Neural-Kalman Schemes for Non-Stationary Channel Tracking and Learning

A dissertation submitted by

Diego Méndez Romero

in partial fulfillment of the requirements for the degree of Doctor of Philosophy in

Multimedia and Communications

Universidad Carlos III de Madrid

Advisor/Tutor:

M. Julia Fernández-Getino García

Leganés, February 2023

I feel very grateful for the opportunity to research this crucial topic: abrupt changes.
Abrupt changes is where miracles come true.
To my little abrupt changes in life, Alonso and Estela.
To my wife Ángela and the abrupt changes we will enjoy together.
To my tutor Mª Julia, for bearing with me through so many abrupt changes, both in our simulations and the wider world.
To my parents, for whom I once was an abrupt change.

Abstract

This Thesis focuses on channel tracking in Orthogonal Frequency-Division Multiplexing (OFDM), a widely-used method of data transmission in wireless communications, when abrupt changes occur in the channel. In highly mobile applications, new dynamics appear that might make channel tracking non-stationary, e.g. channels might vary with location, and location rapidly varies with time. Simple examples might be the different channel dynamics a train receiver faces when it is close to a station vs. crossing a bridge vs. entering a tunnel, or a car receiver in a route that grows more traffic-dense. Some of these dynamics can be modelled as channel taps dying or being reborn, and so tap birth-death detection is of the essence.

In order to improve the quality of communications, we delved into mathematical methods to detect such abrupt changes in the channel, such as the mathematical areas of Sequential Analysis/Abrupt Change Detection and Random Set Theory (RST), as well as the engineering advances in Neural Network schemes. This knowledge helped us find a solution to the problem of abrupt change detection by informing and inspiring the creation of low-complexity implementations for real-world channel tracking. In particular, two such novel trackers were created: the Simplified Maximum A Posteriori (SMAP) and the Neural-Network-switched Kalman Filtering (NNKF) schemes.

The SMAP is a computationally inexpensive, threshold-based abrupt-change detector. It applies the three following heuristics for tap birth-death detection: a) detect death if the tap gain jumps into approximately zero (memoryless detection); b) detect death if the tap gain has slowly converged into approximately zero (memory detection); c) detect birth if the tap gain is far from zero.

The precise parameters for these three simple rules can be approximated with simple theoretical derivations and then fine-tuned through extensive simulations. The status detector for each tap using only these three computationally inexpensive threshold comparisons achieves an error reduction matching that of a close-to-perfect path death/birth detection, as shown in simulations.

This estimator was shown to greatly reduce channel tracking error in the target Signal-to-Noise Ratio (SNR) range at a very small computational cost, thus outperforming previously known systems. The underlying RST framework for the SMAP was then extended to combined death/birth and SNR detection when SNR is dynamical and may drift. We analyzed how different quasi-ideal SNR detectors affect the SMAP-enhanced Kalman tracker's performance. Simulations showed SMAP is robust to SNR drift in simulations, although it was also shown to benefit from an accurate SNR detection.

The core idea behind the second novel tracker, NNKFs, is similar to the SMAP, but now the tap birth/death detection will be performed via an artificial neuronal network (NN). Simulations show that the proposed NNKF estimator provides extremely good performance, practically identical to

a detector with 100% accuracy.

These proposed Neural-Kalman schemes can work as novel trackers for multipath channels, since they are robust to wide variations in the probabilities of tap birth and death. Such robustness suggests a single, low-complexity NNKF could be reusable over different tap indices and communication environments.

Furthermore, a different kind of abrupt change was proposed and analyzed: energy shifts from one channel tap to adjacent taps (partial tap lateral hops). This Thesis also discusses how to model, detect and track such changes, providing a geometric justification for this and additional non-stationary dynamics in vehicular situations, such as road scenarios where reflections on trucks and vans are involved, or the visual appearance/disappearance of drone swarms. An extensive literature review of empirically-backed abrupt-change dynamics in channel modelling/measuring campaigns is included.

For this generalized framework of abrupt channel changes that includes partial tap lateral hopping, a neural detector for lateral hops with large energy transfers is introduced. Simulation results suggest the proposed NN architecture might be a feasible lateral hop detector, suitable for integration in NNKF schemes.

Finally, the newly found understanding of abrupt changes and the interactions between Kalman filters and neural networks is leveraged to analyze the neural consequences of abrupt changes and briefly sketch a novel, abrupt-change-derived stochastic model for neural intelligence, extract some neurofinancial consequences of unstereotyped abrupt dynamics, and propose a new portfolio-building mechanism in finance: Highly Leveraged Abrupt Bets Against Failing Experts (HLABAFEOs). Some communication-engineering-relevant topics, such as a Bayesian stochastic stereotyper for hopping Linear Gauss-Markov (LGM) models, are discussed in the process.

The forecasting problem in the presence of expert disagreements is illustrated with a hopping LGM model and a novel structure for a Bayesian stereotyper is introduced that might eventually solve such problems through bio-inspired, neuroscientifically-backed mechanisms, like dreaming and surprise (biological Neural-Kalman). A generalized framework for abrupt changes and expert disagreements was introduced with the novel concept of Neural-Kalman Phenomena. This Thesis suggests mathematical (Neural-Kalman Problem Category Conjecture), neuro-evolutionary and social reasons why Neural-Kalman Phenomena might exist and found significant evidence for their existence in the areas of neuroscience and finance.

Apart from providing specific examples, practical guidelines and historical (out)performance for some HLABAFEO investing portfolios, this multidisciplinary research suggests that a Neural-Kalman architecture for ever granular stereotyping providing a practical solution for continual learning in the presence of unstereotyped abrupt dynamics would be extremely useful in communications and other continual learning tasks.

Published and Submitted Content

The following papers are included as part of this Thesis:

Journals

- Méndez-Romero, D. and Fernández-Getino García, M. J. (2018). Simpler Multipath Detection for Vehicular OFDM Channel Tracking. *IEEE Transactions on Vehicular Technology*, 67(11):10752-10759. DOI: 10.1109/TVT.2018.2868445

  This work is wholly included in Chapter 5. The material from this source included in this thesis is not singled out with typographic means and references.

Conferences

- Méndez-Romero, D., Fernández-Getino Garcia, M. J., Tonello, A. M., and Dobre, O. A. (2020). Neural-Network-Switched Kalman Filters as Novel Trackers for Multipath Channels. In *2020 IEEE International Conference on Communications Workshops (ICC Workshops)* [7-11 June 2020], pages 1-5, Dublin, Ireland. IEEE. DOI: 10.1109/ICCWorkshops49005.2020.9145198

  This work is wholly included in Chapter 7. The material from this source included in this thesis is not singled out with typographic means and references.

- Méndez-Romero, D. and Fernández-Getino García, M. J. (2020). Death/Birth and SNR Detection for Vehicular Kalman Channel Trackers. In *2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON)* [16-18 June 2020], pages 104-108, Palermo, Italy. IEEE. DOI: 10.1109/MELECON48756.2020.9140497

  This work is wholly included in Chapters 1, 5 and 6. The material from this source included in this thesis is not singled out with typographic means and references.

Pre-prints

- Méndez-Romero, D., Fernández-Getino Garcia, M. J. (2022). Neural-Kalman Channel Trackers and Partial-Tap Hop Detection (in preparation)

  This work is wholly included in Chapter 8. The material from this source included in this thesis is not singled out with typographic means and references.

- Méndez-Romero, D., Fernández-Getino Garcia, M. J. (2022). Neural-Kalman Conjecture and Market Hypothesis (in preparation)

  This work is wholly included in Chapter 9 and Appendix B. The material from this source included in this thesis is not singled out with typographic means and references.

- Méndez-Romero, D., Fernández-Getino Garcia, M. J. (2022). Bets Against Failing Experts: Neurofinancial Consequences of Unstereotyped Abrupt Dynamics (in preparation)

  This work is wholly included in Chapter 9 and Appendix B. The material from this source included in this thesis is not singled out with typographic means and references.

Other Research Merits

The following work has also been published:

Book Chapters

- Juan Carlos Estrada-Jiménez, M. Julia Fernández-Getino García, Diego Méndez-Romero, "Chapter 3. Pilot-Assisted Channel Estimation" in Wiley 5G REF: The Essential 5G Reference Online, Section 1: Radio Technologies, Editors: Rahim Tafazolli, Ching-Liang Wang, Periklis Chatzimisios, Ed. John Wiley & Sons Ltd., USA, pp. 1-19, May 2020. ISBN: 9781119471509. DOI: 10.1002/9781119471509.w5GRef003

- Juan Carlos Estrada-Jiménez, M. Julia Fernández-Getino García, Diego Méndez-Romero, "Chapter 4. Superimposed Training and Blind Channel Estimation " in Wiley 5G REF: The Essential 5G Reference Online, Section 1: Radio Technologies, Editors: Rahim Tafazolli, Ching-Liang Wang, Periklis Chatzimisios, Ed. John Wiley & Sons Ltd., USA, pp. 1-17, October 2019. ISBN: 9781119471509. DOI: 10.1002/9781119471509.w5GRef004

Book Translation

- Voskuil, E. (2020). Cryptoeconomics: Fundamental Principles of Bitcoin. Eric Voskuil. ISBN 1735060801, 9781735060804. Translated into Spanish as [see 1]: Voskuil, E., Taaki, A., Chiang, J. (*Illustrator*) and Méndez-Romero, D. (*Translator*) (2022). Criptoeconomía: Principios fundamentales de Bitcoin. ISBN: 9798986994611.

Contents

List of Figures

List of Tables

Mathematical Notation

General

Symbol	***Description***
$a, b, c...$	Elements in a set
$\mathbf{a}$,$\mathbf{b}$,$\mathbf{c}$...	Random elements
$\underline{a}, \underline{b}, \underline{c}...$	Vectors in a Euclidean space $\mathbb{R}^d$
$A, B, C...$	Sets
$\mathrm{A}, \mathrm{B}, \mathrm{C}...$	Matrices
$\mathcal{A}, \mathcal{B}, \mathcal{C}...$	Collections of sets or collections of collections
$[a, b]$	A closed real interval, $a \leq x \leq b$
(a, b)	An open real interval, $a < x < b$
$[a, b)$	A half-open interval, $a \leq x < b$
$[a, b, c]^T$	A column vector
$\underline{a}^T, \mathrm{A}^T$	The transpose of a vector $\underline{a}$, and a matrix A, respectively
$a \in A$	a is in A.
$A \subset B$	Set A is a subset (not necessarily a proper set) of set B
$A \supset B$	Set A is a superset (not necessarily a proper superset) in B
$A \backslash B$	The set difference $A \cap \complement B$
$\complement A$	The complement of a set A
$\|A\|$	The cardinality of set A (i.e. the number of elements it has)
$a \leftarrow b$	Value a gets replaced by value b
$n!$	Factorial of n, i.e. $\prod_{i=1}^{n} i$
$\mathbb{N}$,$\mathbb{Z}$,$\mathbb{Q}$,$\mathbb{R}$,$\mathbb{C}$	The sets of natural numbers (≥ 0),integers, rational numbers, real numbers and complex numbers, respectively
$\mathbf{x} \sim D$	Random variable $\mathbf{x}$ follows the distribution D

$\biguplus_{t=0}^{T} A_t$	Union of disjoint subsets A_t for $t \in \{0, ..., T\}$
$\int f(x)\mathrm{d}\mu(x)$	The integral of functions f w.r.t. measure μ
$\int_A f(x)\mathrm{d}\mu(x)$	The integral over set A
$\int f(x)\mathrm{d}x$	The Riemann integral over $\mathbb{R}$

Random Variables

Symbol	***Description***
$c(a, b)$	Bayesian cost function representing the cost of (mis)estimating b as a
$\mathbb{E}[\mathbf{x}]$	The expectation of a random variable $\mathbf{x}$
$\mathbb{E}[\mathbf{x}\|\mathcal{G}]$	The conditional expectation of random variable $\mathbf{x}$ conditional to a $\mathcal{G}$-generated σ-algebra
$\mathbb{E}[\mathbf{x}\|\mathbf{y}]$	The conditional expectation of random variable $\mathbf{x}$ conditional to the σ-algebra induced by random variable $\mathbf{y}$
$\mathbb{E}[\mathbf{x}\|\mathbf{y}= y]$	The conditional expectation of a random variable $\mathbf{x}$ given $\mathbf{y}= y$.
$h : \Omega \to X$	A function h maps the values of a source set Ω into a target set X.
$h : (\Omega, \mathcal{M}) \to (X, \mathcal{N})$	A measurable function h maps the values of a source measurable space $(\Omega, \mathcal{M})$ into a target measurable set $(X, \mathcal{N})$
$\mathcal{N}(\mu, \sigma^2)$	The Normal distribution centered on μ, with variance σ^2
P	The probability measure in $(\Omega, \mathcal{M})$.

Communications

Symbol	***Description***
$c(a, b)$	Bayesian cost function representing the cost of (mis)estimating b as a
$\delta[k - l]$	Kronecker's Delta function
$\circledast$	Cyclic convolution operator
$\nwarrow$	Partial-tap left-hop
$\nearrow$	Partial-tap right-hop

Random Sets

Symbol	***Description***
$\int f(Z)\delta Z$	Set integral

Kalman Filtering

Symbol	***Description***
F	Dynamic coefficient matrix
G	Coupling matrix
H	Measurement sensitivity matrix
K	Kalman gain matrix
P	Covariance matrix of the uncertainty in the estimation of the state
Q	Covariance matrix of process noise in the dynamics of system states
R	Covariance matrix of observational uncertainty (of measurements)
$\underline{\mathbf{x}}$	(Random) vector of states of a linear dynamical system
$\underline{\mathbf{z}}$	(Random) vector of measured values
Φ	State transition matrix of a discrete linear dynamical system
κ	Scalar Kalman gain
$\mathbf{x}$	State (random, scalar) of a linear dynamical system
$\underline{\mathbf{x}}_k$	The kth component of the random vector $\underline{\mathbf{x}}$
$\underline{\mathbf{x}}_k$	The kth element in the random vector sequence ..., $\underline{\mathbf{x}}_{k-1}$,$\underline{\mathbf{x}}_k$,$\underline{\mathbf{x}}_{k+1}$
$\hat{\underline{\mathbf{x}}}$	An estimate for $\underline{\mathbf{x}}$.
$\hat{\underline{\mathbf{x}}}_k(-)$	*Prior* estimate for $\underline{\mathbf{x}}_k$, conditional on all previous measurements except the one at time t_k.
$\hat{\underline{\mathbf{x}}}_k(+)$	*Posterior* estimate for $\underline{\mathbf{x}}_k$, conditional on all previous measurements, including the one at time t_k.
$\dot{\mathbf{x}}$	Derivative of $\mathbf{x}$ with respect to t (time).
$\underline{\mathbf{w}}$	Process noise $(1 \times r)$ vector
$\underline{\mathbf{z}}$	Measurement $(1 \times l)$ vector
$\underline{\mathbf{v}}$	Measurement noise $(1 \times l)$ vector
$\Delta(k-l)$	Kronecker's delta function

List of Abbreviations

Abbreviation	***Meaning***
4G	Fourth-Generation Mobile Communications
5G	Fifth-Generation Mobile Communications
B5G	Beyond-Fifth-Generation Mobile Communications
6G	Sixth-Generation Mobile Communications
ADSL	Asymmetric Digital Subscriber Line
AR	Auto-regressive
AVE	Alta Velocidad Española (Spain's High Speed Rail)
BBG	Bloomberg
BEM	Basis Expansion Model
BER	Bit Error Rate
BoE	Bank of England
BPSK	Binary Phase Shift Keying
CDL	Clustered Delay Line
CP	Cyclic Prefix
CPI	Consumer Price Index
CTMSE	Channel Tracking Mean Squared Error
CUSUM	CUmulative SUM
DB	Deutsche Bahn (Germany's Rail Corporation)
DFT	Discrete Fourier Transform
ETSI	European Telecommunications Standards Institute
EVT	Extreme Value Theory
KF	Kalman Filtering (or Kalman Filter)
FMA	Finite Moving Average algorithm

FFT	Fast Fourier Transform
FISST	FInite Set STatistics
GLR	Generalized Likelihood Ratio
GMA	Geometric Moving Average algorithm
GMAP	Global Maximum A Posteriori
GMAP-I	Global Maximum A Posteriori of the first kind
GMAP-II	Global Maximum A Posteriori of the second kind
GMAP-III	Global Maximum A Posteriori of the third kind
GPD	Generalized Pareto Distribution
HKF	Hypernetwork Kalman Filter
HLABAFEO	Highly Leveraged Abrupt Bet Against Failing Expert Opinions
HMM	Hidden Markov Model
HSR	High-Speed Rail
ICE	InterCity Express
ICI	Inter-Carrier Interference
IDFT	Inverse Discrete Fourier Transform
IEEE	Institute of Electrical and Electronics Engineers
IFFT	Inverse Fast Fourier Transform
i.i.d.	independent and identically distributed
ISD	Ideal SNR Detector
ISI	Inter-Symbol Interference
ISS	Ideal Switching System
LAN	Local Area Network
LGM	Linear Gauss-Markov Model
LOS	Line Of Sight
LQE	Linear Quadratic Estimator
LQG	Linear Quadratic Gaussian
LSTM	Long Short-Term Memory
MAP	Maximum A Posteriori
MIMO	Multiple Input, Multiple Output

MKF	Modified Kalman Filter
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MPC	Multipath Component
MSE	Mean Squared Error
NN	Neural Network
NNKF	Neural-Network-switched Kalman Filter
OFDM	Orthogonal Frequency-Division Multiplexing
ONS	Office for National Statistics (United Kingdom)
PF	Particle Filtering
PSAM	Pilot Symbol Assisted Modulation
PSK	Phase Shift Keying
PTC	Partial Tap Component
QAM	Quadrature Amplitude Modulation
QISD	Quasi-Ideal SNR Detector
QISE	Quasi-Ideal SNR Estimator
QoS	Quality of Service
QRD	QR Decomposition
RNN	Recurrent Neural Network
r.v.	random variable
RST	Random Set Theory
SAR	Seasonally Adjusted Rate
SMAP	Simplified Maximum A Posteriori
SMC	Sequential Monte Carlo
SNR	Signal-to-Noise Ratio
SPRT	Sequential Probability Ratio Test
SSS	Stochastic State Stereotype
TDL	Tapped Delay Line
UAV	Unmanned Aerial Vehicle
UMTS	Universal Mobile Telecommunications System

URLLC	Ultra-Reliable Low Latency Communications
UWB	Ultra-wideband
WLAN	Wireless Local Area Network

Part I

Review of the Problem

Chapter 1

Introduction

This Thesis focuses on channel tracking in Orthogonal Frequency-Division Multiplexing (OFDM), a widely-used method of data transmission in wireless communications, when abrupt changes occur in the channel. In order to improve the quality of communications, we will delve into mathematical methods to detect such abrupt changes in the channel.

1.1 Why Are Abrupt Changes Important?

"When experts are wrong, it's often because they're experts on an earlier version of the world." [Graham, 2014] In these words written by serial entrepreneur & essayist Paul Graham, we grasp our first intuition of an abrupt change: a change that suddenly renders expert opinions wrong, despite the fact that those experts (whether human or machine systems) used to be right before the change.

It's not difficult to find examples of abrupt changes that made previously succesful expert frameworks obsolete or in need of a serious update. Among human social affairs, the sudden outbreak of a pandemic, an economic event (sudden inflation, sudden stop in international trade or any other Black Swan [Taleb, 2007]), the sudden outbreak of war or the invention of a game-changing weapon [Betts, 1987] force an update on social, economic and geopolitical expectations.

This notion also applies to purely technical systems, especially when those systems are required to produce a correct 'worldview' (in the form of predictions) so they can take the right course of action. An example of such technical system is a pacemaker. The human user's life depends on the pacemaker correctly predicting whether the heart will stop beating or not.

When the heart is not beating, the pacemaker needs to be careful not to overreact. But if the user really had a heart attack, the pacemaker needs to detect such *abrupt change* and make it beat again. Similar crucial examples can be found in many other applications in biomedicine, chemical plants, etc. and, more generally, across signal processing, pattern recognition and time series analysis.

While in the previous intuition-forming examples we've mentioned changes that arguably were noticeable due to their large magnitudes, abrupt changes don't require large magnitudes to be considered as such. In fact, many performance-relevant changes are low-magnitude in the tracked metric either in an initial stage (the first 100 patients in a pandemic that will affect billions) or over the whole lifespan of the change (a small but crucial difference in average temperature in a chemical process).

This topic of unquestionable interest also bears relevance to communication channels, particularly non-stationary communication channels, such as those involving fast-moving vehicles where communication paths can get hidden or reappear.

In such channels, the wireless environment might change abruptly, rendering the expectations based on immediately previous channel tracking obsolete. This might be caused, e.g. by the appearance or disappearance of channel taps due to sudden shadowing or new reflection paths. The correct detection of abrupt changes in such cases is crucial to avoid performance degradation.

1.2 Non-Stationarity and Abrupt Changes in Communication Channels

In highly mobile applications, new dynamics appear that might make channel tracking non-stationary [Matolak, 2014, Wu, 2012], e.g. channels might vary with location, and location rapidly varies with time. Simple examples might be the different channel dynamics a train receiver faces when it is close to a station vs. crossing a bridge vs. entering a tunnel [He et al., 2016]. A similar example is a car receiver in a route that grows more traffic-dense [Hassan et al., 2020]. Such non-stationary examples can be modeled in some cases as abrupt changes in statistical properties of the underlying stochastic dynamic [Wu, 2012]. Although the methods and conclusions of this Thesis can be generalized to other problems in communication engineering, this Thesis will focus on such non-stationarity/abrupt changes in the context of vehicular OFDM communication channels.

Why is this problem currently relevant? Because the incessant growth in wireless communication demand, including new scenarios such as high-speed rail (HSR) [Guo et al., 2017], is increasing requirements for communication technologies. Orthogonal Frequency-Division Multiplexing (OFDM) has become the base for many of them, including Long-Term Evolution (LTE) systems, and it has also been put forward as a prime candidate in broadband data networks for high-mobility applications [Sheng et al., 2017]. OFDM's advantages include frequency diversity and a manageable complexity [Tse and Viswanath, 2005]. Therefore, in the wake of the 5G mobile generation and its deployment in several countries, OFDM has been considered in the first release of 5G New Radio (5G-NR) [Dahlman et al., 2020]. Vehicular applications usually require the tracking of time-variant channels. Kalman filtering (KF) and its many adaptions and extensions [Grewal and Andrews, 2015] have been proposed for such task. KF's advantage is its optimality as an estimator for linear problems: it minimizes the mean-squared estimation error for a linear stochastic system using noisy linear sensors. Thus, when a channel tracking problem can be modeled as approximately linear, the KF is the optimal solution to the linear problem. However, this approximation is not always valid. More powerful channel tracking techniques will be demanded as mobile communications spread into more dynamic channels, such as those of high-speed railways [He et al., 2016] in rugged terrain [Chen et al., 2015] or Unmanned Aerial Vehicles (UAVs) [Matolak and Sun, 2015] in suburban/ hilly terrain environments [Sun and Matolak, 2017]. In these situations, random intermittent multipath components (MPCs) appear, rendering the linear approximation for tap dynamics invalid: phenomena such as the appearance of a new tap or the disappearance of a previous tap [Mahler et al., 2017] are catastrophic for estimation performance when KF-based techniques are used. Tap birth-death conditions can be reconciled with linearity during non-birth, non-death periods by using the framework of Random Set Theory (RST) models. This makes it possible to use powerful RST-based estimators. In [Angelosante et al., 2007], [Angelosante et al., 2009], three such possible estimators were compared. However, they required an impracticable computational cost for any practical applications.

1.3 Motivation, Goals and Contributions

Kalman filters in a wide variety of variants and implementations have been proposed to improve channel tracking in wireless systems, particularly OFDM systems. Unfortunately, such filters have been shown to degrade catastrophically in performance in the presence of abrupt changes, such as those appearing in birth/death dynamics and fast drift of path contributions to taps.

Our main goal in the current research is analyze how such abrupt changes can be detected and mitigated by analyzing the body of work reserch in the Theory of Abrupt Changes (a self-sustained branch of mathematics), Random Set Theory (another branch of mathematics) and recent developments in Neural Networks, then devise and implement novel solutions to this problem in OFDM channel tracking, obtain performance metrics in communication simulations coded in Matlab, and evaluate their feasibility for practical implementation in real-world scenarios. As such, the solutions we create must be low-complexity; an optimal or asymptotically optimal high-complexity algorithm that doesn't fit this description wouldn't do the job in the real world.

Once we have successfully found novel, low-complexity solutions to abrupt change dynamics in OFDM communications, we will leverage our experience by reflecting on the nature of the problem and extracting relevant lessons to communication engineering, as well as other engineering and non-engineering problems.

Thus, our research intends to fulfill the following goals:

1. Carefully choose a channel for OFDM transmission that is suitable for the purposes of testing detection schemes for abrupt change dynamics.

2. Choose and implement a suitable Kalman filtering scheme for the purpose of tracking taps in the context of OFDM channel tracking.

3. Critically review the current literature in the mathematical areas of Sequential Analysis/Abrupt Change Detection and Random Set Theory, as well as the engineering advances in Neural Network schemes, to the extent that knowledge in such areas can help us find a solution to the problem of abrupt change detection, either directly through known algorithms or indirectly by informing and inspiring the creation of low-complexity implementations for real-world channel tracking.

4. Devise novel channel trackers that exploit the advantages of Kalman filtering while preventing degradation in the face of abrupt changes, thus providing superior performance in channel tracking. Such trackers must necessarily be low-complexity for applicability to real-world communication purposes.

   In particular, we intend to devise novel trackers for tap birth/death detection, as well as combined Signal-to-Noise Ratio (SNR) change and birth/death dynamics in OFDM systems, and novel trackers for lateral half-tap hops, i.e. path-related energy contributions that switch abruptly from one tap to an adjacent tap. Again, such novel trackers need to be low-complexity and provide superior performance.

5. Obtain performance metrics for our proposed trackers through Matlab simulations.

6. Critically review the performance, complexity and feasibility of our proposed solutions.

7. Propose a generalized mathematical framework to help problem-solvers think about tractable abrupt change dynamics in communication engineering, as well as other engineering and non-engineering problems where abrupt changes in tracked magnitudes might appear.

1.3.1 Research Contributions

A suitable channel model for OFDM transmission (Goal 1) and its corresponding Kalman tracker (Goal 2) were implemented in order to test novel channel trackers against. After a critical review of the literature (Goal. 3), two such novel trackers were created: the Simplified Maximum A Posteriori (SMAP) and the Neural-Network-switched Kalman Filtering (NNKF) schemes (Goal 4). Moreover, other related schemes were designed and implemented: a combined Quasi-Ideal SNR Detector (QISD)/SMAP scheme; and a neural detector for lateral hops. Those schemes were simulated in Matlab to obtain performance metrics (Goal 5) and their performance, complexity and feasibility were critically reviewed (Goal 6).

The contributions fulfilling the first 6 goals correspond to the following works:

- Méndez-Romero, D. and Fernández-Getino García, M. J. (2018). Simpler Multipath Detection for Vehicular OFDM Channel Tracking. *IEEE Transactions on Vehicular Technology*, 67(11):10752-10759.

  Results: A computationally inexpensive, threshold-based abrupt-change detector called Simplified Maximum A Posteriori (SMAP) is presented. Simulations show this estimator greatly reduces channel tracking error in the target SNR range at a very small computational cost, thus outperforming previously known systems.

- Méndez-Romero, D., Fernández-Getino Garcia, M. J., Tonello, A. M., and Dobre, O. A. (2020). Neural-Network-Switched Kalman Filters as Novel Trackers for Multipath Channels. In *2020 IEEE International Conference on Communications Workshops (ICC Workshops)* [7-11 June 2020], pages 1-5, Dublin, Ireland. IEEE.

  Results: Neural-Network-Switched Kalman Filters (NNKFs) are proposed as novel trackers for multipath channels. The core idea is similar to the SMAP switch in the previous paper, but now the tap birth/death detection will be performed via an artificial neuronal network (NNs).

  The proposed Neural-Kalman scheme is robust to wide variations in the probabilities of tap birth and death. Such robustness suggests a single, low-complexity NNKF could be reusable over different tap indices and communication environments.

- Méndez-Romero, D. and Fernández-Getino García, M. J. (2020). Death/Birth and SNR Detection for Vehicular Kalman Channel Trackers. In *2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON)* [16-18 June 2020], pages 104-108, Palermo, Italy. IEEE.

  Results: This paper considers combined death/birth and SNR detection when SNR is dynamical and may drift. Simulation results compared different schemes combining SNR detection and SMAP and suggest that SMAP, while being robust to SNR drift, benefits from an accurate SNR detection.

- Méndez-Romero, D., Fernández-Getino Garcia, M. J. (2022). Neural-Kalman Channel Trackers and Partial-Tap Hop Detection (in preparation)

  Results: This paper proposes a different kind of abrupt change: energy shifts from one tap to adjacent taps (partial tap lateral hops) and discusses how to model, detect and track such changes. A geometric justification for this and additional non-stationary dynamics is provided. A neural detector for lateral hops with large energy transfers is introduced. Simulation results suggest the proposed NN architecture might be a feasible lateral hop detector, suitable for integration in NNKF schemes.

The ambitious 7th goal, namely, "a generalized mathematical framework to help problem-solvers think about tractable abrupt change dynamics in (...) engineering and non-engineering problems", was addressed by analyzing the abrupt-change detection problem from different perspectives: a generalized mathematical perspective (Neural-Kalman Problem Category Conjecture), a neuroscientific perspective (bio-inspired AI mechanisms for the Changepoint - Oddball - Reversal model) and a neurofinancial perspective (Neural-Kalman Market Hypothesis, Neural-Kalman Phenomena and Highly Leveraged Bets Against Failing Expert Opinions). The combination of all three perspectives[1] provide valuable insights for Communications Engineering, among other areas.

The contributions fulfilling this last goal correspond to the following works:

- Méndez-Romero, D., Fernández-Getino Garcia, M. J. (2022). Neural-Kalman Conjecture and Market Hypothesis (in preparation)
- Méndez-Romero, D., Fernández-Getino Garcia, M. J. (2022). Bets Against Failing Experts: Neurofinancial Consequences of Unstereotyped Abrupt Dynamics (in preparation)

1.4 Structure of this Thesis

This paper is organized as follows.

Part I describes the contextual basics of the considered communications, exposes the problem of abrupt changes and provides a critical analysis of the state-of-the-art solutions proposed by other authors.

- Chapter 2 describes OFDM techniques and discusses channel models in the OFDM context.
- Chapter 3 describes the integration of Kalman Filtering (KF) in channel tracking schemes so as to enhance OFDM communications provided that channels are stationary (i.e. not subject to abrupt changes).
- In Chapter 4, different solutions are reviewed for abrupt change detection either in communications (Random Set Theory, Particle Filters) or in other areas (Threshold-Based Online Changepoint Detection, Advanced Sequential Analysis).

Once the problem and its context has been fully described in Part I, new proposals are introduced to solve the problem of abrupt changes in communications, as well as the consequences that abrupt changes themselves might have for neural intelligence.

[1] The multidisciplinary approach to the 7th goal of this Thesis led me to translate a 303-page treatise on Rational Economics [Voskuil, 2020] and critically review its logical consistency with its author. This effort was instrumental in my creation of the Neural-Kalman Phenomena concept; therefore, this commercially available translation [Voskuil and Taaki,] was included as a research merit for this Thesis on Engineering despite its rational-economic nature.

- Chapter 5 proposes a simplified framework for the channel tap birth-death problem and derives a computationally inexpensive, threshold-based estimator: the Simplified Maximum a Posteriori (SMAP) abrupt-change detector.
- Chapter 6 extends the RST channel models in Chapter 5 to abrupt increases/decreases of SNR for combined death/birth & SNR detection.
- Chapter 7 presents Neural-Network-Switched Kalman Filters (NNKFs) as novel trackers for multipath channels.
- Chapter 8 proposes a different kind of abrupt change: energy shifts from one tap to adjacent taps (partial tap lateral hops). A novel channel model for lateral partial-hop dynamics is presented, as well as a neural detector to help track such changes.
- In Chapter 9, we leverage our newly found understanding of abrupt changes and the interactions between Kalman filters and neural networks to analyze the neural consequences of abrupt changes and briefly sketch a novel, abrupt-change-derived stochastic model for neural intelligence, extract some neurofinancial consequences of unstereotyped abrupt dynamics, and propose a new portfolio-building mechanism in finance: Highly Leveraged Abrupt Bets Against Failing Experts (HLABAFEOs). We discuss some communication-engineering-relevant topics in the process, such as a Bayesian stochastic stereotyper for hopping Linear Gauss-Markov (LGM) models.
- Finally, Chapter 10 sums up our work and provides our conclusions and some guidelines for future work to expand knowledge in this research area.

Chapter 2

OFDM System & Channel Models

Orthogonal Frequency-Division Multiplexing (OFDM) is a method of modulating digital data on multiple carrier frequencies. OFDM has become a popular system for broadband digital communication, used in applications such as digital television, digital audio broadcasting, Asymmetric Digital Subscriber Line (ADSL) Internet access, wireless networks, communications networks by conventional electrical cables (Powerline Communications), as well as fourth generation (4G) and fifth generation (5G) mobile communications [Dahlman et al., 2020].

2.1 Intuitive Grasp of OFDM

OFDM is a frequency division multiplexing technique used as a digital multicarrier modulation method. In this technique, a large number of orthogonal subcarrier signals, with a small distance between them, are used to transmit the data in several parallel data streams (sometimes called channels). In that sense, two images can describe the difference between an OFDM system and the traditional simple communication system. As can be seen in figure 2.1, if the traditional communication has been carried out with a single carrier signal that will carry all the information (resembling the single and mighty jet of a faucet), an OFDM system is more similar to the multiple jets produced by shower head or diffuser head. When comparing models, it must be assumed that the flow rate of the tap and the shower are approximately identical, although in the case of the shower/OFDM it is divided into several trickles/subcarriers.

Each subcarrier is modulated with a conventional modulation scheme, such as Quadrature Amplitude Modulation (QAM) or Phase Shift Keying (PSK), at a low symbol rate (which does not prevent the total flow of the shower from being equal to or similar to that of the tap). The main advantage of OFDM over single-carrier systems is its ability to withstand very difficult channel conditions (for example, high-frequency attenuation in a long copper conductor, narrow-band interference, and frequency-selective fading due to to the multipath character of the channel) without complex equalization filters. Channel matching is simplified because OFDM can be thought of as using multiple narrowband signals with slow modulation instead of one broadband signal with fast modulation. The low symbol rate makes it feasible to use a guard interval between symbols, which makes it possible to eliminate the InterSymbol Interference (ISI) and to use echoes and time spreading to obtain a diversity gain, that is, an improvement in the Signal-to-Noise Ratio (SNR). This mechanism also facilitates the design of single-frequency networks in which several adjacent transmitters send the same signal simultaneously on the same frequency, since signals from multi-

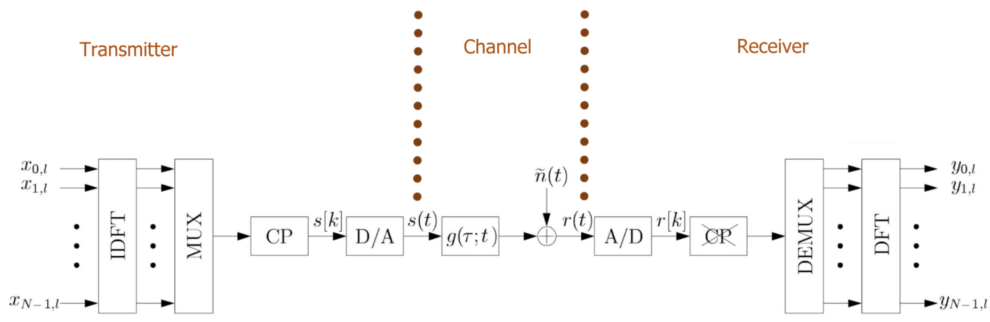

Figure 2.1: Digital implementation of a baseband OFDM system. Adapted from [Edfors et al., 1996].

ple distant transmitters can be combined constructively, rather than interfering as they typically would in the case of a traditional single-carrier system. In this sense, OFDM can be combined with other forms of spatial diversity, such as antenna arrays and MIMO channels, as realized in the IEEE802.11 wireless LAN standard.

2.2 Generic Modelling for OFDM Systems

The idea behind OFDM is the division of the available spectrum in several subchannels (that we will call subchannels, channels or subcarriers in this document). Once you have narrow-band channels, you can benefit from the fact that each of them faces flat fading, which eases equalization. To get a high spectral efficiency, subchannels' frequency responses are orthogonal in frecuency (as its own name, Orthogonal-Frequency Division Multiplexing, suggests). This means that whichever frequency a subchannel peaks, adjacent subchannels have a local bottom.

We will introduce the generic modelling for OFDM systems in its continuous-time version in the first place, and later in its discrete-time version.

We will make several starting assumptions:

- A cyclic prefix (CP) is used.
- The impulse response of the channel is shorter than the length of the CP.
- Transmitter and receiver are perfectly synchronized.
- Channel noise is white, additive, and Gaussian.
- Fading is slow enough to be considered constant over the interval of one OFDM symbol.

Please notice that these assumptions are simplifications of the problem, e.g. in reality, the channel estimate is more accurate for the beginning of the OFDM symbol than for the end, but it is a common approximation to consider the fading to be constant in a first approach to the problem. "Every model is wrong, some models are useful", and this generic OFDM model is certainly *very* useful.

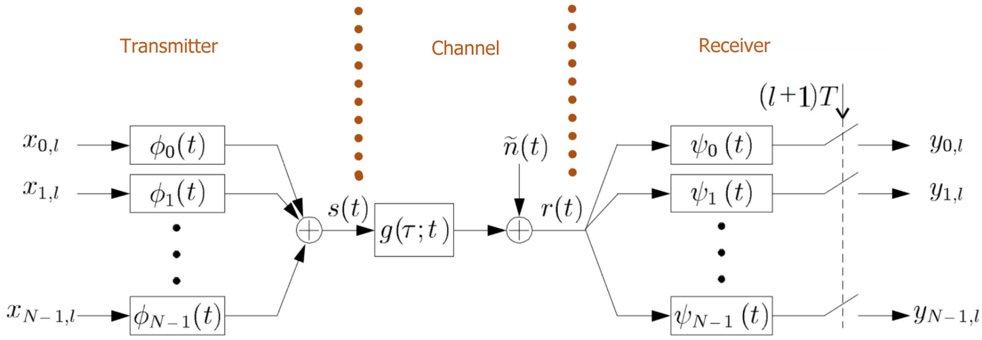

Figure 2.2: Baseband OFDM system model. Adapted from [Edfors et al., 1996].

The system in figure 2.1, which digitally implements OFDM in baseband, allows the parallel transmission of a set of N M-ary symbols. Such symbols could have been formed from the data stream (which can be simulated as a random binary stream).

After what is known as a mapper in the English literature (which could be translated as an applicator), the binary stream (M-PSK or M-QAM) is converted from serial to parallel, so that we will get a set of N complex numbers x_k.

Thus, the data x_k are coded into N carriers by applying the Inverse Discrete Fourier Transform (IDFT), so that the complex values a_{k} are obtained. This signal enters a serializer that copies the last L samples as a preamble or cyclic prefix[1]. The result is the OFDM symbol to be transmitted.

On the receiver, the steps are followed in reverse. Thus, the cyclic extension is first extracted, then the resulting signal is demodulated by applying the Discrete Fourier Transform (DFT).

Performing a complete theoretical analysis of an OFDM system is very difficult. It is easier to use simplified models that allow us a simpler analysis without great loss of resolution. Simplified models are usually classified according to whether they are continuous-time or discrete-time.

2.3 Continuous-Time Model

Let us study the ideal OFDM case, that is, a continuous OFDM that does not use modulation or demodulation[2]. Let us break it down case by case, starting with the waveforms used at the transmitter and moving module by module to the receiver. The baseband model is shown in Fig. 2.2.

Before we begin, let us formalize the concept of orthogonality. Suppose we have a set of signals ϕ, where ϕ_p is the pth element. Signals are said to be orthogonal if their integral over a period satisfies this condition:

$$\int_a^b \phi_p(t)\phi_q^*(t)\mathrm{d}t = \begin{cases} 1 & \text{if } p = q \\ 0 & \text{if } p \neq q \end{cases} \tag{2.1}$$

[1]The insertion of the cyclic prefix is commonly accepted in order to avoid ISI and preserve the orthogonality between the subchannels.

[2]That is the approach of [Edfors et al., 1996], which we follow closely in this section.

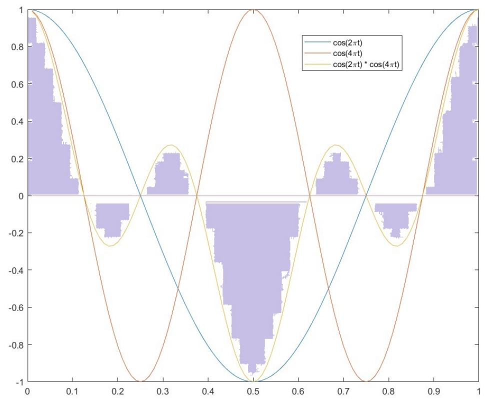

Figure 2.3: Two orthogonal signals, $cos(2\pi x)$ and $cos(4\pi x)$, and their product, whose integral over the fundamental period is zero, as per the graphic proof.

Some typical examples of orthogonal signals are combinations of sines and cosines. Figure 2.3 shows a pair of orthogonal signals with graphic proof of their orthogonality.

Once the concept of orthogonality has been clarified, let us analyze how the signals are transformed throughout each stage of the continuous model of the OFDM system.

2.3.1 Transmitter

Assuming an OFDM system with N subcarriers, a bandwidth of W Hz and a symbol length of T seconds, of which T_{cp} is the length of the cyclic prefix; the transmitter uses the waveforms indicated in the following equation [Edfors et al., 1996]:

$$\phi_k(t) = \begin{cases} \frac{1}{\sqrt{T-T_{cp}}} e^{j2\pi \frac{W}{N} k(t-T_{cp})} & \text{if } t \in [0, T] \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

where $T = N/W + T_{cp}$. Please note $\phi_k(t) = \phi_k(t + N/W)$ where t lies within the cyclic prefix $[0,\ T_{cp}]$. Given that $\phi_k(t)$ is a rectangular pulse coded in the carrier frequency kW/N, the usual interpretation for OFDM states it uses N subcarriers, and each of them transmits a low bit rate. The waveforms $\phi_k(t)$ are used in the modulation and the transmitted baseband signal for OFDM symbol index l is

$$s_l(t) = \sum_{k=0}^{N-1} x_{k,l} \phi_k(t - lT) \tag{2.3}$$

where $x_{0,l}, x_{1,l}, ..., x_{N-1,l}$ are complex numbers of a set of signal constellation points. When an infinite sequence of OFDM symbols is transmitted, the transmitter output is a juxtaposition of individual OFDM symbols:

$$s(t) = \sum_{l=\infty}^{\infty} s_l(t) = \sum_{l=\infty}^{\infty} \sum_{k=0}^{N-1} x_{k,l} \phi_k(t - lT) \tag{2.4}$$

2.3.2 Physical channel

If we assume that the support of the impulse response[3] $g(\tau;t)$ of the physical channel is restricted to the interval $\tau \in [0, T_{cp}]$, that is, to the length of the cyclic prefix, the received signal becomes:

$$r(t) = (g * s)(t) = \int_0^{T_{cp}} g(\tau;t)s(t-\tau)d\tau + \tilde{n}(t) \tag{2.5}$$

where $\tilde{n}(t)$ is a complex Gaussian white additive channel noise.

2.3.3 Receiver

The OFDM receiver is composed of a filter bank, adapted to the last part $[T_{cp}, T]$ of the waveforms $\phi_k(t)$ of the transmitter, that is,

$$\psi_k(t) = \begin{cases} \phi_k^*(T-t) & \text{if } t \in [0, T-T_{cp}] \\ 0 & \text{otherwise} \end{cases} \tag{2.6}$$

Effectively, this means that the cyclic prefix is removed in the receiver. Since the cyclic prefix contains all of the ISI of the previous symbol, the sampled output of the receiver's filter bank does not contain any ISI. Therefore, we can ignore the time index l when computing the sampled output of the kth matched filter. From the equalities (2.4), (2.5) and (2.6), we get

$$y_k = (r * \psi_k)(t)|_{t=T} = \int_{-\infty}^{\infty} r(t)\psi_k(T-t)dt =$$

$$= \int_{T_{cp}}^{T} (\int_0^{T_{cp}} g(\tau;t)[\sum_{k'=0}^{N-1} x_{k'}\phi_{k'}(t-\tau)]d\tau)\varphi_k^*(t)dt + \int_{T_{cp}}^{T} \tilde{n}(T-t)\phi_k^*(t)dt \tag{2.7}$$

Here we incorporate the approximation that the channel is fixed (quasi-constant) over the interval of an OFDM symbol, and denote it with $g(\tau)$, which gives us:

$$y_k = \sum_{k'=0}^{N-1} x_{k'} \int_{T_{cp}}^{T} (\int_0^{T_{cp}} g(\tau)\phi_{k'}(t-\tau)d\tau)\phi_k^*(t)dt + \int_{T_{cp}}^{T} \tilde{n}(T-t)\phi_k^*(t)dt \tag{2.8}$$

The integration intervals are $T_{cp} < t < T$ y $0 < \tau < T_{cp}$, which implies $0 < t - \tau < T$ and the inner integral can be written as:

$$\int_0^{T_{cp}} g(\tau)\phi_{k'}(t-\tau)d\tau = \int_0^{T_{cp}} g(\tau)\frac{e^{j2\pi k'(t-\tau-T_{cp})W/N}}{\sqrt{T-T_{cp}}}d\tau =$$

$$= \frac{e^{j2\pi k'(t-T_{cp})W/N}}{\sqrt{T-T_{cp}}} \int_0^{T_{cp}} g(\tau)e^{-j2\pi k'\tau W/N}d\tau \tag{2.9}$$

[3]The impulse response could be time-variant.

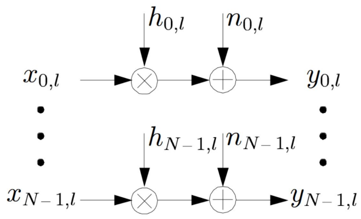

Figure 2.4: The continuous-time OFDM system, interpreted as parallel Gaussian channels.

where $T_{cp} < t < T$. The last part of the expression (2.9) is the channel's frequency response, sampled at frequency $f = k'W/N$, that is, thek'th carrier frequency:

$$h_{k'} = G(k'\frac{W}{N}) = \int_0^{T_{cp}} g(\tau)e^{-j2\pi k'\tau W/N} d\tau \tag{2.10}$$

where $G(f)$ is the Fourier transform of $g(\tau)$. Using this notation, and following at all times the theoretical derivation presented in [Edfors et al., 1996], the receiver filter bank output can be simplified to:

$$y_k = \sum_{k'=0}^{N-1} x_{k'} \frac{e^{j2\pi k'(t-T_{cp})W/N}}{\sqrt{T-T_{cp}}} h_{k'} \phi_k^*(t)dt + \int_{T_{cp}}^{T} \tilde{n}(T-t)\phi_k^*(t)dt = \sum_{k'=0}^{N-1} x_{k'} h_{k'} \int_{T_{cp}}^{T} \phi_{k'}(t)\phi_k^*(t)\mathrm{d}t + n_k \tag{2.11}$$

where $n_k = \int_{T_{cp}}^{T} \tilde{n}(T-t)\phi_k^*(t)dt$. Since the filters $\phi_k(t)$ in the transmitter are orthogonal[4],

$$\int_{T_{cp}}^{T} \phi_{k'}(t)\phi_k^*(t)\mathrm{d}t = \int_{T_{cp}}^{T} \frac{e^{j2\pi k'(t-T_{cp})W/N}}{\sqrt{T-T_{cp}}} \frac{e^{-j2\pi k'(t-T_{cp})W/N}}{\sqrt{T-T_{cp}}} dt = \delta[k-k'] \tag{2.12}$$

where $\delta[k]$ is the Kronecker delta function [Oppenheim et al., 1997], we can simplify equation (2.12) to get:

$$y_k = h_k x_k + n_k \tag{2.13}$$

where n_k is the additive white Gaussian noise.

The advantage of a cyclic prefix is twofold: it avoids the ISI (since it acts as a guard space) and the ICI (since it maintains the orthogonality of the subcarriers). If we re-introduce the time index l, we can now see the OFDM system as a set of parallel Gaussian channels, according to Figure 2.4.

2.4 Discrete-Time Model

Another way to approach the (simplified) modeling of an OFDM system is by considering all its stages in discrete time. Unlike the continuous-time model, the modulation and demodulation are replaced by discrete transforms (IDFT and DFT) and the channel effect will be computed

[4] For the definition of orthogonality, see equation (2.1).

by discrete-time convolution. The cyclic prefix operates in the same way in this system and the calculations can be performed analogously. The main difference is that all integrals are replaced by summations.

A discrete-time OFDM system model is shown in Figure (2.4).

From the receiver's point of view, using a longer cyclic prefix than the channel will transform the linear convolution of the channel into a cyclic convolution. If we use the symbol "$\circledast$" to denote cyclic convolution, we can write the entire OFDM system as:

$$\underline{\mathbf{y}}_l = \mathrm{DFT}(\mathrm{IDFT}(\underline{\mathbf{x}}_l) \circledast \underline{\mathbf{g}}_l + \underline{\tilde{\mathbf{n}}}_l = \mathrm{DFT}(\mathrm{IDFT}(\underline{\mathbf{x}}_l) \circledast \underline{\mathbf{g}}_l + \underline{\mathbf{n}}_l \tag{2.14}$$

where $\underline{\mathbf{y}}_l$ contains the N received data points, $\underline{\mathbf{x}}_l$ the N transmitted points of the constellation, $\underline{\mathbf{g}}$ the impulse response of the channel (padded with zeroes to make it N-long), and $\underline{\tilde{\mathbf{n}}}_l$ the noise of the channel. Since channel noise is assumed white and Gaussian, the term $\underline{\mathbf{n}}_l = \mathrm{DFT}(\tilde{\mathbf{n}}_l)$ represents uncorrelated Gaussian noise. Furthermore, we can take advantage of the result that the DFT of two signals to which a cyclic convolution is applied is equivalent to the product of their respective DFTs. Denoting element-by-element multiplication by "$\cdot$", the above expression can be written as:

$$\underline{\mathbf{y}}_l = \underline{\mathbf{x}}_l \cdot \mathrm{DFT}(\underline{\mathbf{g}}_l) + \underline{\mathbf{n}}_l = \underline{\mathbf{x}}_l \cdot \underline{\mathbf{h}}_l + \underline{\mathbf{n}}_l \tag{2.15}$$

where $\underline{\mathbf{h}}_l = \mathrm{DFT}(\underline{\mathbf{g}}_l)$ is the frequency response of the channel. Thus, the above equation represents the same type of parallel Gaussian channels as the one obtained in the previous section for the continuous-time model. The only difference is that the channel attenuations $\underline{\mathbf{h}}_l$ are given by the N-point DFT of the discrete-time channel, instead of the sampled frequency response that appeared in (2.10).

2.5 Tap Death-Birth Dynamics in Ideal Channels

2.5.1 Distinction between paths and taps

Suppose we transmit a waveform that's bandlimited to W. That means its baseband equivalent is limited to W/2 and can be represented by [Oppenheim et al., 1997]:

$$x_b(t) = \sum_n x[n] \cdot sinc(Wt - n) \tag{2.16}$$

where $x[n]$ is given by $x_b(n/W)$ y $sinc(t) = sin(\pi t)/(\pi t)$.

Suppose that, for a short time, the channel consists of N invariant paths. Each of these paths has a gain a_i and a delay τ_i. The number of paths N is finite (imagine we could physically force such situation with N totally isolated/independent circuits). How many taps are there?

The number of taps depends on two properties: 1) the sampling frequency at the receiver (and, therefore, the transmission bandwidth), and 2) the delay spread:

$$T_d \doteq max_{i,j} \, |\tau_i(t) - \tau_j(t)| \tag{2.17}$$

For higher transmit bandwidth (or, alternatively, higher receive sample rate), higher number of taps. In fact, for a non-zero T_d and an arbitrarily high transmission bandwidth W, arbitrarily high numbers of taps can be obtained. Hence, hundreds of taps can be obtained in Ultra-wideband

(UWB) communications (the exact number also depends on the area served, in the form of "delay spread").

Therefore, depending on the transmission, you can have 3 taps or 300 taps on the same channel, even though you might still have the same (finite) N physical paths. Each of these physical paths can also contribute to more than one tap. In effect, let us visualize each of these paths as a sinc function. Each path will contribute energy to several taps; obviously, its main contribution will be provided to the tap that is closest to the top of the sinc function. However, since the sinc isn't zero at all points outside that tap, it will continue to make minor contributions to subsequent taps. This "multi-contribution" will be less important if a raised cosine is used instead of a sinc, due to its lower permanence in time.

Once the distinction between trajectory (path) and channel tap is made, previous works such as [Angelosante et al., 2007, Angelosante et al., 2009] might seem a bit confusing and strange. What, exactly, was "being born" or "dying"? And what physical sense does that have? Virtually all tap variation is due to phase changes (which occur in a matter of ms); and only a part of that variation is due to a path not contributing to one tap and contributing more to another tap (this change happens in a matter of seconds). The changes in amplitude, produced by the birth and death of paths, seem to correspond more to a scale of seconds than to one of ms.

However, in the next few pages we will provide evidence that taps do indeed die and come back to life very frequently in real-world situations.

2.5.2 Statistical model of taps

The Rayleigh model is totally random, that is, autocorrelation $R_l[n] = 0$. Therefore, it is not suitable for modeling taps that should have a very high autocorrelation from Kalman instant to Kalman instant.

Another option is to use the Rician model, which assumes the line-of-sight (specular) path is large and has a known magnitude, while there are also many other independent (Rayleigh-like) paths. In this case, $h_l[m]$ can be modeled as (eq. 2.54 in [Tse and Viswanath, 2005]):

$$h_l[m] = \sqrt{\frac{\kappa}{\kappa+1}} \cdot \sigma_l \cdot e^{j\theta} + \sqrt{\frac{\kappa}{\kappa+1}} N_c(0, \sigma_l^2) \tag{2.18}$$

If we compute the autocorrelation of the tap, we obtain (for all $n \neq 0$):

$$R_l[n] \triangleq E\{h_l^*[m]h_l[m+n]\} = \frac{\kappa}{\kappa+1}\sigma_l^2 \tag{2.19}$$

Let us normalize $\sigma_l = 1$. What autocorrelation could we expect from a Rician model? For $\kappa = 10$ dB $= 10$, $R_l[n] = 0.91$. Thus, one tap would have an autocorrelation of 0.91, while the other taps would have an autocorrelation of 0. In our work (and those co-authored by Biglieri), we are using autocorrelations of 0.999 for all taps.

2.5.3 Taps in a Rician channel

What do we need to fully characterize a reasonably simplified channel model? According to the measurements that I have found, the channel in HSR scenarios is approximately of the Rician type, that is, each tap would be given by [Tse and Viswanath, 2005]:

$$h_l[m] = \sqrt{\frac{\kappa}{\kappa+1}}\sigma_l e^{j\theta} + \sqrt{\frac{1}{\kappa+1}}\mathcal{CN}(0,\sigma_l^2) \tag{2.20}$$

Thus, to characterize the channel, at least during sections of significant length and uniform behavior, we would need for each tap: the Rice factor (called c_R, K or κ in the literature) and the average energy of each tap, σ_l^2. In the Zhengxi measurements, consistent results were obtained with Rice factors of mean μ_κ=6 dB and σ_κ=2 dB. It is our understanding that this would allow the channel to be characterized for the first time (that is, if we do not have previous references: information relative to its value at previous instants) and that, furthermore, it can be assumed for most applications that this channel remains stationary during short enough periods of time.

We can write that for any researcher there are two possibilities:

a) Maintain the assumption that the Rice system is static in the very short run.

b) Keep the assumption that the Rice system is stationary in a wide sense (WSS) but not static.

Option a) could make sense if the objective of the investigation is not to track each tap. However, since our primary goal is to track the linear drift of each tap (using Kalman) and the non-linear jumps of the channel (using an appropriate abrupt change detection system), we would prefer option b) and therefore we'd need to choose how to model the drift of the Rice system.

2.5.3.1 Drift model for the Rician Channel

The small short-term changes of the Rice channel will be called "Rice drift" from this point on. There are two parameters associated with Rice drift that are very important for mobile communications and particularly for this research:

1) Level crossing rate: the level crossing rate, $N_\xi(r)$, describes how many times per second a stochastic process $\xi(t)$ crosses on average a given signal level r from top to bottom (or vice versa). During World War II, Mr. Rice obtained the integral expression to calculate it. In 1948 he obtained the particular expression for the rate of level crossings in Rice processes [Rice, 1948]:

$$N_\xi(r) = \sqrt{\frac{\beta}{2\pi}} \cdot p_\xi(r) \tag{2.21}$$

for $r \geq 0$, where β is a shorthand approximation of the negative curvature of the mean autocorrelation functions at the origin. This equation is Pätzold's (2.119) [Pätzold, 2011].

2) Fade duration: the fade duration, $T_{\xi_-}(r)$, is the expected value of the duration of the time intervals in which the stochastic process $\xi(t)$ is below a signal level r. It comes to be something similar to an inverse of the rate of level crossings, but with a much more complex expression. For Rice processes, it can be calculated with a very complicated integral:

$$T_{\xi\text{-}}(r) = \sqrt{\frac{2\pi}{\beta}} \cdot \frac{e^{\frac{r^2}{2\sigma_0^2}}}{rI_0(\frac{r\rho}{\sigma_0^2})} \cdot \int_0^r x e^{-\frac{x^2}{2\sigma_0^2}} I_0(\frac{x\rho}{\sigma_0^2})dx \tag{2.22}$$

for $r \geq 0$, which corresponds to equation (2.123) in [Pätzold, 2011]. This integral expression can only be evaluated numerically.

In practice, since these expressions are very complicated, for all the purposes of this line of research researchers could generate Rice channels with the required parameters N_ξ and $T_{\xi-}$ by means of their synthesis and subsequent fine-tuning. Thus, instead of presenting exact theoretical parameters, they would present the corresponding approximate sample parameters (% of time under a certain threshold level during long simulations, etc.).

It can be proved that both parameters depend exclusively on the first term of the autocorrelation function of the Rice process (section 3.4.3 of [Pätzold, 2011]) and, furthermore, that for low r the average duration of fading is approximately proportional to r (equation 3.78 of [Pätzold, 2011]).

2.5.4 Probability of tap non-birth and non-death

In the recent technical literature, abrupt changes where a tap disappears (gets very close to zero) or re-appears (gets away from zero) are usually characterized by parameters P_{00}, or probability of non-birth (i.e. probability of staying approximately at zero), and P_{11}, or probability of non-death (i.e. probability of staying away from zero), as will be shown in full detail in the empirical research summarized in Section 2.6. This information can also be presented through the probability of birth, P_{birth}:

$$P_{birth} = 1 - P_{00} \tag{2.23}$$

and the probability of death, P_{death}:

$$P_{death} = 1 - P_{11} \tag{2.24}$$

Such parameters can be easily connected to N_ξ and $T_{\xi-}$, as will be shown in the next Subsection.

2.5.4.1 How N_ξ and $T_{\xi-}$ relate to P_{birth} and P_{death}

If we make the approximation that the threshold level r is so low as to consider that it is only within that threshold when the path has died, the parameters of Section 2.5.3 can be easily related to the probabilities of birth and death of the tap.

For a given r, the mean fade duration would be related to the birth probability by:

$$\frac{1}{P_{birth}} = T_{\xi-} \tag{2.25}$$

Symmetric models, i.e. P_{birth}=P_{death}, can be found in the literature (e.g. [Angelosante et al., 2007, Angelosante et al., 2009]). However, a more precise approximation for P_{death} could be reached via this equation:

$$\frac{1}{P_{death}} = T_{\xi+} \tag{2.26}$$

where $T_{\xi+}$ is the *average connecting time interval*, i.e. the expected value of the length of the time intervals in which the stochastic process $\xi(t)$ is above a given level r. Thus,

$$T_{\xi+}(r) = \frac{F_{\xi+}(r)}{N_\xi(r)} \tag{2.27}$$

where $F_{\xi+}(r) = P\{\xi(t) > r\}$ is called the *complementary cumulative distribution function* [Pätzold, 2011].

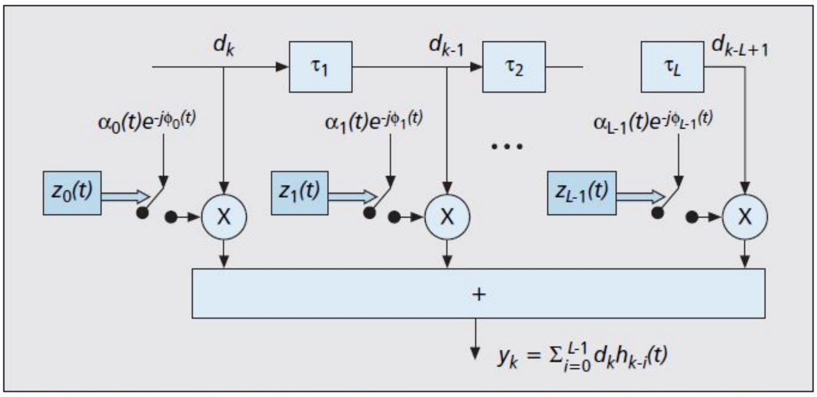

Figure 2. *Tapped delay line model for a wireless channel, showing explicit multipath persistence (on/off) processes* $z_k(t)$. *Notation* $h_k(t) = \alpha_k(t)\exp[-j\phi_k(t)]z_k(t)$; *kth switch is closed for* $z_k = 1$, *open for* $z_k = 0$.

Figure 2.5: Explicit persistence process (Fig. 2 in [Matolak, 2008]).

2.6 Empirical Tap Death-Birth Dynamics in Vehicular Communications

On the basis of different measurement campaigns on highways in the United States and Germany, several authors have developed empirical channel models with so-called "persistence processes" (i.e. the birth and death of channel taps) within the framework of intervehicular communications (V2V) [Hassan et al., 2020, Wu, 2012, Matolak and Wu, 2008]. In addition, they have published all kinds of parameters for these models, including the birth and death probabilities of each tap for different frequencies, bandwidths and channel types. In this section, the analyzed models and the feasibility of birth and death detection systems for them are studied.

2.6.1 Persistence processes

The idea that the taps have a "life time" or a finite "life cycle" ("MPC lifetime") or, equivalently, the idea of including a function $z(k)$ with values 0 or 1, which acts as a switch on a tap, so that taps can be born or die, appears in [Chong et al., 2005]. We have not found any previous reference, so it could be the first paper that uses it; [Matolak, 2014] doesn't cite any earlier reference for the concept of MPC lifetimes, either.

These birth and death processes were called "persistence processes", assigned a $z(k)$ function and modeled as discrete-time Markov chains [Sen and Matolak, 2008, Matolak and Wu, 2008].

2.6.2 Empirical probabilities of tap birth and death

Matolak, Sen and Wu published [Wu et al., 2010] the transition probabilities (birth-death) of each tap for different bandwidths and situations, in the framework of V2V communication channels. These parameters had been obtained from measurements under actual driving conditions on US highways (e.g. the I-71 highway from Cleveland to Ohio) [Matolak, 2008]. Birth/death odds depend on tap index; high-index taps have a much more intense birth/death dynamic than low-index taps; an example can be seen in figure 2.6, which reproduces fig. 5 of [Matolak, 2008].

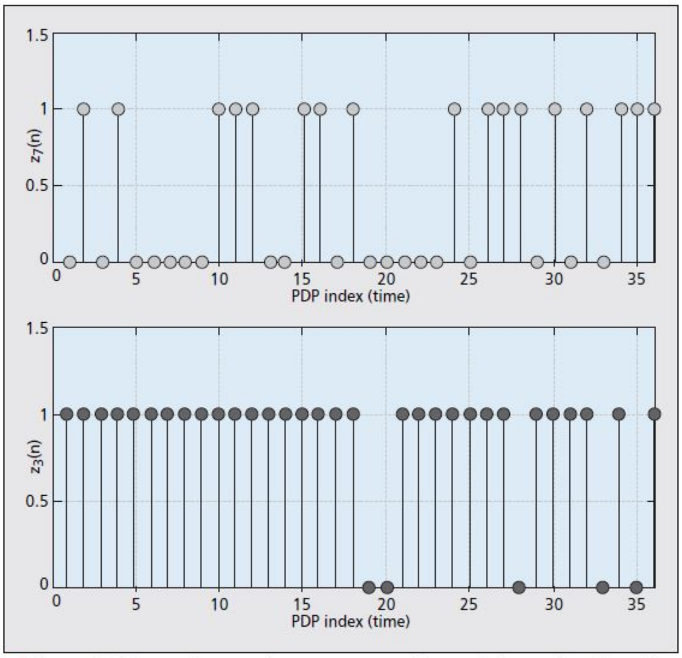

Figure 5. *Measured tap persistence processes for tap 3 (bottom) and tap 7 (top) for UOC V2V channel.*

Figure 2.6: Measured persistence processes, i.e. tap birth and death in real-world channels [Matolak, 2008].

A detailed study with probabilities of birth and death for f=5 GHz can be found in [Wu et al., 2010]. It considers 5 different scenarios: 1) Urban-Antenna outside the car (UOC); 2) Urban-Antenna inside the car (UIC); 3) Small City (Small City, S); 4) Highway-Low Traffic Density (OLT); and 5) Highway-High Traffic Density (OHT). Figures 2.7 and 2.8 show tables published in these studies. As you can see, all the necessary parameters to simulate a birth/death dynamic are presented, including the empirical transition probabilities for each tap.

Note that the transition probabilities are not symmetric ($P_{birth} \neq P_{death}$), that the probability of death is higher at a higher tap index, and that the transition probabilities are sometimes much higher ($P > 0.20$) than those used in theoretical models such as [Angelosante et al., 2007, Angelosante et al., 2009], which are adopted in the simulations of this Thesis (typically, $P = 0.05$). Therefore, in some cases, abrupt dynamics have been proven to be even more important than some theoretical models assumed.

2.6.2.1 Applicability to MIMO

In the last couple of years, a similar procedure has been applied to data from roads in Cologne (Germany). The corresponding probabilities of birth and death have been published in [Hassan et al., 2020]. Fig. 2.9 shows the resulting empirical channel model; the odds of abrupt changes are given as P_{11} (probability of non-death) and P_{00} (probability of non-birth). Again, we can see that abrupt dynamics depend on tap index.

Same authors have also analyzed MIMO channels ([Hassan et al., 2021]). In the case of MIMO ($M_T \times M_R$ antennas), transition probabilities can be applied to the $M_T \times M_R$ equivalent SISO channels, perhaps also incorporating information about the correlation between those channels.

TAP index K	Energy	Weibull shape factor (β_K)	$P_{00,K}$	$P_{11,K}$	P_1
UIC					
1	0.756	2.49	NA	1.0000	1
2	0.120	1.75	0.0769	0.9640	0.9625
3	0.051	1.68	0.3103	0.8993	0.8732
4	0.034	1.72	0.3280	0.8521	0.8199
5	0.019	1.65	0.5217	0.7963	0.7017
6	0.012	1.6	0.6429	0.7393	0.5764
7	0.006	1.69	0.6734	0.6686	0.4971
Small city					
1	0.90	3.95	NA	1.0000	1.0000
2	0.08	2.0	0.4839	0.9446	0.9034
3	0.02	2.0	0.3452	0.7712	0.7383
OHT					
1	0.95	4.3	NA	1.0000	1.0000
2	0.04	1.64	0.3625	0.8366	0.7960
3	0.01	2.0	0.5999	0.6973	0.5696

Table 1. *Channel models for 10 MHZ UIC, S, and OHT channels.*

Figure 2.7: Probabilities of non-birth and non-death for each inactive or active tap, respectively [Matolak, 2008].

TABLE III
CHANNEL PARAMETERS FOR 33.33-MHz-BANDWIDTH
V2V CHANNELS (MODEL-1)

Tap Index k	Energy	Weibull Factor (β_k)	$P_{00,k}$	$P_{11,k}$	P_1
UOC					
1	0.7858	3.26	Na	1.0000	1.0000
2	0.0804	1.68	0.2707	0.9233	0.9048
3	0.0394	1.73	0.3744	0.8727	0.8309
4	0.0270	1.67	0.4405	0.83	0.7670
5	0.0178	1.77	0.5419	0.7686	0.6646
6	0.0142	1.71	0.6200	0.7242	0.5796
7	0.0114	1.71	0.6251	0.6856	0.5440
8	0.0092	1.68	0.7038	0.6563	0.4630
9	0.0075	1.76	0.7116	0.6318	0.4394
10	0.0073	2.00	0.7283	0.6214	0.4180
Small City					
1	0.7608	4.53	na	1.0000	1.0000
2	0.0987	2.17	0.0833	0.9643	0.9626
3	0.0638	2.00	0.1	0.9069	0.9065
4	0.0378	1.77	0.24	0.8593	0.8442
5	0.0181	2.00	0.2976	0.7542	0.7383
6	0.0116	2.00	0.4874	0.6965	0.6293
7	0.0097	2.17	0.4833	0.695	0.6262
OHT					
1	0.8910	4.40	na	1.0000	1.0000
2	0.0684	1.54	0.2465	0.8994	0.8820
3	0.0210	1.68	0.4417	0.7874	0.7244
4	0.0111	1.57	0.6392	0.6258	0.4908
5	0.0085	1.62	0.6407	0.5868	0.465
OLT					
1	0.9013	5.29	na	1.0000	1.0000
2	0.0681	1.27	0.2857	0.8902	0.8668
3	0.0190	1.61	0.4468	0.7839	0.7196
4	0.0116	1.66	0.6384	0.6246	0.4913
UIC					
1	0.5991	2.69	na	1.0000	1.0000
2	0.1068	1.76	0.1333	0.9808	0.9784
3	0.0621	2.00	0.1373	0.9315	0.9265
4	0.0405	1.71	0.2841	0.8959	0.8732
5	0.0289	1.74	0.3739	0.8754	0.8343
6	0.0221	1.73	0.3605	0.8278	0.7882
7	0.0204	2.00	0.425	0.7647	0.7104
8	0.0175	2.00	0.5405	0.7834	0.6801
9	0.0143	2.00	0.4623	0.7651	0.6945
10	0.0145	1.67	0.5126	0.7451	0.6571
11	0.0113	2.00	0.5434	0.7196	0.6182
12	0.0112	2.00	0.5668	0.7115	0.6009
13	0.0100	2.00	0.5942	0.6753	0.5562
14	0.0092	1.56	0.6503	0.6513	0.5
15	0.0070	1.66	0.6967	0.6151	0.4395
16	0.0077	1.68	0.6992	0.6401	0.4539
17	0.0061	2.00	0.7063	0.5730	0.4063
18	0.0064	1.57	0.7512	0.5970	0.3804
19	0.0049	1.74	0.7397	0.5569	0.3689

Figure 2.8: Probabilities of non-birth and non-death for different vehicular settings [Wu et al., 2010]

TABLE I
TDL TABLE PARAMETERS FOR $M2$ AFTER MERGING THE DATA [5]

NLOS1											
Taps	Delay (μ sec)	Energy(lin)	SS_1	SS_0	P_{11}	P_{00}	σ	μ	a	β	σ_R
1	0.95	0.9818	0.9670	0.033	0.9981	0.9439	0.9344	-15.7	$1.99 * 10^{-7}$	0.86	$2.3 * 10^{-7}$
2	1.35	0.0175	0.9897	0.0103	0.9974	0.75	0.697	-17.4	$3.9 * 10^{-8}$	1.46	$3.15 * 10^{-8}$
3	2.1	0.0005	0.277	0.723	0.9558	0.9831	0.6744	-19.1216	$7.02 * 10^{-9}$	1.533	$5.5 * 10^{-9}$
4	2.35	0.0001	0.133	0.867	0.9	0.9847	0.4619	-19.61	$3.8 * 10^{-9}$	2.22	$2.6 * 10^{-9}$

NLOS2											
Taps	Delay (μ sec)	Energy(lin)	SS_1	SS_0	P_{11}	P_{00}	σ	μ	a	β	σ_R
1	1	0.8864	0.8346	0.1654	0.9919	0.9591	1.3016	-18.7	$1.16 * 10^{-8}$	0.6866	$1.7 * 10^{-8}$
2	1.5	0.0671	0.9596	0.0404	0.9965	0.9168	1.0681	-19.505	$5.8 * 10^{-9}$	0.9684	$7.4 * 10^{-9}$
3	1.85	0.0197	0.8089	0.1911	0.9802	0.9161	0.9874	-19.9532	$3.6 * 10^{-9}$	1.0892	$4 * 10^{-9}$
4	2.35	0.0093	0.4635	0.5365	0.9643	0.9692	0.903	-20.167	$2.8 * 10^{-9}$	1.2304	$2.7 * 10^{-9}$
5	2.65	0.0109	0.2615	0.7385	0.9444	0.9803	1.0255	-20.3244	$2.7 * 10^{-9}$	1.179	$3.01 * 10^{-9}$
6	2.95	0.0066	0.1642	0.8358	0.9438	0.989	0.7604	-20.1	$2.7 * 10^{-9}$	1.3674	$2.3 * 10^{-9}$

Figure 2.9: Empirical channel model. Among its parameters, non-birth and non-death odds for each tap [Hassan et al., 2020].

2.6.2.2 Higher-order and inhomogenous Markov models

Wu, author of [Wu et al., 2010, Wu, 2012], first studied a homogeneous (i.e., stationary: transition probabilities did not change over time) first-order Markov chain and published the transition probabilities (birth-death) of each tap for different bandwidths and situations, within the framework of V2V communication channels. Later he decided to try more complicated models: higher-order Markov models, non-homogeneous (i.e., non-stationary) Markov models and, on the other hand, Markov models that included strong correlations between taps; as an example of what could cause these correlations, in his thesis [Wu, 2012] he talks about:

> when large objects (buildings, trucks, etc.) cause multiple adjacent taps to be "on" or "off" at the same time. Thus, a more realistic persistence Markov model should incorporate correlations especially for large channel bandwidth (or small delay resolutions).

After testing different models and comparing them with measurements based on entropy measures, he concluded that the 2nd order Markov model was practically optimal for modeling the persistence of taps and, in case a complexity was admissible higher, the Markov model with correlations could be used to obtain maximum fidelity.

Note that, in the case of non-homogeneous Markov models, it might make sense to use some kind of intelligent system that tries to stereotype the birth/death dynamics in real time[5].

However, all published transition probabilities found in this literature review refer to the simplest Markov model (i.e., 1st order and with no correlations).

2.6.2.3 Unsolvable Subcomponents and Delay Drift

Some of the V2V papers referred above (e.g. [Matolak, 2014]) mention the fact that in each tap there are non-solvable subcomponents that are combined by complex addition. There is also some chatter that, over approx. 1 second, delay drift can cause a tap to pass to the neighboring tap. This happens too slowly to be a problem in most applications [Matolak, 2014].

[5]To deal with such particularly difficult channel tasks, this Thesis presents a novel proposal for a Bayesian stereotyper in Section 9.4.

Later in this Thesis, a novel model for tap subcomponents will be presented (Chapter 8), whereby tap subcomponents will be able to “hop” laterally to adjacent taps, i.e. change tap index.

No in-depth description for such possible subcomponents or lateral hopping has been found in the cited studies (for a formal definition of subcomponents and lateral hopping, see Section 8.1; for a geometric justification why such lateral dynamics would exist, see Subsection 8.1.1). Finally, it should be added that in models such as Saleh-Valenzuela [Saleh and Valenzuela, 1987], the dynamics of a cluster can sometimes be resolved without the different taps included being resolvable ([Zhang, 2016], p. 27/44). For channel tracking purposes, our aim is not to resolve each individual tap subcomponent, but the channel-relevant effects of hopping dynamics: it might be the case that some subcomponent delay drift/lateral hopping dynamics could be resolvable without the individual subcomponents themselves being resolvable.

2.6.3 Main Takeaways from our Empirical Literature Review

- There are empirical models of the birth/death of taps.
- The most widely used model is a first-order Markov model like the one considered by Angelosante et al. and adopted by this Thesis.
- There are more complex Markov models.
- The dynamics of birth/death in intervehicular communications are very important. Indeed, they have been demonstrated in different measurement campaigns and they have a significantly higher intensity than that assumed in some theoretical works, especially for high-index taps.
- Mentions have been found of the correlation in the birth/death of adjacent taps (because they share the same reflector, such as a large truck or a building), but this literature review has found no explicit models of “partial tap lateral hopping”.

Chapter 3

Kalman Filtering for Channel Tracking

3.1 What Is a Kalman Filter?

The Kalman filter (KF), also known as Linear Quadratic Estimator (LQE) is a statistically optimal estimator for the linear quadratic problem, that is, to estimate the instantaneous state of a dynamic linear system with a function of quadratic cost[1], using measures linearly related to the state but corrupted by white noise.

It is one of the greatest advancements in the history of statistical estimation theory and possibly the most important discovery of the 20th century in this field [Grewal and Andrews, 2015]. It is a mathematical tool, implementable as a computer program, which constitutes a complete statistical characterization of an estimation problem.

Therefore, at a practical level it is not just an estimator, since it propagates the entire probability distribution of the variables that it has to estimate; in other words, it fully characterizes the current state of knowledge of the dynamical system [Chui and Chen, 2017], including the influence of all past measurements.

It was first introduced in [Kalman, 1960] by Rudolf E. Kálmán[2], who is considered one of the main developers of the underlying theory, and soon this invention became popular among engineering faculties [Grewal and Andrews, 2015].

3.1.1 Applications and intuitive explanation of the algorithm

The Kalman filter has numerous applications in technology. A common application is for the guidance, navigation and control of vehicles [Marchthaler and Dingler, 2017], specifically aircraft and spacecraft. Another frequent use case is target detection in radar applications [Kim, 2011]. Furthermore, the Kalman filter is a widely applied concept in time series analysis, as it is used in fields such as signal processing and econometrics [Grewal and Andrews, 2015]. Kalman filters are

[1] The typical estimation problem is Linear Quadratic Gaussian (LQG). In that case, the dynamical systems are linear, the cost functions applied to the estimation quality are quadratic and, as an added condition, the random processes are Gaussian.

[2] Kálmán was born in Hungary; the correct way to write this Hungarian surname is with two accents. To refer to the Kalman filter, we will use the anglicized version (without accents) of the surname, as it is the most common in technical literature.

also one of the main topics in robotic motion planning and control, and are sometimes included in path optimization algorithms [Grewal and Andrews, 2015].

The algorithm works in a two-stage process: a predictive one and an update one. In the predictive stage, the Kalman filter produces estimates of the current state variables as well as their uncertainties. Once the next measurement (necessarily corrupted by some random error or noise) has been obtained, these estimates are updated using a weighted average, with more weight given to estimates with greater certainty. Due to the recursive nature of the algorithm, it can work in real time using only the current input measurements and the previously computed state and its uncertainty matrix; you don't need any additional information from the past [Chui and Chen, 2017].

3.1.2 The basic equation

The equations from which the Kalman filtering manuals are based are excessively complex for a first explanation; fortunately, often (and certainly in the case we are concerned with in the tracking tasks in this PhD Thesis), the state transition matrix disappears, and you can come up with the following, much easier, equation to start with:

$$\underline{\hat{\mathbf{x}}}_k(+) = (\mathrm{I} - \overline{\mathrm{K}}_k)\underline{\hat{\mathbf{x}}}_k(-) + \overline{\mathrm{K}}_k\underline{\mathbf{z}}_k \tag{3.1}$$

where

$\underline{\hat{\mathbf{x}}}_k(+)$ The estimate of the signal $\underline{\mathbf{x}}$ at time t_k after incorporating the information of a new reading (i.e. the *posterior* estimate, a.k.a. the estimate conditioned by the reading $\underline{\mathbf{z}}_k$). It is represented by a random vector of dimensions $n \times 1$.

$\overline{\mathrm{K}}_k$ The Kalman gain matrix at time t_k, with dimensions $n \times l$.

I The identity matrix

$\underline{\hat{\mathbf{x}}}_k(-)$ The estimate of the signal $\underline{\mathbf{x}}$ at time t_k before incorporating the information of a new reading (i.e. the *prior* estimate, unconditioned by the reading $\underline{\mathbf{z}}_k$). It is represented by a random vector of dimensions $n \times 1$.

$\underline{\mathbf{z}}_k$ The reading (i.e. measurement) obtained at time t_k, represented as a random vector with size $l \times 1$.

Remember that the subscripts k and $k-1$ refer to the epoch (the time instant). For example, in our case, they will represent discrete-time intervals, so that $t_k = t_{k-1} + \tau$, for all $k > 0$ and a given time interval τ.

Although the variables in equation (3.1) are random vectors and matrices, there are often situations where scalar values need to be estimated based on scalar-valued measurements. Specifically, that will be our case in this PhD Thesis. Recall that, in the current work, bold notation indicates randomness and underlining indicates that the element is a vector (such as $\underline{\mathbf{z}}_k$), while uppercase and unitalicized (such as $\overline{\mathrm{K}}_k$) denote matrices[3]. To obtain the scalar form of the equation (3.1), we will use the corresponding variables without underlining and in italics, such as $\mathbf{z}_k$ and $\overline{\kappa}_k$[4], so

[3] For any questions about notation, you can consult the extensive list of symbols on "Mathematical Notation" and, in particular, the section "Kalman filtering" .

[4] Do not confuse the Latin letter ka (k) and the Greek letter kappa (κ). The symbol $\overline{\mathrm{K}}$ denoting the Kalman gain matrix is the uppercase kappa (with a dash above it). Therefore, the scalar Kalman gain will always be $\overline{\kappa}$ (lowercase kappa with a dash above it). In the current work, to avoid confusion, $\overline{\mathrm{K}}$ will never be used with any other meaning. The notation of the dash above is used exclusively because it is common in the literature, cf. [Anderson and Moore, 1979, Brown and Hwang, 2012, Catlin, 1989, Gelb et al., 1974].

that we'll write:

$$\hat{\mathbf{x}}_k(+)=(1-\overline{\kappa}_k)\cdot\hat{\mathbf{x}}_k(-)+\overline{\kappa}_k\cdot\mathbf{z}_k \tag{3.2}$$

Our goal is to find $\hat{\mathbf{x}}_k(+)$, the updated estimate of the "signal"[5] $\mathbf{x}_k$. Furthermore, the algorithm can be applied online, that is, by recursively finding the estimate for each subsequent instant (i.e. the estimate for the state at the kth instant for each subsequent k). Note that $\mathbf{z}_k$ is the value of the measurement, but the measurement is neither perfectly accurate nor perfectly distortion-free (otherwise we would not need the Kalman filter). The measurement value is typically distributed assuming a Gaussian error. We are also familiar with $\hat{\mathbf{x}}_k(-)$, often called the *a priori* estimation, or *prior* estimation; in reality, it is the prediction of the signal calculated at the previous instant.

Therefore, from the expression to the right of the equality in equation (3.2), the only unknown component is the Kalman gain, $\overline{\kappa}_k$. The Kalman filter needs to calculate the Kalman gain which is, so to speak, the trustworthiness the filter grants to each of its sources of information: one such source of information would be the measurement (i.e. the readings it receives in each instant) and another source of information would be the underlying theoretical model (e.g. the equation that would theoretically govern a magnitude to be estimated, such as a single tap gain in a multipath channel.

To better understand how the Kalman filter works, just think what would happen if we assume the following starting condition: $\overline{\kappa}_k = 0.5$. In such a case, the scalar equation (3.2) would take the trivial form:

$$\hat{\mathbf{x}}_k(+)=0.5\cdot\hat{\mathbf{x}}_k(-)+0.5\cdot\mathbf{z}_k \tag{3.3}$$

The estimate for $\mathbf{x}_k$ made after receiving the reading $\mathbf{z}_k$ would be the average of both the reading and the prediction made in t_{k-1} (based on previous readings and information about the underlying theoretical model of the evolution of $\mathbf{x}$). This can be interpreted as follows: as we have no experience with these sources of information, we do not consider any more reliable than the other, but rather we combine them to obtain the average. No doubt this is a very reasonable initial decision; in fact, mathematically optimal in the linear-quadratic case.

As the Kalman filter gains experience, i.e. as it works with these two information sources and learns which one is more reliable, it finds the optimal weighting factor to combine these information at each subsequent instant. In addition, as we already indicated in the previous section, the Kalman filter can *remember* a little bit, the most important thing, the essential thing, about past instants. In mathematical terms, it propagates the probability distribution of the variable(s) to estimate.

In the following sections, the linear estimation problem and its optimal solution based on the Kalman filter will be presented using a rigorous mathematical approach.

3.2 Statement of the Linear Estimation Problem

The problem is how to estimate the state of a linear stochastic system using measurements that are linear functions of the state. We assume that stochastic systems can be represented by channel

[5]Strictly speaking, what we are estimating might be any magnitude, e.g. speed, position within a map, height and position of an aircraft or, as will be our case and get explained soon, any channel tap gain within OFDM communications. In Estimation Theory, the physical nature of the magnitude is not considered, but only its mathematical nature: for nomenclature purposes, everything is "signals". Our ignorance about the signal is mathematically represented as the randomness of the signal, hence the bold notation of $\mathbf{x}_k$.

generation (plant[6]) and measurement models according to the following equations (3.4) and (3.7) in discrete time. Important note: for any questions about the notation, consult the section on "Kalman filtering" in the "List of symbols" located immediately after the Index of this PhD Thesis.

3.2.1 Plant model

$$\underline{\mathbf{x}}_k = \Phi_{k-1}\underline{\mathbf{x}}_{k-1} + \underline{\mathbf{w}}_{k-1} \tag{3.4}$$

where

$\underline{\mathbf{x}}_k$ The signal $\underline{\mathbf{x}}$ at time t_k, represented as a random vector with size $n \times 1$.

Φ_{k-1} The state transition matrix of the (discrete) dynamic linear system at time t_{k-1}, with dimensions $n \times n$.

$\underline{\mathbf{x}}_{k-1}$ The signal $\underline{\mathbf{x}}$ at time t_{k-1}, represented as a random vector with size $n \times 1$.

$\underline{\mathbf{w}}_{k-1}$ The process noise $\underline{\mathbf{w}}$ at time t_{k-1}, represented as a random vector with size $r \times 1$.

Plant noise (sometimes called process noise) $\underline{\mathbf{w}}$ is assumed to be a zero-mean Gaussian process; therefore, each realization of that process is a random vector[7]. In addition, we will assume that the process and plant noises are uncorrelated ($\mathbb{E}\left\langle \underline{\mathbf{w}}_k \underline{\mathbf{v}}_j^T \right\rangle = 0$ for all k and j). Process $\underline{\mathbf{x}}$ is a stochastic process whose realizations $\underline{\mathbf{x}}_k$ are random vectors. The initial value $\underline{\mathbf{x}}_0$ is a Gaussian variable with known mean $\underline{\mathbf{x}}_0$ and known covariance matrix P_0.

3.2.1.1 Scalar equation for the plant model

By simply setting $n = 1$ and $r = 1$, the random vectors for the signal and the process noise, as well as the state transition matrix, would become scalars and the model collapses into the following scalar plant model[8]:

$$\mathbf{x}_k = \varphi_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1} \tag{3.5}$$

where

$\mathbf{x}_k$ The signal $\mathbf{x}$ at time t_k, represented as a (scalar) random variable.

φ_{k-1} The state transition (scalar) variable for the (discrete) dynamic linear system at time t_{k-1}.

$\mathbf{x}_{k-1}$ The signal $\mathbf{x}$ at time t_{k-1}, represented as a (scalar) random variable.

$\mathbf{w}_{k-1}$ The process noise $\mathbf{w}$ at time t_{k-1}, represented as a (scalar) random variable.

The simplest application of Kalman filtering is obtained when the plant model has a constant scalar transition $\varphi_k = \varphi$.

[6] Kalman filters were initially used to estimate states in industrial plants, such as chemical plants. That's why the underlying generating model is called 'plant' model. In our case, this 'plant' model is a channel generator.

[7] In this PhD Thesis, bold always indicates randomness (cf. "List of symbols").

[8] The reasoning is applicable to both real and complex random variables.

3.2.2 Measurement model

$$\underline{\mathbf{z}}_k = \mathrm{H}_k \underline{\mathbf{x}}_k + \underline{\mathbf{v}}_k \tag{3.6}$$

where

$\underline{\mathbf{z}}_k$ The measurement $\underline{\mathbf{z}}$ at time t_k, represented as a random vector with size $l \times 1$.

H_k The measurement sensitivity matrix, which defines the linear relationship between the state of the dynamical system and the measurements that can be made at time t_k. It has dimensions $l \times n$.

$\underline{\mathbf{x}}_k$ The signal $\underline{\mathbf{x}}$ at time t_k, represented as a random vector with size $n \times 1$.

$\underline{\mathbf{v}}_k$ The measurement noise $\underline{\mathbf{v}}$ at time t_k, represented as a random vector with size $l \times 1$.

The measurement noise $\underline{\mathbf{v}}_k$ is a zero-mean Gaussian random vector that represents the error produced by an imperfect measurement subject to distortions.

3.2.2.1 Scalar equation for the observation model

Again, by simply setting $n = 1$ y $l = 1$, we can state the following scalar measurement model:

$$\mathbf{z}_k = h_k \mathbf{x}_k + \mathbf{v}_k \tag{3.7}$$

siendo

$\mathbf{z}_k$ The measurement $\mathbf{z}$ at time t_k, represented as a (scalar) random variable.

h_k The (scalar) value for the sensitivity of the measurement.

$\mathbf{x}_k$ The signal $\mathbf{x}$ at time t_k, represented as a (scalar) random variable.

$\mathbf{v}_k$ The measurement noise $\mathbf{v}$ at time t_k, represented as a (scalar) random variable.

The simplest application of Kalman filtering is obtained when the measurement model has a scalar, constant measurement sensitivity $h_k = h$.

3.2.3 Equations for plant noise

$$\mathbb{E}\left\langle \underline{\mathbf{w}}_k \right\rangle = \underline{0} \tag{3.8}$$

$$\mathbb{E}\left\langle \underline{\mathbf{w}}_k \underline{\mathbf{w}}_i^T \right\rangle = \Delta(k-i)\mathrm{Q}_k \tag{3.9}$$

where $\Delta(k-i)$ is the Kronecker delta function (a scalar) and Q_k is the covariance matrix (with size $n \times n$) of the process noise used in the system state dynamics.

In simple words, this means that we require that the process $\underline{\mathbf{w}}$ to be zero-mean (mean is a zero vector) and realizations $\underline{\mathbf{w}}_k$ of the process $\underline{\mathbf{w}}$ at each instant t_k are uncorrelated with each other (although at each time t_k, there may be a non-zero correlation between different elements of the random vector $\underline{\mathbf{w}}_k$, as required by Q_k).

The simplest case is obtained when the process noise covariance matrix does not vary over time; in that case, $\mathrm{Q}_k = \mathrm{Q}$. If we are working with scalar equations (according to the transformation described in the previous sections), the simplest case occurs when $\sigma_{\mathbf{w},k}$, the noise variance of the process, is constant: $\sigma_{\mathbf{w},k} = \sigma_{\mathbf{w}}$.

3.2.4 Equations for observation noise

$$\mathbb{E}\langle \underline{\mathbf{v}}_k \rangle = \underline{0} \tag{3.10}$$

$$\mathbb{E}\left\langle \underline{\mathbf{v}}_k \underline{\mathbf{v}}_i^T \right\rangle = \Delta(k-i)\mathrm{R}_k \tag{3.11}$$

where $\Delta(k-i)$ is the Kronecker delta function (a scalar) and R_k is the covariance matrix (with size $l \times l$) of the observation or measurement noise.

Again, in simple words this means that we require the process $\underline{\mathbf{v}}$ to be zero-mean (i.e. mean is a zero vector) and realizations $\underline{\mathbf{v}}_k$ of the process $\underline{\mathbf{v}}$ at each instant t_k are uncorrelated with each other (although at each time t_k, there may be a non-zero correlation between different elements of the random vector $\underline{\mathbf{v}}_k$, as required by R_k).

The simplest case is obtained when the observation noise covariance matrix does not vary over time; in that case, $\mathrm{R}_k = \mathrm{R}$. If we are working with scalar equations (according to the transformation described in the previous sections), the simplest case occurs when $\sigma_{\mathbf{v},k}$, the noise variance of the observation, is constant: $\sigma_{\mathbf{v},k} = \sigma_{\mathbf{v}}$.

3.2.5 Optimization goal

With the models proposed in the previous sections, the objective of the Kalman filter will be to find an estimate of the n-state vector $\underline{\mathbf{x}}_k$; that estimate will be represented as $\underline{\hat{\mathbf{x}}}_k$ and will be a linear function of the readings or measurements $\underline{\mathbf{z}}_i$,..., $\underline{\mathbf{z}}_k$ that minimizes the mean square error:

$$E[\underline{\mathbf{x}}_k - \underline{\hat{\mathbf{x}}}_k]^T \mathrm{M}[\underline{\mathbf{x}}_k - \underline{\hat{\mathbf{x}}}_k] \tag{3.12}$$

where M is any nonnegative, definite, symmetric weighting matrix.

3.3 Solution to the Linear Estimation Problem

It can be shown [Grewal and Andrews, 2015] that the linear estimation problem raised in the previous sections is solved using the Kalman estimator in discrete time, a procedure that can be implemented by computer and consists of the following basic steps:

1. Error covariance extrapolation
2. State estimate gets updated with the new reading
3. Error covariance update
4. Computation of the Kalman gain

In the following sections, each stage will be described as well as the computations and parameters involved.

3.3.1 Step One: Error covariance extrapolation

The Kalman filter is based on a covariance matrix, P, which stores information about the uncertainty of the estimates, the degree of "insecurity" that the Kalman filter itself computes to 'feel' about its own estimates (predictions). This uncertainty will increase, for example, if the new readings lead to the assumption that previous estimates (predictions) were more wrong than expected.

The first filtering step is the extrapolation of $\mathrm{P}_{k-1}(+)$, which is the covariance of the *posterior* estimation error at time t_{k-1}, and update it (by incorporating the expired time) to obtain $\mathrm{P}_k(-)$, that is, the covariance of the *prior* estimation error at time t_k. In this context, *prior* (or *a priori*) and *posterior* (or *a posteriori*) mean before (respectively after) incorporating the information related to the reading at a given time. The computation can be done using the following equation:

$$\mathrm{P}_k(-) = \Phi_{k-1}\mathrm{P}_{k-1}(+)\Phi_{k-1}^T + \mathrm{Q}_{k-1} \tag{3.13}$$

where

$\mathrm{P}_k(-)$ — *A priori* value of the covariance matrix of estimation uncertainty at time t_k, with size $n \times n$.

Φ_{k-1} — State transition matrix at time t_{k-1}, with size $n \times n$.

$\mathrm{P}_{k-1}(+)$ — *A posteriori* value of the covariance matrix of estimation uncertainty at time t_k, with size $n \times n$.

Q_{k-1} — Covariance matrix for the process noise $\underline{\mathbf{w}}$ at time t_{k-1}, with size $n \times n$.

3.3.2 Step Two: State estimate update

In this step, the state estimate gets updated by incorporating the information of the last reading (the measurement at the current time instant). To do so, we first have to compute the Kalman gain, $\overline{\mathrm{K}}_k$, which gives us, so to speak, the weighting factor (reliability) of each source of information. To compute $\overline{\mathrm{K}}_k$, several variables are used: $\mathrm{P}_k(-)$ (computed in step 1), H_k and R_k, via the following equation [Grewal and Andrews, 2015]:

$$\overline{\mathrm{K}}_k = \mathrm{P}_k(-)\mathrm{H}_k^T[\mathrm{H}_k\mathrm{P}_k(-)\mathrm{H}_k^T + \mathrm{R}_k]^{-1} \tag{3.14}$$

where

$\overline{\mathrm{K}}_k$ — Kalman gain matrix at time t_k, with size $n \times l$.

$\mathrm{P}_k(-)$ — *A priori* error covariance matrix at time t_k, with size $n \times n$.

H_k — Measurement sensitivity matrix, which defines the linear relationship between the state of the dynamic system and the measurements that can be made at time t_k. Its size is $l \times n$.

R_k — Noise covariance matrix for measurement noise $\underline{\mathbf{v}}$ at time t_k, with size $l \times l$.

3.3.3 Step Three: Covariance error update

Since $\overline{\mathrm{K}}_k$ has already been computed in step 2 and $\mathrm{P}_k(-)$ was computed in step 1, we can update the error covariance matrix (the prediction uncertainty) using the following equation [Grewal and Andrews, 2015]:

$$\mathrm{P}_k(+) = [\mathrm{I} - \overline{\mathrm{K}}_k\mathrm{H}_k]\mathrm{P}_k(-) \tag{3.15}$$

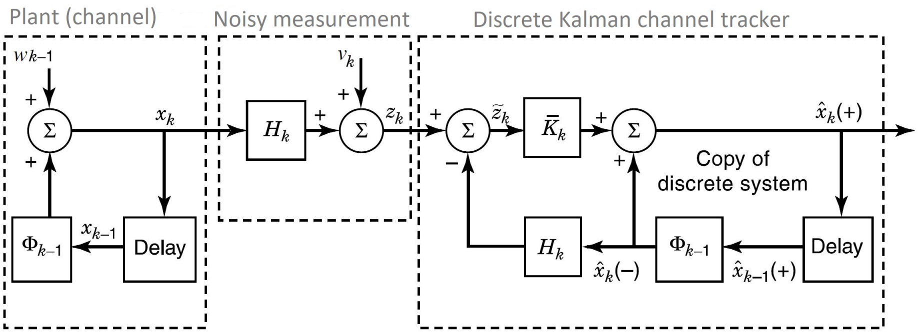

Figure 3.1: Block diagram for the linear system, the measurement and the discrete KF. Adapted from [Grewal and Andrews, 2015].

The parameters in this equation have already been defined in previous sections. It is interesting to note that the factor that updates the matrix P is sometimes denoted by $\mathrm{K}_k^1 = \mathrm{I} - \overline{\mathrm{K}}_k \mathrm{H}_k$ and generates a projection that is orthogonal to the Kalman gain.[9]

3.3.4 Step Four: Estimation and recursion

In this step, successive values for $\underline{\hat{\mathbf{x}}}_k(+)$ are computed recursively using the computed values for $\overline{\mathrm{K}}_k$ (from step 3), the given initial estimate $\underline{\hat{\mathbf{x}}}_0$, and the input data $\underline{\mathbf{z}}_k$[10], using the following equation:

$$\underline{\hat{\mathbf{x}}}_k(+) = \underline{\hat{\mathbf{x}}}_k(-) + \overline{\mathrm{K}}_k[\underline{\mathbf{z}}_k - \mathrm{H}_k \underline{\hat{\mathbf{x}}}_k(-)] \tag{3.16}$$

By iteratively performing steps 1 to 4 for each instant, an online estimate is obtained using Kalman filtering. Figure 3.1 shows a block diagram of the linear system, the measurement and the discrete Kalman filter.

3.4 Treatment of Measurement Vectors with Uncorrelated Errors as Scalars

In many (if not most) applications with measurement vectors $\underline{\mathbf{z}}$, the corresponding measurement noise covariance matrix R is a diagonal matrix, which means that the individual components of $\underline{\mathbf{v}}_k$ are uncorrelated. For those applications, it is advantageous to consider the components of $\underline{\mathbf{z}}$ as independent scalar measurements rather than vector measurements. In this manner, two important advantages are obtained: a much shorter computation time (the number of arithmetic computations is significantly less) and a greater numerical precision (by avoiding rounding in matrix inversions).

The implementation of the filter, in those cases, would require l iterations of the update equations with the new observation, using the rows of H as measurement "matrices" (in reality, of

[9] In a way, Kalman filtering is as if one were to create a product space with two projections: one projection is "what can be known about the actual state from the measure"; the other projection would be "what cannot be known about the real state from the measurement, but from the model". The first projection would be $\mathrm{K}_k^1 \underline{\hat{\mathbf{x}}}_k(-)$ and the second projection would be $\overline{\mathrm{K}}_k \underline{\mathbf{z}}_k$; and their sum would be $\underline{\hat{\mathbf{x}}}_k(+)$.

[10] Note that $\underline{\mathbf{z}}$ is random as measured by random (Gaussian) noise, but in the Kalman filter the vector $\underline{\mathbf{z}}$ is already known to us. Bold only emphasizes that it contains random noise.

dimension 1, and, therefore, scalars) and the diagonal elements of R as the corresponding (scalar) measurement noise variance.

In that sense, such a matrix filter would be equivalent to a bank of scalar filters, each of which could be implemented independently as if doing Kalman tracking of totally unrelated scalar variables (mathematically they are). The plant and noise model associated with each filter would be given by the equivalent scalar equations (3.5) and (3.7) that we derived in the previous sections.

In the framework of this Thesis, matrix-scalar reduction will be useful when we want to follow multipath channels. If we assume that the gain of each path can be considered independent in its evolution, the Kalman tracking can be done by means of a bank of Kalman filters, each of which will follow one of the paths (or, to be more precise, *taps*) of the channel.

3.5 Kalman Filter Applications in OFDM Channel Tracking

Kalman filtering applied to channel tracking in OFDM systems has been proposed in the technical literature through multiple and diverse approaches. Below we briefly present some of these applications.

3.5.1 KF tracking outperformance vs. training-sequence-based estimate

A novel channel estimation and tracking method for OFDM wireless systems based on pilots and Kalman filtering was proposed in [Yuanjin, 2003]. This paper looked at coherent burst OFDM systems, such as those adopted by the IEEE 802.11a (802.11g) and ETSI HiperLAN/2 standards, using a Kalman filtering algorithm to overcome the AWGN and ICI, as well as pilots to track variations in the impulse response of the channel (including a pilot-based phase compensation system). This technique was simulated in a channel with Rayleigh fading and a 5.0 GHz 802.11 wireless LAN system. The entire algorithm could be processed in real time with significantly better performance than the channel estimation method (without KF) based on training sequences or pilots.

3.5.2 KF for tracking subchannels over time according to Jakes model

When the Doppler frequency is known, KF is widely used for tracking [Barbieri et al., 2009, El Husseini et al., 2019, Kashyap et al., 2017]. An auto-regressive (AR) model is assumed for the transition dynamics, and the parameters are chosen either based on a Doppler dependent model, e.g., Jakes model or by fitting the parameters to the data [Pratik et al., 2021].

In [Chen and Zhang, 2004] and later in [Zhang and Chen, 2004], an attempt was made to optimize a method unrelated to Kalman filtering: MMSE channel estimation applied to time of frequency. This technique did not address time-domain dynamics, which the authors solved by presenting a KF method for time- and frequency-selective fading channels. Based on an autoregressive model of the Jakes model, MMSE estimation in the frequency domain was combined with KF tracking of subchannel dynamics over time. This two-stage system offered comparable performance to a much more complicated joint-Kalman (time-frequency) estimator. The authors concluded that the good performance derived from this simplification of the system could be due to the fact that the time and frequency components of the Jakes model presented in [Van De Beek et al., 1995] and [Chen and Liu, 2000] were of a separable nature.

3.5.3 An application of the predictive character of KF

A very original use of KF appears in [Simeone and Spagnolini, 2004]. It is an adaptive pilot pattern for OFDM systems. The underlying idea is that adapting the link across fading channels based on information (available at the receiver) about the state of the channel increases the spectral efficiency and reliability of the link. Typically, that adaptation has referred to power, adaptive modulation, etc. The authors propose that the placement of the pilot subcarriers in the time-frequency grid is also adapted to the prediction that, by means of a Kalman filter (specifically, the calculation of the error covariance matrix), is obtained from the error of the channel estimate. Thus, based on the Kalman prediction of the foreseeable error, the number of pilot subcarriers that guarantees a sufficiently reliable channel estimate based on the Quality of Service (QoS) requirements can be minimized. The simulations showed the efficiency of the algorithm with respect to a periodic re-training.

3.5.4 The limits of KF: theory vs. practice

Published simulations are not always practical for implementation on real systems. The Kalman filter proposed in [Al-Naffouri, 2007] makes collective use of the data and channel constraints inherent in the communications problem, with the disadvantage that this entails excessive computational cost. That is, other approaches focus on a subset of the possible constraints (the statistical properties of the channel, the pilots, etc.) while [Al-Naffouri, 2007] used all the constraints and integrated them into a Kalman filter (a forward-backward KF, although a simplified implementation is also proposed as a forward-type KF) that essentially makes a potentially unlimited number of recursive approximations that improve accuracy. Specifically, the simulations showed that more signal processing always produced better performance in terms of Bit Error Rate (BER). This work shows us that, theoretically, by means of Kalman filtering, unexpected[11] improvements can be achieved in exchange, albeit, for a computational cost totally outside the range that can be implemented in current practice.

3.5.5 Combining KF with other powerful algorithms

Some of the most interesting KF techniques are those that are proposed in combination with some other algorithm that has shown its power separately in previous works. Thus, for example, in [Kim et al., 2005], it is proposed to use a QR decomposition algorithm (QRD)[12] adapted to MIMO-OFDM, the QRD-M algorithm, having verified the power of QR decomposition in channel estimation and data joint detection for CDMA in a previous study. The decision rule of Maximum Likelihood (ML) based on the QRD corresponds to a complete search in a tree, a search whose computational complexity can be reduced considerably by combining it with the algorithm M. The Kalman filter is used for the joint estimation of channel coefficients in a similar way to [Komninakis et al., 2002]. The QRD-M stage uses the channel estimate calculated during the previous symbol interval, that is, the prediction that in this PFC we denote with $\hat{\underline{\mathbf{x}}}_k(-)$. Simulations showed that this combined QRD-M-KF system outperformed other techniques and had

[11] Unexpected to the extent that the Kalman filter is optimal in linear estimation problems, and it is not obvious that there are linear relationships between aspects as diverse as the correlation between frequency and time or the finite alphabet of symbols (to name two examples used) and the best estimate.

[12] In linear algebra, the QR decomposition or factorization of a matrix is its decomposition as the product of an orthogonal matrix Q by a triangular upper R. The QR decomposition is often used to solve the linear least squares problem and is the basis of the QR algorithm used for computing the vectors and eigenvalues of a matrix.

significantly lower computational cost, especially for larger constellations and a higher number of transmitting antennas. Furthermore, it was robust even for large (normalized) Doppler delays. In summary, by combining KF with another proven algorithm, the authors achieved a good candidate for implementation in MIMO-OFDM systems.

Another combination with the KF is presented in [Banelli et al., 2007]. In this case, data-assisted Kalman tracking is performed for channel estimation in Doppler-affected OFDM systems. The Kalman filter is used to estimate the (fast) variation of the channel from one OFDM symbol to another, taking advantage of information from a Basis Expansion Model (BEM). The Kalman filter estimate is iteratively refined by the aid of pilots and/or data in Pilot Symbol Assisted Modulation (PSAM) [Jiménez et al., 2019], whether real or virtual (V-PSAM). In this way, by combining KF and PSAM, it is possible to improve the global BER, although the authors suggest that this technique could be further studied.

These studies suggest that combining KF with other proven techniques can lead to significant improvements in BER.

3.5.6 Unknown-parameter channels and trend channels

The most common applications of KF are based on autoregressive or very similar models (such as BEM) that approximate[13] well-defined channel models such as Jakes or Rayleigh.

A peculiar case is that in which the channel model is not perfectly characterized. For example, an attempt is made in [Han et al., 2004] to estimate fast-fading channels for OFDM using a Modified Kalman Filter (MKF), but although the channel can be described using an autoregressive (AR) process, its parameters are unknown and it is proposed to estimate the parameters of the AR process by minimizing the Mean Square Error (MSE). The result is that the simulated channels are random walks and the performance of the MKF is comparable to that of the KF that does know the parameters of the AR process. In practice, this means that KF can be used even when there is uncertainty about the fading parameter (or Doppler frequency).

Instead, in [Shu et al., 2014], the typical autoregressive model is discarded in favor of an integrated random walk model.[14]. Their simulations support the idea that, in this case, the resulting one-path KF outperforms the KFs based on autoregressive models proposed in the literature, such as some of those explained above.

3.5.7 Practical conclusions

From the background review in the literature, the following practical conclusions can be drawn:

- The application of KF to OFDM systems **can improve channel estimation and reduce BER**. Among the studies cited, for example, simulations of [Kashyap et al., 2017, Yuanjin, 2003, Chen and Zhang, 2004, Al-Naffouri, 2007, Kim et al., 2005, Banelli et al., 2007, Shu et al., 2014] demonstrate improvements in channel estimation and reductions in bit error rate. In the case of [Simeone and Spagnolini, 2004], the improvement consists of meeting the QoS conditions with the minimum number of pilots, and in [Han et al., 2004] the added

[13]The process for creating these models typically consists of synthesizing the desired model with a filter and then obtaining the process noise variance (or covariance matrix, if applicable) from the Yule-Walker equation as described, among others, in [Porat, 2008].

[14]"Integrated random walk" means that the speed plot is a random walk. The plot of the position (that is, the graph of the complex amplitude of the subchannel) is the integral of the speed, the integral of the random walk. Therefore, the complex amplitude is not a random walk, but rather shows trends.

improvement of being able to work with unknown Doppler fading/frequency channels is presented.

- The typical proposal consists of **applying KF to the monitoring of the temporal variation of the subchannels**. Although there are numerous different proposals to apply KF to OFDM, practically all of them (including all the papers explained above) use KF to take advantage of time-domain correlation, independently of, or combined with, other systems that take advantage of time-domain correlation. the frequency domain. In that sense, some of the most interesting proposals combine KF with other proven techniques, as in [Kim et al., 2005] with QR decomposition, or as pilot assistance (PSAM) in [Banelli et al., 2007].

- **Simplification of matrix KF using a bank of scalar Kalman filters** may be accurate under certain circumstances. Using scalar Kalman filters for each subchannel is equivalent to assuming that the time evolution of each subchannel is uncorrelated with that of the other subchannels. In [Shu et al., 2014] it is shown that, when the temporal variation is slow or moderate, an estimate of the multipath complex amplitude of the channel can be obtained using the integrated Random Walking (RW) model. The proposed singlepath RW-KF filter bank is shown to be as efficient as the exact RW-KF (that is, the joint multipath RW-KF) and outperforms the KFs based on autoregressive models proposed in the literature. This supports our decision (explained later) to perform the multipath simulations according to uncorrelated singlepath Gauss-Markov models.

Therefore, the intention to analyze Kalman filtering in OFDM systems (filtering that aims to improve channel estimation and tracking, taking advantage of the correlation of each subchannel in the time domain, and simplifying the matrix KF by means of Kalman filter banks scalars), fits perfectly with current knowledge and the most recent proposals published in the technical literature. Furthermore, the popularity and strength of the KF underscore the need to study how it behaves in non-stationary environments; it is also worth noting the paucity of detailed studies on how the birth and death of paths affect KF: only [Angelosante et al., 2007] and our own work [Méndez-Romero, 2015] have shown KF degrades catastrophically in the presence of tap birth/death dynamics.

Chapter 4

Abrupt Channel Tracking

Since the purpose of this Thesis is the enhancement of wireless communications by better tracking channels when they change abruptly (a problem we might call *abrupt channel tracking*), we need to review the different solutions for abrupt change detection proposed either for communications or in other areas.

We will start by providing a historical overview of the Theory of Abrupt Changes and other abrupt-change detection methods (Section 4.1), followed by an intuitive explanation of Random Set Theory (RST) and RST-based solutions for abrupt channel tracking (Section 4.2). Then we will focus on classical (threshold-based) online changepoint detection algorithms (Section 4.3); while mostly developed in the 20th century, they will give us some ideas about how to develop our own threshold-based solutions for abrupt channel tracking. Finally, other recent developments and advanced online changepoint detection algorithms will be explored in Section 4.4.

4.1 Historical Overview of Abrupt Change Detection

4.1.1 Early research

Interest in on-line change detection first arose in the area of quality control: control charts were introduced in [Shewhart, 1931] and cumulative sum charts in [Page, 1954]. The study of slow, drift-like changes in system states, systematized in Kalman's development of Kalman filters, advanced in parallel to the study of abrupt changes.

Regarding abrupt changes, two main classes of statistical problem statements emerged: the Bayesian and non-Bayesian approaches.

The first Bayesian abrupt change detection problem was an on-line quality control problem stated in [Girshick and Rubin, 1952]. The first optimality results concerning Bayesian change detection algorithms were obtained in [Shiryaev, 1961, Shiryaev, 1963, Shiryaev, 1965].

On the other hand, the first investigation of non-Bayesian abrupt change detection algorithms was made in [Page, 1954]. Cumulative sum algorithms were proven to be asymptotically optimal in [Lorden, 1971].

4.1.2 Theory of Abrupt Changes

Apart from its first applications in quality control [Shewhart, 1931], these methods found an early application in the analysis of biomedical signals such as electroencephalograms, where many contributions to the problem of automatic segmentations of signals were made [Jones et al., 1970, Borodkin and Mottl, 1976, Mathieu, 1976, Sanderson et al., 1980]. Segmentation algorithms aided in geophysical signal processing, e.g., diagraphy [Basseville and Benveniste, 1983] and seismology [Nikiforov and Tikhonov, 1986], and introduced the first step towards continuous speech recognition in [Andre-Obrecht, 1988].

Bayesian and non-Bayesian algorithms were described in detail in [Basseville and Nikiforov, 1993], the essential late 20th century compendium for abrupt-change detection techniques. This compendium was updated in [Tartakovsky et al., 2014] and expanded upon in [Tartakovsky, 2019]. Another important work is [Brodsky, 2016], where not only retrospective and sequential change-point analysis is performed, but also two novel methods are introduced: early change-point detection (i.e. detection of a change that begins gradually) and a method to detect structural changes; these change detection methods are shown to be either optimal or asymptotically optimal and can be successfully applied to financial time series (e.g. early detection of periods of higher volatility in stock exchanges).

4.1.3 State of the art

Interestingly, machine learning (ML) techniques weren't included in the early papers nor in the most recent textbooks. The development of Neural Networks (NN) and related efficient ML algorithms arguably has the potential to revolutionize the detection of abrupt changes, provided that such abrupt changes can be trained beforehand – since abrupt changes sometimes deal with unknown future distributions, the solid mathematical framework behind the self-sustained area of Sequential Analysis, as the sequential detection of abrupt changes is sometimes called, is in heavy demand and irreplaceable.

More recent developments in online change-point detection, including particle filtering, message passing, machine learning, variational learning and advanced sequential learning, are discussed in Section 4.4.

4.1.4 Taleb's Incerto

No historical overview of abrupt changes can be complete if we don't mention Taleb's Incerto [Taleb, 2012, Taleb, 2007, Taleb, 2001, Taleb, 2018, Taleb, 2020] and its via-negativa/non-predictive view of stochastical problems as the best risk management approach [Taleb, 2012, Taleb, 2020], as well as its main idea that Black Swans (i.e. model incompleteness with very relevant rare effects on performance) are the main drivers of performance degradation in what he called "Extremistan", as opposed to "Mediocristan" (ergodic environments where average error matters more than a huge single-point error) [Taleb, 2007, Taleb, 2001, Taleb, 2012]. This categorization of problems can be traced back to multifractal geometry [Mandelbrot and Hudson, 2008].

Most systems considered in the current Thesis aren't financial non-ergodic applications with risk of ruin[1], but communication channels where, although there might be a jump between two states, the relevant metric is an average over time, whether tracking error or BER, i.e. our communication systems usually live in "Mediocristan". However, channel tracking schemes must allow for a potential future where Ultra-Reliable Low Latency Communications (URLLC) become standard for widespread applications, including autonomous vehicle communications (which might

[1] Actually, we will briefly consider financial non-ergodic applications with risk of ruin in Chapter 9.

occasionally require ultra-low-latency remote driving) or military drone swarms; URLLC are subject to the engineering equivalent for risk of ruin, and Extremistan stochastics such as Extreme Value Theory (EVT) and channel models based on the Generalized Pareto Distribution (GPD) arise [Mehrnia and Coleri, 2022]. Extremistan effects and their relevance to abrupt changes are discussed in Appendix Subsection B.3.4.

4.1.5 Random-Set Theory

A recent branch of maths, Random Set Theory, has been proposed to tackle several problems in communications and signal processing. Particle filters for random set models have been proposed for multi-track tracking, estimation using imprecise measurements and calibration of tracking systems [Ristic, 2013], as well as detection of different voices in multi-user speech and detection of tap birth-death in multipath channel tracking. Random Set models were discussed in Section 4.2.

The main disadvantage of Random Set models is that they lead to very difficult-to-handle integrals that require particle filtering to solve. Particle filters (Subsection 4.4.1) involve a huge computational cost that's unfeasible in real conditions for some applications. Thus, a simplification was required.

4.2 Random Set Theory for Channel Tracking

In the previous chapter, we studied the Kalman filter, which is the optimal estimator in Gaussian, linear problems. What happens when the problem involves non-linear evolutions (i.e. non-stationarity through "jumps"? In such non-stationary environments, the Kalman filter loses its optimality. In the field of communications, the non-stationarity of the channel due to abrupt changes causes the channel estimation based on Kalman filtering to lose performance. Performance can be degraded catastrophically in such non-stationary conditions, as shown in [Méndez-Romero, 2015].

One of the approaches proposed to track non-stationary channels in OFDM, when such non-stationarity corresponds to the appearance and disappearance of new paths, is based on Random Set Theory (RST) [Jiménez et al., 2020]. This chapter explains what Random Set Theory is, how it has been introduced in different environments associated with Telecommunications Engineering, and, specifically, how it can help track non-stationary channels in OFDM environments.

4.2.1 Concept and history of Random Set Theory

Random sets are random elements that take values as subsets of some space, serving as general mathematical models for observations whose values are sets, as well as for irregular geometric patterns. Random sets generate the traditional concept of ordinary random points/vectors.

Let's explain it in a more understandable way. $\{3, \pi, 7.23\}$ is a set of three real numbers. Similarly, a set of vectors can be defined without difficulty. Actually, a set can be defined in any space of elements by simply taking one or more of those elements. Now, just as you can define a random variable that takes a real number as its value, e.g., a r.v. $\mathbf{x}$ that takes real values following a uniform probability distribution $\mathrm{U}[0, 1]$, and just as a random entity $\underline{\mathbf{y}}$ can be defined that takes a vector as its value (and in that case we speak of a "random vector"), generalized probabilities can also be defined on any other type of sets, and then such sets can be called "random sets".

In fact, the usual random variables in communications (such as those that take values in $\mathbb{R}$, in $\mathbb{C}$ or in $\mathbb{C}^n$, are just a specific type of *random* sets, just like the natural numbers or the intervals of $\mathbb{R}^n$ are concrete examples of *sets*.

Think, for example, of a set where not only each element is random but also the number of elements itself; for example, the set of paths[2] of a multipath channel. Not only do we not know the complex gain associated with each path, but we may also not know if there are two, three, or four active paths at any given time. The natural way to model this is by a random set where the number of elements is random (and so is the value of each element). This is the type of generalized random set that can be applied for non-stationary channel tracking and other communication engineering problems.

Considering the fact that they are such a natural generalization of the concepts of probability and statistics applied to the usual random variables (and especially to random vectors), it is perhaps surprising that, although they have been used since the middle of the 20th century (for example, for statistical sampling designs in [Hájek, 1981], statistical geometry in [Kendall, 1974], and statistics in [Robbins, 1944]), its first formalized treatment (as random sets are understood today[3]) was not done until 1975 in [Mathéron, 1975].

In the foreword to Mathéron's book on random sets, G. Watson expressed his view of statistics in these words [Mathéron, 1975]:

> "Modern statistics must be defined as the applications of computers and mathematics to data analysis. It must grow as new types of data are considered and as computing technology advances."

That quote is perhaps the best way to express what has been happening (and possibly will continue to happen in the future) around the relationship between random sets and engineering. Thus, out of sheer necessity, engineering (and finance, etc.) have had to resort to advances in mathematics to find new solutions in data analysis and this, in turn, has propelled mathematical research into Random Set Theory, with important recent contributions in [Molchanov, 2017].

4.2.2 Incorporation of Finite Set Statistics to multi-target tracking

Random Set Theory was first applied to telecommunications to solve multitarget tracking problems, becoming popular as Finite Set Statistics (or FISST, from FInite Set STatistics) with the excellent work by Mahler, Goodman and Nguyen [Goodman et al., 2013].

After all, Random Set Theory is nothing more than a probability theory of sets (for all our purposes, we can consider them finite) that exhibit randomness not only in each of the elements, but also in the number of elements [Biglieri et al., 2012]; thus, they make up the ideal mathematical setting to study problems in which not only do we not know which element is there, but we also do not know how many elements there are. This approach is common in radar problems (you don't know who the targets are, but sometimes you don't know how many targets there are and how many signals are red herrings), and also, as we will see in the next section, in wireless communications problems.

Applying the mathematical techniques developed in [Goodman et al., 2013], once a posterior density function $f(\mathcal{H}_p|\mathbf{y}_{1:p})$ is obtained in the entire set space (for example, the plane/case where

[2] There is a distinction between taps and paths. Paths can contribute to different taps, as explained in 2.5. However, for the purposes of the intuitive explanation in this Section, they will be treated as the same concept, as some RST authors do.

[3] Although Kolmogorov already treated random sets in [Kolmogorov, 1950], [Mathéron, 1975] is credited with their first formalized treatment as random variables in locally compact separable Hausdorff spaces. For more details on Mathéron's original contributions, please refer to [Molchanov, 2005].

there is one target, and for the planes/cases where there are two, three, four, five targets, etc.), conditioned to the samples obtained, a pair of Bayesian estimators can be defined, the so-called Global Maximum A Posteriori (GMAP) estimators: GMAP-I (or “Marginal Multi-Target Estimator”) and GMAP-II (or “Joint Multi-Target Estimator”). The difference between them lies in the fact that GMAP-I is a two-stage estimator in which the cardinality of the set is first estimated (that is, first it is decided whether there are 3 objectives or 7; and, later, it is decided which targets those 3 or 7 are), while GMAP-II performs the estimation in a single stage.

The practical implementation of these algorithms in radar applications involves the programming of particle filters. In general, they are very computationally expensive but that is not too much of a problem in some military applications, at least compared to what is common in wireless communications. Bear in mind that it is not the same to learn, with a one second delay, how many enemies there are and where they are (especially if there is no technological alternative that does it before and better), than to delay any communication between a base station and a mobile by an additional second, for instance.

Other military applications without a direct translation in the field of communications are those related to mapping all the possible positions of the enemy, taking into account the inaccuracy of the underlying probabilistic model [Ristic, 2011]. The same idea has merit in wireless communications, namely: the randomization of the stochastic model used, in order to incorporate the uncertainty about the starting hypothesis.

4.2.3 RST for OFDM Channel Tracking

The application of RST to multi-target tracking problems aroused the interest of Biglieri and Lops, who decided to apply the same ideas to wireless communication problems. First they applied it to the problem of multi-user detection:

> In mobile multiple-access communications, not only the lo cation of active users, but also their number varies with time. In typical analyses, multiuser detection theory has been developed under the assumption that the number of active users is constant and known at the receiver, and coincides with the maximum number of users entitled to access the system. This assumption is often overly pessimistic, since many users might be inactive at any given time, and detection under the assumption of a number of users larger than the real one may impair performance.
>
> – [Biglieri and Lops, 2006]

The authors therefore decide to resort to RST because it makes no practical sense to assume that all users are active at the same time. (This contrasts with the practical utility of resorting to RST in OFDM channel tracking problems, where the advantage is another: to avoid the great degradation that occurs in Kalman filtering when there are birth and death paths). Later, the authors turned their sights on another aspect of communications: tracking on multipath channels and, with the team expanded to a new author (Angelosante), managed to create an RST model for tracking multipath in OFDM [Angelosante et al., 2007] and in MIMO-OFDM [Angelosante et al., 2009].

The RST model for multipath tracking is not trivial in complexity. The reader who wants to get started with these techniques can find a detailed exposition and with a guide of recommendations in Appendix A. The basic idea is to define random sets of elements with two “dimensions”[4]: one

[4] The term “dimension” is used here in an explanatory manner and without any rigor. The terms “facets”, “sides” or, in a much more rigorous way, “projections” could be used as synonyms.

dimension is the tap index, a number which defines it as tap (tap 1, tap 5, etc.); the other dimension is the value of the complex gain (e.g. $0.8 + 0.2i$). A very intuitive way to understand the problem is to imagine some good portions of cured cheese; these portions, cut into cubes[5] or slices[6], can be placed one after the other. Thus, we could speak of a first cheese or first tap, a fifth cheese or fifth tap, etc. Inside each piece of cheese, there is a hole; exactly one hole. That hole represents the point in $\mathbb{R}$ or $\mathbb{C}$, respectively, that coincides with the true tap gain at a given instant. For the thought experiment to succeed, the slices must not be folded and must only have a single hole.

In front of the cheese slices there is a person with his eyes covered, which is us (the estimator!); just by feeling for a few seconds with gloved fingers, we have to detect, by gloved touch, the location of those holes. Unfortunately, because we do not have much time and gloved touch isn't very reliable, our situation is precarious: we have to decide on a strategy to follow, where to touch first, and what conclusions to draw.

And that is where the different possible estimators come into play when applying RST. One could, for example, think of touching the center of the slice first, to detect if that slice or path has the hole right in the center (zero gain, i.e. idle path). Or, on the contrary, you could think of first touching the place where you remember the hole was in the previous instant, anticipating that now it is not very far from there. You can decide first which taps are dead (which slices have the hole right in the middle), or maybe you can decide first how many taps are dead (quickly tapping all the slices without noticing their order), etc.

These are, explained in a very easy and intuitive way[7], the different estimation techniques available. If one quickly touches all the slices without noticing their order (their tap index), for example, one is calculating the first stage of the GMAP-I estimator (the second consists of calculating the location of the holes in each slice). The mathematical expression of this estimator has the following form:

$$\text{GMAP-I:} \begin{cases} \hat{n}_p = \arg\max_{n_p \in 0,\ldots,L_{max}} f_{n_p|\mathbf{y}_{1:p}}(n_p, \mathbf{y}_{1:p}), \\ \widehat{H}_p = \arg\max_{\mathcal{H}_p : |\mathcal{H}_p| = \hat{n}_p} f_{\mathcal{H}_p|\mathbf{y}_{1:p}}(\mathcal{H}_p, \mathbf{y}_{1:p}), \end{cases} \tag{4.1}$$

where the details of the parameters can be consulted in Appendix A. As can be seen, these are quite convoluted expressions. This convoluted character is probably one of the biggest shortcomings of RST, as it is a significant barrier for unfamiliar engineers.

"What is the optimal solution?", the reader might ask. The GMAP-III estimator, a technique proposed in [Angelosante et al., 2007]. Following the allegory of the cheese slices, it would consist of feeling the center of the slices and establishing which of the slices have the hole in the center (e.g. slice 2 and slice 4). Afterwards, one concentrates on feeling the remaining slices (the other paths) to locate the hole (calculating the true gain from probability densities). The mathematical expression is:

$$\text{GMAP-III}: \begin{cases} \widehat{\pi(\mathcal{H}_p)} = \arg\max_{\pi(\mathcal{H}_p)} f_{\pi(\mathcal{H}_p)|\mathcal{Y}_{1:p}}(\pi(\mathcal{H}_p)|\mathcal{Y}_{1:p}), \\ \widetilde{\mathbf{h}}_p = \int_{\mathbb{R}^{2|\widehat{\pi(\mathcal{H}_p)}|}} \mathbf{h}_p f_{\mathbf{h}_p|\mathcal{Y}_{1:p}}(\mathbf{h}_p|\mathcal{Y}_{1:p}) d\mathbf{h}_p \end{cases} \tag{4.2}$$

where

[5] If they are strip-shaped, it is easier to see the real line. They are ideal for imagining real-valued tap gains.

[6] A sufficiently large slice is the ideal metaphor for the complex plane or a bounded portion of it.

[7] For a full, formal derivation of GMAP estimators, see Appendix A.

$$f_{\pi(\mathcal{H}_p)|\mathcal{Y}_{1:p}}(\pi(\mathcal{H}_p)|\mathcal{Y}_{1:p}) = \int_{\pi'(\mathcal{H}_p)} f(\mathcal{H}_p|\mathcal{Y}_{1:p})\delta\mathcal{H}_p \tag{4.3}$$

and the definition of each parameter can be consulted in the next Subsection. The main advantage of applying RST (with respect to the pre-existing alternatives) lies in the fact that simulations in [Angelosante et al., 2007, Angelosante et al., 2009] showed a great advantage of their estimation with respect to the unmodified Kalman filter and with respect to LS estimation. For a more detailed discussion, see Appendix A.

4.3 Threshold-Based Online Changepoint Detection

4.3.1 Elementary algorithms

Let us consider a sequence of independent random variables (y_k) with a probability density $p_\theta(y)$ depending upon only one scalar parameter. Before the unknown change time t_0, the parameter θ is equal to 0, and after the change it is equal to $\theta_1 \neq \theta_0$. Our goal is to detect and estimate this change in the parameter.

This problem statement is usually accompanied with some other assumptions, such as the Gaussian nature of the sequence (e.g. detection of a change in the mean of an independent Gaussian sequence) or assuming that the parameter θ_0 before change is known.

The classical, closed-form solutions for such problem statements considered in this section are based on the concept of log-likelihood ratio, i.e. the logarithm of the likelihood ratio, defined by:

$$s(y) = ln\frac{p_{\theta_1}(y)}{p_{\theta_0}(y)} \tag{4.4}$$

where ln denotes the natural logarithm. Typically, you'd expect θ_1 to be the value of θ in the ranges where $p_{\theta_1}(y) > p_{\theta_0}(y)$, and viceversa. This is the key statistical property of this ratio, which is very important in mathematical statistics. More formally, let $\mathbb{E}_{\theta_0}[s]$ and $\mathbb{E}_{\theta_1}[s]$ denote the expectations of the log-likehood ratio under the distributions p_{θ_0} and p_{θ_1}. Then, $\mathbb{E}_{\theta_0}[s] < 0$ and $\mathbb{E}_{\theta_1}[s] > 0$. In other words, you can detect a change in the parameter θ as a change in the sign of the mean value of the log-likelihood ratio.

4.3.1.1 Sufficient statistic

Given a random sample $(X_1, X_2, ..., X_n)$ from a statistical population with a theoretical distribution in a parametric family $\{F_\theta|\theta \in \Theta\}$, a statistic $T(X_1, X_2, ..., X_n)$ is called *'sufficient'* if the sample distribution, conditional to the value for the statistic T, does not depend on θ [García and Vélez, 2012].

The log-likehood ratio for the observations y_j to y_k:

$$s_i = ln\frac{p_{\theta_1}(y_i)}{p_{\theta_0}(y_i)} \tag{4.5}$$

is the sufficient statistic [Basseville and Nikiforov, 1993] for elementary changepoint detection problems.

4.3.2 Shewart control chart

The idea of using Shewart control charts emerged in quality control (continuous inspection). The Shewart control chart is basically a sliding window algorithm, i.e. a moving-average control chart. The size of the window (the size of the samples taken) N is fixed. At the end of each sample we have to decide between the two following hypotheses about the parameter θ:

$$\begin{cases} H_0 & \theta = \theta_0 \\ H_1 & \theta = \theta_1 \end{cases} \tag{4.6}$$

As long as the decision is taken in favour of no change/no defect in quality control (i.e. the null hypothesis H_0), new samples are taken and tested. Sampling/production would be stopped after a change/defect is detected (i.e. the hypothesis H_1).

How do we take a decision? For a fixed sample size N, the optimal decision rule d is given by [Basseville and Nikiforov, 1993]:

$$d = \begin{cases} 0 & if\, S_1^N < h;\, H_0\, is\, chosen \\ 1 & if\, S_1^N < h;\, H_1\, is\, chosen \end{cases}$$

where

$$S_j^k = \sum_{i=j}^{k} s_i \tag{4.7}$$

4.3.3 Geometric Moving Average algorithm

The Geometric Moving Average (GMA) algorithm is based on two ideas: the fact that a change in the parameter is reflected as a change in the sign of the mean value of the log-likelihood ratio, and the idea of weighting observations exponentially. The GMA decision function can be rewritten in a recursive manner as [Basseville and Nikiforov, 1993]:

$$g_k = (1-\alpha)g_{k-1} + \alpha s_k \tag{4.8}$$

where $g_0 = 0$ and the coefficient α acts as a forgetting factor.

4.3.4 Finite Moving Average algorithm

The Finite Moving Average (FMA) algorithm replaces the exponential forgetting operation in Eq. 4.8 by a finite memory one [Nikiforov and Tikhonov, 1986]. Thus, a finite set of weights γ_i is now required:

$$g_k = \sum_{i=0}^{N} \gamma_i ln \frac{p_{\theta_1(y_{k-i})}}{p_{\theta_0(y_{k-i})}} \tag{4.9}$$

This algorithm requires tuning some parameters, namely the size N of the sliding window, the weights γ_i, which are any weights for causal filters, and the threshold h that determines the stopping rule: $t_a = min\{k : g_k \geq h\}$.

4.3.5 CUSUM algorithm

The CUmulative SUM (CUSUM) algorithm, first proposed in [Page, 1954], can be obtained through different derivations, leading to slightly different presentations. Two of those derivations follow.

4.3.5.1 Derivation of the CUSUM algorithm as typical change behaviour detector

The typical behaviour of the log-likehood ratio S_k shows a negative drift before change and a positive drift before change [Basseville and Nikiforov, 1993]. Therefore, the corresponding rule is based on comparing the difference between the value of the log-likehood ratio and its current minimum value, to a threshold:

$$g_k = S_k - m_k \geq h \tag{4.10}$$

where

$$S_k = \sum_{i=1}^{k} s_i \tag{4.11}$$

$$s_i = ln \frac{p_{\theta_1}(y_i)}{p_{\theta_0}(y_i)} \tag{4.12}$$

$$m_k = min_{1 \leq j \leq k} S_j \tag{4.13}$$

The stopping condition would be

$$t_a = min\{k : S_k \geq m_k + h\} \tag{4.14}$$

or equivalently, in the case of change in the mean of a Gaussian sequence, as an integrator compared to an adaptive threshold [Basseville and Nikiforov, 1993].

4.3.5.2 CUSUM as a repeated SPRT

[Page, 1954] derived the CUSUM algorithm by using the repeated testing of two simple hypotheses:

$$\begin{cases} H_0 & \theta = \theta_0 \\ H_1 & \theta = \theta_1 \end{cases} \tag{4.15}$$

with the aid of the Sequential Probability Ratio Test (SPRT). Full derivations can be found in [Page, 1954, Basseville and Nikiforov, 1993]. The resulting decision rule can be recursively written as

$$g_k = (g_{k-1} + s_k)^+ \tag{4.16}$$

where $(x)^+ = sup(0, x)$, or equivalently as

$$g_k = (S_{k-N_k+1})^+ \tag{4.17}$$

where

$$N_k = N_{k-1} \cdot \mathbf{1}_{\{g_{k-1}>0\}} + 1 \tag{4.18}$$

and $1_{\{x\}}$ is the indicator of event x, and t_a is defined as

$$t_a = min\{k : g_k \geq h\} \tag{4.19}$$

The CUSUM algorithm can thus be seen as a randomly-sized sliding-window algorithm.

4.3.6 Bayes-type algorithms

Abrupt change detection algorithms can be derived with a Bayesian approach in which *a priori* information about the distribution of the change time is available, e.g. for the abrupt-change time t_0. A full derivation can be found in [Girshick and Rubin, 1952, Basseville and Nikiforov, 1993]. The main idea: deciding that a change has occured when the prior probability of a change exceeds a conveniently chosen threshold. For example, the application of this Bayesian approach to a change in mean, where mean values μ_0, μ_1 and constant variance σ^2 are known, would lead to the following decision function:

$$g_k = ln(\rho + e^{g_{k-1}}) - ln(1-\rho) + ln\frac{p_{\theta_1(y_k)}}{p_{\theta_0(y_k)}} \tag{4.20}$$

The tuning parameters of this Bayes-type algorithm are the prior probability ρ of a change, the initial probability π implicit in g_0, and the threshold h [Basseville and Nikiforov, 1993].

4.3.7 Unknown parameter after change

Abrupt-change detection algorithms exist for the case where the parameter θ_1 after change is unknown, while $\theta_0$is known.

4.3.7.1 χ^2-CUSUM algorithm

Let us consider the problem of detecting a change in the mean of a Gaussian sequence with known variance σ^2, in the special case where the distribution $F(\theta) = F(\mu)$ is concentrated on two points, $\mu_0 - v$ and $\mu_0 + v$. The stopping time can be proven [Tartakovsky et al., 2014] to be given by Eq. 4.19, where

$$g_k = max_{1\leq j\leq k}[ln\, cosh(b\bar{S}_j^k) - \frac{b^2}{2}(k-j+1)] \tag{4.21}$$

where $b = v/\sigma$.

4.3.7.2 Generalized Likelihood Ratio algorithm

The Generalized Likelihood Ratio (GLR) algorithm exploits prior information about the parameter θ after change. If a minimum magnitud v_m of the changes in the parameter θ is known, then the following decision function can be derived [Basseville and Nikiforov, 1993]:

$$g_k = max_{1\leq j\leq k} sup_{\theta_1} S_j^k(\theta_1) \tag{4.22}$$

For the previous example of change in the mean of an independent Gaussian sequence, if we assume $v_m = 0$, then the following decision function can be derived:

$$g_k = max_{1 \leq j \leq k} \sum_{i=j}^{k} [\frac{\bar{v}_j(y_i - \mu_0)}{\sigma^2} - \frac{\bar{v}_j^2}{2\sigma^2}] \tag{4.23}$$

$$\bar{v}_j = \frac{1}{k-j+1} \sum_{i=j}^{k} (y_i - \mu_0) \tag{4.24}$$

4.4 Recent Developments in Online Change-Point Detection

4.4.1 Particle Filters

Particle filtering (PF), also known as Sequential Monte Carlo (SMC) uses a set of particles (samples) to represent the posterior distribution of a stochastic process given the noisy and/or partial observations. Particle filter techniques provide a well-established methodology for generating samples from the required distribution without requiring assumptions about the state-space model or the state distributions. In signal processing, a popular genetic particle filtering method called 'bootstrap filter' was first proposed in [Gordon et al., 1993]; this method is easy to implement and particularly flexible, since it does not require any assumption about that state-space or the noise of the system.

An excellent tutorial on particle filtering is [Arulampalam et al., 2002]. PF has been proposed for the kind of abrupt change detections studied in this Thesis, under the name "Jump Markov Linear Systems", in [Doucet et al., 2001]. Random Set Theory models for abrupt change detection in channel tracking have been proposed to be solved through particle filtering [Angelosante et al., 2007, Angelosante et al., 2009]. Unfortunately, the computational cost of such methods is too high [Méndez-Romero and Fernández-Getino García, 2018].

An excellent resource for particle filtering in the Random-Set-Theoretical context is [Vihola, 2004] and, for the most advanced particle-filtering applications over random sets, [Ristic, 2013]. A formal derivation of GMAP estimators, including the application of RST particle filters in the channel tracking problem with tap birth/death dynamics, is provided in Appendix Section A.3.

4.4.2 Message Passing

The intuition behind message passing [Adams and Mackay, 2007] is computing the exact Bayesian probability that the last change-point happened long ago or recently. To be more precise, it considers the history of change-points up to time t is a binary sequence, e.g. $c_{1:t} = 1, 0, 0, 1, 0, 1, 1$, where the value 1 indicates an abrupt change in the corresponding time step. Thus, a random variable $R_t = minn \in \mathbb{N} : C_{t-n+1} = 1$ can be defined in order to describe the time since the last change-point, which takes values between 1 to t. The algorithm assumes the probability of the next change-point can be inferred from history, and so it can generate an accurate distribution of the next unseen datum in the sequence, given only data already observed [Adams and Mackay, 2007]. Probabilities are updated after each sample in a manner resembling particle filtering. Unfortunately, the computational complexity and memory requirements of the complete message passing algorithm increase linearly with time t [Liakoni et al., 2021], which reduces its applicability to the tracking problem analyzed in this Thesis. However, it might be useful, in combination with other schemes, in the context of unstereotyped dynamics such as those described in Section 9.4, particularly if a surprise mechanism is considered, such as the “Bayes Surprise Factor” [Liakoni et al., 2021].

4.4.3 Machine Learning

Our first proposal to solve the abrupt-change detection in the tap birth-death problem was a simplified framework [Méndez-Romero and Fernández-Getino García, 2018] that combined the insights from Random Set Theory and Kalman filtering with a set of thresholds. We called this proposal the "Simplified Maximum a Posteriori" or SMAP approach (see Chapter 5) and can be considered a threshold-based machine-learning approach.

The advent of neural networks opens new possibilities and challenges (Subsection 7.2.1). They can be explored by combining NN detectors of tap birth and death with Kalman filters for channel tracking [Mendez-Romero et al., 2020]. The result, called "Neural-Network-Switched Kalman Filters", is described in detail in Chapter 7.

4.4.4 Variational Learning

Variational learning, also known as variational Bayesian methods [Beal, 2003], intends to construct an analytical approximation to the posterior probability of the set of unobserved variables (parameters and latent variables), given the data. Both types of unobserved variables (namely, parameters and latent variables) are treated as random variables.

By analytical approximation, a formula is meant that usually consists of a product of well-known probability distributions; such a formula is an approximation. This approximation has the basic property that it is a factorized distribution, i.e. a product of two or more independent distributions over disjoint subsets of the unobserved variables.

Mathematical manipulations on this approximation help identify the probability distributions of the factors, and mutually dependent formulas for the parameters of these distributions. The actual values of these parameters are computed numerically, through an alternating iterative procedure.

Variational learning can be combined with a surprise mechanism, such as the “Bayes Surprise Factor” [Liakoni et al., 2021], for learning in volatile environments (see also Section 9.4).

4.4.5 Advanced Sequential Analysis

The Theory of Abrupt Change Detection in [Nikiforov and Tikhonov, 1986] has been further developed in recent years [Brodsky, 2016], particularly in the so-called “Change Detection and Isolation” [Pergamenchtchikov et al., 2022] problem. That is, the scenario where the quickest change detection problem is generalized to the case of multiple post-change hypotheses (diagnosis), which can be formulated as joint change detection and identification.

Multihypothesis CUSUM-type and SR-type procedures with some minimax optimality properties are proposed by [Tartakovsky et al., 2014]. Some proposals are based on the assumption that the prior distribution of the change point is geometric also in the i.i.d. case. For non-i.i.d. data and composite post-change hypotheses, a general non-Bayesian asymptotic multistream change detection identification theory (minimax and pointwise) was proposed in [Pergamenchtchikov et al., 2022]. These methods are mathematically optimal but, from an engineering point of view, computationally unfeasible for the channel tracking problems central to this Thesis. Nevertheless, for well-defined abrupt change problems where solutions are not required to be computationally inexpensive (such as some problems explained in Chapter 9), these methods should be considered. An example of their power: the ability to detect COVID-19 in Italy much earlier than standard methods [Pergamenchtchikov et al., 2022].

4.5 Chapter Summary

Different solutions are reviewed for abrupt change detection either in communications (Random Set Theory, Particle Filters) as well as in general cases (Threshold-Based Online Changepoint Detection, Advanced Sequential Analysis). For the channel tracking problem central to this Thesis, this review concludes a lower-complexity abrupt change detection algorithm is needed for practical channel tracking applications. Nevertheless, threshold-based methods might inspire one such algorithm (see Section 5.4) and some advanced methods might also be useful for advanced channel tracking problems and generous complexity budgets (see Section 9.4).

Part II

Novel Contributions

Chapter 5

Simplified Maximum A Posteriori Multipath Detection

This Thesis' primary goal is to track channels with significant tap birth/death dynamics[1] through computationally feasible (i.e. computationally inexpensive) mechanisms. Thus, instead of using computationally expensive, Random-Set-Theory-based (RST) particle filtering[2], this Thesis proposes to use a switched Kalman filtering[3] (KF) system with computationally inexpensive tap birth/death detectors.

To be more precise, the proposed tracking scheme is a KF bank with pre-connected switches. The purpose of those switches is the activation/deactivation of stationary tracking (KF) as a new tap is born or dies. Thus, those switches need to be controlled by a tap birth/death detection mechanism (Fig. 5.1). This Thesis proposes two such mechanisms:

- A threshold-based system called 'Simplified Maximum a Posteriori' (SMAP), which is described in this Chapter 5.
- A trained Neural Network (NN) system that receives feedback from the KF bank; this combined system, called Neural-Network-switched Kalman Filtering (NNKF), is described in Chapter 7.

Both systems could be potentially extended to include detection of other kind of abrupt changes, such as abrupt changes in Signal-to-Noise Ratio (SNR) or abrupt lateral shifts in tap energy. An extension of the SMAP system detecting abrupt changes in SNR is proposed in Chapter 6; abrupt changes in partial tap components (*lateral partial tap hopping*) are proposed, modelled and geometrically justified in Chapter 8, which also includes a discussion of potential NNKF solutions for them.

5.1 Statement of the Problem

We consider an OFDM system employing N orthogonal subcarriers, transmitting OFDM symbols with a time duration T_{symb} and a sampling T_{samp}. It is assumed there is no out-of-band interference.

[1] For a discussion of tap birth/death dynamics, see Sections 2.5 and 2.6.
[2] For a discussion of Random-Set-Theory-based solutions, see Section 4.2.
[3] Kalman filters for channel tracking are described in Chapter 3.

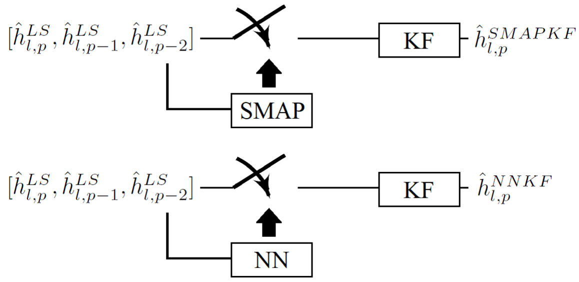

Figure 5.1: Tap birth/death tracking schemes using SMAP (top) and NNKF (bottom).

Each subcarrier is used both for data transmission (through data blocks of fixed length M_{inf}) and for pilot symbols enabling channel estimation. Each data block is preceded by K pilot symbols, so that the first K sent symbols are pilot symbols used to estimate the channel (by averaging over K); this channel estimation will be considered "valid" and will be used for the whole subsequent M_{inf}-symbol data block transmission. Channel changes will be tracked only once every estimation period $T_{est} = (K + M_{inf}) \cdot T_{symb}$, since a certain channel stationarity within each estimation period can be assumed due to its coherence time and, accordingly, channel changes can be reasonably modeled as occurring at the beginning of each estimation period.

If $k = \{1, ..., K\}$ is the pilot symbol index in the pth estimation period, then the received signal, which is the input for channel estimation, can be given in vector form as:

$$\mathbf{y}_{p,k} = \mathbf{D}_{p,k}\mathbf{F}_p\mathbf{h}_p + \mathbf{z}_{p,k} \tag{5.1}$$

where $\mathbf{y}_{p,k} = [y_{p,1,k}, ..., y_{p,N,k}]^T$, each $y_{p,n,k}$ (for $n = \{1, ..., N\}$) represents the nth subcarrier observation sample at time $t_{p,k} = (p-1) \cdot T_{est} + (k-1) \cdot T_{symb}$; $\mathbf{D}_{p,k} = diag(d_{p,1,k}, ..., d_{p,N,k})$, with $d_{p,n,k}$ the training data on the nth pilot subcarrier at time $t_{p,k}$; $\mathbf{z}_{p,k} = [z_{p,1,k}, ..., z_{p,N,k}]$, with $z_{p,n,k}$ representing the zero-mean complex Gaussian additive noise having variance σ_z^2; if $l = \{1, ..., L(pT_{est})\}$ and $L(pT_{est})$ is the number of active taps, then $\mathbf{h}_p = [h_1(pT_{est}), ..., h_{L(pT_{est})}(pT_{est})]^T$, with $h_l(pT_{est})$ the complex gain of the lth tap at time p, which, as previously explained, will be assumed constant for the duration of the estimation period; and

$$\{\mathbf{F}_p\}_{n,l} = e^{-j2\pi n \cdot \frac{\tau_l(pT_{est})}{NT_{samp}}} \tag{5.2}$$

where $\tau_l(pT_{est})$ is the delay of the lth tap during the pth interval.

Clearly, since each data block is preceded by K pilot symbols, the received signal including reception of all K pilot symbols will be the matrix $\mathbf{Y}_p = [\mathbf{y}_{p,1}, ..., \mathbf{y}_{p,K}]$. A guard time, named cyclic prefix, T_{CP} is reserved between OFDM symbol transmissions, so that $T_{symb} = NT_{samp} + T_{CP}$. It is assumed that the multipath delay spread is smaller than T_{CP}.

Moreover, the active tap gains are assumed to follow an underlying Linear Gauss-Markov (LGM) model (see Subsection 5.3.1), so that, if $a_p^{(l)}$ is to represent the lth tap gain at time p provided that lth tap is active, then the probability density function for $a_p^{(l)}$ would be given by both following equations:

$$f(a_1^{(l)}) = \mathcal{N}(a_1^{(l)}; 0, \sigma_{h_l}^2) \tag{5.3}$$

$$f(a_p^{(l)}|a_{p-1}^{(l)}) = \mathcal{N}(a_p^{(l)}; \lambda a_{p-1}^{(l)}, (1-\lambda^2)\sigma_{h_l}^2) \tag{5.4}$$

where $\sigma_{h_l}^2$ is the average energy of the lth tap and λ is the temporal self-correlation of each active tap gain. However, notice that, if the lth tap is not active (because it has "died"), then $h_l(pT_{est})$=0, i.e. tap gain equals zero. Each active tap has a probability P_{death} of becoming inactive; each inactive tap has a probability P_{birth} of becoming active.

The problem under consideration can now be formulated as follows: given the observations (1), determine a computationally inexpensive causal estimator $\hat{\mathbf{h}}_p$ for $\mathbf{h}_p$, relying upon $\{\mathbf{Y}_{1:p}\}$. Since "computationally inexpensive" may be too vague a term, a slightly different, more precise formulation of the problem would be: determine the *simplest* causal estimator for $\mathbf{h}_p$ exploiting tap self-correlation to the extent of getting *most* of the theoretically maximum possible reduction in Channel Tracking Mean Squared Error (CTMSE), defined as:

$$CTMSE \triangleq ||\sum_p \hat{\mathbf{h}}_p - \mathbf{h}_p||_2^2 \tag{5.5}$$

5.2 Birth-Death Dynamics in Realistic Channels

Several measurement campaigns have found evidence for, and quantified, abrupt change events in real communication scenarios, as well as realistic abrupt change models and the theoretical underpinning of abrupt changes in mobile radio [Hassan et al., 2020, Hassan et al., 2021, Matolak, 2008, Sen and Matolak, 2008, Wu et al., 2010] channels. Thus, tap birth-death dynamics are a realistic feature of some non-stationary channels, such as those in vehicle-to-vehicle (V2V) communications. For a detailed survey of such works, see Section 2.6.

5.3 Channel Tracking

5.3.1 Linear Gauss-Markov (LGM) models

Currently, the digital nature of modulation and demodulation, as well as the ease of simulation with matrix-based algorithms (such as those of the MATLAB program), make the use of the discrete-time model the ideal solution for simulations of OFDM systems. In this Chapter, we have used a model similar to the one in Chapter 2, but with a few more simplifications that would allow us to focus more closely on the key aspects to be tested.

In particular, the Linear Gauss-Markov (LGM) channel model will be used. LGM constitutes the probabilistic model underlying the Kalman filter and is therefore of central importance in recent developments around channel tracking in OFDM (including the current Thesis).

5.3.1.1 Definition and brief theoretical explanation

Consider the scalar state variables[4] $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_t$, $\mathbf{x}_{t+1}$, where bold indicates that they are random variables, and consider that $\mathbf{z}_1$, $\mathbf{z}_2$, ..., $\mathbf{z}_t$, $\mathbf{z}_{t+1}$ are the sequence of corresponding observations. As in hidden Markov models, conditional independences (see Figure 5.2) state that past and future states are uncorrelated given the current state, $\mathbf{x}_t$, at time t. This means that if, for example, we

[4]The linear Gauss-Markov model can, of course, also be applied to vectors and complex one-dimensional state variables. The reasoning would be analogous and its derivation would trivially follow from the presented derivation.

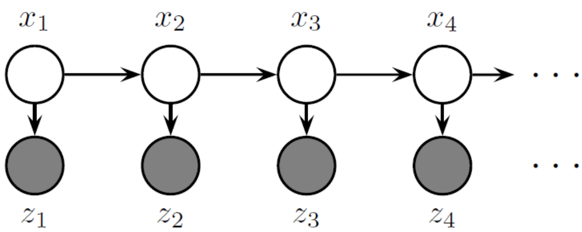

Figura 5.2: Independence diagram in a Markov model

know the value of $\mathbf{x}_2$, then no information about $\mathbf{x}_1$ could help us reason what value $\mathbf{x}_3$ should have.

Then $\{\mathbf{x}_t\}$ follows a linear Gauss-Markov model if:

$$\mathbf{x}_{t+1} = \lambda \mathbf{x}_t + \epsilon \tag{5.6}$$

where

λ is a scalar parameter (in our case, a real-valued constant); and

ϵ is a random variable of errors, with mean $E(\epsilon) = 0$ and $Var(\epsilon) = \sigma_\epsilon^2$. Furthermore, typically (and also for all practical purposes of this paper) the error is assumed to follow a normal distribution: $\epsilon \sim N(0, \sigma_\epsilon^2)$.

Let us now consider a channel with several paths and we will see how to use this expression. For a given time[5] p, let us denote the gain of the kth tap by $\mathbf{h}_p^{(k)}$. We want the tap gain to vary with a given speed with respect to what it was at time $p-1$. A linear Gauss-Markov model that satisfies this condition would have the form:

$$\mathbf{h}_p^{(k)} = \lambda \mathbf{h}_{p-1}^{(k)} + \epsilon \tag{5.7}$$

where λ is a scalar parameter and the error would follow a normal distribution $\epsilon \sim N(0, \sigma_\epsilon^2)$ with variance σ_ϵ^2, which will be a measure of how fast the channel varies. The parameters used in our own simulations[6], as well as in various reference papers [Angelosante et al., 2007, Angelosante et al., 2009], are: $\lambda = 0.999$ and $\sigma_\epsilon = (1-\lambda^2)\sigma_h^2$, where σ_h^2 is the mean energy of a single tap.

Depending on the goal of the model, you could initialize $\mathbf{h}_1^{(k)}$ to a given value (e.g. $\mathbf{h}_1^{(k)} = 0$) or you could randomize your initial value. In such a case, for example, the kth path gain at time p could follow the following distributions:

$$\mathbf{h}_1^{(k)} \sim \mathcal{N}(0, \sigma_h^2) \tag{5.8}$$

$$\mathbf{h}_p^{(k)} | \mathbf{h}_{p-1}^{(k)} \sim \mathcal{N}(\lambda \mathbf{h}_{p-1}^{(k)}, (1-\lambda^2)\sigma_h^2) \tag{5.9}$$

Note that the Gauss-Markov model is a very rough approximation to the real behavior of any physical channel, to the point that it is always possible to obtain gain values $|\mathbf{h}_p^{(k)}| > 1$, which represent a channel that amplifies the signal instead of being lossy. Obviously, the usual physical channels

[5] The notation "time p" or "time t_p" is used interchangeably to refer to the same brief time interval.
[6] Our simulations in [Méndez-Romero, 2015] also included other values, namely $\lambda = \{0.9; 0.95; 0.99\}$.

do not amplify. Depending on the parameters, that situation can be made relatively unlikely, but this anomaly shows that the advantage of the linear Gauss-Markov model is not that it realistically models the usual physical channels (it doesn't); its power lies in the fact that the Kalman filter provides the mathematically optimal solution for estimating Gaussian and linear channels. Therefore, when comparing the performance of a Kalman filter with that of a new algorithm, it makes sense to rely, even if partially, on the linear Gauss-Markov model (as the most favorable extreme or "quasi-optimal" scenario for the Kalman filter). This is the implicit reason for using linear Gauss-Markov that emerges from the works [Angelosante et al., 2007, Angelosante et al., 2009] that compare standard Kalman filtering with different methods, such as those resulting from applying Finite Random Set Theory.

In summary, for the purposes of this Chapter we will consider a LGM model where all taps are active and they follow Eq. (5.3) and (5.4). These LGM models have been previously used in the literature (e.g. in [Angelosante et al., 2007, Angelosante et al., 2009]) and they make it possible to derive a computationally inexpensive, optimal estimation through Kalman filtering. However, they are ideal channels whose behaviour may or may not approximate specific real channels. In this regard, a severe disadvantage is that, since they are perfectly linear, they don't allow for jumps.

5.3.2 KF-based approaches

Under a LGM model, a KF-based approach can be easily implemented. Kalman filtering is an algorithm weighting optimally two information sources: a theoretical one (in our case, the LGM channel model) and another one based on noisy measurements. The Least-Squares (LS) channel estimation, $\hat{\mathbf{a}}_p^{LS}$, can be interpreted as a noisy measurement of the true channel $\mathbf{a}_p = [a_p^{(1)}, ..., a_p^{(L)}]^T$:

$$\mathbf{u}_p \triangleq \hat{\mathbf{a}}_p^{LS} = \mathbf{a}_p + \mathbf{v}_p \tag{5.10}$$

where $\mathbf{v}_p = [v_p^{(1)}, ..., v_p^{(L)}]^T$ is the measurement noise and $\mathbf{u}_p = [u_p^{(1)}, ..., u_p^{(L)}]^T$ will be used for ease of notation. Since tap independence is being assumed, this translates into L scalar equations. Thus, a bank of independent Kalman filters could be used to track the whole channel, whereby each Kalman filter would receive the LS-estimated tap gain and would compute the corresponding KF tap estimation, and also the expected tap gain at time $p+1$, the so-called Kalman prediction, $\hat{u}_p^{(l)}(-)$. This KF approach is optimal when non-linearities are absent. However, the disappearance and reapparance of taps (i.e. the real channel "jumping" in a way the perfect LGM model cannot fit) introduce a severe nonlinear distortion and KF performance degrades catastrophically.

5.3.3 RST-based approaches

RST-based approaches have proven that, under birth-death conditions, a very good estimator is the so-called GMAP-III, which essentially adds a death-birth detector before the KF step. Thus, you first detect which taps are active, and then you estimate active taps' gains. The GMAP-III estimator was proposed in [Angelosante et al., 2007] with the following definition:

$$\text{GMAP-III}: \begin{cases} \widehat{\pi(\mathcal{H}_p)} = \arg\max_{\pi(\mathcal{H}_p)} f_{\pi(\mathcal{H}_p)|\mathcal{Y}_{1:p}}(\pi(\mathcal{H}_p)|\mathcal{Y}_{1:p}), \\ \widetilde{\mathbf{h}}_p = \int_{\mathbb{R}^{2|\widehat{\pi(\mathcal{H}_p)}|}} \mathbf{h}_p f_{\mathbf{h}_p|\mathcal{Y}_{1:p}}(\mathbf{h}_p|\mathcal{Y}_{1:p}) d\mathbf{h}_p \end{cases} \tag{5.11}$$

where

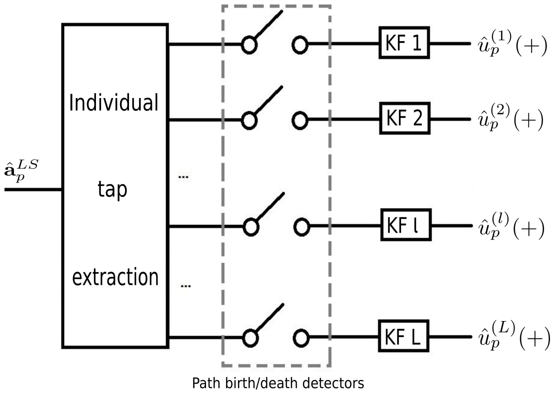

Figure 5.3: A bank of tap death/birth detectors as switches for single-tap KF.

$$f_{\pi(\mathcal{H}_p)|\mathcal{Y}_{1:p}}(\pi(\mathcal{H}_p)|\mathcal{Y}_{1:p}) = \int_{\pi'(\mathcal{H}_p)} f(\mathcal{H}_p|\mathcal{Y}_{1:p})\delta\mathcal{H}_p \tag{5.12}$$

(For details on notation, see the full derivation of GMAP estimators in Appendix A). However, death-birth detection under the RST framework can be difficult to handle and computationally prohibitive; in [Angelosante et al., 2007], death-birth detection is done through a 10,000-particle filter that approximates Eq. (5.11). A much simpler, suboptimal estimator is required. Such an estimator will be proposed in Section 5.4, but, firstly, a measurement of its quality will be introduced in the following Subsection.

5.3.4 Measuring the quality of simpler estimators

Our objective is to reduce CTMSE as defined in Eq. (5.5). What is the theoretically maximum possible reduction in CTMSE when using some given birth-death information?

First, let us consider Fig. 5.3 to see how tap birth-death information could be used in practice. Following the notation for Eq. (5.10), Fig. 5.3 starts with the LS channel estimation, $\hat{\mathbf{a}}_p^{LS}$, having its individual tap components extracted and fed into individual tap birth-death detectors. These detectors are like switches and are represented as such in Fig. 5.3: on birth detection, the ith LS tap estimation is fed into the ith KF, which starts tracking (thus, $\hat{u}_p^{(i)}(+)$ is the ith estimated tap gain); on death detection, however, the ith estimated tap gain is put to zero until the next time the ith tap is detected to be reborn. Since the problems of birth-death detection, on the one hand, and active-tap tracking, on the other hand, are separable in nature, the maximum reduction in CTMSE is obtained when, for each tap, an optimal birth-death detector is connected to a KF.

Now, let us suppose perfect information on the active/inactive status of each tap was available. What would the theoretically maximum possible reduction in CTMSE be when using this perfect birth-death information? Death-birth detectors would then be always right, and looking into this ideal case, henceforth the Ideal Switching System (ISS), could provide useful information about the problem and the quality of different solutions to it.

Thus, the ISS is the system drawn in Fig. 5.3 when death/birth detectors detect 100% of births and 100% of deaths, with no false birth/death detection. It is also possible to easily define an x%-degraded Ideal Switching System (IIS-x%), which is a switching system detecting x% of births and x% of deaths, with no false birth/death detection. These degraded ISS were simulated for the problem at hand (for more details, see Section 5.5) and Table 5.1 shows the % reduction

Table 5.1: % reduction in CTMSE for differently degraded ISS vs. LS estimation.

SNR [dB]	IIS	IIS-99.9%	IIS-99.5%	IIS-99%
8.5	72.84	72.46	69.35	66.52
11	70.16	68.86	64.81	59.14
13.5	65.40	63.75	56.81	45.77
16	56.93	53.83	40.59	23.77
18.5	41.81	35.29	8.72	-16.40

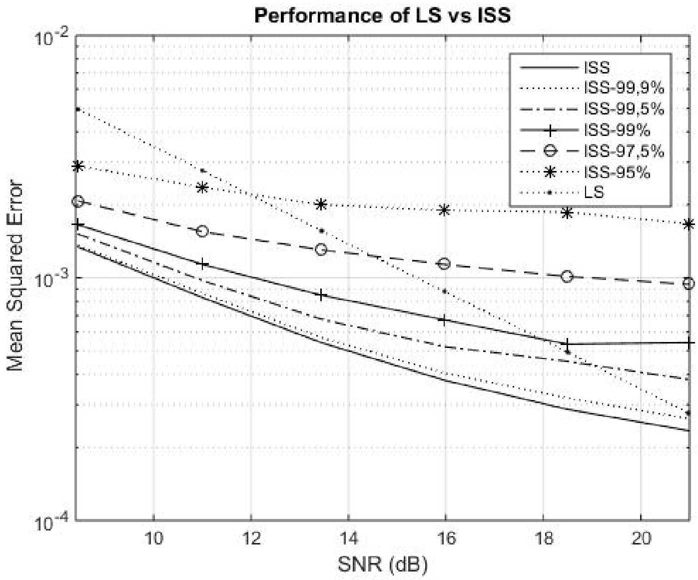

Figure 5.4: CTMSE performance of LS vs. ISS and different x%-degraded ISS

(vs. the LS method, i.e., no KF) these near-optimal ISS devices could achieve, as measured in CTMSE, for different SNR values. A quick look at Table 5.1 makes it possible to conclude that ISS performance in terms of CTMSE degrades significantly even with minor reductions in the % of birth/death detections. For higher SNR, the degradation is so catastrophic that the conventional LS estimation performs better than a slightly degraded ISS (e.g. a 99%-ISS, negative value in the bottom-right corner of Table 5.1). Do note, however, that the degraded ISS greatly improves estimation over conventional LS estimation for low-to-medium SNR levels.

These conclusions are also backed by Fig. 5.4, which plots the degraded systems' performance in terms of CTMSE. This kind of plots let us establish a natural measure for an estimator's fitness in the context of the previously defined problem. Some estimator could be, for example, better than ISS-97% and worse than ISS-98%, meaning that it would be slightly better than a degraded ISS where detectors detected 97% of births and 97% of deaths, but not as good as a system detecting 98% of both.

.

5.4 Simplified Maximum A Posteriori (SMAP) Estimator

A Bayesian-inspired estimator of $\mathbf{h}_p$ can be defined by using KF in combination with the three following heuristics for tap birth-death detection:

1. detect death if the tap gain jumps into approx. zero;
2. detect death if the tap gain has slowly converged into approx. zero;
3. detect birth if the tap gain is far from zero.

The precise parameters for these three simple rules can be either obtained empirically (massive simulations) or approximated with simple theoretical derivations as explained in the three following subsections.

5.4.1 Memoryless detection of large leaps into a narrow, zero-centered range

When the lth tap is dead, its true gain is, by definition, zero. Thus, the observed gain is just the (Gaussian) noise. Hence, tap gain measurement $u_p^{(l)}$ follows a Gaussian distribution:

$$u_p^{(l)} | l\, is\, dead \sim N(0, \sigma^2_{v^{(l)}}) \tag{5.13}$$

where $\sigma^2_{v^{(l)}}$ is the variance of tap (measurement) noise $v_p^{(l)}$. On the other hand, if lth tap is active, then its gain follows a Gaussian centered on the previously expected gain, $\hat{u}_p^{(l)}(-)$, the Kalman prediction computed at time $p-1$ (this is a feature of KF). For the purposes of tracking problems like the one in consideration, it has been found that $\sigma^2_{v^{(l)}}$ makes a good approximation for the corresponding variance. Thus,

$$u_p^{(l)} | l\, is\, alive \sim N(\hat{u}_p^{(l)}(-), \sigma^2_{v^{(l)}}) \tag{5.14}$$

Let us assume that the tap was alive at time $p-1$. Then, either tap death happens at time p (with prior probability P_{death}) or the tap will still be alive (with prior probability $1 - P_{death}$). Thus, a maximum a posteriori criterium leads to detecting "death" whenever

$$P_{death} \cdot f(u_p^{(l)} | l\, is\, dead) > f(u_p^{(l)} | l\, is\, alive) \cdot (1 - P_{death}) \tag{5.15}$$

where the events "l is dead/alive" mean that the lth tap is dead/alive, respectively. Note that this detector compares two scaled Gaussian pdfs, so an equivalent expression to Eq. (5.15) would be:

$$P_{death} \cdot \frac{1}{\sqrt{2\pi\sigma^2_{v^{(l)}}}} \cdot e^{-\frac{(u_p^{(l)})^2}{2\sigma^2_{v^{(l)}}}} > \frac{1}{\sqrt{2\pi\sigma^2_{v^{(l)}}}} \cdot e^{-\frac{(u_p^{(l)} - \hat{u}_p^{(l)}(-))^2}{2\sigma^2_{v^{(l)}}}} \cdot (1 - P_{death}) \tag{5.16}$$

By taking logarithms and simple rearrangement, two other equivalent expressions can be found:

$$ln(\frac{P_{death}}{1 - P_{death}}) - \frac{(u_p^{(l)})^2}{2\sigma^2_{v^{(l)}}} > -\frac{(u_p^{(l)} - \hat{u}_p^{(l)}(-))^2}{2\sigma^2_{v^{(l)}}} \tag{5.17}$$

$$ln(\frac{P_{death}}{1 - P_{death}}) > \frac{\hat{u}_p^{(l)}(-) - 2u_p^{(l)}\hat{u}_p^{(l)}(-)}{2\sigma^2_{v^{(l)}}} \tag{5.18}$$

5.4.2 Memory detection of a sequence at a close range of zero

Let the active tap gain be close to zero (but active, so never exactly zero!) at time p; since large leaps are highly improbable [Peebles Jr, 2001] as Gaussian outliers, let's assume that tap gain will always stay close to zero at time $p+1$. (This is a reasonable assumption if done only for very short sequences, typically 2 or 3 succesive samples). Thus, the active tap is assumed to be close to zero, continously active and static, and the observed tap gain is assumed to be just measurement noise (a reasonable assumption when tap gain is very close to zero). Under these assumptions, the probability of having a sequence of length s at any range $q_B \cdot \sigma_v$ from zero, such that the tap is continously active and $\{|z_p|, ..., |z_{p+s}|\} < q_B \cdot \sigma_v$, is:

$$(1 - P_{death})^s \cdot (1 - 2Q(q_B))^s \tag{5.19}$$

where $Q(\cdot)$ is the Q-function and q_B is any arbitrary threshold consistent with the aforementioned assumptions. Eq. (5.19) follows directly from the properties of the Gaussian distribution [Peebles Jr, 2001]. On the other hand, the probability of a tap dying right before the sequence or during the sequence (and not being reborn afterwards) would be:

$$\sum_{i=0}^{s} (1 - P_{death})^i \cdot P_{death} \cdot (1 - P_{birth})^{s-i} \tag{5.20}$$

Thus, if the system is continuously monitoring the presence of sequences in this range, tap death can be decided following a simplified maximum a posteriori criterium whenever:

$$\sum_{i=0}^{s} (1 - P_{death})^i \cdot P_{death} \cdot (1 - P_{birth})^{s-i} > (1 - P_{death})^s \cdot (1 - 2Q(q_B))^s \tag{5.21}$$

Thus, after setting a single appropiate threshold q_B not too far from zero and a short enough sequence length s, tap death can be decided trivially when an $(s+1)$-long sequence $\{|u_p^{(l)}|, ..., |u_{p+s}^{(l)}|\} < q_B \cdot \sigma_v$ has been detected.

For example, if a sequence monitor for $q_B = \sigma^2_{v^{(l)}}/3$ and $s = 2$ is set up, then this detector would decide "dead" after a 2-long sequence $\{|u_p^{(l)}|, ..., |u_{p+s}^{(l)}|\} < q_B \cdot \sigma_{v^{(l)}} = \sigma^3_{v^{(l)}}/3$ per Eq. (5.21). If assumptions are reasonable and hold true, the probability of having made the wrong decision can be shown [Méndez-Romero and Fernández-Getino García, 2018] to be:

$$P_{err} = \frac{(1 - P_{death})^s \cdot (1 - 2Q(q_B))^s}{P_{aux} + (1 - P_{death})^s \cdot (1 - 2Q(q_B))^s} \tag{5.22}$$

where

$$P_{aux} = \sum_{i=0}^{s} (1 - P_{death})^i \cdot P_{death} \cdot (1 - P_{birth})^{s-i} \tag{5.23}$$

5.4.3 Birth detection

Birth detection could easily be implemented as a threshold detector that decides "alive" whenever $|u_p^{(l)}| > q_C \cdot \sigma_{v^{(l)}}$, for a certain q_C. Even if P_{err} per Eq. (5.22) were too high (close to 0.5), a very sensitive birth detector (a very low q_C) would correct potential errors very early. Thus, detecting a tap birth is a correct decision not only when a new tap is created, but also when a tap was wrongly detected as "dead".

Let's assume our birth detection happens immediately after having decided "tap is dead" in the scenario described in Subsection 4.2. Since P_{err} in Eq. (5.22) is the probability of that previous decision having been wrong, and since noise can make an LS estimate of a dead tap deviate outside a q_C range from zero with a probability $2Q(q_C)$, then the probability of correcting a previous error would be:

$$P_{corr} = P_{err} \cdot (1 - P_{death}) \cdot 2Q(q_C) \tag{5.24}$$

On the other hand, a new tap could have been created (with probability P_{birth}) if the previous decision ("dead tap") was right (with probability $1 - P_{err}$). For simplicity purposes, let's assume that newly created taps start at $|u_{p+1}^{(l)}| > q_C \cdot \sigma_{v^{(l)}}$. Thus, the probability of correctly detecting a new birth would be:

$$P_{birth\,det} = (1 - P_{err}) \cdot P_{birth} \tag{5.25}$$

A birth decision would be wrong if the tap was erroneously detected as dead at time p but it has recently become dead at time $p+1$ and noise puts it outside the detection limits, or when it was correctly detected as dead at time p and is still dead, but noise makes the observation sample shoot outside the detection limits:

$$P_{false\,det} = P_{err} \cdot P_{death} \cdot 2Q(q_C) + (1 - P_{err}) \cdot (1 - P_{birth}) \cdot 2Q(q_C) \tag{5.26}$$

Thus, a simplified maximum a posteriori criterium would decide "birth" whenever:

$$P_{corr} + P_{birth\,det} > P_{false\,det} \tag{5.27}$$

These expressions looks cumbersome but, in fact, you only need them to obtain any appropiate q_C for which Eq. (5.26) holds true. Once a single appropiate threshold q_C is set, tap birth can be decided trivially when $|u_p^{(l)}| > q_C \cdot \sigma_{v^{(l)}}$ has been detected. When birth is detected, the KF is restarted and the first estimate is the LS estimate.

The system resulting from the three detection heuristics shown above and a connected KF block is called [Méndez-Romero and Fernández-Getino García, 2018] "Simplified Maximum a Posteriori" (SMAP) estimator.

5.4.4 Practical optimization issues

SMAP parameters $\{q_i\}$ need to be optimized for each specific SNR level (or each specific set of SNR levels) to make full use of the SNR estimate provided by the QISD. Several optimization strategies were attempted: our experience suggests that the search space is dense with local minima (in the sense of local environments making gradient-descent algorithms stop). Thus, algorithms similar to gradient descent are unsuitable. Instead, we recommend the following optimization Search Steps (SS) for each SNR level:

- SS1: q_i values are randomly initialized following a normal distribution over the space where SMAP assumptions are mostly reasonable, then CTMSE for the SMAP+KF scheme is computed and the N_{SS} best values are chosen. N_{SS} is a parameter to be chosen taking the scope of the simulation (i.e. number of random initializations, parameter diversity) into account, e. g. $N_{SS} = 4$ for 50,000 random initializations.

- SS2:q_i values are randomly initialized following a normal distribution centered on the N_{SS} best SS1 values, but with a fraction (e.g. a quarter) of the variance.

- SS3: values are randomly initialized following a normal distribution centered on the N_{SS} best SS2 values, but variance will be a fraction of each central parameter, e.g. $\mathcal{N}([q_A\ q_B\ q_C];\ [0.1q_A\ 0.1q_B\ 0.1q_C])$.

The aforementioned illustrative values for N_{SS} and variances are heuristics backed by extensive simulations. Due to the random nature of the search, it may happen that the best parameters for an SNR level work better at a different SNR level than its own optimized parameter set does. To avoid this "better-finder" effect, a final step (SS4) can be added whereby each SNR level is tested with all optimal parameter sets, including those for other SNR levels.

5.5 Simulation Results

A system with $N_p = 3$ pilot subcarriers is considered, with $K = 8$ pilot symbols per subcarrier before data transmission. The average energy per pilot symbol, σ_s^2, is uniform, and a BPSK modulation scheme is assumed. Channel assumptions include a uniform multipath delay profile, multipath spread smaller than the guard time, and uncorrelated path gains. The overall channel energy is normalized to one. OFDM symbols were transmitted through a channel with $L_{max} = 3$, $P_{birth} = 0.05$ and $P_{death} = 0.05$. Threshold parameter values $q_A = 0.23$, $q_B = 0.30$ and $q_C = 0.62$ were obtained through simulation-based MSE optimization for $s = 1$ and form the basis for the simulation results shown here. Note that this prior optimization is always performed offline and, thus, it does not increase online computational complexity. This channel was then tracked for 200,000 estimations periods ($2 \cdot 10^5 \cdot T_{est}$). Individual paths were assumed to have the same average energy σ_h^2 over long periods and $\lambda = 0.999$. This choice of parameters makes it possible to compare the performance of the SMAP vs. the computationally heavier methods outlined in [Angelosante et al., 2007] and the KF system advocated in [Yuanjin, 2003].

Fig. 5.5 shows the performance (in terms of CTMSE, as per (5.5)) of our proposed SMAP estimation vs. a scenario with perfect information (ISS), almost perfect information (ISS-99%) and a KF system. It can be seen that the SMAP estimator is very similar in performance to ISS-99%. This means the SMAP estimator gets most of the reduction attainable by an ISS, but with a trivially low computational cost, for low-to-medium SNR.

A measurement of SMAP's robustness in the face of model uncertainty can be provided by setting random thresholds q_A, q_B and q_C extracted from Gaussians centered on the previously determined optimal values. Fig. 5.6 shows simulation results for several standard deviation σ values, e.g. $\sigma_{\%} = 40\%$ means SMAP is not used now with optimal thresholds, but rather with several different random threshold triplets taken from Gaussians: $q_C \sim \mathcal{N}(0.62, \sigma = 0.62 \cdot 0.40 = 0.248)$, etc. In this particular simulation, 10 such random triplets were extracted for each σ and SNR level. It is apparent from Fig. 5.6 that large deviations result in smaller changes in CTMSE, especially when compared to bare Kalman. Thus, the use of a SMAP structure is shown to be a larger contributor to error reduction than the use of perfect thresholds in that SMAP structure. This low sensitivity to threshold deviations means SMAP is robust to some extent when facing some uncertainty in model parameters.

Fig. 5.7 shows the bit-error-rate (BER) performance of SMAP vs. the ideal system (ISS) and a KF system. This considers all samples when at least one path is active (channel energy

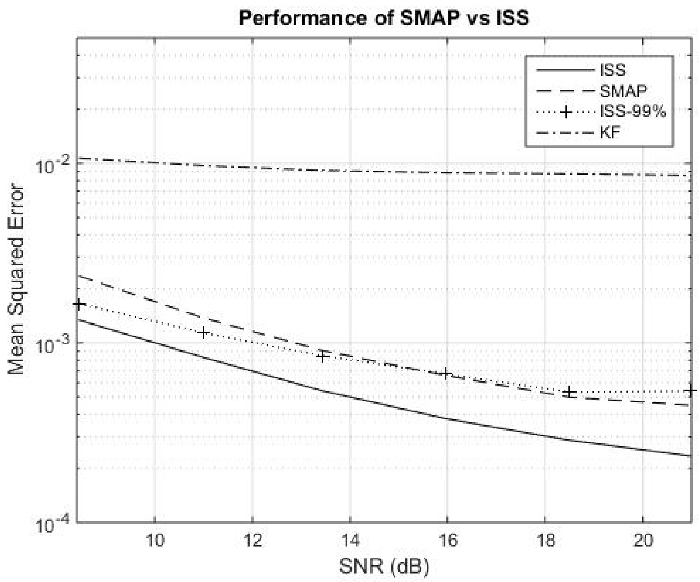

Figure 5.5: Average CTMSE performance of proposed SMAP vs. ideal systems and KF

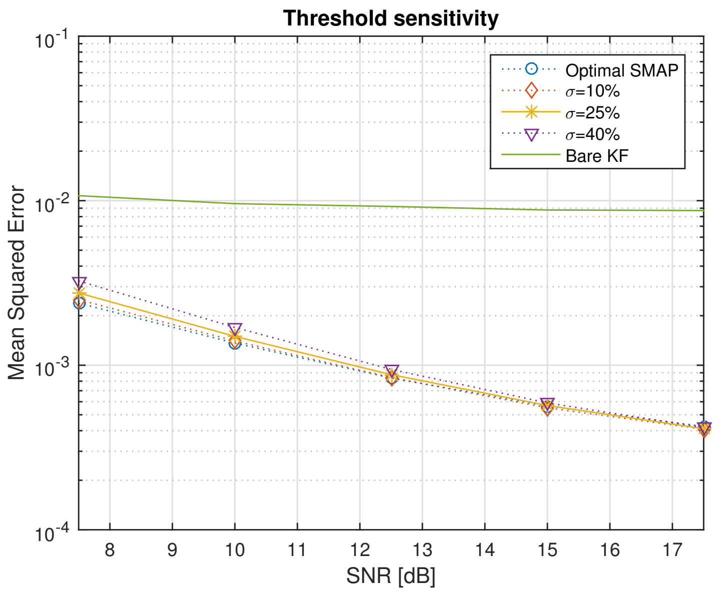

Figure 5.6: Average CTMSE performance of SMAP with differently-deviated random thresholds vs. KF and optimal-threshold SMAP

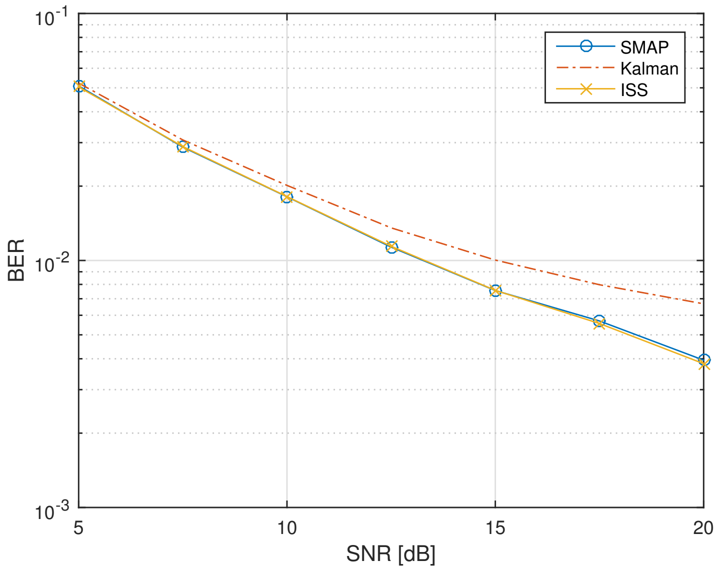

Figure 5.7: BER performance of proposed SMAP vs. ideal system (ISS) and KF

at least 5% of average channel energy), thus ignoring samples when all channel taps are zero and communication must be unfeasible. It must be noted that BER values were obtained without considering any channel coding scheme which would improve system performance. Results are unambiguous: the BER curve for SMAP matches the one for the ideal system (ISS) and both of them have a significant positive gap vs. bare KF.

Now let us take a closer look at Figs. 5.8-5.9, which show the estimated path gain in SMAP and KF systems, respectively, for a signal-to-noise ratio SNR $\triangleq \sigma_s^2/\sigma_z^2 = 15$ dB. These figures support the thesis, already advanced in [Angelosante et al., 2007], that a KF operating on L_{max} paths suffers from a transient effect from tap death/birth. The proposed SMAP estimation adapts more quickly to those non-linearities. This superior performance when a tap disappears or a new tap is born can also be seen for SNR = 9 dB in Figs. 5.10-5.11 and for SNR = 21 dB in Figs. 5.12-5.13.

Moreover, this superior performance does not come at the expense of a prohibitively high computational cost. On the contrary, the status detector for each tap uses only the three computationally inexpensive threshold comparisons outlined in Section 5.4 to achieve an error reduction matching that of a close-to-perfect path death/birth detection (ISS-99%).

5.6 Conclusions

A simplified framework for the birth-death nonlinearity problem has been introduced. A computationally inexpensive, threshold-based estimator was derived. Simulations have shown this estimator greatly reduces channel tracking error in the target SNR range at a very small computational cost, thus outperforming previously known systems.

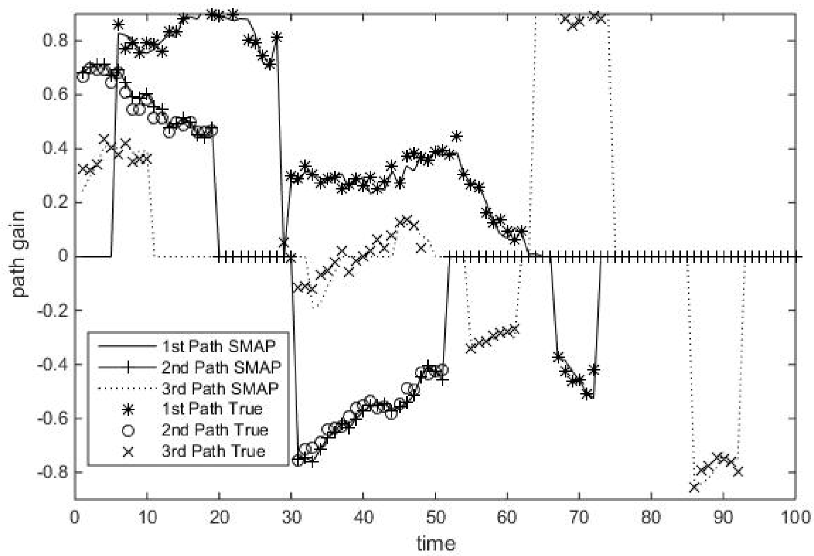

Figure 5.8: SMAP estimation for L=3, SNR=15 dB.

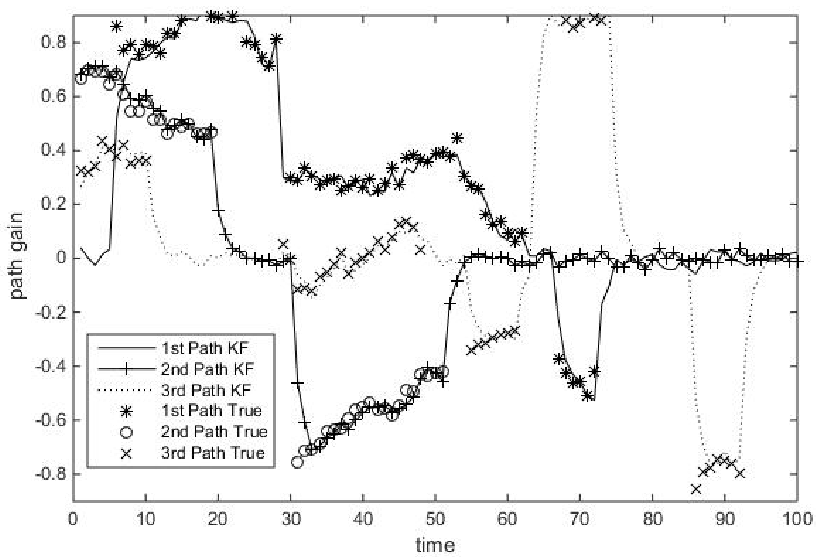

Figure 5.9: KF estimation for L=3, SNR=15 dB.

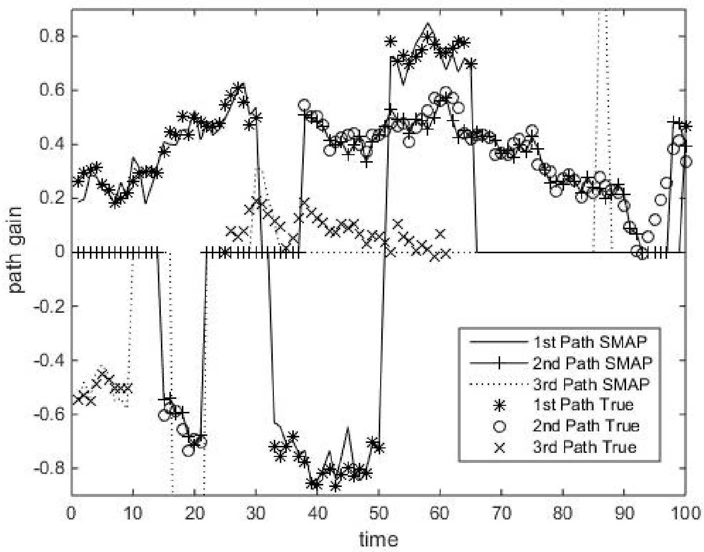

Figure 5.10: SMAP estimation for L=3, SNR=9 dB.

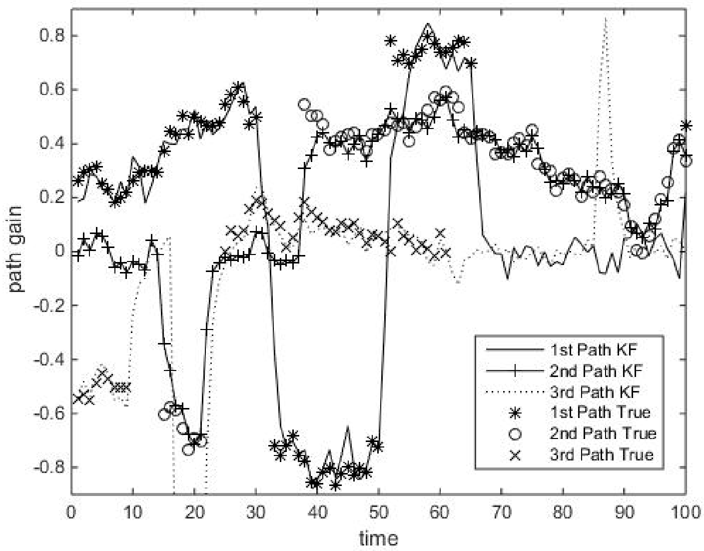

Figure 5.11: KF estimation for L=3, SNR=9 dB.

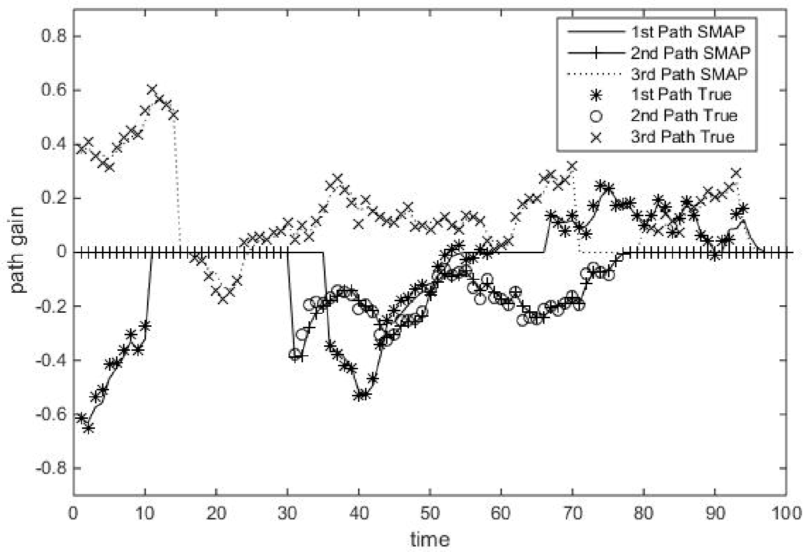

Figure 5.12: SMAP estimation for L=3, SNR=21 dB.

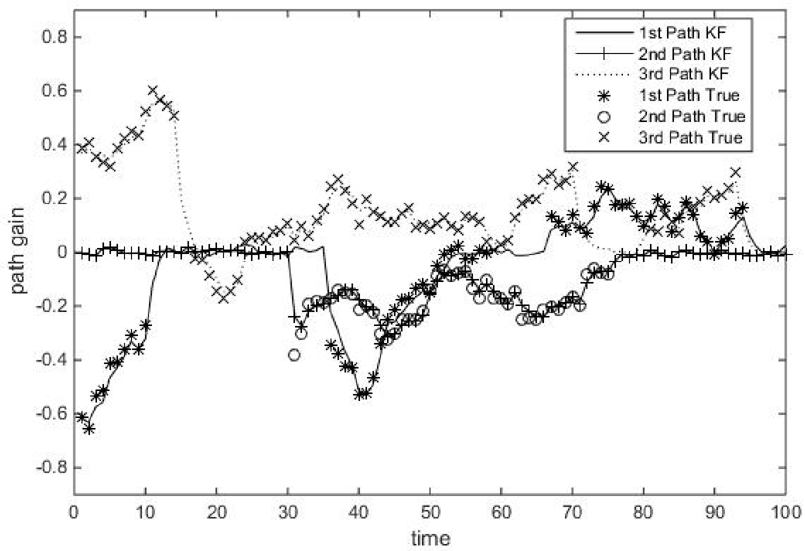

Figure 5.13: KF estimation for L=3, SNR=21 dB.

Chapter 6

Death/Birth & SNR Detection for Kalman Trackers

In previous chapters, we have analized some techniques for detection of death/ birth change-points in non-stationary channels, such as those emerging in fast-moving vehicular environments. The framework of Random Set Theory (RST) made it possible to use powerful RST-based Particle Filtering (PF) estimators (see Section 4.2 and Appendix A), such as those in [Angelosante et al., 2007, Angelosante et al., 2009]. However, they required an impracticable computational cost for any practical applications.

A simpler algorithm was derived in Chapter 5, the so-called Simplified Maximum a Posteriori (SMAP) death/birth detector [Méndez-Romero and Fernández-Getino García, 2018], based on three heuristics with dynamic threshold rules. However, both the SMAP and RST-based PF proposals failed to consider a scenario where SNR levels drifted abruptly.

This chapter extends the RST framework to combined death/birth and SNR detection when SNR is dynamical and may drift. Additionally, it analyzes how different quasi-ideal SNR detectors [Méndez-Romero and Fernández-Getino García, 2020] affect the SMAP-enhanced Kalman tracker's performance.

6.1 An RST Model for Abrupt SNR Changes

Consider a Tapped-Delay-Line (TDL) channel model with K_{max} possible, randomly alternating SNR levels and up to L_{max} different equispaced[1] path delays. Let us thus assume that SNR may change during transmission between a finite number (K_{max}) of known levels. Then, the multipath channel can be denoted by the $\mathcal{H}_p^{(k,l)}$, the following singleton-or-empty random set:

$$\mathcal{H}_p^{(k,l)} = \begin{cases} \{\emptyset\} & if\, path\,(k,l)\, absent \\ \{\mathbf{h}_p^{(k,l)}\} = \{[SNR_p^{(k)}, l, a_p^{(l)}]^T\} & if\, path\,(k,l)\, present \end{cases} \tag{6.1}$$

where path (k,l) means the lth path, under the condition that all active paths are subject to $SNR^{(k)}$ at time p; and $a_p^{(l)}$ is the lth multipath gain at time p. The multipath channel state at time p is completely described by

[1] If path delays are not equispaced, the analysis is somewhat more complex, but equally feasible.

$$\mathcal{H}_p = \bigcup_{k=1}^{K_{max}} \bigcup_{l=1}^{L_{max}} \mathcal{H}_p^{(k,l)} \tag{6.2}$$

which is a random set on the hybrid space $\{1, ..., K_{max}\} \times \{1, ..., L_{max}\} \times \mathbb{C}$. Please notice that we can define the projections of $\mathcal{H}_p$ onto $\{1, ..., K_{max}\}$, onto $\{1, ..., L_{max}\}$ and onto $\mathbb{C}$, respectively:

$$\pi(\mathcal{H}_p) = \bigcup_{(k,l):\mathcal{H}_p^{(k,l)} \neq \emptyset} \{SNR_p^{(k)}\} \tag{6.3}$$

$$\pi'(\mathcal{H}_p) = \bigcup_{k \in \pi(\mathcal{H}_p), l:\mathcal{H}_p^{(k,l)} \neq \emptyset} \{l\} \tag{6.4}$$

$$\pi''(\mathcal{H}_p) = \bigcup_{k \in \pi(\mathcal{H}_p), l:\pi'(\mathcal{H}_p)} \{a_p^{(l)}\} \tag{6.5}$$

If $\mathcal{U}_p$ denotes the set of paths whose parameter SNR is unmodified from time $p-1$ to time p, i.e. $\pi(\mathcal{U}_p) \subseteq \pi(\mathcal{H}_{p-1})$ and $\mathcal{M}_p$ is the set of newly modified paths (i.e. $\pi(\mathcal{H}_{p-1}) \cap \pi(\mathcal{M}_p) = \emptyset$), we thus have

$$\mathcal{H}_p = \mathcal{U}_p \cup \mathcal{M}_p \tag{6.6}$$

Similarly, we can define a set of surviving paths, $\mathcal{S}_p$, such that $\pi'(\mathcal{S}_p) \subseteq \pi'(\mathcal{H}_{p-1})$, and a set of newly born paths $\mathcal{B}_p$ such that $\pi'(\mathcal{H}_{p-1}) \cap \pi'(\mathcal{B}_p) = \emptyset$. Thus,

$$\begin{aligned} \mathcal{H}_p = (\mathcal{U}_p \cup \mathcal{M}_p) \cap (\mathcal{S}_p \cup \mathcal{B}_p) = \\ = (\mathcal{U}_p \cap \mathcal{S}_p) \cup (\mathcal{U}_p \cap \mathcal{B}_p) \cup (\mathcal{M}_p \cap \mathcal{S}_p) \cup (\mathcal{M}_p \cap \mathcal{B}_p) \end{aligned} \tag{6.7}$$

On the basis of this newly defined framework, a problem can be characterized by defining the probability of transition between the four joining subsets in (7).

Please notice that if the probability of transition from $\mathcal{U}_p$ to $\mathcal{M}_p$ is 0, the problem reduces to one where SNR and the probability of tap birth or death, $P_{birth/death}$, are constant, as in [Angelosante et al., 2009, Méndez-Romero and Fernández-Getino García, 2018], solvable with a Rao-Blackwellized Particle Filter (RBPF) (see Appendix Section A.3). In the style of GMAP-III [Angelosante et al., 2007], we hereby propose the following three-step estimator for the general problem:

$$\begin{cases} \widehat{\pi(\mathcal{H}_p)} = arg\,max_{\mathcal{H}_p} f(\pi(\mathcal{H}_p)|\mathcal{Y}_{1:p}), \\ \widehat{\pi'(\mathcal{H}_p)} = arg\,max_{\mathcal{H}_p} f(\pi'(\mathcal{H}_p)|\mathcal{Y}_{1:p}, \widehat{\pi(\mathcal{H}_p)}), \\ \tilde{\mathbf{h}_p} = \int_{\mathbb{R}^{2|\widehat{\pi'(\mathcal{H}_p)}}} \mathbf{h}_p f(\mathbf{h}_p|\mathcal{Y}_{1:p}, \widehat{\pi(\mathcal{H}_p)}) d\mathbf{h}_p \end{cases} \tag{6.8}$$

where

$$f(\pi(\mathcal{H}_p)|\mathcal{Y}_{1:p}) = \int_{\pi'(\mathcal{H}_p) \times \pi''(\mathcal{H}_p)} f(\mathcal{H}_p|\mathcal{Y}_{1:p}) \delta\mathcal{H}_p \tag{6.9}$$

$$f(\pi'(\mathcal{H}_p)|\mathcal{Y}_{1:p}, \pi(\mathcal{H}_p)) = \int_{\pi''(\mathcal{H}_p)} f(\mathcal{H}_p|\mathcal{Y}_{1:p}, \widehat{\pi(\mathcal{H}_p)}) \delta\mathcal{H}_p \tag{6.10}$$

The operation performed in Eqs. (6.9) and (6.10) using the differential $\delta\mathcal{H}_p$ are set integrations

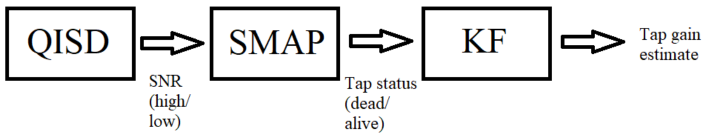

Figure 6.1: Block diagram for the proposed channel estimation scheme.

in the sense specified in [Goodman et al., 2013].

This estimator amounts to first estimating the SNR level at time p, then estimating the identities of the active paths at time p, and then estimating the gains for those active paths deemed active, while inactive paths will be estimated to have zero gain.

The practical implementation of this structure is dependent on the specific problem. A generalized practical implementation is out of the scope of this paper, though it may be suggested that advanced estimation techniques, possibly involving particle filters and/or neural networks, could be used. In the balance of this paper, we will assume that the three steps are somewhat separable problems in nature: the first step will be considered to be solved with quasi-ideal SNR estimators/detectors (QISDs), eventually based on additional information (like track location and speed in HSR); the second step will be performed with the Simplified MAP (SMAP) algorithm proposed in [Méndez-Romero and Fernández-Getino García, 2018] (see Chapter 5); and the third step will be performed in a Kalman filter (KF) bank. A simplified block diagram for this scheme is shown in Fig. 6.1.

6.1.1 Quasi-Ideal SNR Estimators

In Subsection 5.3.4, the Ideal Switching System (ISS) was defined as a system where "death/birth detectors detect 100% of births and 100% of deaths, with no false birth/death detection". It was argued that an x%-degraded Ideal Switching System (IISx%) could be easily defined as "a switching system detecting x% of births and x% of deaths, with no false detection". Such a degraded ideal system could also be called "quasi-ideal".

This concept can be extended to SNR detection. Thus, considering two SNR levels (low, around s dB; and high, around $s + \Delta s$ dB) we can define an Ideal SNR Detector (ISD) where SNR detectors detect 100% of SNR updrifts (i.e. SNR drifting from low to high level) and 100% of SNR downdrifts (i.e. SNR drifting from high to low level), with no false SNR drift detection. (When the probabilities of SNR updrift and SNR downdrift are symmetric, those probabilities will be denoted by P_{switch}.) It can similarly be argued that an x% Quasi-Ideal SNR Detector (QISD-x%) could be easily defined as a system detecting x% of updrifts and x% of deaths, with no false birth/death detection.

Please note this concept can be trivially extended to $n > 2$ SNR levels. If the SNR division is granular enough, they should properly be called "Quasi-Ideal SNR Estimators" (QISE).

Morever, to study the QISE/QISD+SMAP+KF scheme introduced here, it is realistic to consider QISEs/QISDs whose detection is not instantaneous, but has a short delay (5-20 time instants).

6.1.2 Simplified MAP thresholds

As shown in Fig. 6.1, the QISD will feed the SNR level estimate into a Simplified Maximum a Posteriori (SMAP) death/birth detector. Such detector makes use of 3 detection heuristics [Méndez-Romero and Fernández-Getino García, 2018]: a) memoryless detection of large leaps into a narrow, zero-centered range; b) memory detection of a sequence at a close range of zero; and c) birth detection. These heuristics are based on dynamical thresholds that can be fine-tuned through the help of a set of parameters: q_A, q_B, q_C. For example, the SMAP heuristic for large leaps into death (zero) leads to detecting "death" whenever $P_{death} \cdot f(u_p^{(l)} | l\, is\, dead) > f(u_p^{(l)} | l\, is\, alive) \cdot (1 - P_{death})$ where $u_p^{(l)}$ is the Least-Squares-estimated tap gain and the events "l is dead/alive" mean that the lth path is dead/alive, respectively. Since this detector is comparing two scaled Gaussian pdfs, an equivalent expression to Eq. (5.15) would be:

$$P_{death} \cdot \frac{1}{\sqrt{2\pi\sigma_{v^{(l)}}^2}} \cdot e^{-\frac{(u_p^{(l)})^2}{2\sigma_{v^{(l)}}^2}} > \frac{1}{\sqrt{2\pi\sigma_A^2}} \cdot e^{-\frac{(u_p^{(l)} - \hat{u}_p^{(l)}(-))^2}{2\sigma_A^2}} \cdot (1 - P_{death}) \tag{6.11}$$

where $\sigma_A = q_A \cdot \sqrt{\sigma_{v^{(l)}}^2 + (1 - \lambda^2)\sigma_h^2}$. For more details, refer to Section 5.4.

6.2 Higher-Dimensional SMAP

The SMAP can be applied to higher-dimensional tracking problems, such as tracking problems in $\mathbb{R}^n$, $n \geqslant 2$; or $\mathbb{C}^n$, $n \in \mathbb{N}$. Such tracking problems may occur not just in communications, but also in other engineering[2] and non-engineering[3] applications. The simplest way to apply the SMAP method to higher-dimensional problems is by searching for the optimal SMAP thresholds q_A, q_B, q_C over the whole higher-dimensional space.

Of course, depending on the specific problem, it might also make sense to find explicit ad-hoc equations for the memoryless detection of large leaps into a narrow, zero-centered range; the memory detection of a sequence at a close range of zero; and birth detection, respectively. Such derivations could be inspired by the heuristics in Section 5.4 or closely follow the SMAP derivations in the current PhD Thesis, and then the optimal SMAP-like thresholds would be searched for over the whole higher-dimensional space.

6.3 Simulation Results

The OFDM system under consideration has $N_p = 5$ pilot subcarriers and $K = 8$ pilot symbols per subcarrier. The average energy per pilot symbol, σ_s^2, is assumed to be uniform and the overall channel energy is normalized to one. A BPSK modulation is used. The channel is assumed to have a uniform multipath delay profile, multipath spread smaller than the guard time, and uncorrelated path gains. 10^6 OFDM symbols were transmitted through a channel with $L_{max} = 5$ and $P_{birth/death} = 0.05$. Individual paths were assumed to have the same average energy σ_h^2 over long periods and $\lambda = 0.999$. Channel SNR was assumed to experience a sudden switch (from a high SNR level into a low SNR level, or viceversa) with probability $P_{switch} = 0.01$. Between high and low SNR, a difference $\Delta SNR = 6$ dB was considered.

[2] See Subsection 4.1 for a historical overview of abrupt change problems in engineering.
[3] Such as neurofinance. Refer to Chapter 9 for neurofinancial consequences of unstereotyped abrupt dynamics.

Parameter	SNR = 9 dB	SNR = 15 dB
q_A	1.3543	0.8065
q_B	1.6103	1.3427
q_C	2.8823	3.1613
σ_A	0.1448	0.0533

Table 6.1: Optimized SMAP parameter values for SNR={9,15} dB

Each tap is tracked with a combined scheme consisting of a SNR detector, a birth/death SMAP detector, and a KF. We consider the following QISD accuracies for the SNR detection step: 90, 95, 99 and 100%, with a detection delay of 5, 8, 15 and 20 time instants, respectively. For comparison purposes, we also consider the ideal case of perfect, instantaneous SNR detection (ISD), as a lower bound.

The tap birth/death detection is performed with the SMAP parameters optimized according to steps SS1-3, with 50,000 random initializations, $N_{SS} = 4$ and 6,500 subsequent randomizations for each set. SS4 was intentionally omitted to show the "better-finder" effect (see Subsection 5.4.4) in the results. The optimized parameter values for selected SNRs are shown in Table 6.1.

Fig. 6.2 shows the performance of such combined QISD+SMAP+KF schemes for several SNR values. Please notice that each scheme was simulated in a bilevel SNR environment. All SNR values < 15 were considered to be "low" and to switch to a 6dB-higher SNR with probability P_{switch}.

Several conclusions are evident: KF with no birth/death detection performs catastrophically worse than any SMAP scheme. Even when no SNR detection is available ("No QISD"), the SMAP+KF performance is surprisingly robust to SNR drift; this is especially evident at SNR=13 dB, where the "better-finder" effect (see 5.4.4) takes place. While the effect of an accurate SNR detection is muted in the easier scenario (high SNR), channel tracking does benefits from it at low SNR (see "No QISD" has higher MSE at SNR={9,11} dB in Fig. 6.2). Consequently, after considering both the lapses of high and low SNR, overall CTMSE is reduced (up to 10%, according to our results) when any of the QISD schemes are used.

Figs. 6.3 and 6.4 explain why: differences in assumed noise (KF's prediction error variance) make birth/death detection fail when the "baseline" SNR estimate (provided by QISD) happens to be false (Fig. 6.4), generating a slow KF climb or descent (higher MSE). This is the same behaviour underlying KF's catastrophic performance in the face of abrupt tap gain changes (Fig. 6.3) when no SMAP is used.

It is important to remark that all QISD schemes, no matter whether slow and highly accurate or fast and not so accurate, show a similar performance to each other and to the ideal, instantaneous case (ISD). This suggests there is a pronounced trade-off between SNR accuracy and detection delay that can be exploited during SNR detection design.

6.4 Conclusions

A new RST framework for combined death/birth and SNR detection has been introduced; this framework is fully compatible with the SMAP+KF solution introduced in Chapter 5. Some SMAP optimization issues were tackled. Simulation results compared different QISD+SMAP+KF schemes and suggest that SMAP+KF, while being robust to SNR drift, benefits from an accurate SNR detection. Fig. 6.5 provides a visual summary of SNR/SMAP detectors for Kalman trackers.

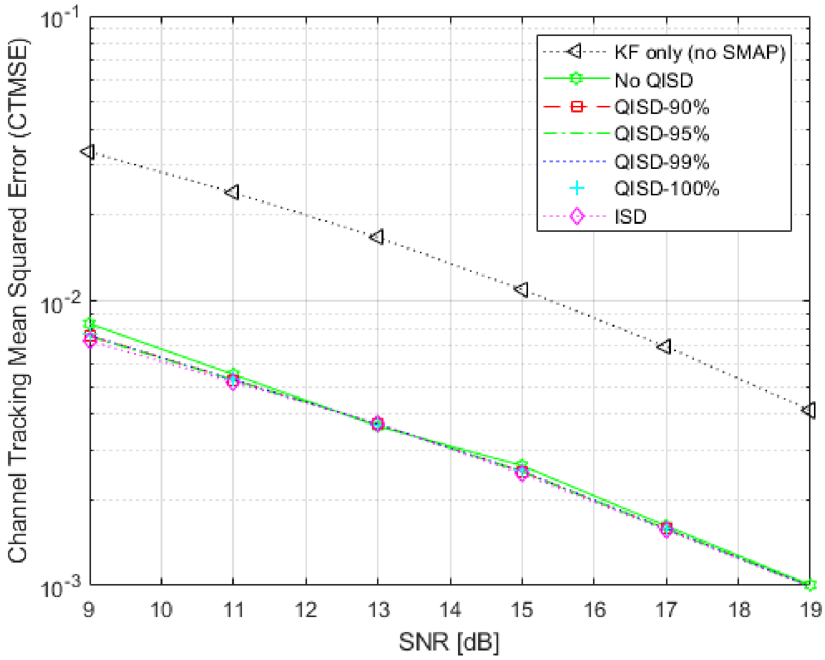

Figure 6.2: Channel Tracking Mean Squared Error vs. Signal-To-Noise Ratio (SNR) for different SNR and birth/death detection schemes.

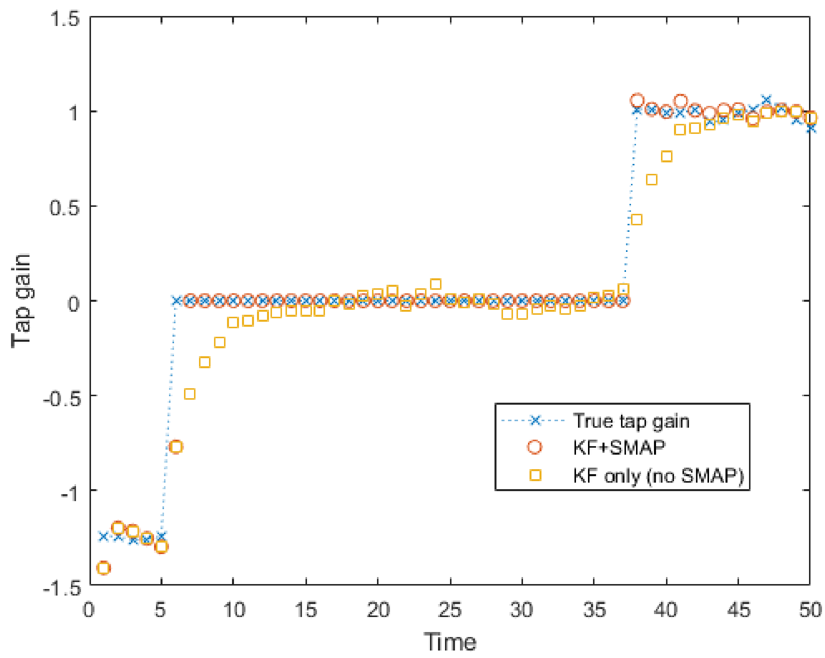

Figure 6.3: Kalman tracker with and without birth/death detector for SNR = 15 dB.

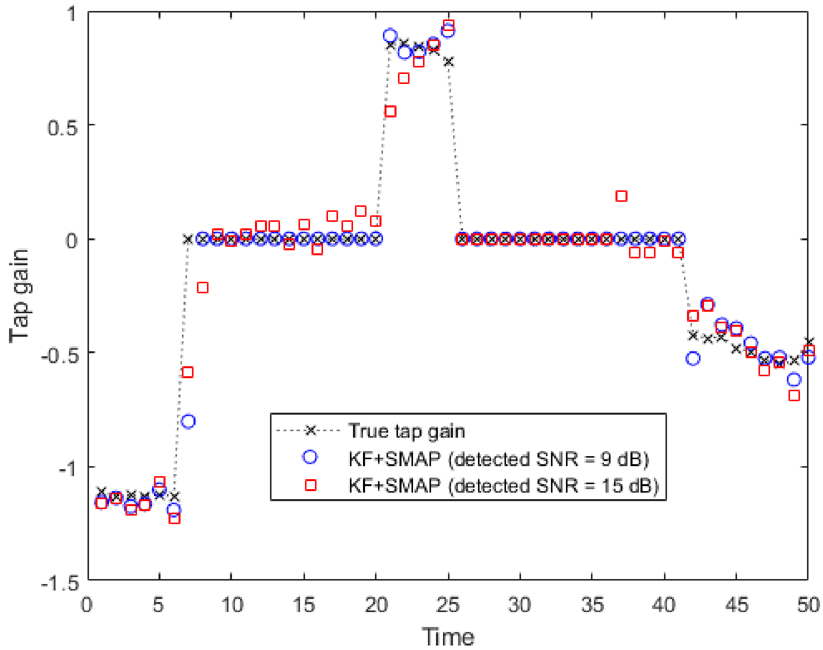

Figure 6.4: Kalman tracker with birth/death detector when SNR detection is correct (9 dB) and incorrect (15 dB).

The information contained in this Chapter 6 may be useful when designing SNR trackers with death/birth detectors for practical smart mobility applications.

Figure 6.5: SNR/SMAP for Kalman trackers

Chapter 7

Neural-Network-Switched Kalman Filters

This Chapter introduces Neural-Network-Switched Kalman Filters (NNKFs) as novel trackers for multipath channels. The core idea is similar to the SMAP switch (Chapter 5), but now the tap birth/death detection will be performed via an artificial neuronal network (NNs).

A discussion of the required Neural-Kalman architecture (combining trainable, low-complexity neural networks with Kalman filters) is presented, as well as simulations showing an excellent performance.

The significance of this proposal goes beyond tap birth/death detection. The NNKF architecture is a neural abrupt change detector that could be expanded to any trainable non-stationarity, i.e. not just tap birth/death, but also more complex non-stationary models that might be developed in the future. An instance of more complex abrupt tap dynamics will be proposed and analyzed in Chapter 8.

7.1 Neural Networks

A neural network (or, more properly, an *artificial* neural network) is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates, i.e. it is inspired by the structure of *biological* neural networks in the brain [Shalev-Shwartz and Ben-David, 2014, Russell et al., 2020]. In this sense, neural networks refer to systems of neurons, either organic (see Fig. 7.1) or artificial in nature.

An artificial neural network is based on a collection of connected units or nodes called artificial neurons, whose connections, like the synapses in a biological brain, can transmit a signal to other neurons. Thus, it can be described as a directed graph (see Fig. 7.2) whose nodes correspond to neurons and edges correspond to links between them [Shalev-Shwartz and Ben-David, 2014]. That is, each artificial neuron/node receives signals and then processes them and can send output signals to neurons/nodes connected to it. Signals (real numbers) travel from the leftmost layer (the input layer, see), to the rightmost layer (the output layer). The middle layer(s) are called hidden layer(s), since the neurons in these layers are neither inputs nor outputs.

The design of the input and output layers is sometimes very intuitive, e.g. if the neuron has to decide whether an object in an image is a "dog", and the image is a 64 by 64 greyscale image, then

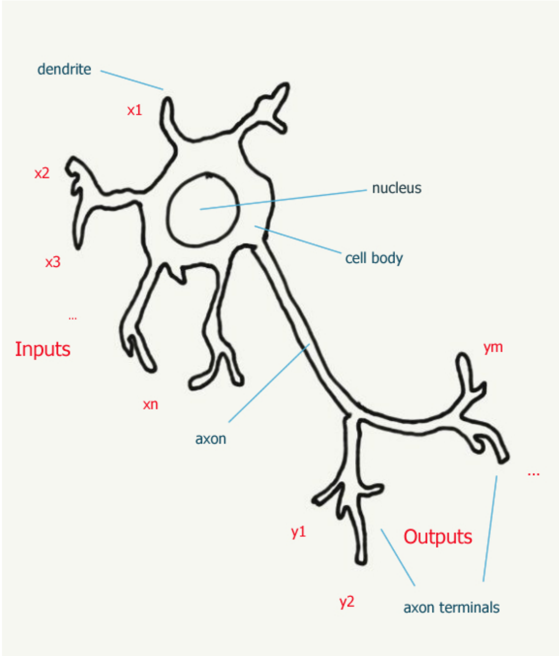

Figure 7.1: A biological neuron as a signal processing system.

we'd have 4,096=64×64 input neurons, with the intensities scaled appropriately between 0 and 1. The output layer will contain just a single neuron, with output values of less than 0.5 indicating "input image is not a dog", and values greater than 0.5 indicating "input image is a dog".

In this Thesis, neural networks will be used for the purpose of identifying abrupt changes, e.g. when a new tap is born (goes from the "inactive" state to the "active" state) or dies.

Learning with neural networks was proposed in the mid-20th century. Learning with NNs provides an effective learning paradigm that, in the recent decades, has been shown to achieve cutting-edge performance on several learning tasks [Russell et al., 2020, Shalev-Shwartz and Ben-David, 2014].

7.1.1 Feedforward vs. recurrent neural networks

This Thesis will consider neural networks where the output from one layer is used as input to the next layer, the so-called feedforward neural networks. This means the signal is always fed forward, never fed back.

Other models of artificial neural networks exist in which feedback loops are possible, the so-called Recurrent Neural Networks (RNNs). Some of our referenced papers use this type of architecture, where neurons fire for some limited duration of time, before becoming quiescent, and that firing can stimulate other neurons. Thus, these other neurons may fire a little while later, also for a limited duration and that causes still more neurons to fire, creating a cascade of neurons firing. Loops don't cause problems in such a model, since a neuron's output only affects its input at some later time, not instantaneously. This kind of NNs do not suit our purposes, so this Thesis will focus on feedforward NNs.

Formally, a feedforward NN [Shalev-Shwartz and Ben-David, 2014] is described by a directed acyclic graph, $G = (V, E)$, and a weight function over the edges, $w : E \to \mathbb{R}$. Nodes of the graph correspond to neurons (see Fig. 7.2). Each single neuron is modeled as a simple scalar function, $\sigma : \mathbb{R} \to \mathbb{R}$, called the *activation* function of the neuron. An intuitive explanation of this function,

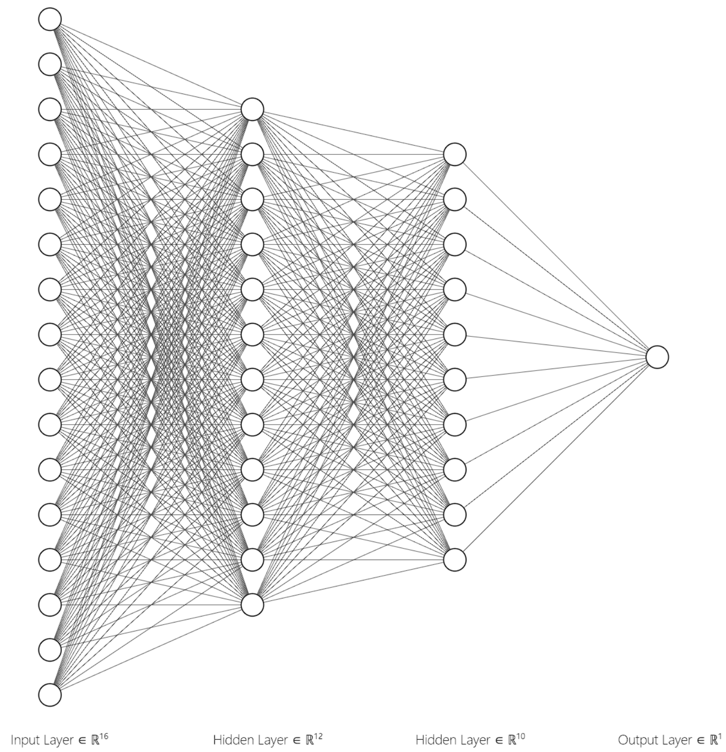

Figure 7.2: An artificial neural network

as well as a list of potential functions that can be used as activation functions in NN-aided learning tasks, is provided in Section 7.1.1.2.

As mentioned above, neural networks are typically assumed to be organized in layers. This simplifies the description [Shalev-Shwartz and Ben-David, 2014] of the calculation performed by the network. Layered organization implies that the set of nodes can be decomposed into a union of (nonempty) disjoint subsets:

$$V = \uplus_{t=0}^{T} V_t \tag{7.1}$$

such that every edge in E connects some node in V_{t-1} to some node in V_t, for some $t \in \{0, ..., T\}$, where T is the number of layers in the network (excluding the input layer, V_0), also known as *depth* of the network.

7.1.1.1 Layers

There are three types of layers: the input layer V_0, the hidden layers $V_1, ..., V_{T-1}$, and the top layer V_T.

The input layer contains $n+1$ neurons, where n is the number of inputs; the first n neurons simply output their input x_i, $i \in \{0, ..., n\}$, and the last neuron in V_0 is a constant neuron which always outputs 1.

In simple estimation problems such as deciding whether a tap is active or inactive, the output layer would contain a single neuron whose output is the output of the network. However, some channel tracking problems might require more complex outputs.

Let us denote by $v_{t,i}$ the ith neuron of the tth layer. When the network is fed with the input vector $\mathbf{x}$, the output of $v_{t,i}$ will be denoted by $o_{t,i}(\mathbf{x})$.

In order to understand how the network's output can be computed in a layer-by-layer manner [Shalev-Shwartz and Ben-David, 2014], let us suppose we already know the outputs of the neurons at layer t, i.e. we have already computed them. Then, we can compute the outputs at layer $t+1$ with the following method. First, we fix some $v_{t+1} \in V_{t+1}$. When the network is fed with the input vector $\mathbf{x}$, the input to $v_{t+1,j}$ will be given by:

$$a_{t+1,j}(\mathbf{x}) = \sum_{r:(v_{t,r},v_{t+1,j})\in E} w((v_{t,r}, v_{t+1,j}))o_{t,r}(\mathbf{x}) \tag{7.2}$$

and the output will be

$$o_{t+1,j}(\mathbf{x}) = \sigma(a_{t+1,j}(\mathbf{x})) \tag{7.3}$$

In other words, the input to $v_{t+1,j}$ is a weighted sum of the outputs of the neurons in V_t that are connected to $v_{t+1,j}$. Such weighted sum is determined by the weight function w. Finally, the activation function σ is applied on the neuron's input to determine its output.

7.1.1.2 Activation function

The role of the Activation Function is to derive output from a set of input values fed to a node (or a layer). The primary role of the Activation Function is to transform the summed weighted input from the node into an output value to be fed to the next hidden layer or as output.

An Activation Function decides whether a neuron should be activated or not, i.e. it will decide whether the neuron's input to the network is important or not in the process of prediction using simpler mathematical operations. Thus, the purpose of an activation function is to add non-linearity to the neural network.

Let's consider a neural network working without the activation functions. In that case, every neuron would only be performing a linear transformation on the inputs using the weights and biases. Such neural network would be simpler, but learning any complex task would be impossible, since that model would be just a linear regression model: all layers (no matter how many hidden layers we add) would behave in the same way because the composition of two linear functions is a linear function itself.

There are three types of activation functions: binary step functions, linear activation functions and non-linear activation functions.

Binary Step Function

$$f(x) = \begin{cases} 0 & for\, x < 0 \\ 1 & for\, x \geq 0 \end{cases} \tag{7.4}$$

Disadvantage: The gradient of the step function is zero. This causes a hindrance in the back-propagation process. Since neural networks require backpropagation (gradient-descent) for optimization, this is an issue.

Linear Activation Function

$$f(x) = x \tag{7.5}$$

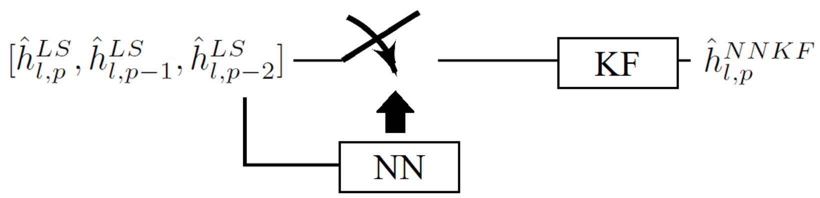

Figure 7.3: Neural-Network-switched Kalman Filter (NNKF).

Non-Linear Activation Functions

- Sigmoid/Logistic

$$f(x) = \frac{1}{1+e^{-x}} \tag{7.6}$$

- Hyperbolic Tangent

$$f(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})} \tag{7.7}$$

- Rectified Linear Unit (ReLU)

$$f(x) = max(0, x) \tag{7.8}$$

- Leaky ReLU Function

$$f(x) = max(0.1x, x) \tag{7.9}$$

- Parametric ReLU Function

$$f(x) = max(ax, x) \tag{7.10}$$

- Exponential Linear Units (ELU) Function

$$f(x) = \begin{cases} x & for\, x \geq 0 \\ \alpha(e^x - 1) & for\, x < 0 \end{cases} \tag{7.11}$$

- SoftMax, Swish, Gaussian Error Linear Unit (GELU), Scaled Exponential Linear Unit (SELU), etc.

7.2 NNKFs as Novel Trackers for Multipath Channels

Proposals like SMAP [Méndez-Romero and Fernández-Getino García, 2018] and PF-solvable estimators such as GMAP-III [Angelosante et al., 2007] have previously suggested that the tracking problem is separable in nature: you can estimate which taps are active at a given time and then estimate the value for those taps deemed active in two separate steps. Following the same philosophy, it is hereby proposed to use a neural network (NN) as birth/death detector for each tap; this This NN would switch on the Kalman filter when the tap is active and it would switch it off in case of death (Fig. 7.3).

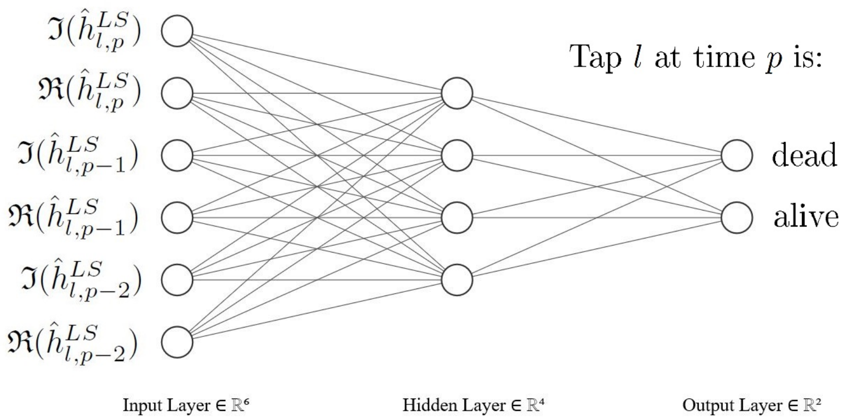

Figure 7.4: Proposed NN tap birth/death detector.

7.2.1 Neural Architecture for NNKF Tracking

The whole tracking scheme is shown in Fig. 7.3, while the detailed 6 x 4 x 2 neural network is shown in Fig. 7.4, where $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary parts of the complex channel tap gains, respectively. The specific dimensions of the NN are not fixed; they must be decided according to the specific problem (this is somehow similar to the parameter s in SMAP, which was not fixed either). For most applications, however, the dimension of the input layer can be reasonably expected to remain very low.

The inputs to the neural network are the LS estimates for the lth tap gain at times $\{p,p-1,p-2\}$; these LS estimates are separated into their real and imaginary parts. This information is used to decide whether there is a dead tap or an active tap in a way that optimizes CTMSE. To do so, the NN birth/death detector needs to be trained with a reasonable amount of labeled noisy tap gain samples.

7.2.1.1 Network Design Decisions

Network design decisions such as the dimensionality of inputs were made through an iterative process of prototyping a suitable design, looking at its performance and trained weights and extracting conclusions.

For example, a 6 x 4 x 2 network design was initially simulated (in a short, preliminary simulation resembling the conditions in Subsection 7.2.4, but using real-valued taps instead of complex-valued taps, as a proof of concept). This design received all 6 latest real-valued tap samples, $\{\hat{h}^{LS}_p, \hat{h}^{LS}_{p-1}, ..., \hat{h}^{LS}_{p-6}\}$, instead of just 3. Matrix W represents the values for the weights[1] $w((v_{t,r}, v_{t+1,j}))$ after the network was trained with a training dataset:

$$W \simeq \begin{bmatrix} 0.57 & -2.32 & -1.34 & 0.05 & -0.39 & -0.48 \\ 5.93 & 7.15 & 2.70 & 0.45 & 0.12 & 0.40 \\ 7.93 & 0.47 & 0.01 & -0.16 & 0.01 & 0.14 \\ 8.89 & 0.87 & 0.23 & 0.15 & -0.17 & 0.11 \end{bmatrix} \tag{7.12}$$

As can be seen, the highest values of the matrix (that is, the highest weights) are found in

[1] For simplicity, the inputs associated to the constant neuron aren't included in W.

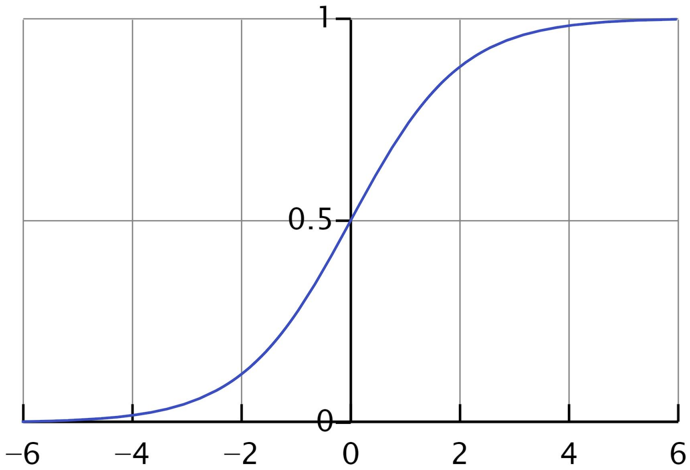

Figure 7.5: The sigmoid/logistic function.

the first 3 columns, and especially the first 2. This result was to be expected; for the purposes of knowing whether the tap has died, it is more important to know $\hat{h}_p^{LS}, \hat{h}_{p-1}^{LS}$ than $\hat{h}_{p-5}^{LS}, \hat{h}_{p-6}^{LS}$.

This suggested the complexity of the NN could be reduced by using a configuration of fewer neurons. By this process we eventually reached the 3 x 4 x 2 network, which was proven to have the same performance as larger sizes, for real-valued taps, and the 6 x 4 x 2 network shown in Fig. 7.4 for complex-valued taps.

Such preliminary simulations also helped discard inputs fed back from the Kalman filter itself (e.g. Kalman estimates in previous epochs), since they didn't provide any improvement in terms of performance.

7.2.1.2 Choice of activation function

The sigmoid/logistic function has been chosen as activation function. Its equation is (7.6). Its characteristic S shape, which is shown in Fig 7.5, provides a smooth gradient, i.e., prevents jumps in output values.

7.2.2 Scheme complexity

The proposed tap birth/death detector is low-complexity. It requires a very low number of neurons (e.g. 12 in Fig. 7.4) for each complex tap. Thus, its complexity grows linearly with the number of taps. In this regards, it is similar to SMAP and different from particle-filter-based methods, where birth/death detection is made for all taps in a single step.

Nevertheless, unlike SMAP, there are several trade-offs you could make when working with the proposed NN switch. For example, you could use either fewer inputs or more of them dependending on the required success rate. This could take the form of using only the real part of the gain (and ignoring the imaginary part) or using only the LS estimate at times $\{p, p-1\}$ instead of a longer sequence.

We hasten to add that the proposed system shown in Fig. 7.4, while low-complexity, shows a perfect detection record when measured in terms of CTMSE for the simulations in Section 7.2.4. This suggests the proposed scheme may provide an excellent trade-off between complexity and detection precision for real applications.

7.2.3 Covering the KF's eyes before death

For this scheme to work properly, it is important to decide what the KF should do at times of death and at times of life. This proposal advocates a strategy which could be described as "covering the KF's eyes before death", in a similar manner to covering a young child's eyes so they do not get traumatised by a shocking view. In fact, if the KF were to process dead tap gains following the standard algorithm, it would get "traumatised" and its estimates would get distorted by non-linear transitions. Therefore, it is proposed that the standard algorithm should be put on hold during death: the KF's prediction variance should not be modified until the tap is activated again. Similarly, once the tap is reborn, the KF should modify its previous predictions to match the new tap gain and avoid distorsions.

7.2.4 Simulation results

A system with $N_p = 10$ pilot subcarriers is considered, with $K = 8$ pilot symbols per subcarrier. The pilot symbols have equal average energy, σ_s^2, and a BPSK modulation scheme is used. The channel follows a LGM model and is assumed to have a uniform multipath delay profile, multipath spread smaller than the guard time, and no path gain correlation across different paths. The overall channel energy is normalized to one. $2.4 \cdot 10^5$ OFDM symbols are transmitted through a channel with $L_{max} = 10$, $P_{birth} = 0.05$ and $P_{death} = 0.05$. This amounts to 5 active taps on average. Individual paths are assumed to have the same average energy σ_h^2 over long periods and $\lambda = 0.999$. This choice of parameters makes it possible to compare the performance of the NNKF tracker (shown in Fig. 7.4, trained with samples over 70,000 estimation periods) to the computationally heavier methods outlined in [Angelosante et al., 2007], the KF system advocated in [Yuanjin, 2003] and the SMAP system in [Méndez-Romero and Fernández-Getino García, 2018].

Fig. 7.6 shows the precision (in terms of active/inactive tap detection mistakes) of the proposed NN in comparison to a ISS and a degraded ISS. Please notice that not all detection mistakes have the same effect on CTMSE. The ISS-99% fails to detect birth and death in 1% of the cases, but sometimes that means a small MSE error (because the tap gain wrongly detected as dead is close to zero, even though not zero), but a greater MSE error under different circumstances (e. g., when tap gain is wrongly detected, while being far from zero).

Fig. 7.7 shows the performance (CTMSE, as given in Eq. (5.5)) of our proposed NN birth/death detector + KF estimation vs. LS estimation and scenarios with perfect information (ISS), almost perfect information (ISS-99%) and a conventional stand-alone KF system with no birth/death detector.

It is evident that the NN+KF estimator provides much better performance when compared to ISS-99% and practically identical to ISS-100% (as shown in Table 7.1). In other words: unlike ISS-99%, the proposed NN's detection errors contribute almost zero to CTMSE. The NN's detection errors are minor in terms of CTMSE.

This is especially remarkable in view of the fact that the weights for the NN are not directly optimized for the MSE reduction, but for the birth/death detection.

This means that the NN+KF estimator achieves the tracking error reduction attainable by perfect death/alive tap state information. Please notice that, due to its very low computational

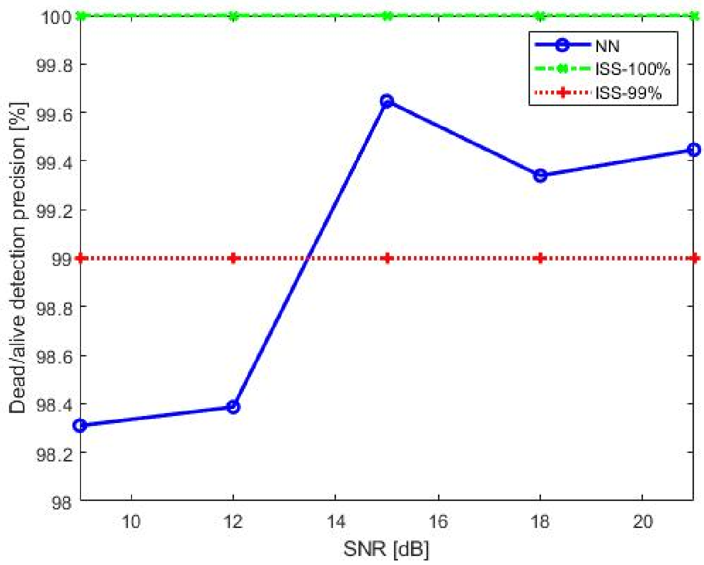

Figure 7.6: Tap activity detection precision [%] for NN and two degraded ISS.

SNR [dB]	CTMSE for NNKF	CTMSE for ISS-100%
9	107.9868	107.9868
12	63.5261	63.5261
15	38.6655	38.6655
18	21.9729	21.9729
21	12.5579	12.5579

Table 7.1: CTMSE for NNKF and ISS-100% in different SNR environments.

cost (as discussed in Subsection 7.2.2), an NN detection bank would be suitable for real applications over the entire SNR range of interest.

7.2.4.1 Sensitivity analysis

To gauge how sensitive vs. assumed death/birth probability the proposed Neural-Kalman scheme is, simulations were repeated for mistrained NNs, i.e. some neural networks were trained for $P_{birth} = P_{death} = P$ values that were different from those of the actual channel they would need to track.

In the first such sensitivity simulation, NNs were trained with a training sequence generated with $P = \{0.025; 0.05; 0.10\}$, then the NNKF had to track a channel where $P_{birth} = P_{death} = 10\%$. Results are shown in Table 7.2 and Fig. 7.8. Despite being trained on different P values, all three NNs had excellent tap status detection precision.

In a second sensitivity simulation, NNs were trained with a training sequence generated with $P = \{0.025; 0.05; 0.10\}$, then the NNKF had to track a channel where $P_{birth} = P_{death} = 2.5\%$. Results are shown in Table 7.2 and Fig. 7.8. Again, all three NNs had excellent tap status detection precision despite training on different P values. No large consistent effect is discernible; while only small differences in performance appear, and they could be reasonably expected to occur due to the randomness in the simulation runs. There is no clear pattern of degradation across both simulations.

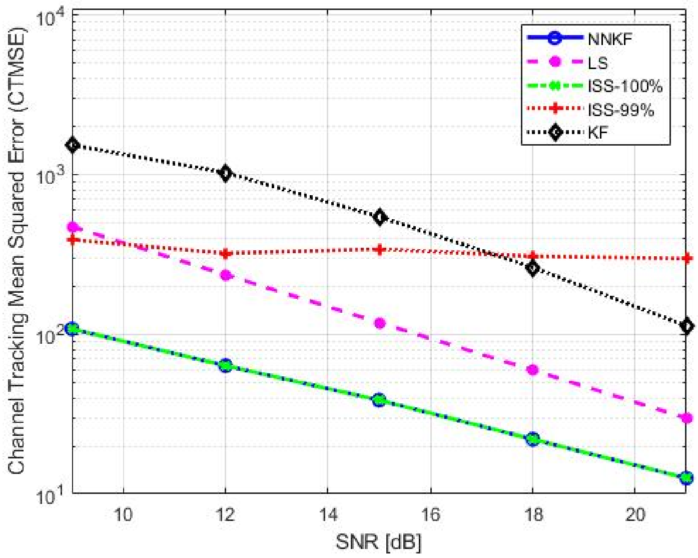

Figure 7.7: Tracking precision (CTMSE) for 3 schemes (NNKF, LS, KF) and 2 ISS.

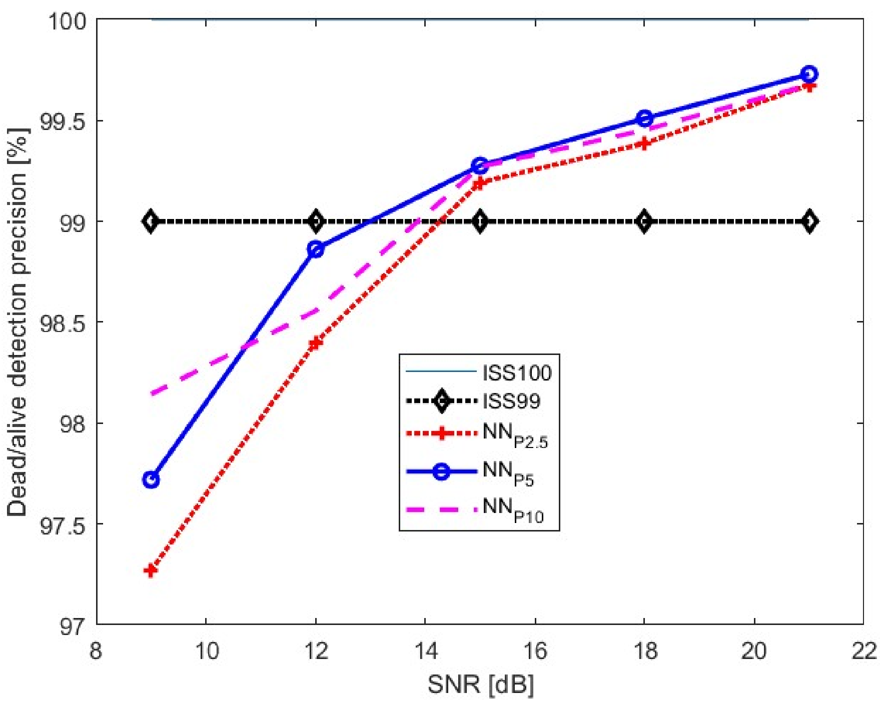

Figure 7.8: Tap activity detection precision [%] for NNs trained for different P, then tested for $P = 10\%$.

SNR [dB]	P=2.5%	P=5%	P=10%
9	97.81	98.43	97.95
12	99.18	99.03	99.02
15	99.21	99.13	99.11
18	99.66	99.56	99.50
21	99.55	99.81	99.85

Table 7.2: Tap activity detection precision [%] over a test sequence with $P_{birth} = P_{death} = 10\%$ when NNs are trained with a training sequence generated with $P_{birth} = P_{death} = P$.

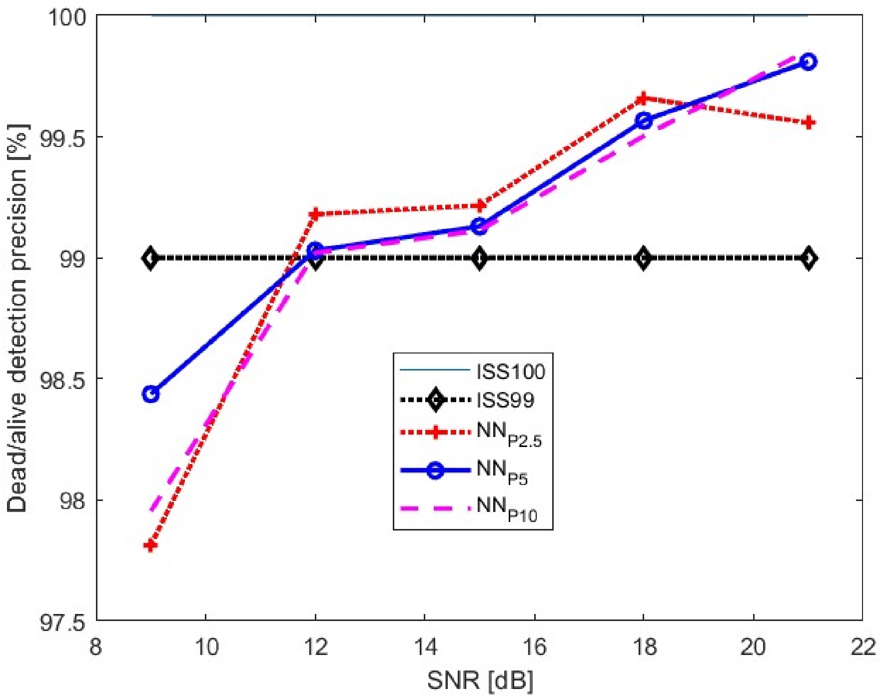

Figure 7.9: Tap activity detection precision [%] for NNs trained for different P, then tested for $P = 2.5\%$.

Therefore, the proposed Neural-Kalman scheme is robust to variations in P.

This statement is relevant for the practical implementation of NNKFs. Empirical channel models [Matolak, 2008, Wu et al., 2010, Hassan et al., 2020] include wide variations in P_{birth} and P_{death} across different taps (see Subsection 2.6.2). Thus, the fact that a single, trained NN could detect death and birth over taps with wide ranges of P suggests such an NNKF (a single, low-complexity scheme) could be reusable over different tap indices and communication environments.

7.3 Later Developments

After the publication of the NNKF proposal in [Mendez-Romero et al., 2020], several other papers have proposed related Neural-Kalman schemes [Revach et al., 2022, Pratik et al., 2021]. One of them, described in the following Subsection, is the Neural-Kalman RNN Hypernetwork.

SNR [dB]	P=2.5%	P=5%	P=10%
9	97.27	97.72	98.14
12	98.40	98.86	98.55
15	99.19	99.27	99.27
18	99.38	99.51	99.45
21	99.67	99.73	99.67

Table 7.3: Tap activity detection precision [%] over a test sequence with $P_{birth} = P_{death} = 10\%$ when NNs are trained with a training sequence generated with $P_{birth} = P_{death} = P$.

7.3.1 A Neural-Kalman RNN Hypernetwork

[Pratik et al., 2021] proposes a Hypernetwork Kalman Filter (HKF) for tracking applications with multiple different dynamics. The problem at hand is channel tracking in time varying channels. Authors recognise that Kalman filtering is the standard tool, but in Kalman proposals, the underlying dynamics of the channel is known, therefore the Kalman parameters can be matched to the dynamics. They assume a multiDoppler scenario where the Doppler is not known a priori.

In this work, Kalman updates are not modeled by an NN, and the model is causal. Instead of keeping a bank of Kalman filters and choosing one based on approximating the actual dynamics, HKF (which combines generalization power of Kalman filters with expressive power of neural networks) adapts itself to each dynamics based on the observed sequence. On the Clustered Delay Line, Type B (CDL-B) channel model, [Pratik et al., 2021] shows that the HKF can be used for tracking the channel over a wide range of Doppler values, matching Kalman filter performance with genie Doppler information. At high Doppler values, it achieves around 2dB gain over genie Kalman filter. The HKF generalizes well to unseen Doppler, SNR values and pilot patterns unlike Long Short-Term Memory (LSTM), which suffers from severe performance degradation.

The prediction is still done by Kalman, thereby enjoying robustness and generalization of Kalman. However, Kalman parameters are updated at each time using an RNN based on the process history. The RNN models the KF parameters in terms of residual around the mean set of parameters ϑ. In other words, the Kalman base parameters are fixed to θ, and the RNN provides corrections.

7.4 Conclusions

The proposed low-complexity NN-switched KF trackers outperform all previously known multipath channel tracking systems for OFDM communications, provided that tap birth/death phenomena are present. Moreover, its performance in terms of CTMSE is identical to that of the ideal case (ISS) where perfect knowledge of tap activations is available (see Fig. 7.10).

The proposed Neural-Kalman scheme is robust to wide variations in $P_{birth}/\ P_{death}$. In the context of the empirical probabilities of tap birth and death studied in Section 2.6, such robustness suggests a single, low-complexity NNKF could be reusable over different tap indices and communication environments.

Figure 7.10: Neural detectors for Kalman trackers

Chapter 8

Partial Tap Components and Lateral Hop Detection

In previous chapters, we have proposed a simplified framework for the channel tap birth-death problem and computationally inexpensive estimators, namely the SMAP and the NNKF abrupt-change channel trackers (Chapters 5 and 7, respectively).

Are there any other abrupt changes, apart from tap birth/death, that we should detect to improve channel tracking performance? Indeed. This Chapter proposes a different kind of abrupt change: energy shifts from one tap to adjacent taps (partial tap lateral hops) and discusses how to model, detect and track such changes.

8.1 A Novel Channel Model for Lateral Partial-Tap Hop Dynamics

Consider a channel with L_{max} possible taps, such that tap index $k \in \{1, 2, ..., L_{max}\}$. Consider a set of known partial tap components (PTC), that we might also indistinctly call semitaps, half-taps, sub-taps or simply partial taps, whose addition at the associated tap index k will determine the energy of the tap [Matolak, 2014]. We will denote each active PTC by:

$$\mathbf{c}_p^{(k,d)} = [k, g_p^{(k,d)}] \tag{8.1}$$

where $d \in \{1, 2, ..., D_{max}\}$ is an index to identify the different partial taps in a single whole tap, D_{max} would be the total maximum quantity of partial taps in a single whole tap, $p \in \mathbb{N}$ is the time interval index (from now on, "epoch" or simply "time") and $g_p^{(k,d)} \in \mathbb{R}$ would be the power contributed by the partial tap.

For some indexes k, p, d, the associated partial tap might be inactive. To account for that possibility, we can define the following singleton-or-void random set [Vihola, 2004]:

$$\mathcal{C}_p^{(k,d)} = \begin{cases} \{\emptyset\} & \text{if partial tap } (k,d) \text{ isn't active at time } p \\ \left\{\mathbf{c}_p^{(k,d)}\right\} = \{[k, g_p^{(k,d)}]^T\} & \text{if partial tap } (k,d) \text{ is active at time } p \end{cases} \tag{8.2}$$

A cautious reader will probably have noticed that we can now relate the partial-tap random sets to the whole-tap random sets $\mathcal{H}_p^{(k)}$ (used e.g. in Chapters 5, 6 and 7), by the following equations:

$$\mathcal{H}_p^{(k)} = \bigcup_{d=1}^{D_{max}} \mathcal{C}_p^{(k,d)} \tag{8.3}$$

$$\mathcal{H}_p = \bigcup_{k=1}^{L_{max}} \mathcal{H}_p^{(k)} = \bigcup_{k=1}^{L_{max}} \bigcup_{d=1}^{D_{max}} \mathcal{C}_p^{(k,d)} \tag{8.4}$$

In previous chapters, we have considered taps could die and be reborn. Similarly, these partial taps feature a specific property: they can hop from one tap to an adjacent tap.

To describe this mathematically, we can define the relationship $\nwarrow$ ("partial-tap left-hop") as:

$$\mathcal{C}_p^{(k,d)} \nwarrow \mathcal{C}_{p+1}^{(k-1,d')} \iff \begin{cases} \mathcal{C}_{p+1}^{(k-1,d')} = \mathbf{c}_p^{(k-1,d')} = [k, ag_p^{(k,d')}] \\ \mathcal{C}_{p+1}^{(k,d)} = \emptyset \end{cases} \tag{8.5}$$

where $a \in \mathbb{R}, a \geq 1$ is the (potential) power increase factor resulting from a relatively shorter path to the receiver, $k > 1$ and d' is any suitable partial tap index such that $\mathcal{C}_p^{(k-1,d')} = \emptyset$.

In simple words, we have defined a lateral hop that is associated to a shorter path to the receiver (thus increasing the contributed partial tap power); this partial tap hop requires an empty subtap at the destination[1], and that's why $\mathcal{C}_p^{(k,d')} = \emptyset$. The nomenclature $\mathcal{C}_p^{(k,d)} \nwarrow \mathcal{C}_{p+1}^{(k-1,d')}$ means the subtap $\mathcal{C}_p^{(k,d)}$ hopped to the adjacent left tap index, $k-1$, at the next epoch, $p+1$, thus leaving its former position void, $\mathcal{C}_p^{(k,d)} = \emptyset$.

Similarly, we can define the relationship $\nearrow$ ("partial-tap right-hop") as:

$$\mathcal{C}_p^{(k,d)} \nearrow \mathcal{C}_{p+1}^{(k+1,d')} \iff \begin{cases} \mathcal{C}_{p+1}^{(k+1,d')} = \mathbf{c}_p^{(k+1,d')} = [k, a^{-1} \cdot g_p^{(k,d')}] \\ \mathcal{C}_{p+1}^{(k,d)} = \emptyset \end{cases} \tag{8.6}$$

where $a \in \mathbb{R}, a > 1$ is the power increase factor resulting from a relatively longer path to the receiver, $k < L_{max}$ and d' is any suitable partial tap index such that $\mathcal{C}_p^{(k,d')} = \emptyset$.

In simple words, we have defined a lateral hop that is associated to a longer path to the receiver (thus reducing the contributed partial tap power); this partial tap hop requires an empty subtap at the destination, and that's why $\mathcal{C}_p^{(k,d')} = \emptyset$. The nomenclature $\mathcal{C}_p^{(k,d)} \nearrow \mathcal{C}_{p+1}^{(k-1,d')}$ means the subtap $\mathcal{C}_p^{(k,d)}$ hopped to the adjacent left tap index, $k+1$, at the next epoch, $p+1$, thus leaving its former position void, $\mathcal{C}_p^{(k,d)} = \emptyset$.

Finally, for a specific subtap $\mathcal{C}_p^{(k,d)}$, we could define the probability of lateral hop for each direction as $P_{lh} = P(\mathcal{C}_p^{(k,d)} \nwarrow \mathcal{C}_{p+1}^{(k-1,d')})$ for the left-hop and $P_{rh} = P(\mathcal{C}_p^{(k,d)} \nearrow \mathcal{C}_{p+1}^{(k+1,d')})$ for the right-hop, and the total probability of lateral hop as:

$$P_h = P_{lh} + P_{rh} = P(\mathcal{C}_p^{(k,d)} \nwarrow \mathcal{C}_{p+1}^{(k-1,d')}) + P(\mathcal{C}_p^{(k,d)} \nearrow \mathcal{C}_{p+1}^{(k+1,d')}) \tag{8.7}$$

8.1.1 Geometric Justification for Partial-Tap Dynamics

Why consider different partial tap components in a single tap, instead of the tap as a whole? Because we might expect different physical paths to contribute to the same tap; as either the transmitter, the receiver or a set of obstacles around them move (relatively to each other), those physical paths might change. They might get shortened or lengthened and, as a consequence, those partical tap components might "jump" to an adjacent tap, i.e. they might change tap index. The

[1]This shouldn't cause any loss of generality, since D_{max} can be chosen from an unbounded set.

result would be a reduction in the energy contributed to the former tap and an increase in the energy contributed to the new tap by this modified physical path [Zhang, 2016].

Furthermore, we can identify five different, geometrically-justified tap dynamics involving either full or partial taps in birth-death or hop dynamics:

1. **Uncorrelated (sub)tap birth-death dynamics** due to the appearance/disappearance of a reflecting obstacle, such as a truck in the same highway you are driving your car (the receiver) in. Once the truck appears on the road, a new tap is born [Hassan et al., 2020]. Once the truck exits the highway, the tap disappears [Hassan et al., 2020]. We can detect this birth/death dynamics through the SMAP and NNKF methods explained in Chapters 5 and 7.

   It might happen that two different trucks contribute two additive subtaps to the same tap index; in that case, we would have an abrupt fall (but not disappearance) of the tap once one of the trucks exits the highway; though not mathematically optimal, this can be reasonably expected to be dealt with non-catastrophically through the KF. However, a mathematical optimal solution (outside the scope of this Thesis) would involve a partial-tap channel model with partial-tap birth/death dynamics.

2. **Correlated (sub)tap birth-death dynamics** due to the appearance / disappearance of a reflecting obstacle. Here we can identify two different sources of correlation: correlated birth/death and post-death correlation.

   For correlated birth/death, consider a very large truck as a reflecting obstacle in the highway (see Fig. 8.1); the truck is large enough for several adjacent taps to appear at once [Hassan et al., 2020]. After the large truck exists the highway, all the associated taps disappear at the same time. A model that assumes uncorrelated birth/death for each tap is not suitable for this phenomenon.

   For post-death correlation, consider the multipath component [Rodríguez-Piñeiro et al., 2021b] dynamics in channels where Unmanned Aerial Vehicles (UAVs, commonly known as drones) are involved, either as transmitters/receivers [Rodríguez-Piñeiro et al., 2021a] or as interfering obstacles. Fig. 8.2 shows one such scenario: an interfering drone swarm (A) gets transitorily shadowed by a building (B). Once it reappears (back to A), (sub)taps correlated to pre-shadowing (sub)taps will reappear. Therefore, such a moving, reflecting obstacle can cause correlated (sub)tap birth-death dynamics.

3. **Gradual tap drift**. Taps (i.e. echoes) aren't really a solid object. Power is assumed to be continuous, i.e. delivered by an infinite number of rays [Tse and Viswanath, 2005]. Taps are a mathematical abstraction resulting from the integration of such assumed infinite number of rays. However, this set of rays/paths does not always get identically divided at the same delay boundary between taps (see Fig. 8.3); there might be some gradual drift to one side, thus delivering slightly more power to an adjacent tap and slightly less power to the tap they are moving away from.

4. **Lateral (sub)tap hop dynamics.** Consider base station-vehicle communications on a highway, similarly to the scenarios in [Hassan et al., 2020, Matolak, 2008, Wu et al., 2010, Hassan et al., 2021], in such a way that there is a tap associated with the LOS and one or more taps are associated with reflections on other vehicles. The closer an interfering vehicle is

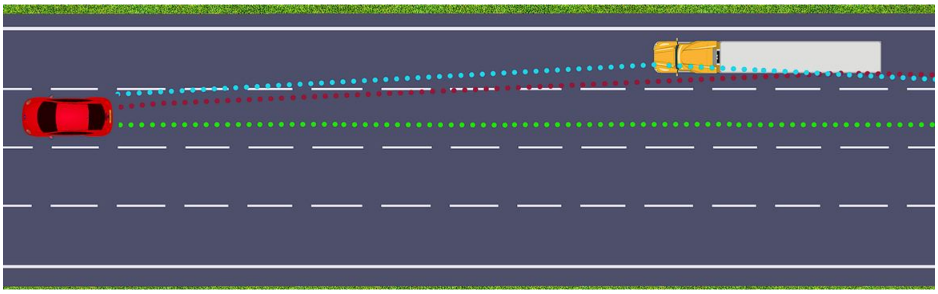

Figure 8.1: A truck in a highway is a reflecting obstacle that can cause correlated (sub)tap birth-death dynamics.

to the receiver vehicle, the more the echo path will resemble the LOS (see height trajectories getting closer in path length as the interfering red vehicles get closer to the orange receiver van in Fig. 8.4).

Let us focus on a scenario with LOS and (at least) two taps associated with reflections in tall vehicles (vans, buses or trucks) placed in the extreme right and left lanes (Fig. 8.5). Therefore, as the interfering vehicle gets closer to the receiver, the closer the tap of such reflecting vehicle will get to the LOS-associated tap (by hops to the left, i.e. to lower tap indexes, see Fig. 8.6).

If the tap index is already occupied by another tap, the hopping (sub)tap will add energy on top of the existing tap (Fig. 8.7). Once the interfering vehicle reaches the (receiver) vehicle, the reflected path becomes about as short as the LOS, i.e. the reflection-associated subtap hops onto the LOS-associated subtap (Fig. 8.8). This example shows how the relative movement of reflectors creates lateral (sub)tap hop dynamics.

5. **Unstereotyped dynamics**. Some dynamics effects might occur as combinations of the previous phenomena or more extreme versions of them. Is there any way to track a channel that is changing abruptly with dynamics that don't fit your previous models? See Section 9.4 for some potential mechanisms to learn such dynamics in real time.

8.2 NNKF Tracking of Channel with Lateral Partial-Tap Hop Dynamics

This Section proposes a neural detector [Shalev-Shwartz and Ben-David, 2014] for lateral hops. Simulations will show this NN detector can detect partial-tap hops in the assumed value ranges. Some potential schemes for practical implementation in wireless systems are discussed.

8.2.1 Partial-Tap Hop Dynamics

Consider the channel model in Section 8.1. The design task at hand would be detecting large, abrupt transfers of energy from a tap index to adjacent tap indexes. For simplicity, a symmetric model is considered, i.e. $P_{lh} = P(\mathcal{C}_p^{(k,d)} \nwarrow \mathcal{C}_{p+1}^{(k-1,d')}) = P(\mathcal{C}_p^{(k,d)} \nearrow \mathcal{C}_{p+1}^{(k+1,d')}) = P_{rh}$ and no power increase/decrease during hops, i.e. $a = 1$. Furthermore, it will be assumed that, at any epoch p, a partial tap with $g_p^{(k,d)} \in (g_{min}, g_{max})$ exists and might hop with probability $P_h = P_{lh} + P_{rh}$.

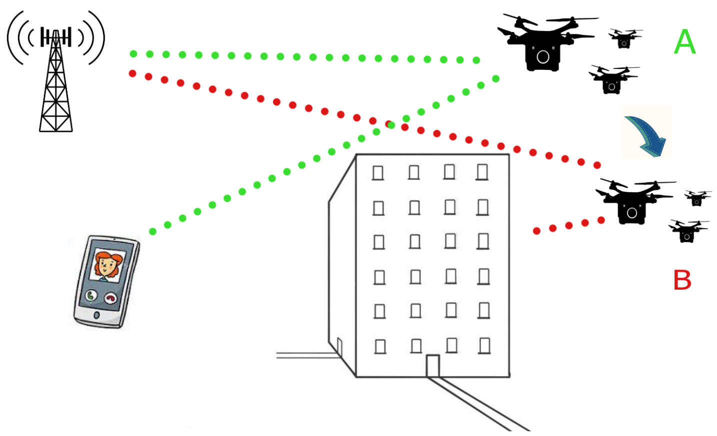

Figure 8.2: A drone swarm transitorily shadowed by a building is a reflecting obstacle that can cause correlated (sub)tap birth-death dynamics.

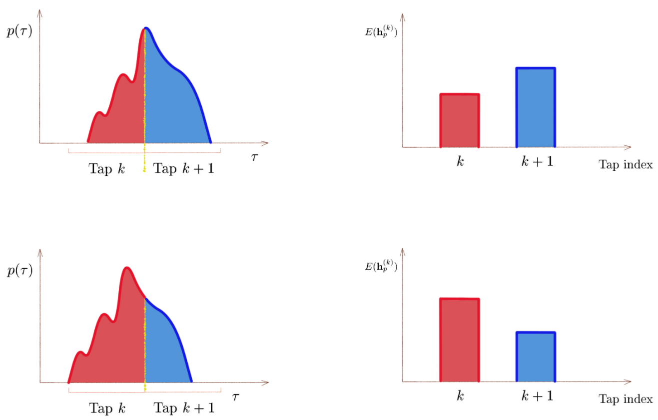

Figure 8.3: Gradual tap drift. Slow relative movements cause gradual drift.

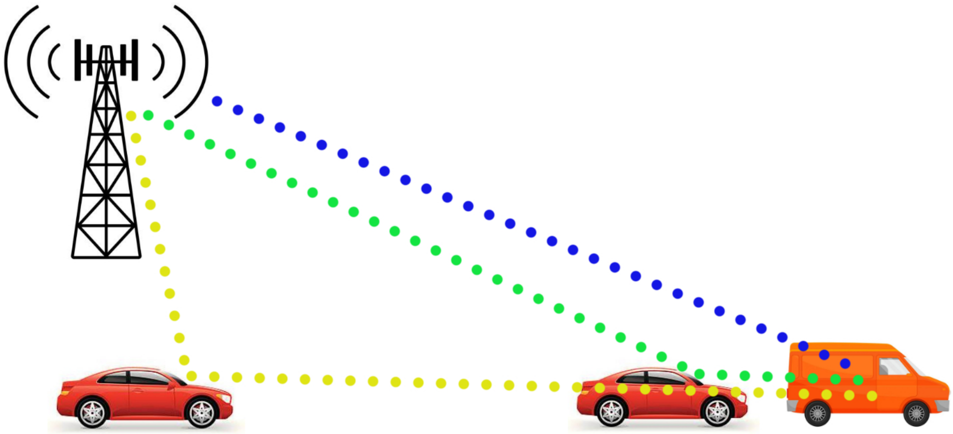

Figure 8.4: Height trajectories of LOS and vehicle reflections on the road.

8.2.2 A Neural Detector for Lateral Hops

A neural detector for such partial-tap hop dynamics is shown in Fig. 8.9. The inputs to the neural network are the LS estimates for the source tap (l_sth) gain and the target tap (l_tth) gain at times $\{p,p-1,p-2\}$; these LS estimates are separated into their real and imaginary parts, represented as $\Re(\cdot)$ and $\Im(\cdot)$, respectively. This information is used to decide whether there has been a large energy transfer between taps ('hop' vs. 'no hop'). To do so, the NN birth/death detector needs to be trained with a reasonable amount of labeled noisy tap gain samples. Since not all energy transfers are of interest, lateral hops that produce a small increase of energy in the target tap will be labeled as 'no hop', i.e.:

$$hop \iff r = \frac{g_{p-1}^{(k_s,d)}}{||h_p^{k_t}||^2} > q_{hop} \tag{8.8}$$

for a chosen threshold $q_{hop} \in \mathbb{R}$, $q_{hop} > 0$. For other NN design guidelines, refer to the similar design task in Subsection 7.2.1.

8.2.3 Simulations

A system with $N_p = 10$ pilot subcarriers is considered, with $K = 8$ pilot symbols per subcarrier. The pilot symbols have equal average energy, σ_s^2, and a BPSK modulation scheme is used. The channel follows a LGM model and is assumed to have a uniform multipath delay profile, multipath spread smaller than the guard time, and no path gain correlation across different paths. The overall channel energy is normalized to one. $2.4 \cdot 10^5$ OFDM symbols are transmitted through a channel with $P_{lh} = P_{rh} = 0.05$. Individual paths are assumed to have the same average energy σ_h^2 over long periods and $\lambda = 0.999$. Samples for NN training and performance testing were labeled as 'hop' only when the hop increased target tap energy by more than 25%, i.e. $q_{hop} = 0.25$.

8.2.3.1 Simulation results

Simulation results are shown in Fig. 8.10, where hopping tap energy is represented as θ for better legibility. Numerical values are shown in Table 8.1. The proposed NN provides a high detection

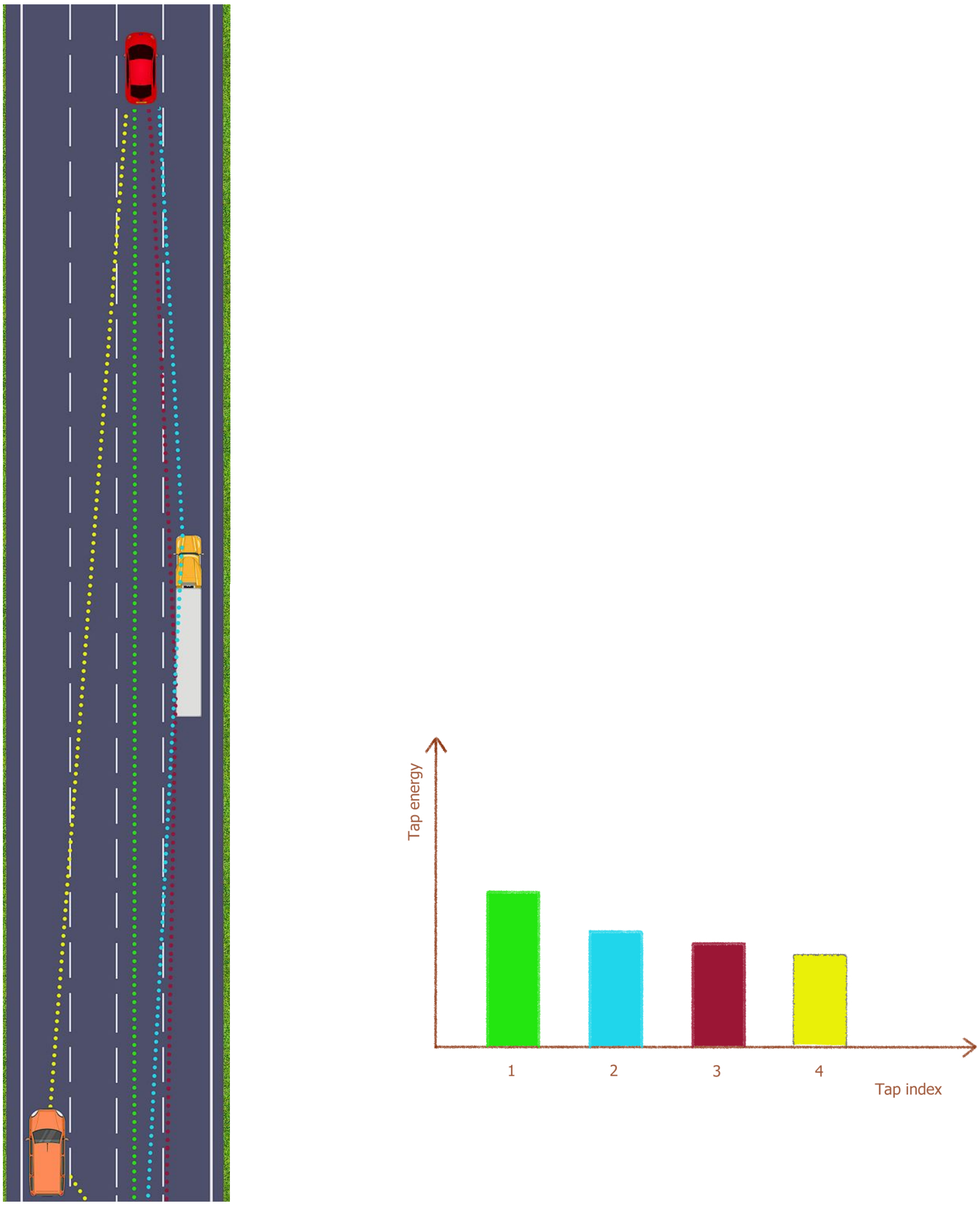

Figure 8.5: Taps in a highway situation: a LOS path (green), reflections over a long vehicle (blue and red taps) and a reflected path on a vehicle further behind (yellow tap).

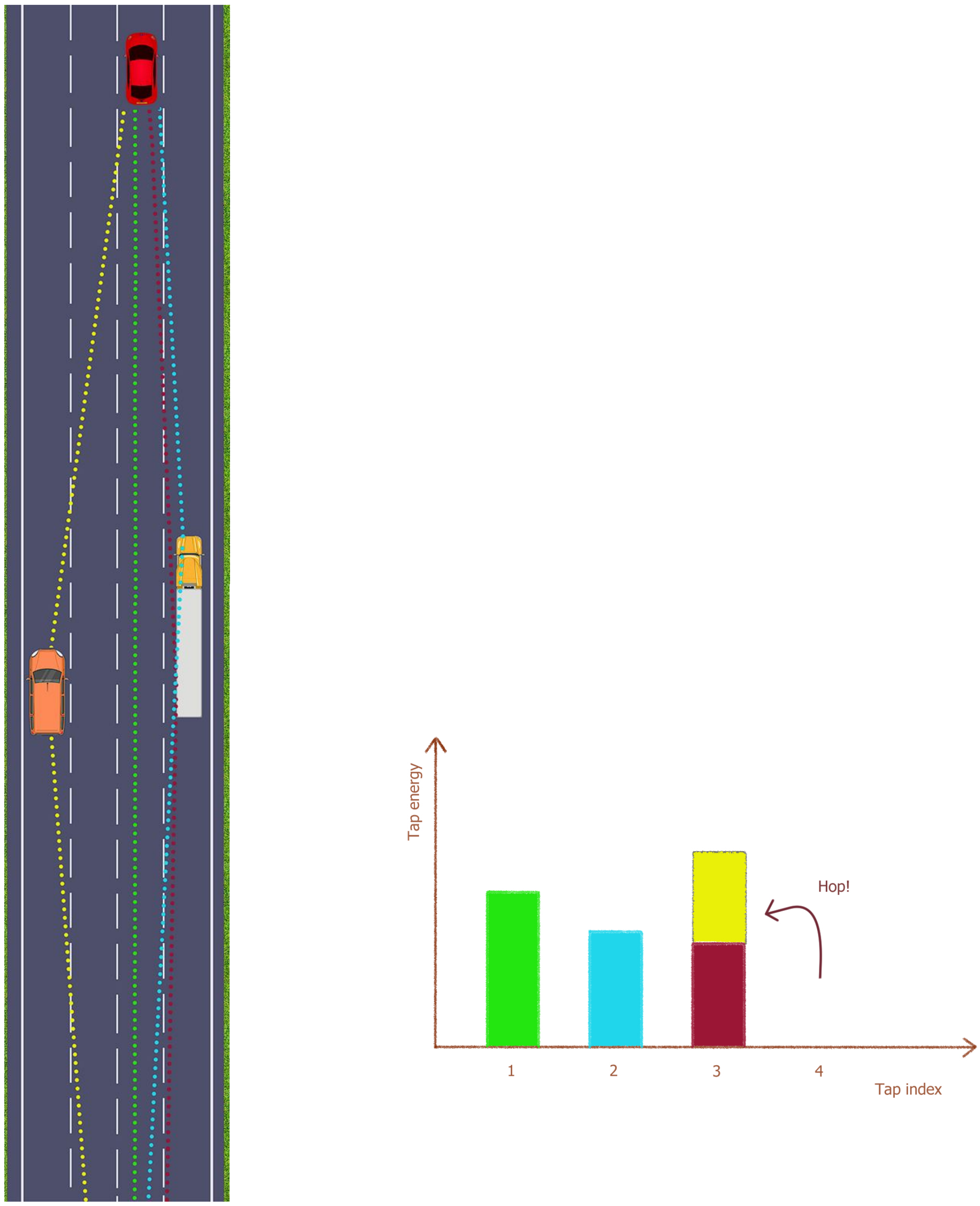

Figure 8.6: Lateral (sub)tap hopping in a highway situation: as the left vehicle advances (relatively), the yellow path becomes as short as the red one, i.e. the yellow tap hops onto the left-adjacent tap.

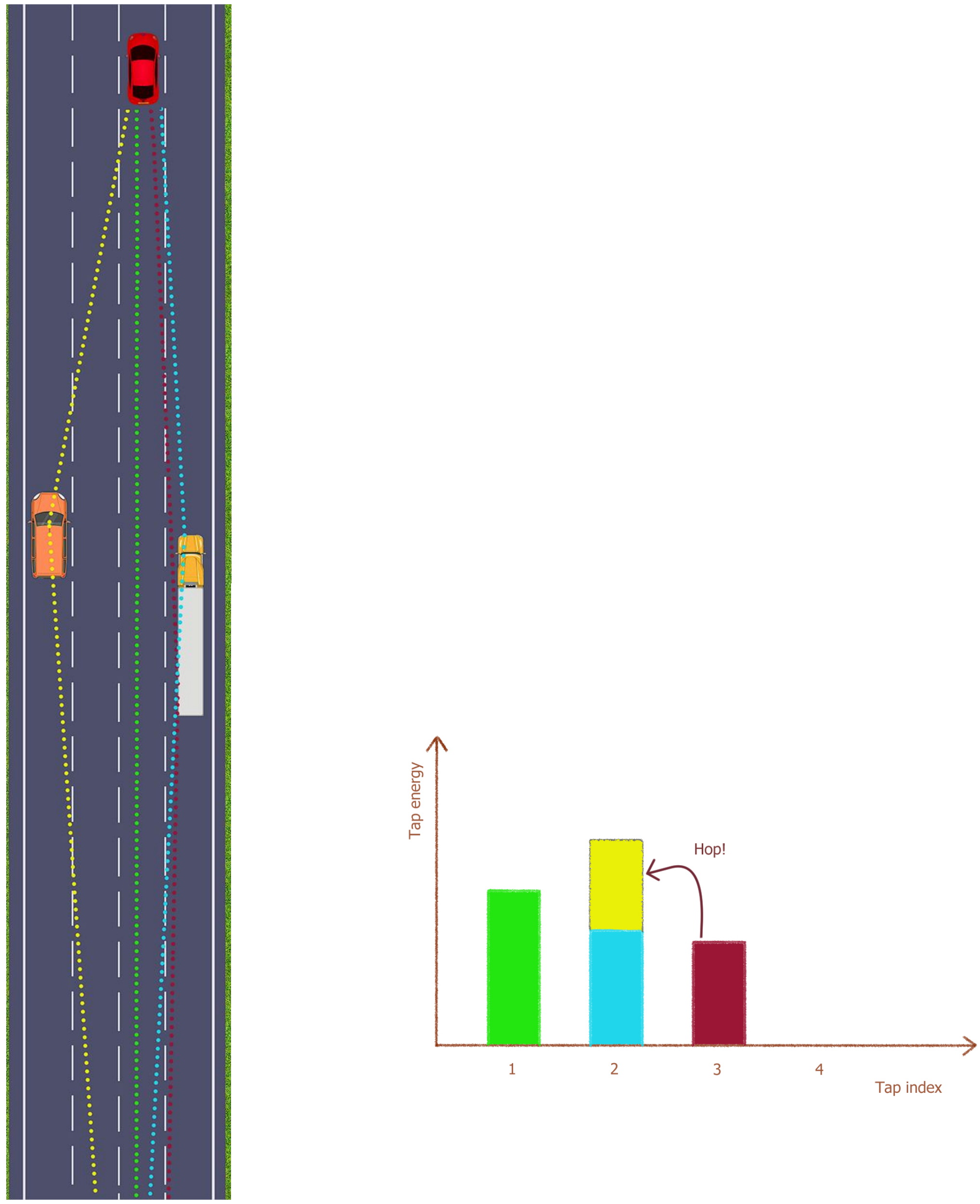

Figure 8.7: Lateral (sub)tap hopping in a highway situation: as the left vehicle keeps advancing (relatively), the yellow path becomes as short as the blue one, i.e. the yellow subtap hops onto the blue subtap.

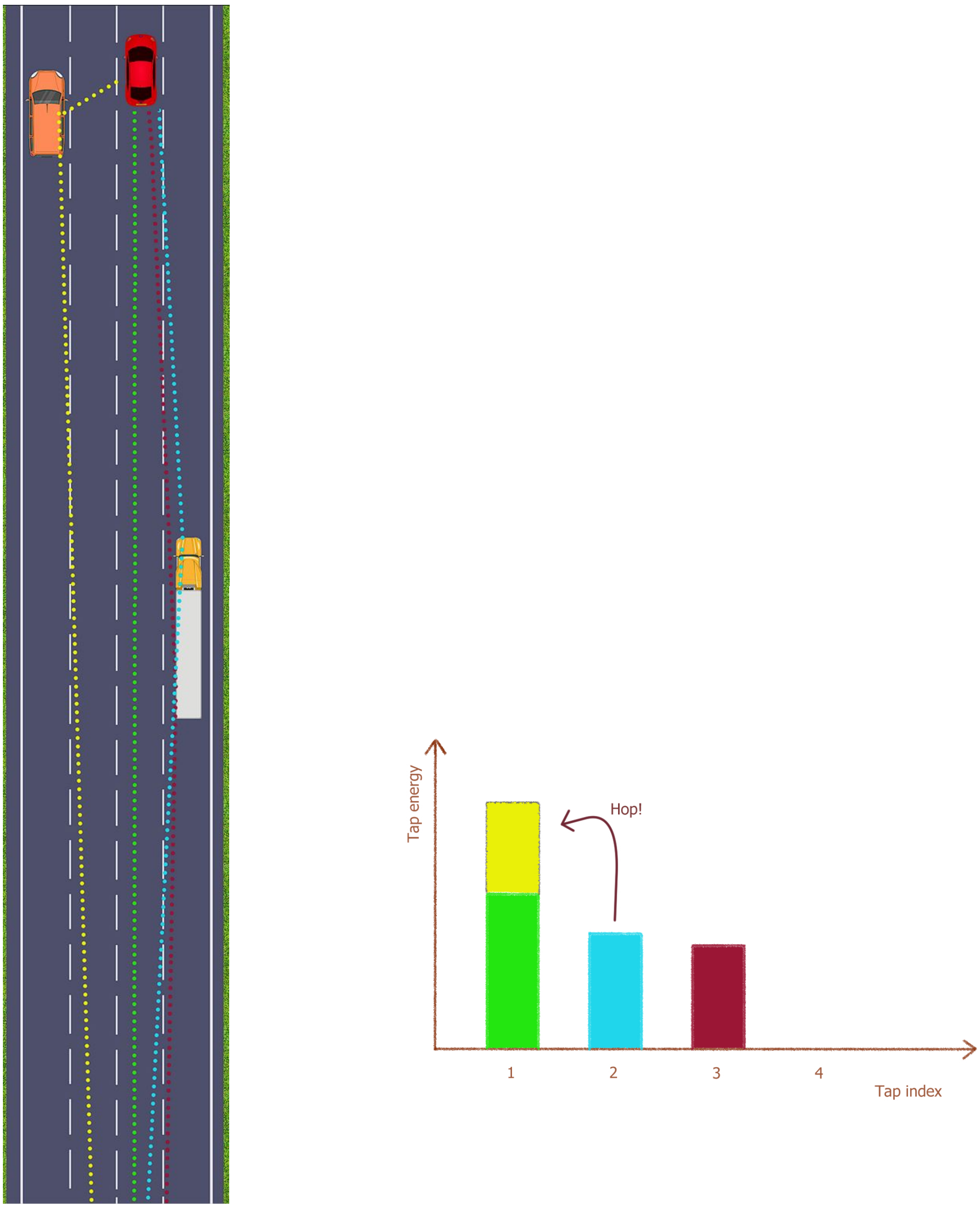

Figure 8.8: Lateral (sub)tap hopping in a highway situation: as the left vehicle reaches the red vehicle (receiver), the yellow path becomes about as short as the LOS, i.e. the yellow subtap hops onto the green subtap.

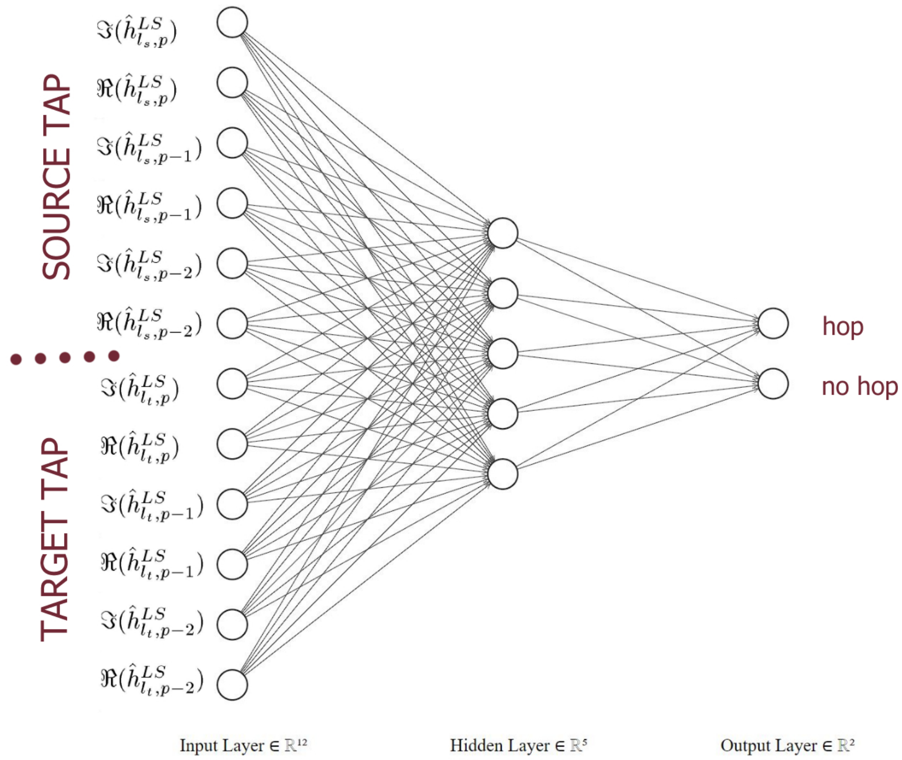

Figure 8.9: A neural detector for lateral hops

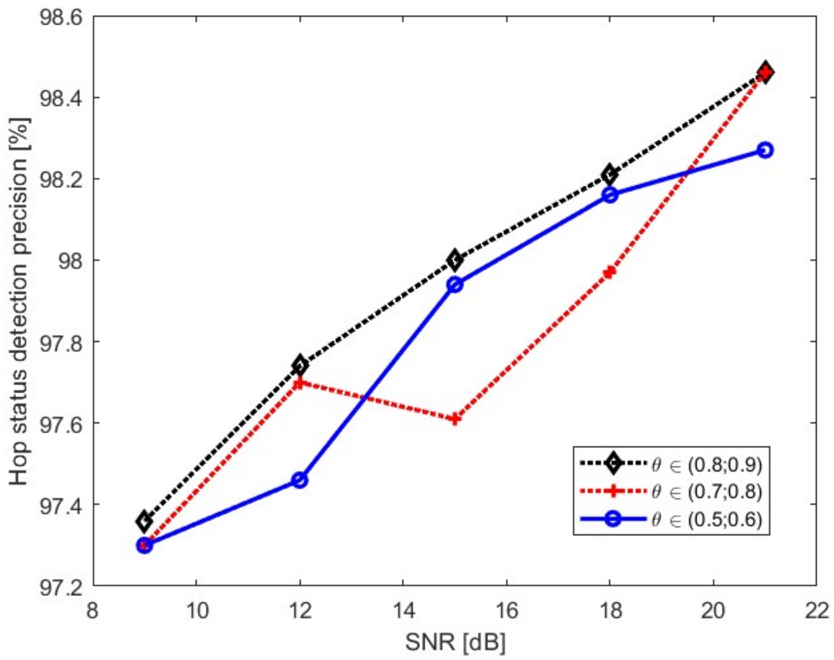

Figure 8.10: Lateral hop detection with NN for different energy transfer sizes $\theta := g_p^{(k,d)}$

SNR	$g_p^{(k,d)} \in (0.8; 0.9)$	$g_p^{(k,d)} \in (0.7; 0.8)$	$g_p^{(k,d)} \in (0.5; 0.6)$
9	97.36	97.30	97.30
12	97.74	97.70	97.46
15	98.00	97.61	97.94
18	98.21	97.97	98.16
21	98.46	98.46	98.27

Table 8.1: Neural detection precision for lateral hopping status

precision for the hop/no-hop classification task. Precision is significantly higher for high SNRs. While there is significant variation due to the randomness inherent in each simulation run, hop/no-hop classification is performed with high precision in all simulated energy transfer sizes with a slight decrease in performance as energy transfer sizes get reduced.

Simulation results suggest the proposed NN architecture might be a feasible lateral hop detector, suitable for integration in NNKF schemes.

8.3 Potential NNKF Integration

While the design of tracking schemes for specific lateral hopping applications goes beyond the purposes of this Thesis, the simulation results in this Chapter support future research in this area. In particular, the neural detector in this Chapter might be coupled with KF to create a tracking Neural-Kalman scheme with potentially superior performance in specific applications. Such a design could include both tap death/birth detection [Mendez-Romero et al., 2020] and lateral hop detection.

8.4 Conclusions

A novel channel model for lateral partial-tap hop dynamics (subtap hopping) was developed, i.e. instead of considering just the death/birth of the whole tap, we considered a single tap could be interpreted as the sum of smaller components, called subtaps or partial taps, that could hop to adjacent tap indexes. We have provided a geometric justification for this and additional non-stationary dynamics (Subsection 8.1.1). Finally, we have designed a neural detector for lateral hops with large energy transfers. Simulation results suggest the proposed NN architecture might be a feasible lateral hop detector, suitable for integration into NNKF schemes.

Chapter 9

Unstereotyped Abrupt Dynamics and Neural-Kalman Phenomena

In previous chapters, abrupt changes and the interactions between Kalman filters and neural networks have been reviewed, and the SMAP and NNKF schemes have been proposed to track channels under tap birth/death dynamics. In this Chapter, we leverage our newly found understanding of abrupt changes and Neural-Kalman schemes to analyze the neural consequences of *unstereotyped* abrupt changes.

Communication-engineering-relevant topics will be discussed in the process, including a Bayesian stereotyper for hopping LGM models that uses bio-inspired, neuroscientifically-backed mechanisms (dreaming, surprise), but such estimation methods will be presented through the prism of finance, where unstereotyped abrupt dynamics can be generalized into remarkable theoretical results.

In particular, a novel Neural-Kalman framework will be proposed to describe how market participants react to failing expert opinions and how to outperform them through a new investing strategy. Moreover, this Chapter will formalize the concept of Neural-Kalman Phenomena and suggest mathematical, neuro-evolutionary and social reasons why Neural-Kalman Phenomena might exist.

Let us begin by describing a disagreement between two financial experts.

9.1 Expert Disagreements

In early 2021, two prestigious economists, Lawrence H. Summers and Paul Krugman, held opposite views on US President Biden's huge stimulus program, including massive infrastructure spending and sending $1,400 checks to most Americans in order to restart the economy after the first Covid pandemic waves. Summers thought the economy would overheat and inflation would skyrocket within months [Summers, 2021, Princeton Bendheim Center for Finance, 2021]. Krugman thought this stimulus program made sense and inflation wouldn't rise to dangerous levels [Princeton Bendheim Center for Finance, 2021].

Both experts are smart and knowledgeable; yet one of these expert opinions would necessarily fail soon afterwards.

If Summers were to be right, inflation expectations would soon become unanchored and the Fed would be forced to raise interest rates in order to combat inflation, possibly causing a hard landing (a recession) in the process [Summers, 2021]. This would impact asset valuations, including stock market valuations. High-growth tech stocks would be particularly affected through lower sales, lower margins, lower earnings, lower short-term growth, higher costs and higher discount rates [Blanchard and Johnson, 2017, de Pablo López, 2001, Authers, 2021]. Thus, if Summers were to be right, high-growth tech stock prices would fall [Authers, 2021, Dale, 2021]. However, if Krugman were to be right, there would be no recession, no high inflation and no need to panic[1].

As this example shows, detecting failing expert opinions early is crucial for investing. It literally means money.

Furthermore, when applied to non-financial consequential matters, such as the emergence of large-scale wars or highly disabling pandemic viruses, it means not just money, but also health and survival.

This chapter will provide a Neural-Kalman framework to understand how market participants react to failing expert opinions and how to outperform them under certain circumstances.

9.2 A Hopping LGM Model for Inflation Forecasting

We can represent the underlying mental model for anchored[2] (Krugman) vs. unanchored (Summers) inflation through a very simple LGM model, so that, if CPI_p is to represent the 3-month seasonally-adjusted (SAR) consumer-price index (CPI) inflation rate at quarter p provided that the 'transitory, anchored' inflation mental model holds true, then the probability density function for CPI_p would be given by the following equations:

$$CPI_p = \mu_{LR} + \Delta CPI_p \tag{9.1}$$

$$f(\Delta CPI_p|\Delta CPI_{p-1}) = \mathcal{N}(\Delta CPI_p; \lambda \cdot \Delta CPI_{p-1}, (1-\lambda^2)\sigma^2_{CPI_{anch}}) \tag{9.2}$$

where μ_{LR} is long-run inflation rate target[3], ΔCPI_p is the CPI inflation gap vs. μ_{LR}, $\sigma^2_{CPI_{anch}}$ is the average energy of the CPI inflation gap (vs. μ_{LR}) when inflation expectations are anchored and $\lambda < 1$ is the temporal self-correlation of quarterly SAR CPI gap data. This would approximate Krugman's 'transitory inflation' hypothesis.

On the other hand, an ever increasing inflation à la Summers could be simplistically modelled by the following equation:

$$CPI_p = \mu_{peak} + \Delta CPI_p \tag{9.3}$$

$$f(\Delta CPI_p|\Delta CPI_{p-1}) = \mathcal{N}(\Delta CPI_p; \lambda \cdot \Delta CPI_{p-1}, (1-\lambda^2)\sigma^2_{CPI_{unanch}}) \tag{9.4}$$

where μ_{peak} is the expected short-term (crisis) peak inflation rate [4], ΔCPI_p is the CPI inflation

[1] This simplified model of asset valuations includes only considerations around liquidity (including inflation/interest rates). Please notice this is an oversimplification that hides non-liquidity-driven risk. A collapse in liquidity was a sufficient condition for the end of the asset price boom, but not a necessary condition. Black Swans could have ended it prematurely (see Appendix Subsection B.3.4). As per legendary investor Jeremy Grantham's view: "A spike in inflation [leading to higher interest rates] may be sufficient, but it is not necessary, to crack this market." [Grantham, 2021]

[2] For a detailed explanation of anchoring, see Appendix, Subsection B.3.3.

[3] The long-run target assumed achievable does not necessarily and will not generally match the long-run empirical average inflation rate.

[4] This could alternatively be replaced by an ever increasing expected peak rate $\mu(p)$ that grows either linearly or

gap vs. μ_{peak}, $\sigma^2_{CPI_{unanch}}$ is the average energy of the CPI inflation gap (vs. μ_{peak}) when inflation expectations are unanchored and $\lambda < 1$ is the temporal self-correlation of quarterly SAR CPI gap data. This would approximate Summers's 'not-that-transitory inflation' hypothesis.

To better illustrate the model, in the Krugman vs. Summer debate, some approximate numbers would be $\mu_{LR} = 4\%$ (a long-term inflation figure Krugman himself had previously defended as benign for the economy [Krugman, 2013]) and, for μ_{peak}, a subjective estimate of the 'inflation pain threshold[5]' at which the Fed would be forced to cause a recession e.g. $\mu_{peak} = 10\%$. Other forecasting models, such as the one approximating Bank of England's forecasts for the UK in the February '21 - August '22 period (see Fig.9.1), would be more optimistic and closer to $\mu_{LR} = 2\%$.

Assuming inflation data is noisy, a KF algorithm could be used to update the current inflation estimation $C\hat{P}I_p$ at epoch p, as well as predictions for the short term. Forecast inflation for epoch $p+n$, $n \in \mathbb{N}$, as forecast at epoch p, would be given by:

$$CPI_{p+n|p} = \mu + (C\hat{P}I_p - \mu) \cdot \lambda^n \tag{9.5}$$

where μ would respectively mean μ_{LR} (anchored inflation) or μ_{peak} (unanchored inflation), respectively.

Our hopping LGM model is convenient for 5 reasons:

1. It is exactly the same model we have been using for channel taps in the previous chapters, so it is straightforward to apply KFs, NNs and our proposed NNKF scheme.

2. It actually has a good overlap with some expert forecasts (e.g. BoE's in Fig. 9.1).

3. One-dimensional dychotomies are frequent in finance and some insights about them are easily generalisable to multi-dimensional, multiple-choice decisions.

4. In this chapter we will explore the possibility that this underlying 'LGM anchor' is generalizable to social and financial affairs as the neurofinancial consequence of our human/social cognitive adaptation to Neural-Kalman Events.

5. This model is simple and intuitive enough to present the neuroscientific advances around the changepoint - oddball - reversal problem.

9.3 The Changepoint - Oddball - Reversal Model

Neuroscientists have proposed that the brain works as a Bayesian estimator. In that model, called the Bayesian Brain, animals (including humans) are continually predicting what is going to happen. If something unexpected happens, the mechanism of surprise gets activated. That mechanism helps you switch from your current model to a new one, whether it is an old model you have already learned or a new model that you will need to learn from scratch. The cognitive task of deciding whether the current experience matches our activated model is arguably common

exponentially for as long as the crisis is not decisively tackled + an inertial period (e.g. interest rate hikes' effects on inflation arguably do not fully show up until 12-18 months later).

[5] To be more precise, μ_{peak} would be the expected peak inflation after a Fed hawkish pivot, i.e. the peak inflation rate that would be reached after a Fed pivot if the 'inflation pain threshold' is reached, forcing the Fed to pivot. The post-hike peak inflation rate would be expected to be higher than such early pain threshold due to the fact that interest rate hikes would only slow down the economy with some delay.

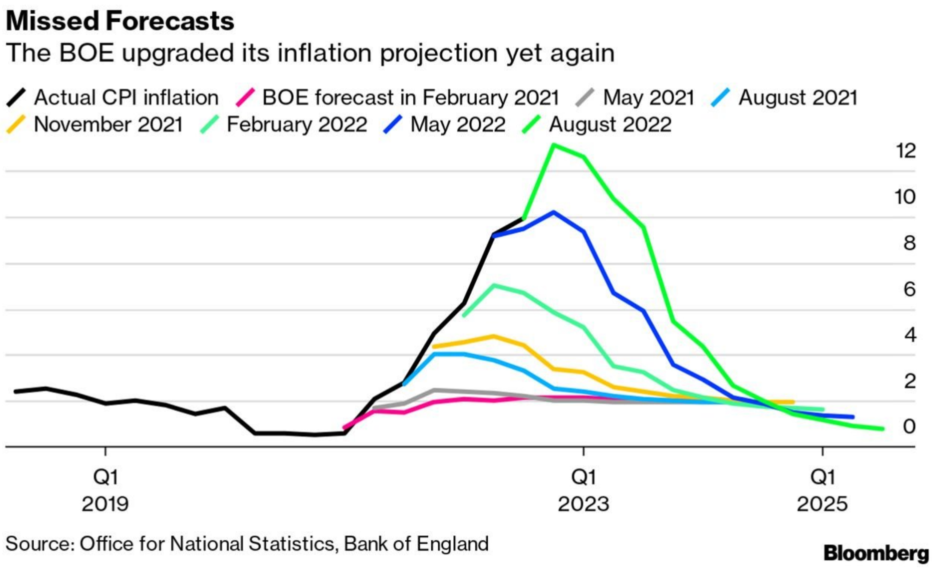

Figure 9.1: Inflation forecasts for the UK by the Bank of England (BoE), reminiscent of a KF prediction based on the wrong underlying plant model. (ONS, BoE, BGG)

to mice in laberynths, investors facing a new potential high-inflation paradigm or channel tracking algorithms under birth/death/partial tap hopping dynamics (Fig. 9.2).

Neuroscientists model this dilemma as the so-called “changepoint - oddball - reversal problem”. This problem is represented in Fig. 9.3 through its three possibilities:

- **Changepoint**. This occurs when a new paradigm starts, forcing you to switch from one model to the next. In Fig. 9.3 (uppermost line) we can see how one state, let us call it state s_1 (denoted by green dots), is followed by state s_2 (denoted by light blue dots) and then state s_3 (denoted by purple dots).

- **Oddball**. Due to the statistical nature of the problem, you can get an outlier at any given moment. If that outlier isn't followed by other similar outliers (which would suggest a changepoint), then it makes sense to discard it. In Fig. 9.3 (central line) we can see how state s_1 (green) does not change into a different state; the highest samples, marked with a yellow star, are simply surprising outliers that must not lead to cognitive overreaction (false detection of a changepoint).

- **Reversal**. This occurs when a new paradigm starts and then reverses after a short while. We have seen this phenomenon in channel tracking, e.g. when a tap is activated and later dies (i.e. returns to the initial state). In Fig. 9.3 (lowermost line) we can see how state s_1 (green) is followed by s_2 (blue) (that would be a changepoint) and then the sequence returns to state s_1 (that would be a reversal).

These dynamics appeared in the channel tracking problems we dealt with in previous chapters, such as tap death/birth, where underlying dynamics were known. An interesting question would be: what happens if those underlying dynamics are unknown? How can you solve the estimation/prediction problem in that case? This problem would have applications both in communications engineering and other areas such as finance. Life itself, after all, is to some extent an

Figure 9.2: Mice, investors and robots face similar neurocognitive challenges.

estimation problem under abrupt unstereotyped dynamics. In our next section we will provide some hints to solve it.

9.4 A Bayesian Stereotyper for Hopping LGM Models

Consider an algorithm (let us call it Stochastic State Stereotyper[6], or simply Stereotyper) that receives noisy samples from a random generator with N possible active states. At each epoch, the active state can jump to (some of) the other possible states with a fixed likelihood or stay at the same active state, i.e. the underlying states follow a Hidden Markov Chain Model (HMM) [Fuh and Tartakovsky, 2019, Satorras et al., 2019]. Each state outputs samples following an LGM model centered on different **unknown** means μ_1, μ_2, μ_3,...,μ_N, etc, for states s_1,s_2,s_3,...,s_N, respectively. For the sake of simplicity, let us assume all LGM models have the same variance and, again, the total number of underlying LGM models is finite (capped at N, but N is unknown to the Stereotyper).

The Stereotyper will receive one element at each time step.

For each time step, the Stereotyper's goal is to produce the best possible approximation of the underlying Markov Chain, ideally through a minimal-computation update algorithm.

Each of the stereotypes generated by the Stereotyper will be called a Stochastic State Stereotype (SSS).

9.4.1 Trivial case (no noise)

In the trivial case[7], where both measurement noise and plant noise are assumed to be zero, you are receiving a time series whose elements belong to a finite set. The time series is distributed as

[6] We are taking an engineering approach to this problem, i.e. fully Bayesian stereotypers will only be acceptable solutions for low-complexity problems such as the trivial case. For the general case we will propose a combination of AI and bio-inspired algorithms with Bayesian features, so, if we take off our engineering hat and we put on our 'mathematics hat', Stochastic Stereotyper or maybe Bayesian-Inspired Stochastic Stereotyper would be more precise terms than 'Bayesian', strictly speaking.

[7] I would like to thank Robert Israel, Associate Professor Emeritus at the University of British Columbia, for suggesting both this interpretation of the problem and its solution to me.

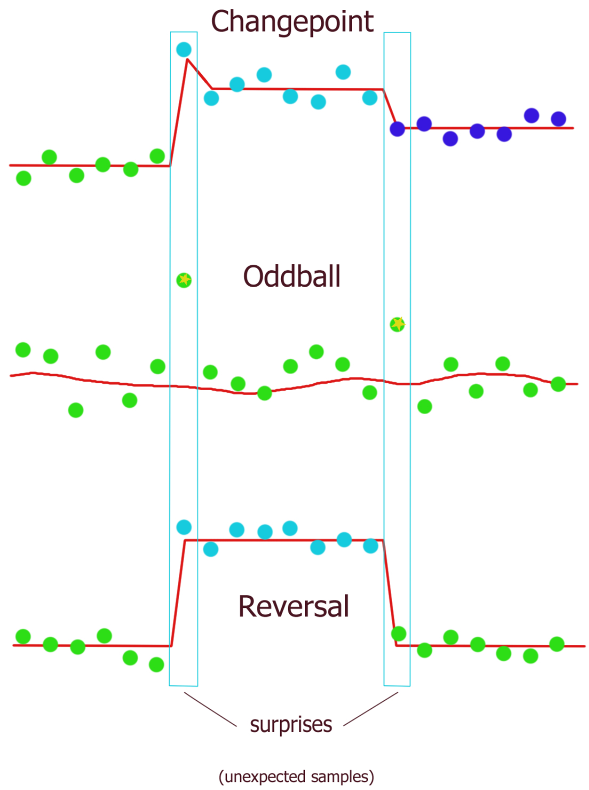

Figure 9.3: Surprising events can be classified as changepoints, oddballs (outliers) or reversals.

a Discrete-Time Markov Chain on the finite set.

For a given state i, row i of the transition matrix $\Theta_{tr} = (P_{ij})$ gives the transition probabilities P_{ij} from i to j, $j = 1..n$ (the number of states). This is a probability distribution, and the minimum variance unbiased estimator of it is the empirical distribution $\hat{\Theta}_{tr} = (\hat{P}_{ij})$, with elements $\hat{P}_{ij} = N_{ij}/N_i$, where N_i is the number of times state i was observed before the latest time and N_{ij} is the number of times state i was directly followed by j.

9.4.2 Bio-inspired AI mechanisms for the non-trivial case

A general, efficient solution for the non-trivial case is beyond the scope of the current PhD Thesis. However, we can provide an overview of how different schemes analyzed in the current Thesis, as well as other bio-inspired AI mechanisms, might fit into the full picture of a general solution.

The following proposals have some commonalities with very recent publications such as LeCun's Autonomous Machine Intelligence [LeCun, 2022] and Yu's Adaptive Learning Structure models for biological brains [Yu et al., 2021], though the basic core ideas in our proposal were initially developed earlier than, and independently from, the publication of such works and are more geared towards problems where Neural-Kalman schemes might be particularly useful.

Let us consider the following bio-inspired AI mechanisms:

1. **Surprise**: once we receive a surprising sample (a sample that looks like an outlier based on our current active stochastic model and could be a changepoint/oddball, see Section 9.3), we consider the possibility that this is a new (potentially unstereotyped) state.

2. **Attention (stereotyping)**: attention is the recognition that something is interesting and deserves further neurocognitive work or, in other words, a non-zero complexity-memory budget, i.e. memory space or computational power are reserved for the task deemed interesting. For the hopping LGM model problem, the task will be: create a better stereotype (a more precise SSS).

3. **Dream**: generation of training samples based on a set of SSSs so as to train an NNKF scheme and adapt it to a specific SSS, or to improve classification when choosing between two or more SSSs. Importantly, dream is an optimization task that can take place remotely (say, if the Bayesian stereotyper is an estimator doing stereotyping/exploitation work as a channel track on a train, the dream stage could be done on a computer on that train; this is feasible to the extent idle computers and the internet are everywhere now, greatly increasing computational power).

4. **Logic (stereotype algebra)**: SSSs' growth is exponential if left unchecked. You need a set of reasoning rules that prunes the set of SSSs deemed valid. To do so, you might need to merge several closely related stereotypes into one (e.g. $\hat{\mu_1}, \hat{\mu_2} = \hat{\mu_1} + \epsilon$, $\hat{\mu_3} = \hat{\mu_1} - \epsilon$, for a sufficiently small ϵ with respect to noise and very similar estimated HMM transition probabilities, would be merged into a single SSS). Moreover, some subtle differences between possible states might not cause enough surprise and, thus, such different states would never be stereotyped into different stereotypes. By defining certain convenient operations on stereotypes (operations that will enable the Stereotyper to speculate on the existence of more granular state divisions), the stereotyping outcome might become more granular and precise.

5. **Neurons and Kalman filters** (KFs, NNs, NNKFs). Kalman filters would provide optimal estimations/predictions for the LGM models (for as long as there is no transition; then surprise will likely be activated). NNs can computationally optimize detections by frontloading the computational cost in an external, computationally-rich device. Once the HMM is properly stereotyped, NNKFs will provide excellent performance for the exploitation phase (as they did in our channel tracking tasks in previous chapters).

9.4.2.1 A Stereotyper thought experiment

Let us imagine how such bio-inspired mechanisms could interact with each other to solve the problem. The Stereotyper would start with 0 stereotypes. Once it receives the first sample, the Stereotyper is **surprised**[8], since it is out of typical range for the currently active SSS (since there is no SSS at all), so it assigns some complexity-memory budget to stereotype it (**attention**), i.e. it locally computes a stereotype. For example, for a random number n_i of samples, the Stereotyper will assume that all such received samples will have been generated by the same HMM state. If the complexity-memory budget is large enough, the Stereotyper might run several competing stereotyping tasks in parallel, i.e. $n \in \mathbb{R}$, $i \in \{1, 2..I_{max}\}$ and later decide which one better fits the data through **logic** (elimination of the SSS with less predictive power, or merging all I_{max} SSS or a subset of them, to the extent they are identical or close to identical).

Why would SSS have predictive power? Because the underlying LGM makes it possible to run a KF over it, and KFs provide predictions. Actually, KF's predictions are optimal assuming the stereotype is perfect, as far as no new state transition occurs due to the HMM dynamics.

Let us continue with our *Gedankenexperiment*. Let us assume the Stereotyper has achieved its first stereotype and it is the only available stereotype left (all $I_{max} - 1$ others were eliminated or merged with it). As far as the new samples are close to the KF's prediction, the KF keeps exploiting the SSS and no other bio-inspired mechanism is activated. However, as soon as a new surprising sample is received, the mechanism of surprise is activated and a second state is stereotyped (again, the Stereotyper might run several random-length stereotyping tasks in parallel).

Soon, the set of valid SSS (i.e. the Stereotyper's estimation of the HMM) becomes large enough that the computations to decide which state is the current state sometimes becomes costly.

Fortunately, an idle computer lies nearby for a while and that gives the Stereotyper the opportunity to **dream**, that is, to temporarily use an external resource in order to train **NNs** that, in combination with the **KFs**, can easily and correctly decide in case of edgy jumps that would require too much computation otherwise.

Once decisions have been computationally optimized through the newly trained **NNKFs**, explotation is much cheaper (locally) and computational shortage issues have been averted.

In conclusion, bio-inspired mechanisms could help solve the Bayesian stereotyping estimation/prediction problem through structures such as the proposed Bayesian stereotyper. Our next section focuses on one such mechanism: surprise.

[8] There are many possible ways to implement surprise. The SMAP and NNKF algorithms arguably implement surprise. Another interesting technique would implement surprise through the "Bayes Surprise Factor" [Liakoni et al., 2021], which can be applied to Particle Filters (Subsection 4.4.1), Message Passing (Subsection 4.4.2) and Variational Learning (Subsection 4.4.4) schemes.

9.5 *Ojipláticos*: Neural-Kalman Schemes in Biological Brains

Human languages have related surprise to the eyes with expressions such as *having eyes like saucers*, *poner los ojos como platos* or *quedarse ojipláctico*[9] (Spanish) or *große Augen machen* (German). Our surprise is very noticeable in our eyes, and not just in those of us who cannot hide our emotions easily.

Adrenaline (or, to be more precise, norepinephrine) is involved in the human mechanism of surprise and the activity of the norepinephrine-dispensing system[10] can be measured by pupil dilations. In the face of an abrupt change, some people dilate their pupils more than others; interestingly, the size of that pupil dilation has been shown to be a measure of how much they have changed[11] their minds [Filipowicz et al., 2020]. Some researchers have speculated that the P300 response, that is, a measured brain response (recorded by electroencephalography), also reflects the extent to which people update their beliefs in response to an abrupt change [Jepma et al., 2016]. Interestingly, when facing an oddball instead of a real changepoint, people get *sauce-eyed* for longer. In those cases, both pupil dilation and the P300 response stay active for longer and they predict less, not more, learning [Nassar et al., 2019, O'Reilly et al., 2013].

Isn't this a contradiction? Not necessarily. You get surprised when your expectations don't get fulfilled; in that case, your brain tries to switch to a new state (consider an NNKF system detecting an abrupt change and switching to a new state). If the state is completely different, you would try to "learn" it (stereotype its statistical behaviour); if it soon resembles something you already know (it comes back to 'normal' after an atypical reading), then you won't need to learn it, but you will need to switch back to the old state. Thus, [Yu et al., 2021] proposed that norepinephrine signals the need for a state transition, rather than the need to learn per se.

In other words, to the extent we primates might occasionally behave in a manner resembling Neural-Kalman schemes[12], our surprise (norepinephrine) signals the KF is wrong and we have to switch into a new (potentially unstereotyped) state.

9.6 The Neural-Kalman Conjecture & Market Hypothesis

9.6.1 Neural-Kalman Problem Category Conjecture

Conjecture 10.1. For a fixed memory requirement, there is a non-empty class of problems that can be solved with Kalman schemes by expending lower complexity than schemes not incorporating the Kalman filtering algorithm.

In particular, this conjecture implies that the algorithm called 'Kalman filter' is computationally optimal for Kalman filtering[13], i.e. you cannot obtain a MMSE estimator in linear problems with

[9] Literally "sauced-eyed".

[10] The so-called locus coeruleus/norepinephrine system.

[11] A measure of how people update their priors after noticing an abrupt change. "Using an auditory adaptive decision-making task, we show that evoked pupil diameter is more parsimoniously described as signaling violations of learned, top-down expectations than changes in low-level stimulus properties. We further show that both baseline and evoked pupil diameter is modulated by the degree to which individual subjects use these violations to update their subsequent expectations, as reflected in the complexity of their updating strategy." [Filipowicz et al., 2020]

[12] For a full discussion of references around this hypothesis, see Section 9.7 and Appendix Sections B.2 (for biological Neural-Kalman schemes in animals) and B.3 (for Neural-Kalman features in human behaviour).

[13] 'Performing an operation' and 'performing an operation optimally' are two different things. Addition performs addition, but addition doesn't always perform addition *optimally* if memory is available, e.g. the computation of 333 x 444 = 147,852 can be performed by adding the number 444 once and again, 333 times, or it can be optimally solved through multiplication.

a different, computationally more inexpensive set of steps *if* there is a sufficiently low *memory requirement*. To the best knowledge of this author, nobody has proven that the Kalman filter is computationally optimal for Kalman filtering in such conditions, so we need to conjecture it (read: assume it). See Appendix Section B.1 for a deeper discussion of Kalman vs. Neural-Kalman complexity.

Definition 10.1. The Kalman Problem Category is the class of problems that, for a fixed memory requirement, can be solved with Kalman schemes by expending lower complexity than schemes not incorporating the Kalman filtering algorithm.

Conjecture 10.2. There is a non-empty class of problems that can be solved with Neural-Kalman schemes by expending lower complexity than schemes not incorporating the Kalman filtering algorithm, such as pure (non-Kalman) Neural Networks, do.

An exhaustive mathematical analysis of this conjecture, or a mathematical proof of it, is well beyond the scope of the current Thesis. However, if such problems did exist and weren't very rare in human contexts, we would find, at the very least, circumstancial evidence compatible with the existence of Neural-Kalman problems in neuroscience and human affairs.

Indeed, we have found significant, wide-ranging evidence. See Appendix Sections B.2 (for biological Neural-Kalman schemes in animals) and B.3 (for Neural-Kalman features in human behaviour).

If Neural-Kalman schemes are indeed superior to non-Kalman schemes in a frequently appearing class of problems, then **a Neural-Kalman architecture for ever granular stereotyping** that provides a practical solution for continual learning in the presence of unstereotyped abrupt dynamics **would be extremely useful in communications** and other continual learning tasks [Hadsell et al., 2020]. This Thesis has proposed some ideas to design its architecture in Section 9.4.

Let us finally add that we believe the concept of Neural-Kalman Stochastics might prove useful for communication engineers and other problem-solvers in a first stage of design *even if* this conjecture were to eventually be proven wrong.

9.6.2 Neural-Kalman Market Hypothesis

The future is uncertain and determined by events whose occurrence is unpredictable for most people [Taleb, 2007], though a small amount of experts will detect such abrupt, consequential changes, draw approximately right conclusions before most and start making surprising predictions that real data will confirm.

Most people, including most investors, might notice the new information but will draw the wrong conclusion.

In terms of Neural-Kalman engineering, a few smart experts will have detected a changepoint and they will have either started learning the new paradigm or recalling and applying a similar past paradigm, i.e. they will behave as if a Neural-Network-switched scheme switched their brains into a new model[14].

The huge majority of the people will not make that switch early. The consequence is their investment decisions will be guided by expert opinions that are no longer grounded on reality;

[14]The Spanish-language expression *cambiar el chip* (literally 'change the chip') perfectly describes this behaviour.

expert predictions that surprisingly fail and keep failing; and possibly biosocial dynamics that reinforce the wrong belief in the face of ever increasing evidence to the contrary [Staw, 1976, Staw, 1996, Schaumberg and Wiltermuth, 2014, Kang and Kim, 2022].

If we assume humans have this bias, then we will necessarily conclude that, when those abrupt, consequential changes happen, most market participants will be pricing assets in a suboptimal manner, and the small minority of market participants who have drawn the right conclusions about the change might not have the resources or risk tolerance to immediately correct that suboptimal pricing in the markets.

We can recapitulate this reasoning in the following formal statement:

- **Neural-Kalman Market Hypothesis.** Financial markets are driven by such abrupt, consequential changes that only a small minority of market participants can correctly interpret them early as the abrupt, consequential changes they are.

- **Local-Domain Expert Hypothesis.** The interpretation of such abrupt changes requires expertise in the specific area where they emerge. A minority of talented experts will correctly detect and interpret abrupt changes that have far-reaching consequences beyond that specific area.

  Such abrupt changes will be called **Neural-Kalman Events** and their resulting neurofinancial dynamics will be called **Neural-Kalman Phenomena**.

- **Corollary (abrupt bets against failed experts).** Market participants with increased focus/belief in local-domain experts whose surprising predictions become true will, ceteris paribus, outperform their peers in the long run.

  Outperformance will be maximized through levered positions against failing expert opinions that take advantage of the relatively short time frame of an abrupt change. Such strategy will be called **Highly Leveraged Abrupt Bets Against Failing Expert Opinions**, or **HLABAFEOs.**

An example would be a hypothetical person who started following the right epidemiologists in January 2020, noticed their surprising predictions were coming true and decided to short the airlines early[15]. Once market participants priced in the Covid pandemic and the Western lockdowns by March 2020, that person would have certainly outperformed them.

In the next four Subsections I will consider several potential counterarguments: the Efficient Market Fallacy, high-frequency trading algorithms, lack of empirical evidence, and the argument that supply eventually meets demand for any strategy.

9.6.2.1 Compatibility with the Efficient-Market Hypothesis

The Efficient-Market Hypothesis states that security prices at any time “fully reflect” available information [Fama, 1970]. This has sometimes been misinterpreted into the fallacy that stocks trade at their fair market value on exchanges, generating above-average returns consistently is impossible, beating the market is impossible, etc.

However, the Efficient-Market Hypothesis does not state that. That is a fallacious interpretation [Siegel, 2009] which deserves to be referred to as the Efficient-Market Fallacy.

[15] Or, for maximum outperformance, long deep out-of-the-money puts on airline stocks with short-term expiry (maximum abrupt leverage).

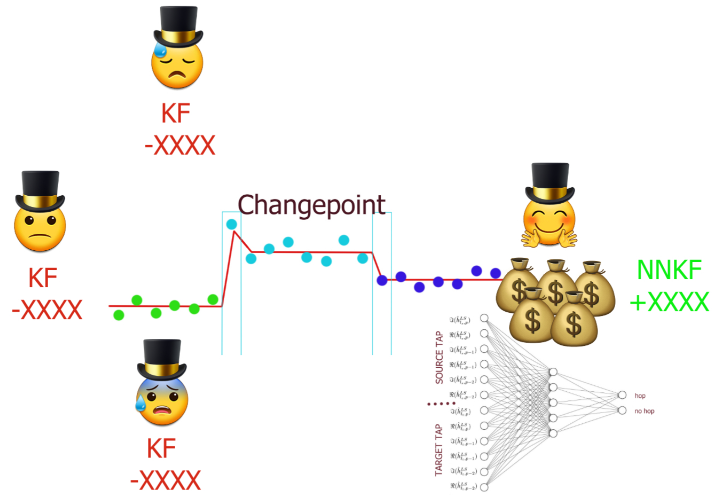

Figure 9.4: Analogy of Efficient Market participants as filtering methods with different performance.

The Efficient-Market Hypothesis only states that investors have incorporated all new available information into the price of the stock, i.e. they have updated their pricing with the new available information. This does not imply that such update is optimal, follows the smartest possible interpretation or cannot be beaten with a smarter interpretation.

Eugene Fama himself [Fama, 2022] recognises two ways to outperform the market: insider trading or being better at interpreting financial conditions than "the market", i.e. all other significant market participants combined. Thus, the Neural-Kalman Market Hypothesis is fully compatible with the Efficient-Market Hypothesis, since any outperformance comes from a superior, scarce, neurocognitively expensive interpretation of the market (better predictions).

In that sense, we could model the fair price of a stock as a channel tracking problem where market participants update their fair price estimate by updating it with a new sample (the new available information). If we assume that all market participants are exclusively using KF models to update that model (e.g. because no better alternatives, such as NNs or NNKFs, have been invented yet), then a new market participant using a superior strategy (e.g. a properly adapted NNKF) would outperform the market while not violating the Efficient-Market Hypothesis, particularly since its initial resources would be very limited in comparison with the market size.

The analogy is far from perfect, but it is no coincidence that the first hedge fund who consistently used neural networks for high-frequency trading (Medallion Fund, a fund established by Renaissance Technologies LLC) has been massively outperforming the market for three decades straight since it was established in 1988, with an average **annual**[16] gross return of **66%** for those 31 pre-Covid years [Zuckerman, 2019].

9.6.2.2 High-Frequency Trading Algorithms

During day-to-day trading, some decisions are automatized by so-called high-frequency trading algorithms. However, the strategic decisions relating to overall risk, such as leverage or exposure to certain assets, are made by humans. Two good examples for this are the decision-making at

[16] To get a grasp of the magnitude of such success, consider the gross compound interest: $1.66^{31} = 6,658,106.37$.

Renaissance Technologies LLC during stock market challenges such as the August 2007 quant crunch or the December 2018 taper tantrum [Zuckerman, 2019]. Thus, human biases are not necessarily reduced by high-frequency trading algorithms in key processes, such as deleveraging.

The Neural-Kalman Hypothesis, and particularly its corollary (HLABAFEOs) are intended for phenomena with completely different timescales, environments, movement range, market bias and decision makers than high-frequency trading algorithms and cannot be reasonably expected to be affected by them.

More generally, the Neural-Kalman Market Hypothesis cannot be expected to be affected by any algorithm that doesn't have complete, human-like neurocognition and complete, human-like decision-making power. In any other case, the human limitations at the core of the Neural-Kalman Market Hypothesis will appear.

9.6.2.3 Empirical Evidence

In the current early phase of research around the Neural-Kalman Market Hypothesis, the empirical evidence supporting this hypothesis is a number of neuroscientific and sociological phenomena (presented in Section 9.7 and Appendix Section B.3) as well as the admittedly anecdotal (but continuously growing) evidence of outperformance in our own HLABAFEO portfolios (Section 9.9) and the related simulations based on such empirical live trading data (Section 9.10).

We expect more empirical evidence for the Neural-Kalman Market Hypothesis to be gathered in future research (see Section 10.2).

9.6.2.4 Chronic Undersupply of HLABAFEO Traders

A counterargument against the Neural-Kalman Market Hypothesis might center on the apparent contradiction that HLABAFEOs traders, if they do indeed significantly outperform, would quickly increase their capital (as well as the popularity of the strategy) and collectively become dominant market participants, at which point the outperformance would disappear or even reverse.

It might be interesting to trace an analogy here with serial unicorn entrepreneurship, i.e. talented entrepreneurs that, after having founded a successful business that grows into a $1bn+ valuation, start other companies that grow into extremely successful businesses, as well.

If creative, talented entrepreneurship were really outperforming uncreative, untalented entrepreneurship, why isn't the market overcrowded with creative, talented entrepreneurs? Why isn't everybody a unicorn founder[17]?

Some of the important factors why not everyone is a unicorn founder likely include: a shortage in natural talent (intelligence, emotional resilience), the impossibility to train people for 'unicorn founding' on scale and the tendency of the few very talented people to exchange money for free time as they grow wealthier.

The same arguments can be applied to HLABAFEOs. There might exist a chronic, irreversible undersupply of talented HLABAFEO traders. Consider the requirements for such a task: you would need to know both about finance and about a wide array of potentially consequential topics[18]; if something with abrupt potential seemed to emerge, you would need to run a learning marathon around that topic, identifying key experts, understanding who makes sense and who doesn't, tracking short-term predictions and directly talking to the local-domain experts with the best

[17] Notice that non-founding human workers are not strictly necessary for this to be true in an era of robotics and AI, e.g. once everyone had launched its own successful automatization start-up.

[18] Ideally, you would also keep a high-value network of contacts in finance and all those consequential areas.

recent prediction track record to make sure you understand the range of probable outcomes in their minds, i.e. you basically need to become better at predicting local-domain outcomes than most local-domain experts, and you should be able to do this overnight (within weeks). Then you will bet a significant part of your net worth (or managed capital) on your recently acquired expertise on a highly-levered (i.e. mostly binary, “10x-or-nothing”) financial trade.

Moreover, this job isn’t divisible, since learning, judgement or guts are not really divisible. You will be the one who needs to learn it all, make the right call and assume the risk. These core tasks cannot be delegated.

If that is the job description, why would we expect an oversupply of talented HLABAFEO traders? Why wouldn’t we expect talented HLABAFEO traders to outperform the market?

9.6.3 Separability of ’business as usual’ vs. ’leap into new model’

For many purposes, including finance, the mathematical conjecture and neurosocial hypothesis are functionally equivalent. The approximate separability of ’business as usual’ vs. ’leap into new model’, well represented as the KF vs. NN dychotomy, might be backed by mathematical means (complexity), neuro-evolutionary arguments (energy efficiency, fitness, biological neural structures) or social dynamics (groupthink, incentives). It is the author’s belief that all three underlying phenomena combine into the emerging neurofinancial structures and events analyzed in this chapter.

9.7 Neurosocial Evidence for Neural-Kalman Phenomena

A review of mental models among successful (abrupt-change) investors and research papers around abrupt-change empirically-backed sociological phenomena has provided the following (non-exhaustive) list of empirical Neural-Kalman-compatible phenomena:

- “Normalcy bias” in emergency research [Ripley, 2009]
- “When Experts Fail” [Graham, 2014]
- Anchoring (cognitive bias) [Tversky and Kahneman, 1974, Yasseri and Reher, 2022]
- “Black Swans” in “Extremistan” [Taleb, 2007, Taleb, 2020]
- “Gray Rhinos” [Wucker, 2016]
- Escalation of commitment to a failing proposition [Staw, 1976, Staw, 1996]
- Mirror neuron system [Keysers and Fadiga, 2009] and mimetic desire

All these empirical phenomena are analyzed in depth in Appendix B.3, including an extensive review of references. For the sake of brevity, I will summarize here the first of them, “normalcy bias” during catastrophes (Fig. 9.5), since all Neural-Kalman phenomena are arguably variations on this same theme: a consequential abrupt change has occured, such as a fire in the plane you are seating on, and only a small minority reacts and runs away. Most people stay on their seats until it is too late and die like badly scripted robots (or maladapted KFs without a supporting NN). It certainly sounds like an exaggeration, but it is not. Catastrophe researchers have found this bias once and again when analyzing behaviour during catastrophes: only a minority flees early to

Figure 9.5: "Normalcy bias" as represented by the "This is fine" Internet meme [Green, 2013]. Reproduced with permission from the author and copyright-holder. © 2013 KC Green.

avoid death [Ripley, 2009, Wucker, 2016, Klein, 1994, Proulx, 2002]! Most of the people need to see others flee in big numbers before they make the decision to flee[19]. Flight is mimetic.

Please notice the inherent contradiction of acknowledging humans are empirically shown to behave with this deadly bias when it matters most (their own lives and the lives of their loved ones) while believing these same humans will behave rationally and with no normalcy bias in financial markets[20].

9.8 Our Proposal: Highly Leveraged Bets Against Failing Expert Opinions (HLABAFEOs)

Our proposal to exploit the Neural-Kalman Phenomena in a market assumed to follow the Neural-Kalman Market Hypothesis is by using Highly Leveraged Bets Against Failing Expert Opinions (HLABAFEOs).

What is a HLABAFEO? In simplified terms, it is *a bet on the fact that models proven to be likely wrong will still be used for a while.* Remember that KFs degraded catastrophically when tracking a channel tap that had different underlying stochastics than assumed, different statistical behaviour than the built-in plant model (see Chapters 3 and 5). If we assume the same thing happens in financial market participants (Neural-Kalman Market Hypothesis), then a most profitable trade would be to bet against such failing models as they start to fail. By using options with high leverage[21], we can maximize the profit while limiting the possible downside to a tolerable loss that would not prevent multi-year compounding. This crucial feature is known as barbell strategy (see Appendix Subsection B.3.4.2) and it is simulated in Section 9.10.

How can you detect an opportunity for a HLABAFEO? Ideally you would like to find a situation

[19] The reverse is also true. When people see others run away, they decide to flee, too (without knowing why everyone is running away); this happened recently when a group of CrossFitters passed by a restaurant [News.com.au, 2022].

[20] The potential argument that you only need a few smart traders to bring the price to its fair price does not consider the fact that changepoints might be associated with periods of high volatility where even making 'the right choice' might be extremely risky (e.g. shorting an extremely overvalued market, even if the bubble turns bust, has the risk of timing the bear market rallies [Schultz, 2003]), so that smart market participants might be forced to abstain (this is particularly true in times of deleveraging, functionally equivalent to fleeing the plane once everyone has woken up to the reality that the fire in the airplane might kill them all, i.e. when fire escape are crowded and flight is slowest and unfeasible for most.

[21] Some highly volatile assets, such as some tech stocks or cryptocurrencies, can be considered functionally equivalent to high leverage for the purposes of the HLABAFEO strategy.

where the following properties apply:

1. Uncertainty: There is uncertainty about the future (i.e. experts are failing, but not 100% failed yet), with some (consensus) expert frameworks predicting completely different short-term values/facts around Neural-Kalman Events than other (fringe) expert frameworks.

2. Highly Consequential Abrupt Change[22]: The occurrence of the values/events produced by these competing expert frameworks lead to abrupt, consequential developments in a specific industry or asset category.

3. Asymmetry: high risk-reward ratio.

4. Cognitive deficits due to social, psychological, biological or emotional issues[23], e.g. a widespread situation in that specific security such that most investors in it have personal reasons to believe the failing expert opinion will eventually be proven right (pandemic fatigue, inability to recognize unrealized losses, national bias, sense of belonging to a financial cult).

5. Uncorrelated (ideally). While the best HLABAFEO opportunities will necessarily be infrequent, a simple strategy based on uncorrelated, simultaneous HLABAFEO streams would reduce the probability of total loss and increase long-term return.

9.9 Real-World HLABAFEO Portfolio Performance

Two HLABAFEO portfolios have been built in real time, based on the Neural-Kalman Market Hypothesis (Section 9.6.2) and the failing expert opinion about inflation explained in Sections 9.1 and 9.2, and have been managed in real financial markets for a total combined duration of 22 months. This must be interpreted as a preliminary proof of concept for HLABAFEO portfolios in real-world conditions.

These are real-world portfolio returns, not backtests.

Exhaustive trading records will be provided by the author upon justified demand. Trades in Portfolio A can also be cryptographically verified (gladly through a cryptographic message by the author, if applicable).

9.9.1 Portfolio A: Liquidity-Driven Asset Price Boom

Live returns for a HLABAFEO Portfolio based on the final parabolic impulse of cryptocurrencies during the liquidity-driven asset price boom in 2021 are shown in Table 9.1. This Table shows the 10 top investments by sales proceeds.

9.9.2 Portfolio B: Liquidity-Driven Asset Price Bust

Live returns for a HLABAFEO Portfolio based on the collapse of growth tech stocks during the liquidity-driven asset price bust in 2022 are shown in Table 9.2. Description follows the format "ticker + expiry date + strike price", i.e. "PUT BLI 04/14/22 12.50" means a put against Berkeley Lights (ticker: $BLI) with expiry date on April 14, 2022, and a strike price of 12.50 US dollars.

[22]See Appendix B.3 for detailed examples of such highly consequential abrupt changes.
[23]See Appendix B.3 for detailed examples of such cognitive deficits.

Name	Symbol	First Date Acquired	Last Date Sold	% Gain/Loss
Terra	LUNA	03/03/2021	01/12/2022	516.30
Bitcoin	BTC	01/04/2021	05/22/2021	28.47
Solana	SOL	02/20/2021	01/30/2022	280.56
Ethereum	ETH	01/04/2021	02/13/2022	31.07
Cardano	ADA	01/05/2021	02/25/2021	305.56
yearn.finance	YFI	01/05/2021	02/23/2021	27.61
Helium	HNT	02/20/2021	01/27/22	166.38
Hathor	HTR	03/19/2021	12/17/21	-6.78
ADADOWN	ADADOWN	03/15/2021	03/16/2021	-40.53
Zcash	ZEC	01/07/2021	02/23/2021	92.00
Total Portfolio Gain	–	–	–	**162.58%**

Table 9.1: HLABAFEO Portfolio A. Live returns during the 2021 liquidity-driven boom. **Not a backtest.**

Description	Days Held (Median)	First Date Acquired	Last Date Sold	% Gain
PUT BLI 04/14/22 12.50	197	09/23/2021	04/14/2022	258.73
PUT TSLA 01/20/23 250	182	04/05/2022	10/04/2022	32.15
PUT AAPL 01/20/23 130	76	04/06/2022	06/21/2022	93.79
PUT BLI 10/21/22 2.50	132	04/11/2022	10/20/2022	49.45
PUT CVNA 01/20/23 85	176	04/11/2022	10/04/2022	274.18
PUT NVDA 01/20/23 157.50	169	04/12/2022	10/04/2022	158.93
PUT META 01/20/23 160	173	04/13/2022	10/04/2022	196.56
PUT AAPL 01/20/23 120	51	08/03/2022	10/07/2022	107.74
Total Portfolio Gain	–	–	–	**175.44%**

Table 9.2: HLABAFEO Portfolio B. Live returns during the 2022 liquidity-driven bust. **Not a backtest.**

9.9.3 Investing style and tools

The basic investing strategy can be summarized as follows:

1. Detect KF-like behaviour in market participants, such as underlying prediction models that do not get discarded in the face of abrupt changes in recent real-world data. An example of such KF-like behaviour in the face of persistent inflation was provided in Fig. 9.1.

2. Bet against such underlying prediction model. Properly sizing the bet is extremely important; see relevant simulations and discussion in Section 9.10.

3. Profit and compound.

The following quantitative and qualitative tools have been used in the implementation of HLABAFEO Portfolios A and B: macroeconomic indicators, including leading indicators such as surveys; public expert opinions, including non-financial-domain expert opinions; conversations with industry-specific experts (such as investors, CEOs and analysts with excellent track record); composite indexes to aggregate expert opinions from different domains; as well as patterns and models implicit or explicit in successful experts' opinions, and their comparison with leading or other incoming data.

9.10 Simulation: HLABAFEO Sizing and Median Long-Term Returns

HLABAFEOs, as highly leveraged investments, are quasi-binary in nature: in each cycle you get either a large gain or a total loss (e.g. your stock options or *altcoins* become worthless). Therefore, a size bet of 100% of your net worth/investment portfolio eventually leads to the complete loss of capital.

The goal of this simulation is to determine what HLABAFEO bet sizes produce optimal median long-term returns. The following (percentage) profit/loss (PnL) model is considered:

$$PnL = \begin{cases} -100 & with\, probability\, P_L \\ r_n & with\, probability\, 1 - P_L \end{cases} \tag{9.6}$$

where P_L is the probability of total loss and

$$r_n \sim \mathcal{N}(\mu_H, \sigma_H) \tag{9.7}$$

where the average profit μ_H in case of no loss and its variance σ_H are assumed, for the purposes of this simulation, to be fully consistent with the real-world HLABAFEO portfolio returns in Section 9.9, i.e. $\mu_H = (175.44 + 162.58)/2 = 169.01$, etc.

9.10.1 Simulation results

Median long-term returns were simulated over 30 HLABAFEO cycles for different P_L values, as well as different bet sizes and luck percentiles. Each scenario was simulated 1001 times, then the median (P50) and percentiles 3 (P3) and 97 (P97) were obtained. Tables 9.3 and 9.4 show long-term performance as final net worth for different bet sizes (as percentage of net worth) and performance percentiles.

Bet size	P3	**P50**	P97
8%	66	191	557
12%	48	234	1143
16%	33	265	2205
20%	21	**276**	3687
24%	12	269	5918
28%	6	242	9032

Table 9.3: HLABAFEO long-term performance as final net worth (initial = 100) for different bet sizes and luck percentiles, assuming total loss in 50% of cycles.

Bet size	P3	**P50**	P97
8%	84	249	800
12%	68	335	1972
16%	53	554	3442
20%	37	635	9396
24%	24	768	13133
28%	14	**920**	21443
32%	8	549	56771

Table 9.4: HLABAFEO long-term (30-cycle) performance as final net worth (initial = 100) for different bet sizes and luck percentiles, assuming total loss in 45% of cycles.

Assuming total loss in 50% of cycles, the HLABAFEO strategy is still profitable (176% growth in net worth) over 30 cycles with a maximal drawdown of 20% (total loss of bet size) per cycle. The respective trajectories for P3, P50 and P97 are shown in Fig. 9.6.

Obviously, the strategy is not risk-free. However, the median trajectory is profitable. The overall risk can be reduced (i.e. the distribution of outcomes can be made more uniform) by reducing bet size.

On the other hand, a low bet size would greatly reduce the profit in favourable outcomes if the probability of total loss happened to be lower than 50%. Assuming total loss in just 45% of cycles, the HLABAFEO strategy is far more profitable (794% growth in net worth) over 30 cycles with a maximal drawdown of 28% (total loss of bet size) per cycle. The respective trajectories for P3, P50 and P97 are shown in Fig. 9.6.

Thus, bet size presents a trade-off between lower risk (more uniformity in outcomes) and higher median wealth growth.

9.10.2 Limitations

Taxes are not considered in this analysis due to the wide variety of diverging tax regulations around the world. For example, trades in HLABAFEO Portfolio A (Subsection 9.9.1) would be subject to short-term capital gains tax in the US (50% tax rate), while those same trades would be tax-free in Portugal [Selkis, 2020].

Changes in the EURUSD exchange rates are not considered. If portfolio gains were to be computed in euros, this would lead to a significant increase in reported profit for HLABAFEO Portfolio B (Subsection 9.9.2).

Nothing in this chapter should be construed as financial advice. The Neural-Kalman Market Hypothesis is just an unproven hypothesis as of yet. This analysis considers what the optimal bet size would be, in terms of long-term median wealth generation, provided that the Neural-Kalman

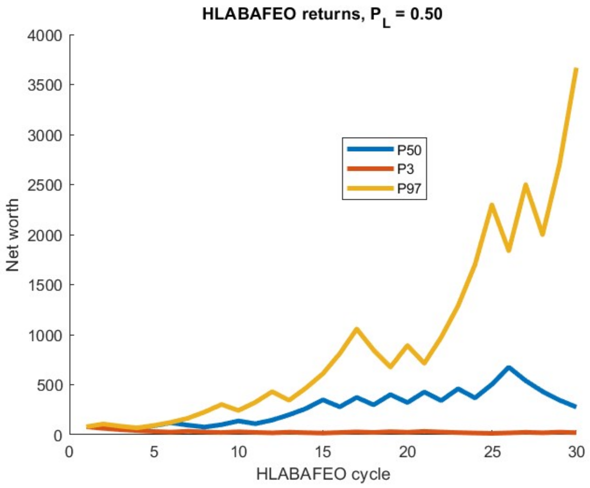

Figure 9.6: HLABAFEO returns for different performance percentiles, $P_L = 0.50$.

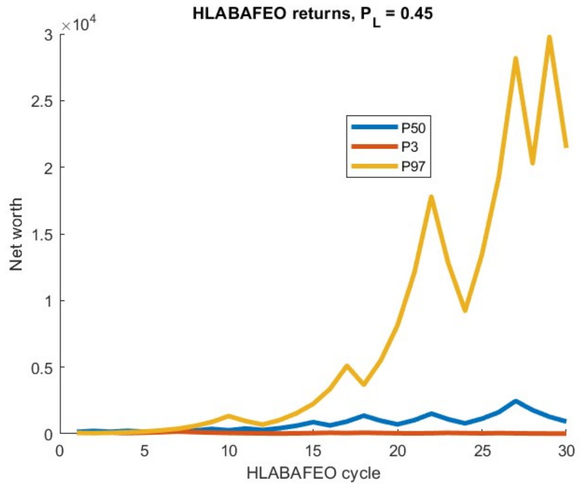

Figure 9.7: HLABAFEO returns for different performance percentiles, $P_L = 0.45$.

Market Hypothesis held true and the the model based on empirical HLABAFEO returns could realistically be assumed. Past performance is no guarantee of future results. In particular, an argument could be made that the assumed probability of loss could be exceedingly high; if lower odds are assumed, optimal bet size would increase.

No risk mitigation strategies are assumed. A simple strategy based on uncorrelated, simultaneous HLABAFEO streams would reduce the probability of total loss and increase long-term returns; accordingly, optimal bet size would increase. However, finding uncorrelated HLABAFEO streams might not always be possible or might require an unfeasible cognitive effort.

High-volatility bets are reported to increase stress and lead to bad, emotionally-induced decisions, such as panic selling or not following your strategy consistently (e.g. not cutting your losses according to your pre-loss plan) [Schultz, 2003, Taleb, 2007]. In this Thesis' author's experienced opinion, the emotional and cognitive effort demanded by HLABAFEO strategies is hard to overstate; they are major hindrances for the practical, successful implementation of any HLABAFEO strategy[24].

9.11 Main Takeaways for Communications Engineering

Topics in neuroscience are relevant for communications engineering; in particular, the Changepoint-Oddball-Reversal problem studied by neuroscientists is functionally equivalent to the channel tracking problem under abrupt changes central to this Thesis. The SMAP estimator can be reinterpreted in light of neuroscience as a compound detector for changepoints, oddballs and reversals.

Furthermore, by drawing explicit parallelisms between neuroscientific findings around biological neurocognition, such as the Changepoint-Oddball-Reversal problem, and the channel tracking problem under abrupt changes in communications engineering, we have suggested some neuroscientifically backed, bio-inspired mechanisms that could potentially improve channel tracking and learning under unstereotyped dynamics.

In particular, this Thesis has formed the Neural-Kalman Problem Conjecture and has suggested that a Neural-Kalman architecture for ever granular stereotyping providing a practical solution for continual learning in the presence of unstereotyped abrupt dynamics would be extremely useful in communications and other continual learning tasks [Hadsell et al., 2020].

Neuroscience, and particularly neurofinance, have similar problems to channel tracking under abrupt dynamics: mechanisms eventually elucidated in neuroscientific research should be a part of future research in communications engineering.

9.12 Conclusions

A novel Neural-Kalman framework has been proposed to understand how market participants react to failing expert opinions and how to outperform them through a new investing strategy, Highly Leveraged Abrupt Bets Against Failing Experts (HLABAFEOs), under the assumptions of our newly stated Neural-Kalman Market Hypothesis.

The inflation forecasting problem has been illustrated with a hopping LGM model and we have provided a novel structure for a (partly Bayesian) stochastic stereotyper that might eventually solve such problems through bio-inspired, neuroscientifically-backed mechanisms, like dreaming

[24]That explains why HLABAFEO investing is not crowded; otherwise it wouldn't be so profitable (as argued in Subsection 9.6.2.4).

and surprise. This Chapter has suggested mathematical (Neural-Kalman Problem Category Conjecture), neuroscientific and social reasons why Neural-Kalman Phenomena might exist and it has found strong evidence for their existence in the neuroscientific and financial literature. Furthermore, this research has significant consequences for Communications Engineering, as explained in Section 9.11.

Finally, we have provided specific examples, practical guidelines and historical performance for some HLABAFEO investing portfolios. Though this research should be considered in its early stages, the bulk of reviewed evidence is consistent with our proposed hypotheses and investment strategies.

Chapter 10

Conclusions and Further Work

10.1 Summary and Conclusions

Three basic ideas are supported by this Thesis:

1. Some mechanisms work great when there are no abrupt changes. The **Kalman filter is** one of them (**great for channel tracking**!). These mechanisms degrade catastrophically when there is an abrupt change and you do not detect it or do not adapt accordingly.

2. You can mitigate this issue by **detecting abrupt changes** with other schemes (such as our proposed SMAP or our proposed neural detectors, which produce significantly **improved performance** in channel tracking). By combining these detectors with Kalman filters (e.g. the NNKF scheme), we get its advantages while mitigating degradation in the face of abrupt changes.

3. Similar **Neural-Kalman dynamics** can be identified as relevant in communications, as well as one's life or finances. We introduced a long list of examples and specific frameworks **to profit** from them, such as lateral subtap hopping (in channel tracking) or Highly Leveraged Abrupt Bets Against Failed Expert Opinions (HLABAFEOs) in finance.

In other words: we have proposed schemes such as the SMAP, NNKF and HLABAFEOs and we have shown they can outperform competing schemes under specific conditions.

Figure 10.1 provides a mind-map summary for this Thesis. The components in this summary are explained in the following section.

10.1.1 A Mind Map for this Thesis

10.1.1.1 Abrupt changes

The main topic of this Thesis is **abrupt changes** and how to deal with them, particularly in channel tracking. Our channel literature review (Section 2.6) concluded that there are empirical models of the birth/death of taps, typically first-order Markov models. These abrupt changes are not rare: the dynamics of birth/death in intervehicular communications are very important and they have been demonstrated in different measurement campaigns (in fact, they have a significantly higher intensity than that assumed in our simulations). We have expanded the repertoire of abrupt

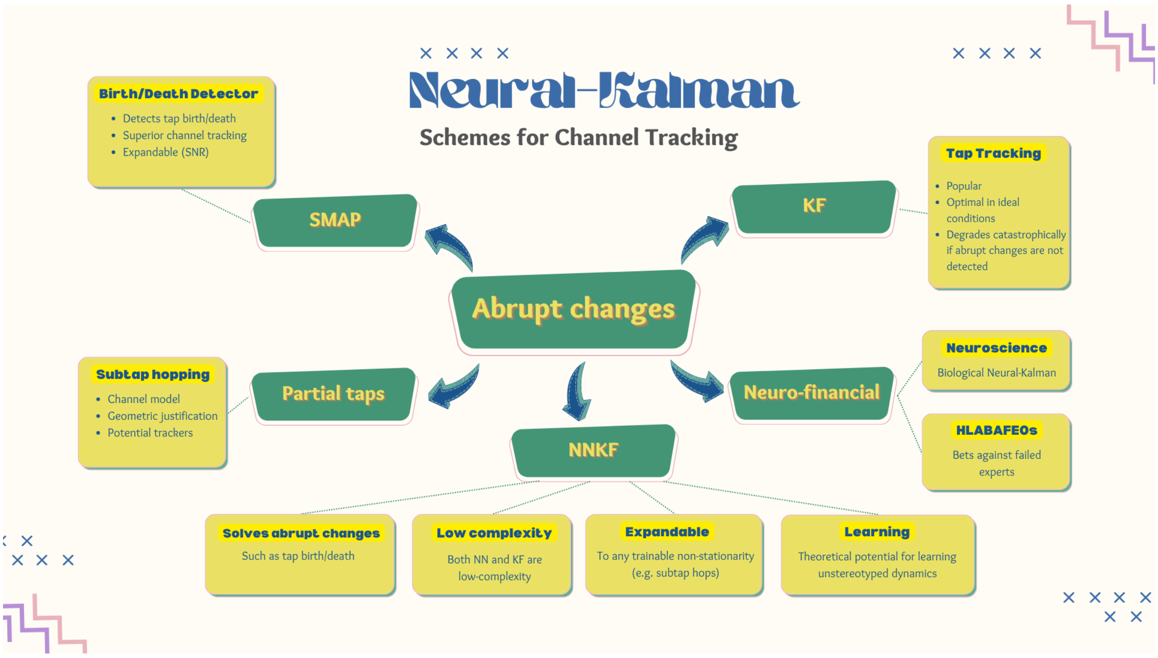

Figure 10.1: Neural-Kalman Schemes for Channel Tracking: a PhD Thesis.

change models in the literature by proposing new geometrically-justified **abrupt changes**, such as lateral partial tap hopping (Subsection 8.1.1).

10.1.1.2 KF (Kalman Filtering)

Abrupt changes makes **KF**'s performance **degrade catastrophically** when tracking channel taps, as our simulations have shown. That is a real setback, since the application of **KF** to OFDM systems can improve channel estimation and reduce BER. Our KF literature review (Section 3.5) has found many proposals using **KF** for channel estimation, typically applying **KF** to the monitoring of the temporal variation of the subchannels. **KFs** are **popular** because they are **optimal** estimators **under ideal conditions**. This leads to the question: what would happen if we could detect such abrupt changes with low computational cost?

10.1.1.3 SMAP (Simplified Maximum A Posteriori)

After reviewing the technical literature about abrupt change detection, we found a simple algorithm was needed. An algorithm that could detect tap birth/death dynamics with a lower computational cost than previous proposals, such as Rao-Blackwellized Particle Filters under the Random Set Theory model. We derived the Simplified Maximum A Posteriori (**SMAP**) **tap birth/death detector**, a computationally inexpensive, threshold-based estimator that **reduces channel tracking error** in combination with KF (Chapter 5).

Moreover, the abrupt change detection stage with a SMAP core is **expandable**. We have shown how to expand abrupt change detection with new features, such as abrupt changes in SNR, by building a new RST framework for combined death/birth and SNR detection (Chapter 6). Our simulations suggest that **SMAP**, while being surprisingly robust to SNR drift, benefits from an accurate **SNR detection**.

10.1.1.4 Partial taps

Are there any other abrupt changes, apart from tap birth/death, that we should detect to improve channel tracking performance? Yes. We developed a novel **channel model** for lateral partial-tap hop dynamics (**subtap hopping**), i.e. instead of considering just the death/birth of the whole tap, we considered a single tap could be interpreted as the sum of smaller components, called subtaps or partial taps, that could hop to adjacent positions. We have provided a geometric justification for this and additional non-stationary dynamics (Subsection 8.1.1). Finally, we have suggested some **potential trackers**, including a neural detector for lateral hops with large energy transfers.

10.1.1.5 NNKF (Neural-Network-switched Kalman Filters)

We later created a neural abrupt change detector that could be expanded to any trainable non-stationarity, i.e. not just tap birth/death, but also more complex non-stationary models, such as **subtaps hopping laterally** to adjacent taps or other dynamics we might encounter in the future. We found a way to combine trainable neural networks with Kalman filters: the proposed **low-complexity** NN-switched KF trackers (**NNKF**) outperform all previously known multipath channel tracking systems for OFDM communications, provided that tap birth/death phenomena are present (Section 7.2). Moreover, its performance in terms of CTMSE is identical to that of the ideal case (ISS) where perfect knowledge of tap activations is available. Furthermore, **NNKF** schemes are not just **expandable to any** known **trainable non-stationarity**, but the combination of NNKF with other bio-inspired computing mechanisms also gives them the **theoretical potential for learning unstereotyped dynamics**, as we proposed in Section 9.4.

10.1.1.6 Neuro-financial

A novel Neural-Kalman framework has been proposed to understand how market participants react to failing expert opinions and how to outperform them through a new investing strategy, Highly Leveraged Abrupt **Bets Against Failing Experts** (**HLABAFEOs**), under the assumptions of our newly stated Neural-Kalman Market Hypothesis.

We have illustrated the inflation forecasting problem with a hopping LGM model and we have provided a novel structure for a Bayesian stereotyper that might eventually solve such problems through bio-inspired, neuroscientifically-backed mechanisms, like dreaming and surprise (**biological Neural-Kalman**). We have suggested mathematical (Neural-Kalman Problem Category Conjecture), neuro-evolutionary and social reasons why Neural-Kalman Phenomena might exist and we have found strong evidence for their existence in the areas of **neuroscience** and finance.

Finally, we have provided specific examples, practical guidelines and historical performance for some **HLABAFEO** investing portfolios.

10.2 Future research

The following ideas for future research should be considered.

1. Applying our proposed NNKF scheme to real environments, such as those reviewed in 2.6, e.g. a V2X Weibull channel with birth-death dynamics based on a road channel measurement campaign [Hassan et al., 2020].

2. Comparing the performance trade-off (channel tracking improvement vs. computational complexity) via simulations of the different available tracking schemes for partial taps (see Section 8.3).

3. Empirical characterization for the prevalence and behaviour of partial tap hops (and similar phenomena) in measurement campaigns.

4. Applying our proposed NNKF to real-world tasks such as those related to 6G and recent advancements in smart mobility [Noor-A-Rahim et al., 2022].

5. Incorporating eventual advancements in KFs, NNs, NNKF and other Neural-Kalman and particle optimisation schemes into the models presented in this PhD Thesis, particularly in the context of unstereotyped abrupt dynamics (Section 9.4).

6. Delve into the neurofinancial research of the HLABAFEOs through a compilation of micro and macroeconomic models and historical examples that can be interpreted within the framework of Neural-Kalman dynamics and HLABAFEOs.

Appendix A

Derivation of GMAP Estimators in an RST Model

A.1 RST Model for Complex Tap Gains

Let $\mathcal{H}_p^{(k)}$ denote the following random set, which can be formed by a single element or an empty set [Goodman et al., 2013, Angelosante et al., 2007]:

$$\mathcal{H}_p^{(k)} = \begin{cases} \{\emptyset\} & \text{if tap } k \text{ is absent} \\ \left\{\mathbf{h}_p^{(k)}\right\} = \{[k, a_p^{(k)}]^T\} & \text{if tap } k \text{ is present} \end{cases} \tag{A.1}$$

where $a_p^{(k)}$ is the complex gain for the kth tap at time p. All taps can be described by the union set

$$\mathcal{H}_p = \bigcup_{k=1}^{L_{máx}} \mathcal{H}_p^{(k)} \tag{A.2}$$

which is a random set in the hybrid space $\{1, ..., L_{máx}\} \times \mathbb{C}$. A hybrid space is the conceptual equivalent to a product space when, instead of metric spaces, what we have are spaces of random sets. For the reader interested in understanding the details about the differences between the concepts associated with metric spaces (measures, integrals, probability, etc.) and the dual concepts generalized to any space of random sets, it is strongly recommended (not without warning of its complexity) to consult the explanations in [Goodman et al., 2013, Nguyen, 2006] and, above all, [Vihola, 2004].

Let us define the projection random sets $\pi(\mathcal{H}_p)$ y $\pi'(\mathcal{H}_p)$, which describe the projections of $\mathcal{H}_p$ over $\{1, ..., L_{máx}\}$ and over $\mathbb{C}$, respectively:

$$\pi(\mathcal{H}_p) = \bigcup_{k:\mathcal{H}_p^{(k)} \neq \emptyset} \{k\} \tag{A.3}$$

$$\pi'(\mathcal{H}_p) = \bigcup_{k \in \pi(\mathcal{H}_p^{(k)})} \{a_p^{(k)}\} \tag{A.4}$$

If $\mathcal{S}_p$ denotes the set of surviving taps, i.e. taps who survive in an active state from time $p-1$

to time p, and $\mathcal{B}_p$ is the set of newly born taps, then:

$$\mathcal{H}_p = \mathcal{S}_p \cup \mathcal{B}_p \tag{A.5}$$

The restrictions on (A.3) and (A.4) as follows:

$$\pi(\mathcal{H}_{p-1}) \cap \pi(\mathcal{B}_p) = \emptyset \tag{A.6}$$

$$\pi(\mathcal{S}_p) \subseteq \pi(\mathcal{H}_{p-1}) \tag{A.7}$$

reflect the fact that no component that is active at time $p-1$ can transfer to the set of new taps, and surviving taps at time p are a subset of active taps at time $p-1$. To simplify the derivation, it might be assumed that only one tap can be born at each epoch. This simplification is made, e.g. in [Ma et al., 2006, Angelosante et al., 2007, Angelosante et al., 2009]; in our case, that implies defining the set of newly born taps as:

$$\mathcal{B}_p = \begin{cases} \{[l, a_p^{(l)}]^T\} & \text{with probability } P_{birth} \\ \emptyset & \text{with probability } 1 - P_{birth} \end{cases} \tag{A.8}$$

where $l \in \{1, ..., L_{máx}\} \backslash \pi(\mathcal{H}_{p-1})$, and P_{birth} is the probability of a new tap being born. By applying such information, the conditional probability density can be calculated:

$$f_{\mathcal{B}_p|\mathcal{H}_{p-1}}(\mathcal{B}_p|\mathcal{H}_{p-1}) = \begin{cases} P_{birth} f_{a_p^{(l)}}(a_p^{(l)}) & \text{if } \mathcal{B}_p = \{[l, a_p^{(l)}]^T\} \\ 1 - P_{birth} & \text{if } \mathcal{B}_p = \emptyset \\ 0 & \text{if } |\mathcal{B}_p| > 1 \end{cases} \tag{A.9}$$

where $l \in \{1, ..., L_{máx}\} \backslash \pi(\mathcal{H}_{p-1})$, and $f_{a_p^{(l)}}(a_p^{(l)})$ is the probability density function for the lth tap gain at time p. Similarly, we can obtain the union set of surviving taps:

$$\mathcal{S}_p = \bigcup_k \mathcal{S}_p^{(k)} \tag{A.10}$$

where

$$\mathcal{S}_p^{(k)} = \begin{cases} \emptyset & \text{with probability } P_{death} \\ \{\mathbf{h}_p^{(k)}\} & \text{with probability } 1 - P_{death} \end{cases} \tag{A.11}$$

where P_{death} is the probability that an active path will disappear (probability of death).

A.2 Properties of the RST Model for Tap Gains

Assuming that the different paths survive or die independently with respect to each other[1], we have the following properties:

- The conditional probability density function for random set $\mathcal{S}_p$ given $\mathcal{H}_{p-1}$ can be derived from the generalized convolution of probability density functions for random sets $S_p^{(k)}$

[1] Notice such independence condition holds true for the channel tracking problem in Chapters 5, 6 and 7, but not for the lateral hopping problem introduced in Chapter 8.

[Ma et al., 2006, Biglieri and Lops, 2006, Biglieri et al., 2012, Angelosante et al., 2007]:

$$f_{\mathcal{S}_p|\mathcal{H}_{p-1}}(\mathcal{S}_p|\mathcal{H}_{p-1}) = P_{death}^{|\mathcal{H}_{p-1}|-|\mathcal{S}_p|}(1-P_{death})^{|\mathcal{S}_p|}\prod_{l\in\pi(S_p)} f_{a_p^{(l)}|a_{p-1}^{(l)}}(a_p^{(l)}|a_{p-1}^{(l)}) \tag{A.12}$$

where $\mathcal{S}_p \subseteq \mathcal{H}_{p-1}$, and $f_{a_p^{(l)}|a_{p-1}^{(l)}}(a_p^{(l)}|a_{p-1}^{(l)})$ is the transition density describing the evolution of surviving tap gains.

- Random set sequences $\mathcal{S}_p$ and $\mathcal{B}_p$ are conditionally independent given $\mathcal{H}_{p-1}$.
- $(\mathcal{H}_p)_{p=1}^{\infty}$ forms a Markov sequence.

Therefore, the transition density $f_{\mathcal{H}_p|\mathcal{H}_{p-1}}(\mathcal{H}_p|\mathcal{H}_{p-1})$ can be determined through the generalized convolution formula, which, when specified for the current scenario, produces the following result [Biglieri et al., 2012]:

$$f_{\mathcal{H}_p|\mathcal{H}_{p-1}}(\mathcal{H}_p|\mathcal{H}_{p-1}) = f_{\mathcal{S}_p}(\mathcal{H}_p\cap\mathcal{H}_{p-1})f_{\mathcal{B}_p}(\mathcal{H}_p\backslash(\mathcal{H}_p\cap\mathcal{H}_{p-1})|\mathcal{H}_{p-1}) \tag{A.13}$$

The basic step to obtain causal estimates of the sequence of random sets $(\mathcal{H}_p)_{p=1}^{\infty}$ based on the observations $\mathbf{y}_{1:p}$ is the implementation of Bayesian recursions [Mahler, 2003] as follows:

$$f_{\mathcal{H}_p|\mathbf{y}_{1:p-1}}(\mathcal{H}_p|\mathbf{y}_{1:p-1}) = \int f_{\mathcal{H}_p|\mathcal{H}_{p-1}}(\mathcal{H}_p|\mathcal{H}_{p-1})f_{\mathcal{H}_{p-1}|\mathbf{y}_{1:p-1}}(\mathcal{H}_{p-1}|\mathbf{y}_{1:p-1})\delta\mathcal{H}_{p-1} \tag{A.14}$$

$$f_{\mathcal{H}_p|\mathbf{y}_{1:p}}(\mathcal{H}_p|\mathbf{y}_{1:p}) \propto f_{\mathbf{y}_p|\mathcal{H}_p}(\mathbf{y}_p|\mathcal{H}_p)f_{\mathcal{H}_p|\mathbf{y}_{1:p-1}}(\mathcal{H}_p|\mathbf{y}_{1:p-1}) \tag{A.15}$$

At this point, we must remember that we are working with generalized random set spaces, and not with conventional metric spaces. For this reason, the integrals are not the conventional ones (they are not integrals over conventional metric spaces) and the differential δH_{p-1} of (A.14) emphasizes the fact that it is a set integral, like the one presented in [Goodman et al., 2013]. For more details on this, it is recommended to read the explanation on set integrals included in [Vihola, 2004].

Can integrals (A.14)-(A.15) be resolved? In general, it does not seem feasible to obtain explicit solutions (in the sense of closed expressions) for these integrals, despite the great simplification implied by the starting parameters (the disappearance and birth of the channels could be gradual, etc.) and despite other simplifications that might be made (such as simulating the temporal variation of active taps as LGM models). We're now facing the biggest stumbling block of this development: the inability to obtain an optimal solution in an explicit and computable way without a high complexity.

A.3 Rao-Blackwellization and Finite Random Set Particle Filters

The solution proposed in [Angelosante et al., 2007] resorts to so-called "particle filtering" or Sequential Monte Carlo (SMC) methods to approximate Bayesian recursions [Arulampalam et al., 2002] (see also Subsection 4.4.1). This finite random set SMC filter is described below. The posterior probability density function is approximated by a set of particles as:

$$f_{\mathcal{H}_p|\mathbf{y}_{1:p}}(\mathcal{H}_p|\mathbf{y}_{1:p}) \approx \sum_{i=1}^{M} \omega_p^{(i)} m_{\mathcal{H}_p}(\mathcal{H}_p^{(i)}) \tag{A.16}$$

where $m_{\mathcal{X}}(\mathcal{Y})$ is the "0-1" measure, defined as follows:

$$\int_C m_{\mathcal{X}}(\mathcal{Y})\delta\mathcal{X} = \begin{cases} 1, & \text{if } Y \subseteq C \\ 0, & \text{otherwise} \end{cases} \tag{A.17}$$

In Eq. A.16, $\mathcal{H}_p^{(i)}$ is the "particle" of the ith set, $\omega_p^{(i)}$ is its "weight" or weighting factor, and M is the total number of particles. The asymptotic convergence properties of the SMC filter on finite random sets have been proved in [Vo et al., 2005], where the authors showed that, for a sufficiently large M, the mean square error of approximation of the SMC filter on finite random sets is inversely proportional to M^{α}, for some constant $0 < \alpha \leq 1$, while the complexity of the implementation is roughly linear with M.

Once the posterior density $f_{\mathcal{H}_p|\mathbf{y}_{1:p}}(\mathcal{H}_p|\mathbf{y}_{1:p})$ is obtained, there are several ways to obtain an estimation for $\mathcal{H}_p$, as explained in [Goodman et al., 2013]. If we follow the choices made in [Angelosante et al., 2007] (whose derivation this Appendix is based on), we can define two Bayesian estimators, known as GMAP-I (or "Marginal Multi-Target Estimator") and GMAP-II (or "Joint Multi-Target Estimator"). GMAP-I is a two-stage estimator: in the first stage, only the cardinality of the set (how many elements the set has) is estimated.

With the following definition:

$$f_{n_p|\mathbf{y}_{1:p}}(n_p|\mathbf{y}_{1:p}) \triangleq \int_{|\mathcal{H}_p|=n_p} f_{\mathcal{H}_p|\mathbf{y}_{1:p}}(\mathcal{H}_p|\mathbf{y}_{1:p})\delta\mathcal{H}_p \tag{A.18}$$

the following GMAP estimators can be obtained:

$$\text{GMAP-I:} \begin{cases} \hat{n}_p = \arg\max_{n_p \in 0,\ldots,L_{m\acute{a}x}} f_{n_p|\mathbf{y}_{1:p}}(n_p, \mathbf{y}_{1:p}), \\ \widehat{H}_p = \arg\max_{\mathcal{H}_p:|\mathcal{H}_p|=\hat{n}_p} f_{\mathcal{H}_p|\mathbf{y}_{1:p}}(\mathcal{H}_p, \mathbf{y}_{1:p}), \end{cases} \tag{A.19}$$

Au contraire, GMAP-II performs a single-stage estimation:

$$\widehat{\widehat{\mathcal{H}}}_p = \arg\max_{\mathcal{H}_p} f_{\mathcal{H}_p|\mathbf{y}_{1:p}}(\mathcal{H}_p|\mathbf{y}_{1:p}) \frac{c^{|\mathcal{H}_p|}}{|\mathcal{H}_p|!} \tag{A.20}$$

where c is a small constant determined by the cost function this estimator intends to minimize [Goodman et al., 2013]. (Clarification: remember again that $|\mathcal{H}_p|$ represents the cardinality of $\mathcal{H}_p$, that is, the number of active taps at time p).

In addition, a third estimation rule is proposed in [Angelosante et al., 2007] that outperforms the two previous ones in simulations [Angelosante et al., 2007, Angelosante et al., 2009]. It is based on first estimating the identities of the active taps at time p (e.g., instead of estimating that there are two active taps, as in GMAP-I, it estimates that the active tap indexes are 1 and 3), and then estimating only the weights of the active taps as the expected value *a posteriori*, while setting the weights for the inactive paths to zero. Formally, the authors defined the GMAP-III estimator as:

$$\text{GMAP-III}: \begin{cases} \widehat{\pi(\mathcal{H}_p)} = \arg\,\text{máx}_{\pi(\mathcal{H}_p)} f_{\pi(\mathcal{H}_p)|\mathcal{Y}_{1:p}}(\pi(\mathcal{H}_p)|\mathcal{Y}_{1:p}), \\ \widetilde{\mathbf{h}}_p = \int_{\mathbb{R}^{2|\widehat{\pi(\mathcal{H}_p)}|}} \mathbf{h}_p f_{\mathbf{h}_p|\mathcal{Y}_{1:p}}(\mathbf{h}_p|\mathcal{Y}_{1:p}) d\mathbf{h}_p \end{cases} \tag{A.21}$$

where

$$f_{\pi(\mathcal{H}_p)|\mathcal{Y}_{1:p}}(\pi(\mathcal{H}_p)|\mathcal{Y}_{1:p}) = \int_{\pi'(\mathcal{H}_p)} f(\mathcal{H}_p|\mathcal{Y}_{1:p})\delta\mathcal{H}_p \tag{A.22}$$

This last integral can be explained intuitively as an integration (addition) of the probability of $\mathcal{H}_p$ (that is, adding up all hypothetical sets of a certain number of active paths) over all its possible continuous values. Therefore, in GMAP-III the discrete parameter is first estimated (in a similar way to GMAP-I) and then the standard estimate of the posterior expectation is calculated.

Appendix B

Additional Evidence for Neural-Kalman Phenomena

B.1 Complexity: Neural vs. Kalman

B.1.1 Could Neural-Only or Kalman-Only Solutions Solve Neural-Kalman Problems?

Yes, approximately, but not necessarily optimally (in terms of either complexity or performance). A KF can be implemented in a NN [Denève et al., 2007, Wilson and Finkel, 2009, Millidge et al., 2021] (see Appendix Subsection B.2.3). A neural representation of a KF can encode NNKF-like cognition (see Appendix Subsection B.2.1). A complex enough network of KFs might encode some abrupt-change-related cognition.

Thus, it is a matter of optimality, not feasibility.

B.1.2 Computational Cost in the Neural-Kalman Problem Category Conjecture

B.1.2.1 Computational Cost of Addition vs. Multiplication

'Performing an operation' and 'performing an operation optimally' are two different things. Addition performs addition, but addition doesn't always perform addition *optimally* if memory is available, e.g. the computation of 333 x 444 = 147,852 can be performed by adding the number 444 once and again, 333 times, or it can be optimally solved through multiplication.

There is a trade-off between memory (frontloaded, memorized operations, such as multiplications tables) and computational cost/implementation complexity.

B.1.2.2 Computational Cost of Kalman Filtering vs. Neural Networks

Similarly to the addition-vs-multiplication example above, we can find a memory-complexity trade-off in Kalman vs. non-Kalman schemes.

The Neural-Kalman Problem Category Conjecture (see Section 9.6.1) refers to Kalman schemes being computationally optimal solutions in a class of problems (Kalman Problem Category), when only Kalman and (non-Kalman) Neural schemes are considered, for a specific memory requirement.

The relative (dis)advantage in computational cost of KF vs. NNs depends on the specific architecture and application. A first reasonable approach to the problem of comparing the general computational advantage of KF vs. NNKF in frequently occurring tasks might be neuroscientific: what are brains doing? Do they encode Kalman and Neural-Kalman Schemes?

Appendix Section B.2 attempts to answer.

B.2 Neuroscientific Evidence for Biological Neural-Kalman Schemes

Neuroscientists noticed that the brain is capable of approximating Bayesian inference in the face of noisy input stimuli, but the neural underpinnings of this computation are still poorly understood [Millidge et al., 2021].

[Wilson and Finkel, 2009] introduced a novel, bio-inspired neural network that this Appendix Section reinterprets as a biological model for Neural-Kalman schemes.

B.2.1 Biological Encoding of NNKF-Like Cognition

The scheme in [Wilson and Finkel, 2009] is derived from a line attractor architecture [Zhang, 1996] (a model for the spatial orientation in mice thanks to biological head-direction cells), whose dynamics map directly onto those of the KF in the limit of small prediction error. When the prediction error is large, this network was shown to respond robustly to changepoints, in a way that is qualitatively compatible with the optimal Bayesian model, i.e., when confronted with a changepoint, the network no longer approximates a KF and instead maintains two competing hypotheses in a way that is qualitatively similar to that of the run-length distribution in [Adams and Mackay, 2007].

In other words: KF-like behaviour when no changepoint is detected, but NNKF-like behaviour when a changepoint is detected.

B.2.2 Neural-Kalman Stochastics as a Model for Biological Cognitive Dynamics

This reinterpretation suggests the bio-inspired architecture mentioned above might explain how the brain encodes Neural-Kalman-like schemes.

Importantly, [Wilson and Finkel, 2009] doesn't prove it does nor does it prove that it mathematically can beyond unidimensional cases or, more importantly, that such a scheme is optimal (i.e., with lower complexity than a pure, lowest-complexity NN changepoint detector connected to a pure KF scheme).

B.2.3 A Biological Gradient-Descent Approximation to the Kalman Filter

> Neurons that fire together, wire together.
>
> – Donald Hebb

Can neurons encode a "pure" Kalman filter? That is plausible if we could prove that KF can be approximated via gradient descent while respecting the Hebbian rule (namely, that each time the stimulus is repeated, the connections between learning neurons grow stronger and stronger).

[Millidge et al., 2021] proved that a gradient-descent approximation to the KF requires only local computations with variance weighted prediction errors. Moreover, it showed that it is possible under the same scheme to adaptively learn the dynamics model with a learning rule that corresponds directly to Hebbian plasticity, and provided a neural implementation of the required equations.

Thus, from a mathematical point of view, it is not unfeasible for current mathematical models of biological neural networks to act as composable, self-tuning KFs. This doesn't preclude the existence of more complex, NNKF-like schemes at higher levels. As [Millidge et al., 2021] concludes:

> (...) it is also possible that in the brain the prior state could be set by, or influenced by feedback connections from higher levels, representing more advanced states of processing which could inform the state estimate. We believe this is an important point about how Kalman filtering may fit into the larger picture. Since Kalman filtering turns out to be relatively straightforward algorithm which can be implemented in a rather small neural circuit, it could perhaps serve as a composable building block of cortical processing. Kalman filters could potentially be implemented at the lowest hierarchical levels to achieve some immediate processing and reduction of sensory noise before passing on improved state estimates to higher levels of processing which can then apply more complex nonlinear filtering algorithms.

B.3 Neurosocial Evidence

The following concepts and works provide neurosocial evidence for NKP.

B.3.1 Normalcy bias in emergency research

> Actual human behaviour in fires is somewhat different from the 'panic' scenario. What is regularly observed is a lethargic response. People are often cool during fires, ignoring or delaying their response.
>
> – [Proulx, 2002]

The consensus view in emergency research includes many statements that are shocking, easily mistakable for jokes, but serious, enlightening and tragic. On Saturday nights, when a fire visibly starts in a bar and firefighters arrive and tell customers to evacuate, many cling to their beers and decide to stay [Ripley, 2009]. When a fire starts an an airplane and flight attendants warn passengers to brace for the emergency landing, passengers laugh [Ripley, 2009]. Similarly widespread denial reactions (silence, laughter, anger, inaction) have also been reported around other disasters, such as 9/11 or Hurricane Katrina [Wucker, 2016, Ripley, 2009], as well as in persecuted minorities when given life-saving information about the extermination camps they would soon be sent to [Klein, 1994]. For additional discussion, see 9.7.

B.3.2 "When Experts Fail" (Paul Graham)

> When experts are wrong, it's often because they're experts on an earlier version of the world.

> (...) I spent almost a decade investing in early stage startups, and curiously enough protecting yourself against obsolete beliefs is exactly what you have to do to succeed as a startup investor. Most really good startup ideas look like bad ideas at first, and many of those look bad specifically because some change in the world just switched them from bad to good. I spent a lot of time learning to recognize such ideas, and the techniques I used may be applicable to ideas in general.
>
> The first step is to have an explicit belief in change. People who fall victim to a monotonically increasing confidence in their opinions are implicitly concluding the world is static. If you consciously remind yourself it isn't, you start to look for change.
>
> (...)
>
> Startup investors have extraordinary incentives for correcting obsolete beliefs. If they can realize before other investors that some apparently unpromising startup isn't, they can make a huge amount of money.
>
> – [Graham, 2014]

Paul Graham didn't use terms like Neural-Kalman Phenomena or HLABAFEO investing, but that is exactly what he is describing [Graham, 2014]. An abrupt change that makes a previous, neurocognitively undemanding strategy ('business as usual', such as KF tracking) obsolete and requires switching to a new model (in tracking terms, it requires an available NN switch to detect that change and, if necessary, switch into a Bayesian stereotyper[1] to identify the novel dynamics); such an abrupt change, if detected, can be extremely profitable through highly leveraged bets. Graham is arguably an expert in NKP and HLABAFEO investing through early-stage start-ups[2].

B.3.3 Anchoring

Anchoring is a heuristic in behavioral finance that describes the subconscious use of irrelevant information, such as the purchase price of a security, as a fixed reference point (or anchor) for making subsequent decisions.

Many studies have found evidence for this effect. In one of them, participants were asked to estimate the percentage of African countries in the United Nations [Yasseri and Reher, 2022].

Before estimating, the participants first observed a roulette wheel that was predetermined to stop on either 10 or 65. Participants whose wheel stopped on 10 guessed lower values (25% on average) than participants whose wheel stopped at 65 (45% on average) [Tversky and Kahneman, 1974].

Anchoring is also a term used in inflation research. A simple explanation for inflation anchoring is that workers are more likely to demand a 10% rise in salaries if they expect inflation to be close to 10% next year than if they expect inflation to be close to 2%. Thus, when inflation expectations rise (i.e. when they become *unanchored*), there is a risk of higher wage demands leading to higher wages and higher prices (the so-called wage-price spiral [Blanchard and Johnson, 2017]). There is significant evidence that unanchored expectations lead to higher inflation, while shock-induced inflation might be contained as far as inflation expectations remain anchored; in that case, inflation reacts by less and returns quickly to its pre-shock level [Bems et al., 2021].

[1] For the Bayesian stereotyper proposed in this Thesis, see 9.4.

[2] Paul Graham is the founder of Y-Combinator, an early stage startup incubator that coached and invested in AirBnB, Stripe, DoorDash, Coinbase, Instacart, Dropbox, Reddit and Gusto.

Anchoring (to an underlying model) is not just a feature of KF, but a KF scheme is a reasonably good model for anchoring, too. Therefore, wherever anchoring as a cognitive bias is significant, an overadapted KF mechanism will be applicable as a model and NKP will appear. That explains why BoE's forecasts (Fig. 9.1) were reminiscent of a KF prediction based on the wrong (anchored) model. The mere existence of anchored inflation as an empirically backed concept is evidence for KF-like behaviour in economic agents; unanchoring processes are NKP.

B.3.4 Black Swans in Extremistan (Nassim Nicholas Taleb)

Black Swans [Taleb, 2007] are events that come as a surprise and have a major effect. The theory of Black Swans was developed in Nassim Nicholas Taleb's Incerto series [Taleb, 2001, Taleb, 2007, Taleb, 2012, Taleb, 2018, Taleb, 2020] to explain the disproportionate role of high-profile, hard-to-predict, and rare events that are beyond the realm of normal expectations in history, science, finance, and technology. The probability of such rare events is non-computable, but such events do not surprise everybody; the same event can be a Black Swan for ordinary investors, but a predictable outcome for some local-domain experts (an example would be the magnitude of Covid when it was just a local event in Wuhan in early January 2020). Thus, what makes an event a Black Swan is our lack of (specific) knowledge about them. Black Swans are subjective, not necessarily universal.

A careful reader might have noticed the similarities between a Black Swan and a Neural-Kalman Event, defined in this Thesis as an "abrupt, consequential change that only a small minority of market participants can correctly interpret early as the abrupt, consequential change it is" (Subsection 9.6.2).

The main difference between Taleb's approach and the NKP/HLABAFEO approach favoured in this Thesis (Section 9.8) is that Taleb generally treats Black Swans as instantaneous, unknowledgeable events, while we try to anticipate some of them by focusing on expert disagreements (Section 9.1).

Black Swans can cause NKP to the extent they are abrupt changes with significant impact and unexpected by most market participants, though typically expected by a minority of market participants who can profit from them. Such market participants, who predict the Black Swans and position for it, are behaving like NNKFs in our analogy of Efficient Market participants as filtering methods with different performance (Subsection 9.6.2.1), while the rest of the participants ignore the early warning signs of the coming Black Swan (such as hidden risk) and behave like KFs (over)adapted to pre-Black-Swan dynamics.

B.3.4.1 Extremistan and Gaussian financial models

Taleb's theory distinguishes two types of random event environments: Mediocristan and Extremistan [Taleb, 2007]. Mediocristan environments, such as Linear Gauss-Markov models, can be safely modeled by using the Gaussian distribution.

However, Extremistan is characterized by frequent abrupt movements (thick tails) that don't fit Gaussian models.

The KFs used in this Thesis, based on underlying LGM models, can be interpreted as suitable estimators in Mediocristan, but they fail in Extremistan. Same can be said about any estimator based on popular Gaussian models in the Modern Theory of Finance [Mandelbrot and Hudson, 2008, Taleb, 2007], including Capital Asset Pricing Model [Sharpe, 1964], Modern Portfolio Theory [Markovitz, 1959], and the Black-Scholes option pricing model [Black and Scholes, 1973].

The widespread use of such Mediocristan models is evidence for KF-like behaviour in market participants (Subsection 9.6.2.1). Both Black Swans and NKP live in Extremistan and destroy their performance (Fig. 9.4).

B.3.4.2 Barbell strategy

The HLABAFEO strategy, based on relatively small bet sizes while the rest of the net worth/portfolio remains uninvested, is our novel implementation of Taleb's barbell strategy [Taleb, 2007]. The focus on median long-term returns in this Thesis' HLABAFEO simulations was inspired by Ergodicity Economics [Peters, 2019].

B.3.5 Gray Rhinos (Michele Wucker)

Gray Rhinos are a similar concept to Black Swans: both are events that cause major impact and which decision-makers, such as market participants, aren't prepared for [Wucker, 2016]. However, for a Gray Rhino to exist, at least some credible experts will have sounded the alarm [Wucker, 2016]. Some Black Swans arguably start as Black Swans (low-probability events) and then, once enough has gone wrong and the event has become an obvious threat[3] for some experts (e.g. high-probability events for them), they become Gray Rhinos. Our concept of NKP is closer to this Gray-Rhino stage of Black Swans; in particular, the fact that credible (local-domain) experts have explicitly pointed to the event is essential for the application of a HLABAFEO strategy (Section 9.8). The real-world examples in [Wucker, 2016] provide neurosocial evidence for NKP.

B.3.6 Escalation of Commitment to a Failing Proposition

Escalation of commitment is a human behavior pattern in which an individual or group facing increasingly negative outcomes from a decision (such as an investment) nevertheless continue the behavior instead of altering course. The actor maintains behaviors that might be interpreted as irrational, but align with previous decisions and actions [Staw, 1996, Staw, 1976].

Related concepts are the sunk-cost fallacy (justification of increased investment of money or effort in a failed decision) in economics or commitment bias in sociology.

The persistence of expert disagreements in the face of definitive data (see Section 9.1) can be explained in light of the escalation of commitment.

Escalation of commitment can cause NKP to the extent it affects everybody, including experts and investors whose behaviour follow the expert opinion. This is particularly true when the expert believes their opinion is prosocial [Schaumberg and Wiltermuth, 2014] and/or is invested emotionally, career-wise, politically, socially or financially on the matter.

An example could be a hypothetical expert in a politically relevant subject matter (such as economics or public health) who has written 12 books backing a specific expert opinion, has made many friends through collaborations in conferences and publications that backed such expert opinion, has defended specific policies by a political party related to that expert opinion and has made investments (and maybe even caused close friends and relatives to make investments) on the basis of that expert opinion. If new data were to be disconfirming of the long-held opinion, the risk

[3] While [Wucker, 2016] focuses on threats, the exact same argument can be made for opportunities.

of escalation of commitment to the failing expert opinion would be significant. Criticized experts often double down with overprecise predictions [Kang and Kim, 2022].

B.3.7 Mirror Neuron System and Mimetic Desire

> It is innate for men to imitate and from childhood they differ from other animals in this capability—that imitation is possible and they first form learning through imitation.
>
> – Aristotle, Poetics, 1448b [Aristotle, 2017]

Mirror neurons are premotor neurons, originally discovered in the macaque brain and later in humans, that discharge both during execution of goal-directed actions and during the observation of similar actions executed by another individual [Keysers and Fadiga, 2009]. They therefore mirror others' actions on the observer's motor repertoire. Mirror-like phenomena have been demonstrated also for domains others than the pure motor one, such as the somatosensory and the emotional systems, possibly providing a neurophysiological basis to phenomena such as embodiment and empathy [Keysers and Fadiga, 2009].

Therefore, there might be a neural basis for learning emotions, including desires. This ties into the mimetic theory of desire [Girard, 2001], which posits that humans desire what others desire because we imitate their desires. Girard's mimetic theory of desire is well-known in some financial circles, such as tech investing, since it reportedly helped one of Girard's students, Peter Thiel, understand the potential of Facebook (everyone wanting to set a profile in Facebook because everyone else would already have one) and make a fortune by becoming one of the first investors in the company at an early stage [Perell, 2019]. The existence of mimetic behaviour can also be interpreted as one of the sources and mechanisms of financial reflexivity [Soros, 2003], which causes manic phases of overvaluation and depressive phases of undervaluation in financial markets.

All this suggests humans' main mode of learning might be mimetic (i.e. social), as Aristotle stated in the quote above. If we assume the consequences of abrupt changes are learned by mimesis, it follows that a core social network of local-domain experts will learn it by mimesis before it spreads widely in social networks of financial experts. This mechanism backs our Local-Domain Expert Hypothesis and its HLABAFEO Corollary under the Neural-Kalman Market Hypothesis (see Subsection 9.6.2).

Bibliography

[Adams and Mackay, 2007] Adams, R. P. and Mackay, D. (2007). Bayesian Online Changepoint Detection. 4.4.2, B.2.1

[Al-Naffouri, 2007] Al-Naffouri, T. Y. (2007). An EM-Based Forward-Backward Kalman Filter for the Estimation of Time-Variant Channels in OFDM. *IEEE Transactions on Signal Processing*, 55(7):3924–3930. 3.5.4, 3.5.7

[Anderson and Moore, 1979] Anderson, B. D. O. and Moore, J. B. (1979). *Optimal Filtering*. Prentice-Hall, Inc., Englewood Cliffs, NJ, USA. 4

[Andre-Obrecht, 1988] Andre-Obrecht, R. (1988). A new statistical approach for the automatic segmentation of continuous speech signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(1):29–40. 4.1.2

[Angelosante et al., 2007] Angelosante, D., Biglieri, E., and Lops, M. (2007). Multipath Channel Tracking in OFDM Systems. In *2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 1–5. 1.2, 2.5.1, 2.5.4.1, 2.6.2, 3.5.7, 4.2.3, 4.2.3, 4.2.3, 4.4.1, 5.3.1.1, 5.3.1.1, 5.3.3, 5.3.3, 5.5, 5.5, 6, 6.1, 7.2, 7.2.4, A.1, A.1, A.2, A.3, A.3, A.3

[Angelosante et al., 2009] Angelosante, D., Biglieri, E., and Lops, M. (2009). Sequential Estimation of Multipath MIMO-OFDM Channels. *IEEE Transactions on Signal Processing*, 57(8):3167–3181. 1.2, 2.5.1, 2.5.4.1, 2.6.2, 4.2.3, 4.2.3, 4.4.1, 5.3.1.1, 5.3.1.1, 6, 6.1, A.1, A.3

[Aristotle, 2017] Aristotle (2017). *The Poetics of Aristotle*. CreateSpace Independent Publishing Platform. B.3.7

[Arulampalam et al., 2002] Arulampalam, M., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188. 4.4.1, A.3

[Authers, 2021] Authers, J. (2021). There's a Long Way Down If Inflation Takes Root. *Bloomberg.com*. 9.1

[Banelli et al., 2007] Banelli, P., Cannizzaro, R. C., and Rugini, L. (2007). Data-Aided Kalman Tracking for Channel Estimation in Doppler-Affected OFDM Systems. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, volume 3, pages III–133–III–136. 3.5.5, 3.5.7

[Barbieri et al., 2009] Barbieri, A., Piemontese, A., and Colavolpe, G. (2009). On the ARMA approximation for fading channels described by the Clarke model with applications to Kalman-based receivers. *IEEE Transactions on Wireless Communications*, 8(2):535–540. 3.5.2

[Basseville and Benveniste, 1983] Basseville, M. and Benveniste, A. (1983). Design and comparative study of some sequential jump detection algorithms for digital signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31(3):521–535. 4.1.2

[Basseville and Nikiforov, 1993] Basseville, M. and Nikiforov, I. V. (1993). *Detection of Abrupt Changes: Theory and Applications. By M. Basseville and I. V. Nikiforov. ISBN 0 13 126780 9. Prentice Hall, Englewood Cliffs, 1993. Xxvi +528 Pp.* Prentice-Hall, Inc., Englewood Cliffs. 4.1.2, 4.3.1.1, 4.3.2, 4.3.3, 4.3.5.1, 4.3.5.1, 4.3.5.2, 4.3.6, 4.3.6, 4.3.7.2

[Beal, 2003] Beal, M. J. (2003). *Variational Algorithms for Approximate Bayesian Inference*. Doctoral, UCL (University College London). 4.4.4

[Bems et al., 2021] Bems, R., Caselli, F., Grigoli, F., and Gruss, B. (2021). Expectations' anchoring and inflation persistence. *Journal of International Economics*, 132:103516. B.3.3

[Betts, 1987] Betts, R. K. (1987). *Nuclear Blackmail and Nuclear Balance*. Brookings Institution. 1.1

[Biglieri et al., 2012] Biglieri, E., Grossi, E., and Lops, M. (2012). Random-set theory and wireless communications. *Foundations and Trends® in Communications and Information Theory*, 7(4):317–462. 4.2.2, A.2, A.2

[Biglieri and Lops, 2006] Biglieri, E. and Lops, M. (2006). Multiuser detection using Random-Set Theory. In *2006 14th European Signal Processing Conference*, pages 1–5. 4.2.3, A.2

[Black and Scholes, 1973] Black, F. and Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3):637–54. B.3.4.1

[Blanchard and Johnson, 2017] Blanchard, O. J. and Johnson, D. R. (2017). *Macroeconomics*. Pearson. 9.1, B.3.3

[Borodkin and Mottl, 1976] Borodkin, L. I. and Mottl, V. V. (1976). Algorithm for finding the jump times of random process equation parameters. *Automation and Remote Control*, 37(6):23–32. 4.1.2

[Brodsky, 2016] Brodsky, B. (2016). *Change-Point Analysis in Nonstationary Stochastic Models*. Chapman and Hall/CRC, New York. 4.1.2, 4.4.5

[Brown and Hwang, 2012] Brown, R. G. and Hwang, P. Y. C. (2012). *Introduction to Random Signals and Applied Kalman Filtering with Matlab Exercises*. Wiley, fourth edition. 4

[Catlin, 1989] Catlin, D. E. (1989). *Estimation, Control, and the Discrete Kalman Filter*. Springer Verlag New York Inc. 4

[Chen et al., 2015] Chen, B., Zhong, Z., Ai, B., Guan, K., He, R., and Michelson, D. G. (2015). Channel Characteristics in High-Speed Railway: A Survey of Channel Propagation Properties. *IEEE Vehicular Technology Magazine*, 10(2):67–78. 1.2

[Chen and Liu, 2000] Chen, R. and Liu, J. S. (2000). Mixture Kalman filters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(3):493. 3.5.2

[Chen and Zhang, 2004] Chen, W. and Zhang, R. (2004). Kalman-filter channel estimator for OFDM systems in time and frequency-selective fading environment. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages iv–iv. 3.5.2, 3.5.7

[Chong et al., 2005] Chong, C.-C., Tan, C.-M., Laurenson, D., McLaughlin, S., Beach, M., and Nix, A. (2005). A novel wideband dynamic directional indoor channel model based on a Markov process. *IEEE Transactions on Wireless Communications*, 4(4):1539–1552. 2.6.1

[Chui and Chen, 2017] Chui, C. K. and Chen, G. (2017). *Kalman Filtering: with Real-Time Applications*. New York, NY. 3.1, 3.1.1

[Dahlman et al., 2020] Dahlman, E., Parkvall, S., and Skold, J. (2020). *5G NR: The next Generation Wireless Access Technology*. Academic Press. 1.2, 2

[Dale, 2021] Dale, D. (2021). Macro Scouting Report. Technical report, 42 Macro. 9.1

[de Pablo López, 2001] de Pablo López, A. (2001). *Manuel Práctico de Matemática Comercial y Financiera*, volume II. Editorial Centro de Estudios Ramón Areces, S. A., second edition. 9.1

[Denève et al., 2007] Denève, S., Duhamel, J.-R., and Pouget, A. (2007). Optimal Sensorimotor Integration in Recurrent Cortical Networks: A Neural Implementation of Kalman Filters. *Journal of Neuroscience*, 27(21):5744–5756. B.1.1

[Doucet et al., 2001] Doucet, A., Gordon, N., and Krishnamurthy, V. (2001). Particle filters for state estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing*, 49(3):613–624. 4.4.1

[Edfors et al., 1996] Edfors, O., Sandell, M., Beek, J. V., Landström, D., and Sjöberg, F. (1996). An introduction to orthogonal frequency-division multiplexing. (document), 2.1, 2.2, 2, 2.3.1, 2.3.3

[El Husseini et al., 2019] El Husseini, A. H., Simon, E. P., and Ros, L. (2019). Second-order autoregressive model-based Kalman filter for the estimation of a slow fading channel described by the Clarke model: Optimal tuning and interpretation. *Digital Signal Processing*, 90:125–141. 3.5.2

[Fama, 2022] Fama, E. (2022). Interview with Nobel laurate in Economics Eugene Fama. 9.6.2.1

[Fama, 1970] Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2):383–417. 9.6.2.1

[Filipowicz et al., 2020] Filipowicz, A. L., Glaze, C. M., Kable, J. W., and Gold, J. I. (2020). Pupil diameter encodes the idiosyncratic, cognitive complexity of belief updating. *eLife*, 9:e57872. 9.5, 11

[Fuh and Tartakovsky, 2019] Fuh, C.-D. and Tartakovsky, A. G. (2019). Asymptotic Bayesian Theory of Quickest Change Detection for Hidden Markov Models. *IEEE Transactions on Information Theory*, 65(1):511–529. 9.4

[García and Vélez, 2012] García, A. and Vélez, R. (2012). *Principios de Inferencia Estadística*. UNED: Barcelona. 4.3.1.1

[Gelb et al., 1974] Gelb, A., Kasper, J. F., Nash, Jr., R. A., Price, C. F., and Sutherland, A. A. (1974). *Applied Optimal Estimation*. MIT Press, Cambridge, MA, USA. 4

[Girard, 2001] Girard, R. (2001). *Des choses cachées depuis la fondation du monde*. Grasset. B.3.7

[Girshick and Rubin, 1952] Girshick, M. A. and Rubin, H. (1952). A Bayes Approach to a Quality Control Model. *The Annals of Mathematical Statistics*, 23(1):114–125. 4.1.1, 4.3.6

[Goodman et al., 2013] Goodman, I. R., Mahler, R. P., and Nguyen, H. T. (2013). *Mathematics of Data Fusion*, volume 37. Springer Science & Business Media. 4.2.2, 6.1, A.1, A.1, A.2, A.3, A.3

[Gordon et al., 1993] Gordon, N. J., Salmond, D. J., and Smith, A. F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings F (Radar and Signal Processing)*, volume 140, pages 107–113. IET. 4.4.1

[Graham, 2014] Graham, P. (2014). How to Be an Expert in a Changing World. 1.1, 9.7, B.3.2

[Grantham, 2021] Grantham, J. (2021). GameStop to Tesla: Investor Jeremy Grantham on 'crazy' markets. 1

[Green, 2013] Green, K. (2013). On Fire [Gunshow comic #648]. (document), 9.5

[Grewal and Andrews, 2015] Grewal, M. S. and Andrews, A. P. (2015). *KALMAN FILTERING-Theory and Practice Using MATLAB. Published by John Wiley & Sons, Inc., Hoboken. New Jersey*. (document), 1.2, 3.1, 3.1.1, 3.3, 3.3.2, 3.3.3, 3.1

[Guo et al., 2017] Guo, W., Zhang, W., Mu, P., and Gao, F. (2017). High-Mobility OFDM Downlink Transmission With Large-Scale Antenna Array. *IEEE Transactions on Vehicular Technology*, 66(9):8600–8604. 1.2

[Hadsell et al., 2020] Hadsell, R., Rao, D., Rusu, A. A., and Pascanu, R. (2020). Embracing Change: Continual Learning in Deep Neural Networks. *Trends in Cognitive Sciences*, 24(12):1028–1040. 9.6.1, 9.11

[Hájek, 1981] Hájek, J. (1981). Sampling from a finite population. Technical report, Marcel Dekker, New York. 4.2.1

[Han et al., 2004] Han, K.-Y., Lee, S.-W., Lim, J.-S., and Sung, K.-M. (2004). Channel estimation for OFDM with fast fading channels by modified Kalman filter. *IEEE Transactions on Consumer Electronics*, 50(2):443–449. 3.5.6, 3.5.7

[Hassan et al., 2021] Hassan, N., Dupleich, D. A., Schneider, C., Thomä, R., and Del Galdo, G. (2021). MIMO-TDL Model Parameter Estimation from V2I Channel Sounding. In *2021 15th European Conference on Antennas and Propagation (EuCAP)*, pages 1–5. 2.6.2.1, 5.2, 4

[Hassan et al., 2020] Hassan, N., Thomä, R., and Matolak, D. W. (2020). In-Stationary Tapped Delay Line Channel Modeling and Simulation. In *2020 14th European Conference on Antennas and Propagation (EuCAP)*, pages 1–5. (document), 1.2, 2.6, 2.6.2.1, 2.9, 5.2, 7.2.4.1, 1, 2, 4, 1

[He et al., 2016] He, R., Ai, B., Wang, G., Guan, K., Zhong, Z., Molisch, A. F., Briso-Rodriguez, C., and Oestges, C. P. (2016). High-Speed Railway Communications: From GSM-R to LTE-R. *IEEE Vehicular Technology Magazine*, 11(3):49–58. 1.2

[Jepma et al., 2016] Jepma, M., Murphy, P. R., Nassar, M. R., Rangel-Gomez, M., Meeter, M., and Nieuwenhuis, S. (2016). Catecholaminergic Regulation of Learning Rate in a Dynamic Environment. *PLOS Computational Biology*, 12(10):e1005171. 9.5

[Jiménez et al., 2019] Jiménez, J. C. E., García, M. J. F.-G., and Méndez-Romero, D. (2019). Superimposed Training and Blind Channel Estimation. In *Wiley 5G Ref*, pages 1–17. John Wiley & Sons, Ltd. 3.5.5

[Jiménez et al., 2020] Jiménez, J. C. E., García, M. J. F.-G., and Méndez-Romero, D. (2020). Pilot-Assisted Channel Estimation. In *Wiley 5G Ref*, pages 1–19. John Wiley & Sons, Ltd. 4.2

[Jones et al., 1970] Jones, R. H., Crowell, D. H., and Kapuniai, L. E. (1970). Change Detection Model for Serially Correlated Multivariate Data. *Biometrics*, 26(2):269–280. 4.1.2

[Kalman, 1960] Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45. 3.1

[Kang and Kim, 2022] Kang, S. and Kim, J. W. (2022). The Fragility of Experts: A Moderated-Mediation Model of Expertise, Expert Identity Threat, and Overprecision. *Academy of Management Journal*, 65(2):577–605. 9.6.2, B.3.6

[Kashyap et al., 2017] Kashyap, S., Mollén, C., Björnson, E., and Larsson, E. G. (2017). Performance analysis of (TDD) massive MIMO with Kalman channel prediction. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3554–3558. 3.5.2, 3.5.7

[Kendall, 1974] Kendall, D. G. (1974). Foundations of a theory of random sets, Stoch. Geometry (eds. E. F. Harding and DG Kendall), 322–376. 4.2.1

[Keysers and Fadiga, 2009] Keysers, C. and Fadiga, L. (2009). *The Mirror Neuron System: A Special Issue of Social Neuroscience*. Psychology Press. 9.7, B.3.7

[Kim et al., 2005] Kim, K. J., Yue, J., Iltis, R., and Gibson, J. (2005). A QRD-M/Kalman filter-based detection and channel estimation algorithm for MIMO-OFDM systems. *IEEE Transactions on Wireless Communications*, 4(2):710–721. 3.5.5, 3.5.7

[Kim, 2011] Kim, P. (2011). *Kalman Filter for Beginners: with MATLAB Examples*. Charleston, SC. 3.1.1

[Klein, 1994] Klein, G. (1994). *Pieta*. MIT Press. 9.7, B.3.1

[Kolmogorov, 1950] Kolmogorov, A. N. (1950). *Foundations of the Theory of Probability*. New York: Chelsea Pub. Co. 3

[Komninakis et al., 2002] Komninakis, C., Fragouli, C., Sayed, A., and Wesel, R. (2002). Multi-input multi-output fading channel tracking and equalization using Kalman estimation. *IEEE Transactions on Signal Processing*, 50(5):1065–1076. 3.5.5

[Krugman, 2013] Krugman, P. (2013). The Four Percent Solution. 9.2

[LeCun, 2022] LeCun, Y. (2022). A Path Towards Autonomous Machine Intelligence Version 0.9.2, 2022-06-27. page 62. 9.4.2

[Liakoni et al., 2021] Liakoni, V., Modirshanechi, A., Gerstner, W., and Brea, J. (2021). Learning in Volatile Environments With the Bayes Factor Surprise. *Neural Computation*, 33:1–72. 4.4.2, 4.4.4, 8

[Lorden, 1971] Lorden, G. (1971). Procedures for Reacting to a Change in Distribution. *The Annals of Mathematical Statistics*, 42(6):1897–1908. 4.1.1

[Ma et al., 2006] Ma, W.-K., Vo, B.-N., Singh, S., and Baddeley, A. (2006). Tracking an unknown time-varying number of speakers using TDOA measurements: A random finite set approach. *IEEE Transactions on Signal Processing*, 54(9):3291–3304. A.1, A.2

[Mahler et al., 2017] Mahler, K., Keusgen, W., Tufvesson, F., Zemen, T., and Caire, G. (2017). Measurement-Based Wideband Analysis of Dynamic Multipath Propagation in Vehicular Communication Scenarios. *IEEE Transactions on Vehicular Technology*, 66(6):4657–4667. 1.2

[Mahler, 2003] Mahler, R. (2003). Multitarget Bayes filtering via first-order multitarget moments. *IEEE Transactions on Aerospace and Electronic Systems*, 39(4):1152–1178. A.2

[Mandelbrot and Hudson, 2008] Mandelbrot, B. B. and Hudson, R. L. (2008). *The (Mis)Behaviour of Markets: A Fractal View of Risk, Ruin and Reward*. London. 4.1.4, B.3.4.1

[Marchthaler and Dingler, 2017] Marchthaler, R. and Dingler, S. (2017). *Kalman-Filter: Einführung in die Zustandsschätzung und ihre Anwendung für eingebettete Systeme*. Wiesbaden; Heidelberg. 3.1.1

[Markovitz, 1959] Markovitz, H. M. (1959). *Portfolio Selection: Efficient Diversification of Investments*. John Wiley. B.3.4.1

[Mathéron, 1975] Mathéron, G. (1975). *Random Sets and Integral Geometry*. J. Wiley & Sons, New York. 4.2.1, 3

[Mathieu, 1976] Mathieu, M. (1976). *Analyse de l'eléctro-Encéphalogramme Par Prédiction Lineaire*. PhD thesis, Université Pierre et Marie Curie. 4.1.2

[Matolak, 2008] Matolak, D. W. (2008). Channel Modeling for Vehicle-To-Vehicle Communications. *IEEE Communications Magazine*, 46(5):76–83. (document), 2.5, 2.6.2, 2.6, 2.7, 5.2, 7.2.4.1, 4

[Matolak, 2014] Matolak, D. W. (2014). Modeling the vehicle-to-vehicle propagation channel: A review. *Radio Science*, 49(9):721–736. 1.2, 2.6.1, 2.6.2.3, 8.1

[Matolak and Sun, 2015] Matolak, D. W. and Sun, R. (2015). Unmanned Aircraft Systems: Air-Ground Channel Characterization for Future Applications. *IEEE Vehicular Technology Magazine*, 10(2):79–85. 1.2

[Matolak and Wu, 2008] Matolak, D. W. and Wu, Q. (2008). Markov models for vehicle-to-vehicle channel multipath persistence processes. In *Proc. 1st IEEE Veh. Tech. Society Wireless Access in Veh. Env.(WAVE) Conf.* 2.6, 2.6.1

[Mehrnia and Coleri, 2022] Mehrnia, N. and Coleri, S. (2022). Extreme Value Theory Based Rate Selection for Ultra-Reliable Communications. *IEEE Transactions on Vehicular Technology*, 71(6):6727–6731. 4.1.4

[Méndez-Romero, 2015] Méndez-Romero, D. (2015). Seguimiento de canales no lineales en comunicaciones móviles OFDM. Proyecto Fin de Carrera [Bachelor's Thesis]. Universidad Carlos III de Madrid. 3.5.7, 4.2, 6

[Méndez-Romero and Fernández-Getino García, 2018] Méndez-Romero, D. and Fernández-Getino García, M. J. (2018). Simpler Multipath Detection for Vehicular OFDM Channel Tracking. *IEEE Transactions on Vehicular Technology*, 67(11):10752–10759. 4.4.1, 4.4.3, 5.4.2, 5.4.3, 6, 6.1, 6.1, 6.1.2, 7.2, 7.2.4

[Méndez-Romero and Fernández-Getino García, 2020] Méndez-Romero, D. and Fernández-Getino García, M. J. (2020). Death/Birth and SNR Detection for Vehicular Kalman Channel Trackers. In *2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON)*, pages 104–108, Palermo, Italy. IEEE. 6

[Mendez-Romero et al., 2020] Mendez-Romero, D., Fernandez-Getino Garcia, M. J., Tonello, A. M., and Dobre, O. A. (2020). Neural-Network-Switched Kalman Filters as Novel Trackers for Multipath Channels. In *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 1–5, Dublin, Ireland. IEEE. 4.4.3, 7.3, 8.3

[Millidge et al., 2021] Millidge, B., Tschantz, A., Seth, A., and Buckley, C. (2021). Neural Kalman Filtering. B.1.1, B.2, B.2.3

[Molchanov, 2005] Molchanov, I. (2005). Random Closed Sets. In Bilodeau, M., Meyer, F., and Schmitt, M., editors, *Space, Structure and Randomness: Contributions in Honor of Georges Matheron in the Field of Geostatistics, Random Sets and Mathematical Morphology*, Lecture Notes in Statistics, pages 135–149. Springer, New York, NY. 3

[Molchanov, 2017] Molchanov, I. (2017). *Theory of Random Sets*. Probability Theory and Stochastic Modelling. Springer, London. 4.2.1

[Nassar et al., 2019] Nassar, M. R., Bruckner, R., and Frank, M. J. (2019). Statistical context dictates the relationship between feedback-related EEG signals and learning. *eLife*, 8:e46975. 9.5

[News.com.au, 2022] News.com.au (2022). Customers flee Brazillian bar in fear as CrossFit group mistaken for robbers. 19

[Nguyen, 2006] Nguyen, H. T. (2006). *An Introduction to Random Sets*. CRC Press. A.1

[Nikiforov and Tikhonov, 1986] Nikiforov, I. V. and Tikhonov, I. N. (1986). Application of change detection theory to seismic signal processing. In Basseville, M. and Benveniste, A., editors, *Detection of Abrupt Changes in Signals and Dynamical Systems*, Lecture Notes in Control and Information Sciences, pages 355–373, Berlin, Heidelberg. Springer. 4.1.2, 4.3.4, 4.4.5

[Noor-A-Rahim et al., 2022] Noor-A-Rahim, M., Liu, Z., Lee, H., Khyam, M. O., He, J., Pesch, D., Moessner, K., Saad, W., and Poor, H. V. (2022). 6G for Vehicle-to-Everything (V2X) Communications: Enabling Technologies, Challenges, and Opportunities. *Proceedings of the IEEE*, 110(6):712–734. 4

[Oppenheim et al., 1997] Oppenheim, A. V., Buck, J., Daniel, M., Willsky, A. S., Nawab, S. H., and Singer, A. (1997). *Signals & Systems*. Pearson Educación. 2.3.3, 2.5.1

[O'Reilly et al., 2013] O'Reilly, J. X., Schüffelgen, U., Cuell, S. F., Behrens, T. E. J., Mars, R. B., and Rushworth, M. F. S. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 110(38):E3660–3669. 9.5

[Page, 1954] Page, E. S. (1954). Continuous Inspection Schemes. *Biometrika*, 41(1-2):100–115. 4.1.1, 4.3.5, 4.3.5.2, 4.3.5.2

[Pätzold, 2011] Pätzold, M. (2011). *Mobile Radio Channels*. John Wiley & Sons. 2.5.3.1, 2.5.3.1, 2.5.4.1

[Peebles Jr, 2001] Peebles Jr, P. Z. (2001). *Probability, Random Variables, and Random Signal Principles*. McGraw-Hill. 5.4.2, 5.4.2

[Perell, 2019] Perell, D. (2019). Peter Thiel's Religion. B.3.7

[Pergamenchtchikov et al., 2022] Pergamenchtchikov, S. M., Tartakovsky, A. G., and Spivak, V. S. (2022). Minimax and pointwise sequential changepoint detection and identification for general stochastic models. *Journal of Multivariate Analysis*, 190:104977. 4.4.5

[Peters, 2019] Peters, O. (2019). The ergodicity problem in economics. *Nature Physics*, 15(12):1216–1221. B.3.4.2

[Porat, 2008] Porat, B. (2008). *Digital Processing of Random Signals: Theory and Methods*. Courier Dover Publications. 13

[Pratik et al., 2021] Pratik, K., Amjad, R. A., Behboodi, A., Soriaga, J. B., and Welling, M. (2021). Neural Augmentation of Kalman Filter with Hypernetwork for Channel Tracking. In *2021 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. 3.5.2, 7.3, 7.3.1

[Princeton Bendheim Center for Finance, 2021] Princeton Bendheim Center for Finance (2021). A Conversation with Lawrence H. Summers and Paul Krugman. 9.1

[Proulx, 2002] Proulx, G. (2002). Cool under fire. *Fire protection engineering*, 16(Fall):33. 9.7, B.3.1

[Revach et al., 2022] Revach, G., Shlezinger, N., Ni, X., Escoriza, A. L., van Sloun, R. J. G., and Eldar, Y. C. (2022). KalmanNet: Neural Network Aided Kalman Filtering for Partially Known Dynamics. *IEEE Transactions on Signal Processing*, 70:1532–1547. 7.3

[Rice, 1948] Rice, S. O. (1948). Statistical properties of a sine wave plus random noise. *The Bell System Technical Journal*, 27(1):109–157. 2.5.3.1

[Ripley, 2009] Ripley, A. (2009). *The Unthinkable: Who Survives When Disaster Strikes - and Why*. New York. 9.7, B.3.1

[Ristic, 2011] Ristic, B. (2011). Bayesian Estimation With Imprecise Likelihoods: Random Set Approach. *IEEE Signal Processing Letters*, 18(7):395–398. 4.2.2

[Ristic, 2013] Ristic, B. (2013). *Particle Filters for Random Set Models*. Springer Science & Business Media. 4.1.5, 4.4.1

[Robbins, 1944] Robbins, H. E. (1944). On the Measure of a Random Set. *The Annals of Mathematical Statistics*, 15(1):70–74. 4.2.1

[Rodríguez-Piñeiro et al., 2021a] Rodríguez-Piñeiro, J., Domínguez-Bolaño, T., Cai, X., Huang, Z., and Yin, X. (2021a). Air-to-Ground Channel Characterization for Low-Height UAVs in Realistic Network Deployments. *IEEE Transactions on Antennas and Propagation*, 69(2):992–1006. 2

[Rodríguez-Piñeiro et al., 2021b] Rodríguez-Piñeiro, J., Huang, Z., Cai, X., Domínguez-Bolaño, T., and Yin, X. (2021b). Geometry-Based MPC Tracking and Modeling Algorithm for Time-Varying UAV Channels. *IEEE Transactions on Wireless Communications*, 20(4):2700–2715. 2

[Russell et al., 2020] Russell, S. J., Russell, S., and Norvig, P. (2020). *Artificial Intelligence: A Modern Approach*. Pearson. 7.1, 7.1

[Saleh and Valenzuela, 1987] Saleh, A. and Valenzuela, R. (1987). A Statistical Model for Indoor Multipath Propagation. *IEEE Journal on Selected Areas in Communications*, 5(2):128–137. 2.6.2.3

[Sanderson et al., 1980] Sanderson, A. C., Segen, J., and Richey, E. (1980). Hierarchical modeling of EEG signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2(5):405–415. 4.1.2

[Satorras et al., 2019] Satorras, V. G., Akata, Z., and Welling, M. (2019). Combining generative and discriminative models for hybrid inference. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, number 1238, pages 13820–13830. Curran Associates Inc., Red Hook, NY, USA. 9.4

[Schaumberg and Wiltermuth, 2014] Schaumberg, R. L. and Wiltermuth, S. S. (2014). Desire for a positive moral self-regard exacerbates escalation of commitment to initiatives with prosocial aims. *Organizational Behavior and Human Decision Processes*, 123(2):110–123. 9.6.2, B.3.6

[Schultz, 2003] Schultz, H. D. (2003). *Bear Market Investing Strategies*. John Wiley & Sons. 20, 9.10.2

[Selkis, 2020] Selkis, R. (2020). Crypto Theses for 2021. 9.10.2

[Sen and Matolak, 2008] Sen, I. and Matolak, D. W. (2008). Vehicle–Vehicle Channel Models for the 5-GHz Band. *IEEE Transactions on Intelligent Transportation Systems*, 9(2):235–245. 2.6.1, 5.2

[Shalev-Shwartz and Ben-David, 2014] Shalev-Shwartz, S. and Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press. 7.1, 7.1, 7.1, 7.1.1, 7.1.1.1, 8.2

[Sharpe, 1964] Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk*. *The Journal of Finance*, 19(3):425–442. B.3.4.1

[Sheng et al., 2017] Sheng, Z., Tuan, H. D., Nguyen, H. H., and Fang, Y. (2017). Pilot Optimization for Estimation of High-Mobility OFDM Channels. *IEEE Transactions on Vehicular Technology*, 66(10):8795–8806. 1.2

[Shewhart, 1931] Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product*. American Society for Quality Control. 4.1.1, 4.1.2

[Shiryaev, 1961] Shiryaev, A. N. (1961). The problem of quickest detection of a violation of stationary behavior. In *Doklady Akademii Nauk*, volume 138, pages 1039–1042. Russian Academy of Sciences. 4.1.1

[Shiryaev, 1963] Shiryaev, A. N. (1963). On optimum methods in quickest detection problems. *Theory of Probability & Its Applications*, 8(1):22–46. 4.1.1

[Shiryaev, 1965] Shiryaev, A. N. (1965). Some exact formulas in a "disorder" problem. *Theory of Probability & Its Applications*, 10(2):348–354. 4.1.1

[Shu et al., 2014] Shu, H., Ros, L., and Simon, E. P. (2014). Simplified Random-Walk-Model-Based Kalman Filter for Slow to Moderate Fading Channel Estimation in OFDM Systems. *IEEE Transactions on Signal Processing*, 62(15):4006–4017. 3.5.6, 3.5.7

[Siegel, 2009] Siegel, J. J. (2009). Efficient market theory and the recent financial crisis. *Wall Street Journal*, page A23. 9.6.2.1

[Simeone and Spagnolini, 2004] Simeone, O. and Spagnolini, U. (2004). Adaptive pilot pattern for OFDM systems. In *2004 IEEE International Conference on Communications (IEEE Cat. No.04CH37577)*, volume 2, pages 978–982 Vol.2. 3.5.3, 3.5.7

[Soros, 2003] Soros, G. (2003). *The Alchemy of Finance*. Wiley Investment Classics. J. Wiley, Hoboken, N.J. B.3.7

[Staw, 1976] Staw, B. M. (1976). Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action. *Organizational Behavior and Human Performance*, 16(1):27–44. 9.6.2, 9.7, B.3.6

[Staw, 1996] Staw, B. M. (1996). The escalation of commitment: An update and appraisal. In Shapira, Z., editor, *Organizational Decision Making*, Cambridge Series on Judgment and Decision Making, pages 191–215. Cambridge University Press, Cambridge. 9.6.2, 9.7, B.3.6

[Summers, 2021] Summers, L. (2021). Summers Disagrees with Krugman on Fed's Inflation Risk. 9.1

[Sun and Matolak, 2017] Sun, R. and Matolak, D. W. (2017). Air–Ground Channel Characterization for Unmanned Aircraft Systems Part II: Hilly and Mountainous Settings. *IEEE Transactions on Vehicular Technology*, 66(3):1913–1925. 1.2

[Taleb, 2001] Taleb, N. N. (2001). *Fooled by Randomness: The Hidden Role of Chance in the Markets and in Life. W. W.* Norton, New York City. 4.1.4, B.3.4

[Taleb, 2007] Taleb, N. N. (2007). *The Black Swan: The Impact of the Highly Improbable*, volume 2. Random house. 1.1, 4.1.4, 9.6.2, 9.7, 9.10.2, B.3.4, B.3.4.1, B.3.4.2

[Taleb, 2012] Taleb, N. N. (2012). *Antifragile: How to Live in a World We Don't Understand*, volume 3. Allen Lane London. 4.1.4, B.3.4

[Taleb, 2018] Taleb, N. N. (2018). *Skin in the Game: Hidden Asymmetries in Daily Life*. Random House. 4.1.4, B.3.4

[Taleb, 2020] Taleb, N. N. (2020). *Statistical Consequences of Fat Tails: Real World Preasymptotics, Epistemology, and Applications.* 4.1.4, 9.7, B.3.4

[Tartakovsky et al., 2014] Tartakovsky, A., Nikiforov, I., and Basseville, M. (2014). *Sequential Analysis: Hypothesis Testing and Changepoint Detection.* CRC Press. 4.1.2, 4.3.7.1, 4.4.5

[Tartakovsky, 2019] Tartakovsky, A. G. (2019). *Sequential Change Detection and Hypothesis Testing: General Non-i.i.d. Stochastic Models and Asymptotically Optimal Rules.* Chapman and Hall/CRC, New York. 4.1.2

[Tse and Viswanath, 2005] Tse, D. and Viswanath, P. (2005). *Fundamentals of Wireless Communication.* Cambridge university press. 1.2, 2.5.2, 2.5.3, 3

[Tversky and Kahneman, 1974] Tversky, A. and Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131. 9.7, B.3.3

[Van De Beek et al., 1995] Van De Beek, J.-J., Edfors, O., Sandell, M., Wilson, S. K., and Borjesson, P. O. (1995). On channel estimation in OFDM systems. In *1995 IEEE 45th Vehicular Technology Conference. Countdown to the Wireless Twenty-First Century*, volume 2, pages 815–819. IEEE. 3.5.2

[Vihola, 2004] Vihola, M. (2004). *Random Sets for Multitarget Tracking and Data Fusion.* PhD thesis. 4.4.1, 8.1, A.1, A.2

[Vo et al., 2005] Vo, B.-N., Singh, S., and Doucet, A. (2005). Sequential Monte Carlo methods for multitarget filtering with random finite sets. *IEEE Transactions on Aerospace and Electronic Systems*, 41(4):1224–1245. A.3

[Voskuil, 2020] Voskuil, E. (2020). *Cryptoeconomics: Fundamental Principles of Bitcoin.* Eric Voskuil. 1

[Voskuil and Taaki,] Voskuil, E. and Taaki, A. *Criptoeconomía: Principios fundamentales de Bitcoin.* 1

[Wilson and Finkel, 2009] Wilson, R. and Finkel, L. (2009). A Neural Implementation of the Kalman Filter. In *Advances in Neural Information Processing Systems 22 - Proceedings of the 2009 Conference*, pages 2062–2070. B.1.1, B.2, B.2.1, B.2.2

[Wu, 2012] Wu, Q. (2012). *Non-Stationary Vehicle-to-Vehicle Channel Characterization.* PhD thesis, Ohio University. 1.2, 2.6, 2.6.2.2

[Wu et al., 2010] Wu, Q., Matolak, D. W., and Sen, I. (2010). 5-GHz-Band Vehicle-to-Vehicle Channels: Models for Multiple Values of Channel Bandwidth. *IEEE Transactions on Vehicular Technology*, 59(5):2620–2625. (document), 2.6.2, 2.8, 2.6.2.2, 5.2, 7.2.4.1, 4

[Wucker, 2016] Wucker, M. (2016). *The Gray Rhino.* New York. 9.7, B.3.1, B.3.5, 3

[Yasseri and Reher, 2022] Yasseri, T. and Reher, J. (2022). Fooled by facts: Quantifying anchoring bias through a large-scale experiment. *Journal of Computational Social Science*, 5(1):1001–1021. 9.7, B.3.3

[Yu et al., 2021] Yu, L. Q., Wilson, R. C., and Nassar, M. R. (2021). Adaptive learning is structure learning in time. *Neuroscience & Biobehavioral Reviews*, 128:270–281. 9.4.2, 9.5

[Yuanjin, 2003] Yuanjin, Z. (2003). A novel channel estimation and tracking method for wireless OFDM systems based on pilots and Kalman filtering. *IEEE Transactions on Consumer Electronics*, 49(2):275–283. 3.5.1, 3.5.7, 5.5, 7.2.4

[Zhang, 1996] Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *Journal of Neuroscience*, 16(6):2112–2126. B.2.1

[Zhang, 2016] Zhang, L. (2016). *Channel Measurement and Modeling in Complex Environments.* PhD thesis, E.T.S.I y Sistemas de Telecomunicación (UPM). 2.6.2.3, 8.1.1

[Zhang and Chen, 2004] Zhang, R. and Chen, W. (2004). A mixture Kalman filter approach for blind OFDM channel estimation. In *Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, 2004.*, volume 1, pages 350–354 Vol.1. 3.5.2

[Zuckerman, 2019] Zuckerman, G. (2019). *The Man Who Solved the Market: How Jim Simons Launched the Quant Revolution.* Penguin UK. 9.6.2.1, 9.6.2.2

Note: the term "(document)" marks bibliographical references Figures in this document are adapted from, or otherwise based on. See List of Figures to identify the exact figure(s) based on them.