *An Introduction to Convolutional Neural Networks*. 2015. eprint: [arXiv:1511.08458](https://arxiv.org/abs/1511.08458).

- 
- [MDM+08] D. Maslov et al. “Quantum Circuit Simplification and Level Compaction”. In: *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 27.3 (2008), pp. 436–444. DOI: [10.1109/TCAD.2007.911334](https://doi.org/10.1109/TCAD.2007.911334).
- [Has70] W. K. Hastings. “Monte Carlo Sampling Methods Using Markov Chains and Their Applications”. In: *Biometrika* 57.1 (1970), pp. 97–109. DOI: [10.2307/2334940](https://doi.org/10.2307/2334940).
- [ROAG17] Jonathan Romero, Jonathan P Olson, and Alan Aspuru-Guzik. “Quantum autoencoders for efficient compression of quantum data”. In: *Quantum Science and Technology* 2.4 (2017), p. 045001. DOI: [10.1088/2058-9565/aa8072](https://doi.org/10.1088/2058-9565/aa8072).
- [YC] E. Younis and L. Cincio. *Quantum Fast Circuit Optimizer (qFactor)*.
- [CRC+17] M. Cerezo et al. “Factorization and Criticality in Finite  $XXZ$  Systems of Arbitrary Spin”. In: *Physical Review Letters* 119.22 (2017), p. 220605. DOI: [10.1103/PhysRevLett.119.220605](https://doi.org/10.1103/PhysRevLett.119.220605).
- [CLKAG20] Alba Cervera-Lierta, Jakob S Kottmann, and Alán Aspuru-Guzik. “The meta-variational quantum eigensolver (meta-vqe): Learning energy profiles of parameterized hamiltonians for quantum simulation”. In: *arXiv preprint arXiv:2009.13545* (2020).
- [MEAG+20a] Sam McArdle et al. “Quantum computational chemistry”. In: *Rev. Mod. Phys.* 92 (1 2020), p. 015003. DOI: [10.1103/RevModPhys.92.015003](https://doi.org/10.1103/RevModPhys.92.015003).
- [MSK+19] Jarrod R. McClean et al. *OpenFermion: The Electronic Structure Package for Quantum Computers*. 2019.
- [MEAG+20b] Sam McArdle et al. “Quantum computational chemistry”. In: *Reviews of Modern Physics* 92.1 (2020), p. 015003. DOI: [10.1103/RevModPhys.92.015003](https://doi.org/10.1103/RevModPhys.92.015003).
- [CHC+22] Matthias C Caro et al. “Generalization in quantum machine learning from few training data”. In: *Nature Communications* 13, 4919 (2022). DOI: [10.1038/s41467-022-32550-3](https://doi.org/10.1038/s41467-022-32550-3).
- [GEZ21] Konstantinos Georgopoulos, Clive Emary, and Paolo Zuliani. “Modeling and simulating the noisy behavior of near-term quantum computers”. In: *Phys. Rev. A* 104 (6 2021), p. 062432. DOI: [10.1103/PhysRevA.104.062432](https://doi.org/10.1103/PhysRevA.104.062432).

## BIBLIOGRAPHY

---

- [LMR+22] Ryan LaRose et al. “Error mitigation increases the effective quantum volume of quantum computers”. In: *arXiv preprint arXiv:2203.05489* (2022).
- [MMM+20] Prakash Murali et al. “Software Mitigation of Crosstalk on Noisy Intermediate-Scale Quantum Computers”. In: *Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems*. ASPLOS ’20. Lausanne, Switzerland: Association for Computing Machinery, 2020, 1001–1016. ISBN: 9781450371025. DOI: [10 . 1145 / 3373376 . 3378477](https://doi.org/10.1145/3373376.3378477).
- [OPG+04] J. L. O’Brien et al. “Quantum Process Tomography of a Controlled-NOT Gate”. In: *Phys. Rev. Lett.* 93 (8 2004), p. 080502. DOI: [10 . 1103/PhysRevLett . 93 . 080502](https://doi.org/10.1103/PhysRevLett.93.080502).
- [YZKK+14] Joel Yuen-Zhou et al. “The process matrix and how to determine it: quantum process tomography”. In: *Ultrafast Spectroscopy*. 2053-2563. IOP Publishing, 2014, 1–1 to 1–9. ISBN: 978-0-750-31062-8. DOI: [10 . 1088/978-0-750-31062-8ch1](https://doi.org/10.1088/978-0-750-31062-8ch1).
- [BPR+20] Carsten Blank et al. “Quantum classifier with tailored quantum kernel”. In: *npj Quantum Information* 6.1 (2020), p. 41. ISSN: 2056-6387. DOI: [10 . 1038/s41534-020-0272-6](https://doi.org/10.1038/s41534-020-0272-6).
- [QBW22] Nils Quetschlich, Lukas Burgholzer, and Robert Wille. *Predicting Good Quantum Circuit Compilation Options*. 2022. eprint: [arXiv : 2210 . 08027](https://arxiv.org/abs/2210.08027).
- [MPR+21] Lorenzo Moro et al. “Quantum compiling by deep reinforcement learning”. In: *Communications Physics* 4.1 (2021), p. 178. ISSN: 2399-3650. DOI: [10 . 1038/s42005-021-00684-3](https://doi.org/10.1038/s42005-021-00684-3).
- [HM22] David A. Herrera-Martí. “Policy Gradient Approach to Compilation of Variational Quantum Circuits”. In: *Quantum* 6 (Sept. 2022), p. 797. ISSN: 2521-327X. DOI: [10 . 22331/q-2022-09-08-797](https://doi.org/10.22331/q-2022-09-08-797).
- [BPLC+19] Carlos Bravo-Prieto et al. “Variational Quantum Linear Solver”. In: *arXiv preprint arXiv:1909.05820* (2019).
- [HBR19] Hsin-Yuan Huang, Kishor Bharti, and Patrick Rebentrost. “Near-term quantum algorithms for linear systems of equations”. In: *arXiv preprint arXiv:1909.07344* (2019).



- 
- [XSE+21] Xiaosi Xu et al. “Variational algorithms for linear algebra”. In: *Science Bulletin* 66.21 (2021), pp. 2181–2188. DOI: [10 . 1016 / j . scib . 2021 . 06 . 023](https://doi.org/10.1016/j.scib.2021.06.023).
- [AOAG+19] Eric Anschuetz et al. “Variational quantum factoring”. In: *International Workshop on Quantum Technology and Optimization Problems*. Springer. 2019, pp. 74–85.
- [KLP+19] Sumeet Khatri et al. “Quantum-assisted quantum compiling”. In: *Quantum* 3 (2019), p. 140. DOI: [10 . 22331 / q - 2019 - 05 - 13 - 140](https://doi.org/10.22331/q-2019-05-13-140).
- [SKC+20] Kunal Sharma et al. “Noise resilience of variational quantum compiling”. In: *New Journal of Physics* 22.4 (2020), p. 043006. DOI: [10 . 1088 / 1367 - 2630 / ab784c](https://doi.org/10.1088/1367-2630/ab784c).
- [BCS+22] Jacob L Beckey et al. “Variational quantum algorithm for estimating the quantum Fisher information”. In: *Physical Review Research* 4.1 (2022), p. 013083. DOI: [10 . 1103 / PhysRevResearch . 4 . 013083](https://doi.org/10.1103/PhysRevResearch.4.013083).
- [KEJ+20] Bálint Koczor et al. “Variational-State Quantum Metrology”. In: *New Journal of Physics* (2020). DOI: [10 . 1088 / 1367 - 2630 / ab965e](https://doi.org/10.1088/1367-2630/ab965e).
- [LTOJ+19] Ryan LaRose et al. “Variational quantum state diagonalization”. In: *npj Quantum Information* 5.1 (2019), pp. 1–10. DOI: [10 . 1038 / s41534 - 019 - 0167 - 6](https://doi.org/10.1038/s41534-019-0167-6).
- [SSP14] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. “The quest for a quantum neural network”. In: *Quantum Information Processing* 13.11 (2014), pp. 2567–2586. DOI: [10 . 1007 / s11128 - 014 - 0809 - 8](https://doi.org/10.1007/s11128-014-0809-8).
- [ASZ+21] Amira Abbas et al. “The power of quantum neural networks”. In: *Nature Computational Science* 1.6 (2021), pp. 403–409. DOI: [10 . 1038 / s43588 - 021 - 00084 - 1](https://doi.org/10.1038/s43588-021-00084-1).
- [VMN+19] Guillaume Verdon et al. “Quantum Hamiltonian-based models and the variational quantum thermalizer algorithm”. In: *arXiv preprint arXiv:1910.02071* (2019).
- [BHH16] Fernando GSL Brandao, Aram W Harrow, and Michał Horodecki. “Local random quantum circuits are approximate polynomial-designs”. In: *Communications in Mathematical Physics* 346.2 (2016), pp. 397–434. DOI: [10 . 1007 / s00220 - 016 - 2706 - 8](https://doi.org/10.1007/s00220-016-2706-8).
-

## BIBLIOGRAPHY

---

- [DCE+09] Christoph Dankert et al. “Exact and approximate unitary 2-designs and their application to fidelity estimation”. In: *Physical Review A* 80.1 (2009), p. 012304.
- [HL09] Aram W Harrow and Richard A Low. “Random quantum circuits are approximate 2-designs”. In: *Communications in Mathematical Physics* 291.1 (2009), pp. 257–302. DOI: [10 . 1007 / s00220 - 009 - 0873 - 6](https://doi.org/10.1007/s00220-009-0873-6).
- [HM18] Aram Harrow and Saeed Mehraban. “Approximate unitary  $t$ -designs by short random quantum circuits using nearest-neighbor and long-range gates”. In: *arXiv preprint arXiv:1809.06957* (2018).
- [Haf22] Jonas Haferkamp. “Random quantum circuits are approximate unitary  $t$ -designs in depth  $O(nt^{5+o(1)})$ ”. In: *Quantum* 6 (Sept. 2022), p. 795. ISSN: 2521-327X. DOI: [10 . 22331 / q - 2022 - 09 - 08 - 795](https://doi.org/10.22331/q-2022-09-08-795).
- [SMM+22] Stefan H Sack et al. “Avoiding barren plateaus using classical shadows”. In: *PRX Quantum* 3.2 (2022), p. 020365. DOI: [10 . 1103 / PRXQuantum . 3 . 020365](https://doi.org/10.1103/PRXQuantum.3.020365).
- [HKP20] Hsin-Yuan Huang, Richard Kueng, and John Preskill. “Predicting many properties of a quantum system from very few measurements”. In: *Nature Physics* 16.10 (2020), pp. 1050–1057. DOI: [10 . 1038 / s41567 - 020 - 0932 - 7](https://doi.org/10.1038/s41567-020-0932-7).
- [KBA+19] Nathan Killoran et al. “Continuous-variable quantum neural networks”. In: *Phys. Rev. Research* 1 (3 2019), p. 033063. DOI: [10 . 1103 / PhysRevResearch . 1 . 033063](https://doi.org/10.1103/PhysRevResearch.1.033063).
- [Bil02c] Matías Bilkis. 202.
- [Bil02d] Matías Bilkis. 202.
- [BC21] Matias Bilkis and John Calsamiglia. 2021.
- [KM18] Alexander Holm Kiilerich and Klaus Mølmer. “Hypothesis testing with a continuously monitored quantum system”. In: *Phys. Rev. A* 98 (2 2018), p. 022103. DOI: [10 . 1103 / PhysRevA . 98 . 022103](https://doi.org/10.1103/PhysRevA.98.022103).
- [LHT22] Yonglong Li, Christoph Hirche, and Marco Tomamichel. “Sequential Quantum Channel Discrimination”. In: *2022 IEEE International Symposium on Information Theory (ISIT)*. Espoo, Finland: IEEE Press, 2022, 270–275. DOI: [10 . 1109 / ISIT50566 . 2022 . 9834768](https://doi.org/10.1109/ISIT50566.2022.9834768).

- [FRT+22] Alessio Fallani et al. “Learning Feedback Control Strategies for Quantum Metrology”. In: *PRX Quantum* 3 (2 2022), p. 020310. DOI: [10.1103/PRXQuantum.3.020310](https://doi.org/10.1103/PRXQuantum.3.020310).
- [Lau78] Alan J. Laub. “A Schur method for solving algebraic Riccati equations”. In: *MIT Libraries* (1978).
- [Abu17] Matthew J. Aburn. “Critical fluctuations and coupling of stochastic neural mass models”. PhD thesis. The University of Queensland, Jan. 2017. DOI: [10.14264/uq1.2017.148](https://doi.org/10.14264/uq1.2017.148).
- [Rö9] Andreas Rößler. “Second Order Runge–Kutta Methods for Itô Stochastic Differential Equations”. In: *SIAM Journal on Numerical Analysis* 47.3 (2009), pp. 1713–1738. DOI: [10.1137/060673308](https://doi.org/10.1137/060673308). eprint: <https://doi.org/10.1137/060673308>.
- [BPK16] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. “Discovering governing equations from data by sparse identification of nonlinear dynamical systems”. In: *Proceedings of the National Academy of Sciences* 113.15 (2016), pp. 3932–3937. DOI: [10.1073/pnas.1517384113](https://doi.org/10.1073/pnas.1517384113). eprint: <https://www.pnas.org/doi/pdf/10.1073/pnas.1517384113>.
- [CDG+10] A. A. Clerk et al. “Introduction to quantum noise, measurement, and amplification”. In: *Rev. Mod. Phys.* 82 (2 2010), pp. 1155–1208. DOI: [10.1103/RevModPhys.82.1155](https://doi.org/10.1103/RevModPhys.82.1155).
- [GG] J. Calsamiglia G. Gasbarri M. Bilkis. <https://github.com/matibilkis/cdisc>.