Marie Skłodowska Curie Action

WINDMILL

Machine Learning for Wireless Communications

H2020-MSCA-ITN-ETN
Grant Agreement Number: 813999

WP4–Prediction Schemes and Anticipatory Optimization for Fast-Varying Processes

D4.1.A–Joint Probabilistic Modelling of Wireless Channels and Hardware Impairments

Contractual Delivery Date:	30.09.2020
Actual Delivery Date:	23.09.2020
Responsible Beneficiary:	Ericsson, KTH
Contributing Beneficiaries:	Ericsson, KTH
Dissemination Level:	Public
Version:	Final

PROPRIETARY RIGHTS STATEMENT
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813999.

PROPRIETARY RIGHTS STATEMENT
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813999.

Document Information

Document ID:	WP4/D4.1.A
Version Date:	30.09.2020
Total Number of Pages:	36
Abstract:	In this report we discuss a joint modelling problem of wireless channels together with hardware impairments of radio front-end. In the first half we discuss necessity of such simulators and then proceed to an overview of hardware impairments which effect performance the most. We also highlight main disadvantages of existing impairment modelling methods with respect to channel modelling. In the core section we propose a solution in a from of a bayesian non-parametric tool of gaussian processes. Its choice is justified by their generative and universal functional approximation properties, and interpretable way of model construction. Same chapter also includes an overview of more advanced techniques, such as automatic kernel construction and deep gaussian processes. We close report with a thorough overview of modern probabilistic modelling frameworks and possible future research directions.
Keywords:	wireless channel model, radio front-end, distortion, gaussian process, probabilistic programming

Authors

Full name	**Beneficiary/ Organisation**	**e-mail**	**Role**
Sergey S. Tambovskiy	Ericsson, KTH	sergey.tambovskiy@ericsson.com	Contributor, Editor

Reviewers

Full name	**Beneficiary/ Organisation**	**e-mail**	**Date**
Gabor Fodor	Ericsson, KTH	gabor.fodor@ericsson.com	07.09.2020
Hugo Tullberg	Ericsson	hugo.tullberg@ericsson.com	15.09.2020

Version history

Version	**Date**	**Comments**
0.1	30.08.2020	Initial draft, 75% of study
0.2	07.09.2020	Final draft, 100% of study
0.3	15.09.2020	Formatting, incorporation of reviewers comments
1.0	23.09.2020	Submitted, final version
1.1	06.10.2020	Added an extra entry in software section

Table of Contents

Lists of Figures **vi**

List of Tables **vi**

List of Acronyms and Abbreviations **viii**

1 Introduction **1**
- 1.1 The twofold problem . . . 1

2 RF impairments and channel models **3**
- 2.1 Antenna array impairments . . . 3
- 2.2 Intermodulation distortions . . . 5
- 2.3 Channel modelling . . . 8

3 Gaussian processes **10**
- 3.1 Universal functional approximation . . . 10
- 3.2 Gaussian processes . . . 10
- 3.3 Kernel selection . . . 13
- 3.4 Deep gaussian processes . . . 14

4 Overview of programming frameworks **18**

5 Conclusions and future research **21**

6 References **22**

List of Figures

2.1 Reciprocity model, figure was taken from [39]. . . . 3
2.2 Digital pre-distortion (DPD) principle and learning architectures, figures were taken from [76] and [77]. . . . 6
2.3 Digital pre-distortion (DPD) behavioural model's terms, figures were taken from [79]. . . . 7

3.1 One dimensional samples drawn from different kernels with their associated covariance matrices, figure was taken from [129]. . . . 13
3.2 Different representations of the Deep GP model, figure was taken from [163]. . . . 15

4.1 Paper implementations grouped by framework, from November 2014 to August 2020, weekly scale. . . . 20
4.2 Percentage of papers with code support, from November 2014 to August 2020, weekly scale. . . . 20

List of Tables

3.1 Gaussian process (GP) properties 12

List of Acronyms and Abbreviations

"ml" machine learning . . . 1

AAC antenna array calibration . . . 3

AAS advanced antenna system . . . 1

ACLR adjacent channel leakage ratio . . . 7

ADC analog-to-digital converter . . . 3

AOA angle-of-arrival . . . 3

AR autoregressive . . . 1

ARD automatic relevance determination . . . 15

BER bit error rate . . . 1

BNP bayesian non-parametric . . . 10

BS base station . . . 1

CSI channel state information . . . 4

D-GP deep gaussian process . . . 12

DAC digital-to-analog converter . . . 3

DL downlink . . . 3

DNN deep neural network . . . 1

DPD digital pre-distortion . . . 5

DUT device under test . . . 8

EVM error vector magnitude . . . 7

FDD frequency division duplexing . . . 4

FITC fully independent training conditional . . . 12

GP gaussian process . . . 2

GP-LVM gaussian process latent variable model . . . 14

GPU graphics processing unit . . . 19

HPC high-performance computing . . . 19

IMD intermodulation distortion . . . 5

JIT just-in-time 19

LUT look-up-table 7

MCMC Markov chain Monte Carlo 12

MIMO multiple-input multiple-output 1

ML maximum likelihood 4

MSE mean squared error 4

NAS neural architecture search 8

PA power amplifier 1

PIM passive intermodulation 5

RF radio frequency 1

RIS reconfigurable intelligent surface 1

RL reinforcement learning 17

RX receiver 3

SLAM simultaneous localisation and mapping 17

SNR signal-to-noise ratio 1

TDD time division duplexing 3

TPU tensor processing unit 19

TX transmitter 3

UE user equipment 1

UL uplink 1

1. Introduction

1.1. The twofold problem

Development and deployment of any wireless system can't be accomplished without modelling of wireless channels between dedicated transmission points. Beyond 5G (6G) networks [1]–[4] are not exceptions. As radio frequency (RF) hardware complexity and scale of systems grow (e.g. massive MIMO [5], reconfigurable intelligent surfaces (RIS) [6] and other variations of advanced antenna systems (AAS) [7]), so do the same properties of radio channel models [8]–[11].
While we use channel simulations [12]–[14] to estimate metrics such as bit error rate (BER), throughput, latency and tune system parameters in different scenarios [15], [16], in some cases however performance evaluation based on models differs from real measurements. These divergences are commonly caused by uncorrected hardware impairments or their residuals [17]–[19], since in reality correction methods can rarely fully mitigate said distortions. Of course problem of modelling wireless systems with hardware impairments taken into account is not new and has been addressed in multiple studies. For example in [20] authors model all RF impairments as a single matrix embedded into pre-coding, which can be suitable for AAS calibration purposes, but not for modelling of non-linear dynamics[1] like intermodulations. In [23] term of residual transceiver hardware impairment is defined as an additive stochastic process. While the idea of impairment generalisation is important, this term doesn't include non-additive or multiplicative distortions. One of the key studies in this area, conducted by Björnson et. al. [24] analyses the capacity and channel estimation accuracy of massive MIMO systems with non-ideal transceiver hardware. There, impairments are modelled by an additive distortion noise that is proportional to the signal power at each antenna. Despite advantages like mathematical tractability, it would be difficult to estimate how individual distortions or their specific combinations effect the system's performance metrics. Demir and Björnson in [17] consider estimation of the effective channels with non-linear distortion characteristics of both base station (BS) and user equipment (UE) taken into account. Characteristics are modelled as quasi-memoryless polynomials and simulation tool is based on deep neural network (DNN) framework. Work is novel, but not without critical disadvantages: model can only support estimations based on fixed channel realisations and behavioural models of power amplifiers (PAs) (used by the authors) do not take into account memory effects. Paper by Mollén et. al. [25] presents a rigorous framework for analysis of non-linear effects caused by PAs in MIMO uplink (UL) with the presence of blockers. Work by Jiang et. al. [26] focuses on calibration of AAS. There, a theoretical analysis is conducted on the impact of calibration matrix on UL channel estimation accuracy. Zhang et. al. [27] study an impact of residual hardware impairments on the capacity of MIMO communication systems, especially on those operating at high signal-to-noise ratio (SNR), like high-rate systems. Authors derive an ergodic capacity expression for a MIMO system with residual transceiver impairments, which applies for any finite number of antennas and the entire SNR range. In works by Ali et. al. and Challita et. al. [28], [29], authors present useful overviews of future challenges in beyond 5G systems and how they can be addressed by machine learning ("ml") techniques. The last pair of studies, however only briefly introduces problem of hardware impairment mitigation, not including it's relation to wireless channel

[1] E.g. non-linear autoregressive (AR) models [21], [22].

modelling.
Many other works relevant to the subject exist and their analysis leads to following conclusion. On the one hand we can see that such solutions are tailored to specific cases and even attempts to generalise distortions are limited to individual classes (e.g. only additive, only memoryless or only stationary models). Thus, further complicating development of new models or extension of existing ones to ever growing number of beyond 5G scenarios [30], [31]. On the other hand, studies focusing on distortion mitigation techniques[2] rarely consider a rigorous analysis of effects on wireless link metrics.
Consequently, we ask – is it possible to treat the twofold problem of simulating both channel and impairments within a single mathematical framework? While supporting a set of conditions[3]. To answer this question in section 2 we first classify types and properties of hardware impairments, together with an overview of recent key correction techniques. Then, part 3 covers modelling methodology based on gaussian processes (GPs), including advanced methods like deep hierarchies and automatic model selection. Additionally, in section 4, we provide an overview of existing programming libraries for GP based modelling. Part 5 concludes this report with a summary and a set of possible future research directions.

[2]These studies are referenced in section 2.
[3]These conditions and reasons behind them are described in section 3.

2. RF impairments and channel models

2.1. Antenna array impairments

Despite relatively novel usage of AAS, problems which accompany them are not new and date back to radar systems from late 1990s [32]–[34]. Errors related to the electrical and geometrical characteristics of the antenna arrays, can severely degrade quality of beamforming, source detection and angle-of-arrival (AOA) estimation algorithms [26]. Electrical errors occur if analog RF connections on paths between radiating elements and analog-to-digital converter (ADC)/digital-to-analog converter (DAC) have different lengths (group delays, phase delays) and manufacturing imperfections resulting in multiplicative complex gain and phase distortions across signal's bandwidth. Array itself contributes to above-mentioned values if element placement differs from original design, causing parasitic mutual coupling effects [35]–[38]. Additionally, in time division duplexing (TDD) systems, reciprocity property of wireless channel breaks, as downlink (DL) and UL channel responses (during the same coherence interval) become different thus complicating wireless channel estimation process [39], [40].
Mathematical model for said impairments (in point-to-point (A and B) communication case) is easily defined as a set of equations 2.1, with visualisation provided in figure 2.1. Where N_A and N_B are antenna numbers, $C(t)$ – reciprocal effective electromagnetic channel and T_A, T_B, R_A, R_B – filters modelling linear RF imperfections of transmitter (TX) and receiver (RX). Also f_A and f'_A denote the up- and down-conversion frequencies at side A, whereas f_B and f'_B are up- and down-conversion frequencies at node B. In the time domain, a similar set of equations is obtained by replacing products by convolutions and matrices by linear filters in the equation set 2.1.

$$\begin{aligned} G(t,f) &= R_B(f)e^{2\pi j f'_B t}C(t,f)T_A(f)e^{-2\pi j f_A t} \\ H(t,f) &= R_A(f)e^{2\pi j f'_A t}C(t,f)^T T_B(f)e^{-2\pi j f_B t} \end{aligned} \tag{2.1}$$

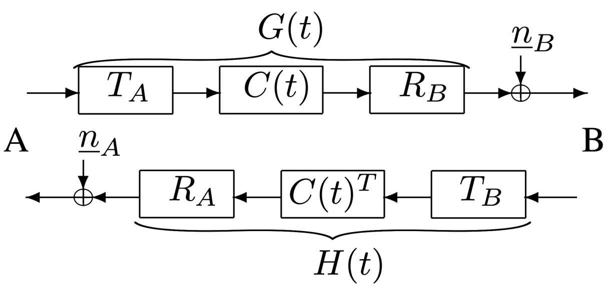

Figure 2.1: Reciprocity model, figure was taken from [39].

Considering a simple mathematical model, what makes antenna array calibration (AAC) a challenging task in modern MIMO systems? Two factors contribute to the answer: absence of feedback connections and non-stationary operating conditions. The first is technically challenging as individual feedbacks and respective ADCs would be required to accompany each antenna port. In massive MIMO cases, where antenna arrays may contain e.g. 512 radiating elements, feedbacks would result in internal RF interferences and unacceptably high hardware cost. The second exists because of changing operating temperature, ageing of RF hardware and varying output power. Thus making pre-recorded "factory" calibration tables [41] and similar "classical" methods [42] almost obsolete.

It is also important to distinguish between so-called relative and absolute calibrations [43]. Conditions of absolute calibration [44] require exact compensation for the imperfections of each RF chain independently. For example it can be achieved via external reference sources or (currently undesired) feedback RF chains. Relative calibration [45], [46], on the other hand, has more relaxed conditions. There parameters of imperfections can be estimated relatively to a chosen channel of the system or a reference source. Such calibration can be done via exchange of channel measurements between devices (BS, UE), self signal transmission or other similar techniques. These conditions (constraints) push researchers and engineers into development of novel solutions. Here we highlight most prominent examples of such studies.

Several of them, despite being old can be still considered relevant. For example methods proposed in studies by Manikas et. al. [47]–[49] are suitable for calibration of TXs, but can be applied to both TDD and frequency division duplexing (FDD) systems. Among other advantages are scalability and absence of hardware requirements for it's operation, like antenna element symmetries [50]. Solution requires knowledge of signal source AOA, given this information expected received signals of an ideal array can be computed. Comparing these with the actual received signal, it is possible to derive the difference due to the uncertainties and correct them. Modern iterations of these algorithms also consider a combined estimation of AOA and imperfections [51].

Another methodology exploits parasitic mutual coupling effects and symmetries in antenna array designs. For example in [52] authors show a fully digital AAS capable of maintaining its initial calibration by monitoring its quadrature imbalance with mutual coupling based measurements. In their case symmetry of antenna array is not required, but only relative calibration can be performed, without array's geometrical imperfections taken into account. A most prominent example in this category is a work done by Vieira et. al. [53], [54]. There, an iterative maximum likelihood (ML) asymptotically efficient algorithm is proposed. It outperforms existing estimators in mean squared error (MSE) and sum-rate capacity sense. Authors also verify performance on massive MIMO test-bench and analyse statical properties of calibration errors.

Rest of the modern methods rely on exploitation of existing system information (e.g. channel state information (CSI)) or modifications of wireless transmission standardisation. In [55] Moon et. al. introduce new online "over-the-air" calibration method which relies on channel estimation in wireless communication in order to measure the magnitude and phase response of the phased array system. Therefore, calibration can be online without a necessity to pause the communication and does not need RF feedback circuitry. While method was verified experimentally it is not clear whenever it can provide absolute calibration or not. In another study, Shan et. al. [56] develop a DNN architecture[1] to diagnose the calibration state of a massive antenna array. There, model is designed to learn an optimized linear DL pilot matrix and a non-linear reconstruction mapping function from measurements to the original sparse RF impairments. Interested reader will get additional insight into array calibration methods from following studies: [58]–[61].

With further literature analysis it becomes apparent, that it is impossible to satisfy all calibration constraints (from the start of the section) with existing methods in pursue of absolute calibration. We can however go back to principles of table-based calibration method as we know that

[1]Original idea to use DNNs for calibration was proposed by Bertrand et. al. [57] but it was applied in a different way, compared to [56].

it provides absolute calibration in stationary setting and analyse how it can be improved. Hypothetically if we measure not only phase and amplitude components of RF channels, but also same parameters during different operating conditions it would result in calibration table with all possible system states. Obviously it would be impractical for massive MIMO systems as a result would form a multidimensional array because of experimental test-bench complexities and number of state measurements. In this case applicability can be achieved with sparsification of measurements, smart interpolation to restore skipped measurements and prediction of temporal (sequential) system states (temperature, output power, ageing effects). Such properties can only be found in hierarchical models with universal functional approximation properties.

2.2. Intermodulation distortions

The largest impact on wireless links' performance is caused by intermodulation distortions (IMDs). These effects are caused by a non-linear processes in analog RF hardware and result in in-band distortions, spectral regrowth, unwanted spurious emissions, increased occupied bandwidth, adjacent channel interference and most importantly reduction of effective throughput [62]–[64]. Two primary sources of IMDs exist: active and passive. The first are caused by PAs [65], [66], in TXs of both TDD and FDD systems. The second occur in RXs of FDD systems and are generated in defective RF components of RX RF chains. Passive intermodulations (PIM) can originate from different sources [67]–[73]:

- shared circuitry (e.g. duplexer);
- materials exhibiting hysteresis upon exposure to reversing magnetic fields;
- components with manufacturing defects for example a cracked solder joints or poorly made mechanical contacts;
- RF connectors or when conductors made of two galvanically unmatched metals come in contact with each other;
- metal flakes or shavings inside RF connections;
- loose mechanical connections effected by mechanical vibrations;
- metallic objects (corrupted by rust, moisture or oxidation) in near field or beam directions (external sources);

Despite the fact that IMD sources and their relations to useful signals are different (PIM is an additive distortion) modelling methods are similar. Considering this, in context of this report we will only cover aspects of PA IMD correction process commonly known as digital pre-distortion (DPD). Operating principles of DPD are based on inverse modelling of PA underlying functional and feedback parameter estimation process. In it a portion of the output signal from PA is fed back and subtracted from the original signal to force the output to be a linear replica of the input signal [74]. Feedforward linearisation is similar, but instead of adding correction signal to a PA input, it is added afterwards [75]. Illustrations of both approaches are given in figure 2.2. From system identification perspective RF front-end with multiple PAs is a non-linear non-stationary dynamical system [22]. For example short-term memory effects are caused by

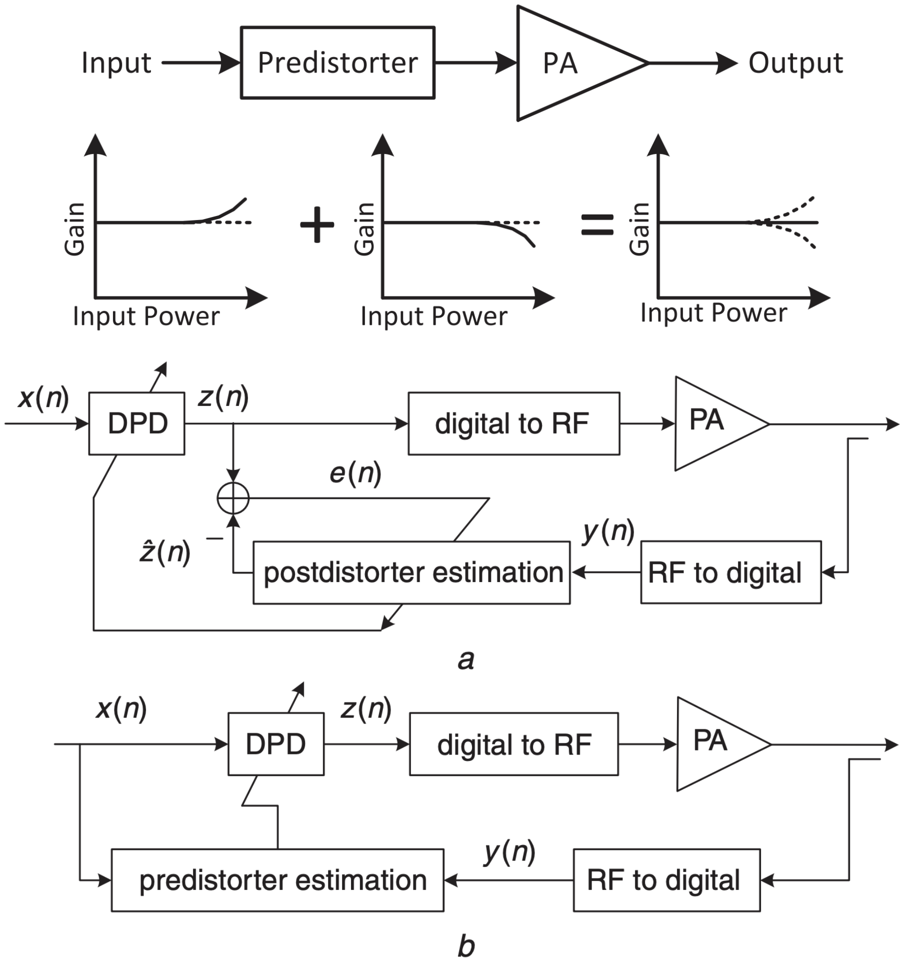

Figure 2.2: DPD principle and learning architectures, figures were taken from [76] and [77].

input and output matching circuitry, long term effects by bias networks and non-stationary non-linearities by active components (transistors) [78], [79].
Since IMDs components produced by PAs of previous wireless systems generations can be derived analytically [80]–[82], it is possible to manually select a structure of parametric model and leave parameter estimation for algorithms. Essentially making DPD a "grey-box" parameter estimation task.
But for beyond 5G more appropriate term would be a "black-box" model selection, as IMD generation processes become analytically intractable. In other words: it is difficult to transform a prior domain knowledge into interpretable models. This statement is supported by increased complexities of modern PAs and challenging operating scenarios. Those include:

- multi-band and ultra wide-band input signals which lead to high order IMDs and their combinations [83]–[88];
- dynamic operation modes, defined by abrupt changes in absolute power and configuration of input signals [76], [89], [90];
- massive MIMO scenarios raise issues of cross-channel leakages, cross-talks and model scalability [91]–[95];

In addition we also have two constraints. Firstly the model and it's adaptation must be done in online fashion or in "ml" terms – support continual learning [96]. Secondly input (excitation) signals can not be directly controlled, limiting number of applicable system identification methods. Of course it is difficult to satisfy all above-mentioned requirements in context of a single model. This is a main reason why DPD has many specialised solutions ranging from behavioural to beamforming methods. While the latter are interesting, goal of spatial (beam domain) DPD [97]–[99] is to form a beam pattern in which IMDs are "sent" in directions free of UEs or other RX equipment. Partly this method relies on existing PA models (not derivation of new ones) because of that we focus on overview of behavioural models in this report.
Throughout the years multiple formulations of behavioural models have been proposed. They can be categorized according to several criteria, such as the inclusion or exclusion of memory effects, types of non-linearities and number of compositional terms in the model [79]. Figure 2.3 contains examples of such terms: look-up-tables (LUTs) (adaptive non-linearities) [100], branching unit delay lines, memory polynomials [87], Volterra series and Wiener-Hammerstein [101] combinations.

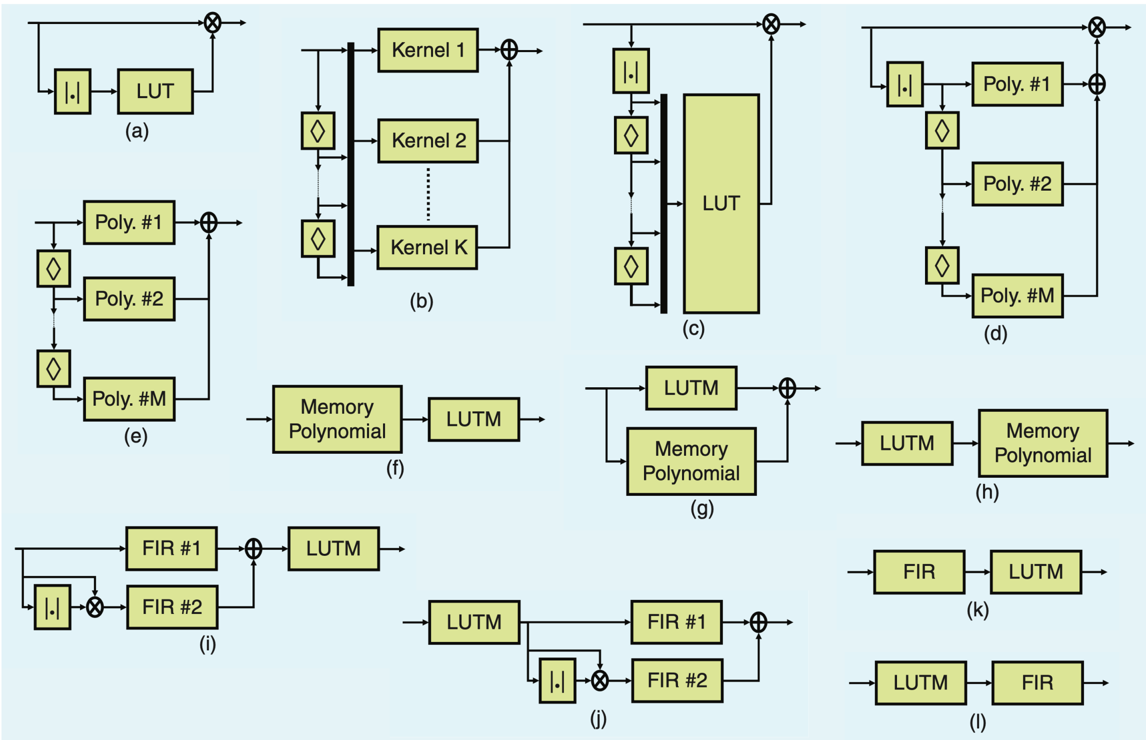

Figure 2.3: DPD behavioural model's terms, figures were taken from [79].

With growing popularity of DNNs researchers began to apply them to DPD tasks. These contributions range from simple experimental studies [102] to proposals of more elaborate architectures [103].
In [104] authors explore complexity and trade-offs between DNN and memory polynomial DPDs based on adjacent channel leakage ratio (ACLR) and error vector magnitude (EVM) metrics. Experimental results also demonstrated that former requires less hardware utilization when compared to a similarly performing polynomial model. Wu et. al. [105] test residual DNN architecture [106] and provide detailed performance comparisons with other models like time delay DNN by Wang et. al. [107]. In short study by Hongyo et. al. [108], authors compare

ACLR for different activation functions and network dimensions. Finally, in the attempt to support MIMO scenario and additional impairments (crosstalk, I/Q imbalance, DC offset) Jaraut et. al. [109] show DNNs having better generalisation properties when compared to traditional models.

Countless other works in this domain exist and while many specialists consider DNN to be generalisations over behavioural models, they are not exactly correct. From statistical "ml" perspective selection of such architectures is a way to select priors on functions [110] or a way to embed domain knowledge about the data. From functional analysis side, both model types are parametric compositional (hierarchical) functional representations and as such they share same disadvantages. Namely: generalisation, scalability and lack of generative properties[2].

All DPD models are capable of compensation IMDs, but each of them is designed to tackle a specific scenario. Consider a multi-band model. Naturally it should be applicable to single-band signals but there are no guarantees of it being useful for dynamic power operation mode. In same way not every MIMO model can be scaled up to a massive MIMO.

Thus process of DPD model selection is inherently redundant, repeating with each new PA model and signal configuration. Eventually RF systems may grow too complex, making it impossible to manually find parametric model structure. Solution to this problem can be found in neural architecture search (NAS) [111], however it will only produce a discriminative model, unsuitable for sampling or random realisations.

If our gaol is to combine channel and impairment models in a context of a single simulator, we have to use a modelling framework designed to address these disadvantages.

2.3. Channel modelling

Before proceeding with selection of a modelling methodology, below we provide a list of relevant wireless channel simulators.

1. QuaDRiGa — [12], [112] — Supports wide range of beyond 5G cases (e.g. vehicular and industrial), three dimensional propagation, continuous channel evolution and transition between propagation scenarios.

2. NYUSIM — [113], [114] — Supports spatial consistency, human blockage and O2I penetration loss. The multiple reflection surfaces method is used to update small-scale parameters in the spatial consistency procedure. Four-state Markov model simulates human blockage events. O2I penetration loss use a parabolic model, useful for indoor scenarios.

3. NYU Real-time Massive MIMO Channel Emulation — [115], [116] — Channel simulation is combined with real operating RF hardware. Test-bench includes both baseband and RF front-end devices under test (DUT)[3].

4. PyLayers — Site specific radio channel simulator. Supports custom antenna patterns and wireless radio standards.

[2]For usage in wireless channel simulators

[3]Researchers who are specifically interested in PA modelling may use RF WebLab (hosted by Chalmers University of Technology) to acquire respective datasets.

5. PyCraf — [117] — Package for functions and procedures related to spectrum-management compatibility studies. Includes implementations of ITU-R recommendations for calculation of path attenuations and interference levels.

None of them contains built-in support for simulated impairment (and correction) models. It can be argued that simulations must stay decoupled or even separated, but from our perspective this eventually leads to unnecessary complexity and unreproducible results. Especially when specialists research effects of RF hardware imperfections on link's performance metrics.
In summary, we see that question of joint modelling of RF impairments with wireless channels remains open, as current academic channel simulators do not include such distortion models and industrial versions are not available for public. To solve this problem we (further) propose a data-driven methodology, capable of solving this task.

3. Gaussian processes

3.1. Universal functional approximation

Variety of distortion classes, wireless channel scenarios and desire to model them in the same framework calls for a modelling tool with a universal functional approximation property. The idea of the universal modelling can be traced back directly to Hilbert's thirteenth problem [118] and it's solution in a form of Kolmogorov-Arnold representation theorem [119], [120]. This notion, together with evidence we've provided in section 3 leads to a belief that DNN is a suitable modelling framework for a "full channel"[1] modelling problem. But are DNNs right modelling tools for our task? Can they satisfy our requirements, are there any alternatives?

- Full channel model must be a controllable generative model from which we can sample new realisations (examples) of wireless channel, whether it is an impulse response or complex valued time series samples.
- If we want to update the model with a new type of distortion – chosen framework must be able to incorporate it in a data-driven fashion.
- Same applies to novel channel scenarios, like dynamic UE or non-stationary environment.
- Due to various constraints it is difficult to obtain large training dataset in domain of wireless communications, so a modelling tool must support small datasets and continual learning [96], [121], [122].
- Consequently model should be prune to overfitting [123].
- Also we would like to assess (and calibrate) model discrepancy, address mis-specification, perform uncertainty quantification and if possible have an interpretable model [124].

With these conditions taken into account it becomes apparent that instead of DNN optimal choice would be to use bayesian non-parametric (BNP) modelling, specifically – GPs [125]–[128][2]. Further we describe basics of GP based modelling and it's advanced extensions: kernel search and hierarchical GP models. As all of them are general purpose "ml" algorithms, meaning that they can be applied to any data-driven problem, including "full channel" modelling.

3.2. Gaussian processes

GPs are a simple and general class of models of functions, in other words a GP is any distribution over functions such that any finite set of function values $f(\mathbf{x}_1), f(\mathbf{x}_2), \dots f(\mathbf{x}_N)$ have a joint

[1]For consistency we will refer to modelling problem of a wireless channel together with RF hardware impairments as a "full channel" modelling.

[2]Extended information from following subsections can also be found in theses of Duvenaud, Daminau, Vafa and Frigola [125], [129]–[131]

Gaussian distribution [123]. A GP model, before conditioning on data, is fully specified by its mean and covariance functions (kernels).

$$\mathbb{E}[f(\mathbf{x})] = \mu(\mathbf{x}) \tag{3.1}$$

$$\operatorname{Cov}\left[f(\mathbf{x}), f\left(\mathbf{x}^{\prime}\right)\right] = k\left(\mathbf{x}, \mathbf{x}^{\prime}\right) \tag{3.2}$$

Usually the mean function is assumed to be zero, since uncertainty about the mean function can be taken into account by adding an extra term to the kernel. After accounting for the mean, the kind of structure that can be captured by a GP model is entirely determined by its kernel. The kernel determines how the model generalizes to new observations. Choice of covariance function is incredibly large, such that a correctly chosen kernel function can specify a wide range of models. For example, linear regression, splines, Kalman filters and even DNNs are examples of GPs with specific kernels or their compositions.
The crucial property of GPs that allows us to automatically construct models is that we can compute the marginal likelihood of a dataset given a particular model, also known as the evidence [132]. The marginal likelihood allows one to compare models, balancing between the capacity of a model and its fit to the data [133], [134]. The marginal likelihood under a GP prior of a set of function values $\left[f\left(\mathbf{x}_1\right), f\left(\mathbf{x}_2\right), \ldots f\left(\mathbf{x}_N\right)\right] := \boldsymbol{f}(\mathbf{X})$ at locations $\mathbf{X}$ is given by:

$$\begin{aligned} p(\boldsymbol{f}(\mathbf{X}) \mid \mathbf{X}, \mu(\cdot), k(\cdot, \cdot)) &= \mathcal{N}(\boldsymbol{f}(\mathbf{X}) \mid \mu(\mathbf{X}), k(\mathbf{X}, \mathbf{X})) \\ &= (2\pi)^{-\frac{N}{2}} \times \underbrace{|k(\mathbf{X}, \mathbf{X})|^{-\frac{1}{2}}}_{\text{controls model capacity}} \\ & \underbrace{\times \exp\left\{-\frac{1}{2}(\boldsymbol{f}(\mathbf{X}) - \boldsymbol{\mu}(\mathbf{X}))^{\top} k(\mathbf{X}, \mathbf{X})^{-1}(\boldsymbol{f}(\mathbf{X}) - \boldsymbol{\mu}(\mathbf{X}))\right\}}_{\text{encourages fit with data}} \end{aligned} \tag{3.3}$$

This multivariate Gaussian density is referred to as the marginal likelihood because it implicitly integrates (marginalizes) over all possible functions values $\boldsymbol{f}(\overline{\mathbf{X}})$, where $\overline{\mathbf{X}}$ is the set of all locations where we have not observed the function.
To make prediction we can "ask the model" which function values are likely to occur at any location, given all (or a finite set of) previous observations. By the formula for Gaussian conditionals the predictive distribution of a function value $f\left(x^{\star}\right)$ at a test point $x^{\star}$ has the form of:

$$\begin{aligned} p\left(f\left(\mathbf{x}^{\star}\right) \mid \boldsymbol{f}(\mathbf{X}), \mathbf{X}, \mu(\cdot), k(\cdot, \cdot)\right) = \mathcal{N}\big(f\left(\mathbf{x}^{\star}\right) \mid & \underbrace{\mu\left(\mathbf{x}^{\star}\right) + k\left(\mathbf{x}^{\star}, \mathbf{X}\right) k(\mathbf{X}, \mathbf{X})^{-1}(\boldsymbol{f}(\mathbf{X}) - \mu(\mathbf{X}))}_{\text{predictive mean follows observations}}, \\ & \underbrace{k\left(\mathbf{x}^{\star}, \mathbf{x}^{\star}\right) - k\left(\mathbf{x}^{\star}, \mathbf{X}\right) k(\mathbf{X}, \mathbf{X})^{-1} k\left(\mathbf{X}, \mathbf{x}^{\star}\right)}_{\text{predictive variance shrinks given more data}}\big) \end{aligned} \tag{3.4}$$

Sampling a function from a GP is also straightforward. A sample from a GP at a finite set of locations is just a single sample from a multivariate Gaussian distribution, given by previous equation. Probabilistic perspective does not necessarily mean that we are assuming the function being learned is stochastic or random, but it helps with uncertainty quantification and calibration.

To conclude this section, in table 3.1 we compare advantages and limitations of GP based modelling.

Table 3.1: GP properties

Advantages	**Limitations**
Analytic inference: predictive posterior distribution can be computed exactly in closed form, rare property, even for BNP models.	**Gaussian predictive distribution**: in some cases we want to use non-gaussian predictive distribution, for example a classification task.
Integration over hypotheses: overfitting isn't an issue, compared to parametric models.	**Slow Inference**: matrix inverse time – $\mathcal{O}(N^3)$, where N is number of samples.
Models are analytical.	**Manual kernel selection**.
Model selection: computation of marginal likelihood of the data given a model allows model comparison.	
Closed-form predictive distribution: GPs can be combined with other models.	
Expressivity: model wide range of functions.	

At first glance it may seem, that limitations 1 and 2 overshadow all advantages, but it's not true, as solution lies in usage of approximate inference methods [135]–[139]. Such methods leave the prior distribution of the GP model unchanged and instead enforce sparse structures in the posterior approximation though variational inference. This gives $\mathcal{O}(M^2N + M^3)$ computation and $\mathcal{O}(MN + M^2)$ storage with M inducing points. Moreover, they allow to perform mini-batch training by sub-sampling data points. Their detailed description can be found in works of Titsias and Hensman et. al. [140]–[143]. Technique also opens the way for new types of models, such as GP models suitable for billion point datasets [144] or convolutional GPs [145], where translation invariance is encoded by summing over GPs that take image patches as inputs. In [146] Keyon Vafa proposed a deep gaussian process (D-GP) sampling algorithm based on Markov chain Monte Carlo (MCMC) sampling to circumvent the intractability hurdle. There predictive means and covariances are sampled to approximate the marginal likelihood, relying on automatic differentiation techniques to evaluate the gradients and optimize given objective. As a part of the procedure, every GP is replaced with the fully independent training conditional (FITC) GP [147], so the time complexity for layers (L) and nodes per layer (H) is $\mathcal{O}(N^2MLH)$ as opposed to $\mathcal{O}(N^3LH)$ Recently, Shi et. al. have introduced a new GP framework, allowing to increase the number of inducing points under a fixed computational budget. It is based on decomposing the GP prior as the sum of a low-rank approximation using inducing points, and a full-rank residual process [148].
In summary, all disadvantages of GP models can be addressed in multiple ways, the choice of solution however, must depend on the modelling problem itself.

3.3. Kernel selection

. . . .

To successfully apply kernel learning algorithms, users have to specify the parametric form of the kernel and this requires a prior knowledge. In other words considerable expertise in the specific task. But even with it, selection may very well become a trial and error process.
Thankfully, it is possible to reformulate the kernel learning problem as one of structure discovery, and automate the choice of kernel form [149]. Given properties of kernels, search space can be defined compositionally in terms of sums and products of a small number of base kernel structures. This provides an expressive modelling language which concisely captures many widely used techniques for constructing kernels. While, main work in this direction is being conducted as a part of the "Automated statistician" project[3], key principles of the kernel search are described in study by Duvenaud et. al. [149]. We summarise them below.
Firstly, four base kernel families are considered: squared-exponential, periodic, linear and rational-quadratic [123] (visualisation of sampled functions, covariance matrices and one composite kernel can be seen in figure 3.1). So, any algebraic expression combining these kernels using sums and products defines a kernel family, whose parameters are the concatenation of the parameters for the base kernel families.

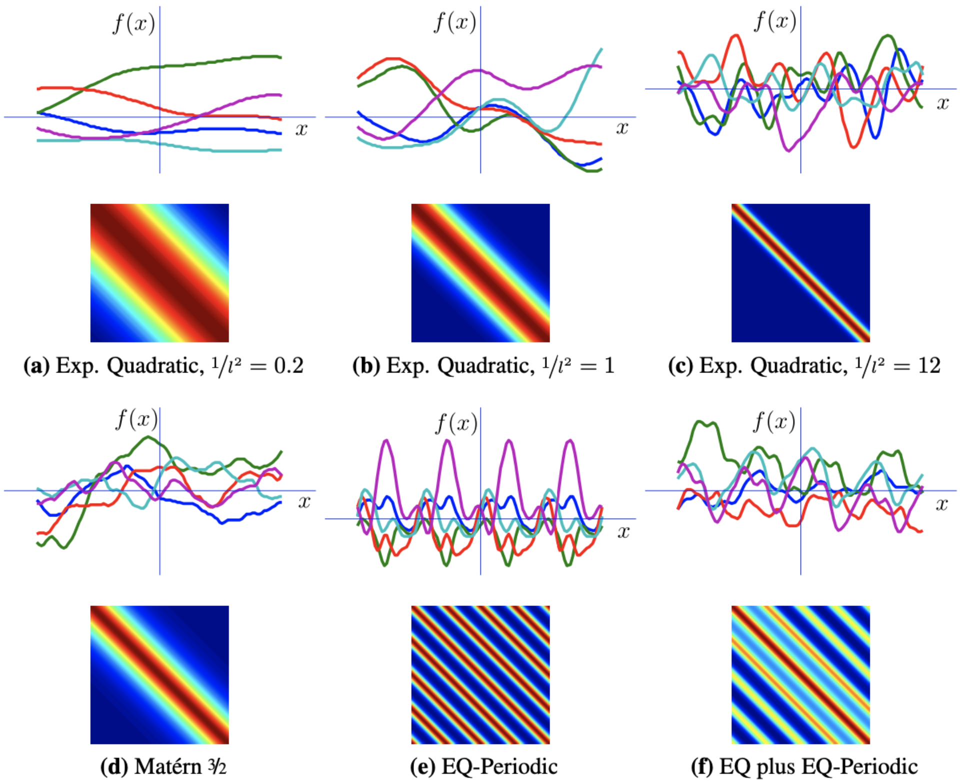

Figure 3.1: One dimensional samples drawn from different kernels with their associated covariance matrices, figure was taken from [129].

Secondly, search procedure begins by proposing all base kernel families applied to all input dimensions. Sub-expression term $\mathcal{S}$ is used to describe any new kernel combination and $\mathcal{B}$

[3] Automated statistician project official website – automaticstatistician.com

denotes any base kernel family.

1. Any $\mathcal{S}$ can be replaced with $\mathcal{S} + \mathcal{B}$.
2. Any $\mathcal{S}$ can be replaced with $\mathcal{S} \times \mathcal{B}$.
3. Any $\mathcal{B}$ may be replaced with any other $\mathcal{B}'$.

Algorithm specified in original authors study, searches over this space using a greedy search: at each stage, it chooses the highest scoring kernel and expands it by applying all possible operators. Search operators are motivated by strategies researchers (engineers) often use to construct kernels, for example:

- Rule 1: One can look for structure, e.g. periodicity, in the residuals of a model, and then extend the model to capture that structure.
- Rule 2: One can start with structure, e.g. linearity, which is assumed to hold globally, but find that it only holds locally.
- Rules 1 and 2: One can add features incrementally, analogous to algorithms like boosting or stepwise regression.

Finally, marginal likelihood is computed as kernel score evaluation criterion [134]. Combining analytical solution for marginal likelihood of a GP and approximate integration over kernel parameters with Bayesian information criterion [150].
Additional information can be found in overview paper by Steinruecken et. al. [151], studies of kernels structured and learnt as DNNs [152]–[154] and recent algorithm scalability research by Hyunjik Kim and Yee Whye Teh [155].

3.4. Deep gaussian processes

Sometimes in cases of non-stationary, non-gaussian data or fitting of a function with complex hierarchical behaviour, simple GPs may underperform. DNNs, however, can learn complex hierarchies, but by using them we loose benefits of BNP framework. Solution which not only combines benefits of both modelling frameworks, but also serves as a generalisation over DNNs[4] [156]–[162] and single GPs is a D-GP.
Idea of D-GPs was originally introduced by Andreas C. Damianou and Neil D. Lawrence [163] as a continuation of works on gaussian process latent variable models (GP-LVMs) [164], [165]. Authors prove that through variational approximations any number of GP models can be sequentially connected to give truly deep hierarchies. The variational approach gives a rigorous lower bound on the marginal likelihood of the model, allowing it to be used for model selection. Same lower bound allows to apply D-GP models even when data is scarce and gives an objective measure from which different structures for deep hierarchy can be selected (e.g. number of layers).
D-GP architecture corresponds to a graphical model with three node types (figure 3.2 (a)). Observations – leaf nodes $\mathbf{Y} \in \mathcal{R}^{N \times D}$, latent spaces – $\mathbf{X}_h \in \mathcal{R}^{N \times Q_h}, h = 1, \ldots, H-1$ and

[4] It is worth noting that proof of GPs model generalisation capabilities are proved by both probability theory and functional analysis.

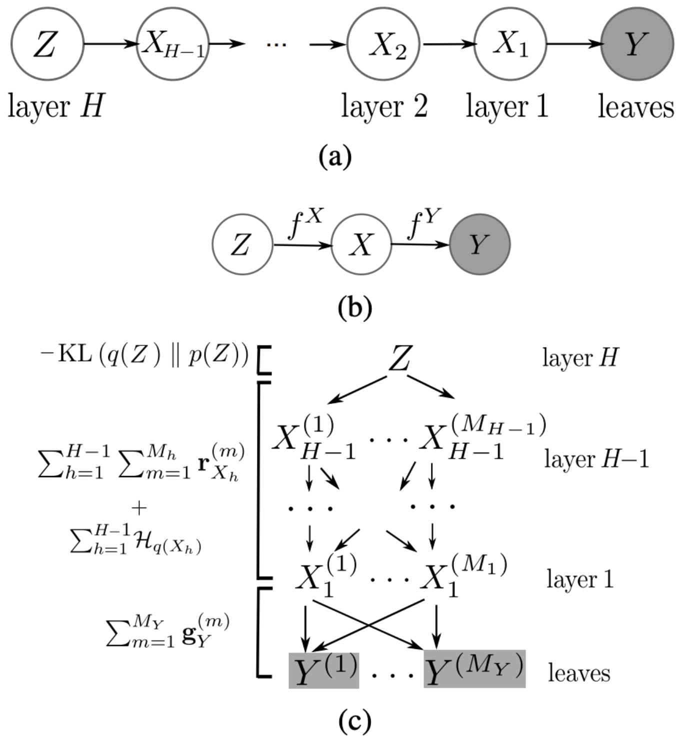

Figure 3.2: Different representations of the Deep GP model, figure was taken from [163].

parent latent nodes – $\mathbf{Z} = \mathbf{X}_H \in \mathcal{R}^{N \times Q_Z}$. For example if we consider a two-layer D-GP, it's generative process can be expressed as (3.5). The intermediate node is involved in two GPs. where $f^Y \sim \mathcal{GP}\left(\mathbf{0}, k^Y(\mathbf{X},\mathbf{X})\right)$ and $f^X \sim \mathcal{GP}\left(\mathbf{0}, k^X(\mathbf{Z},\mathbf{Z})\right)$ are playing roles of an input and an output.

$$\begin{aligned} y_{nd} &= f_d^Y\left(\mathbf{x}_n\right) + \epsilon_{nd}^Y, \quad d = 1,\ldots,D, \quad \mathbf{x}_n \in \mathcal{R}^Q \\ x_{nq} &= f_q^X\left(\mathbf{z}_n\right) + \epsilon_{nq}^X, \quad q = 1,\ldots,Q, \quad \mathbf{z}_n \in \mathcal{R}^{Q_z} \end{aligned} \tag{3.5}$$

To deal with resulting high number of model's hyper-parameters, latent space is marginalised [134]. As part of this process authors introduce automatic relevance determination (ARD) covariance function (3.6). It assumes a different weight w_q for each latent dimension and if weight of corresponding dimension becomes zero, model's size is reduced and structure updated.

$$k\left(\mathbf{x}_i, \mathbf{x}_j\right) = \sigma_{ard}^2 e^{-\frac{1}{2}\sum_{q=1}^{Q} w_q (x_{i,q} - x_{j,q})^2} \tag{3.6}$$

Next, we will describe main points in bayesian training procedure for D-GP models[5], as it is the most important modelling part.

First, we specify model evidence and prior (3.7) which will be optimised. In general case, integral is intractable [163], [164]. Next, Jensen's inequality is applied to find variational lower bound $\mathcal{F}_v$ (3.8), where $\mathcal{Q}$ is variational distribution. Decomposition of joint distribution (3.9) in numerator is still intractable due to non-linearities of $p\left(\mathbf{F}^Y \mid \mathbf{X}\right)$ and $p\left(\mathbf{F}^X \mid \mathbf{Z}\right)$ terms. But from [165], we know that probability space of GP prior $p(\mathbf{F} \mid \mathbf{X})$ can be expanded with pseudo-inputs or inducing points, resulting in augmented probability space (3.10), where $\mathbf{U}^Y \in \mathcal{R}^{K \times D}$ and $\mathbf{U}^X \in \mathcal{R}^{K \times Q}$ are additional values and K denotes a number of inducing points. Finally,

[5]Interested reader will find a more rigorous description in [129], [163].

combination of $\mathcal{Q}$ and (3.10) gives a a tractable variational bound (3.11), where $q\left(\mathbf{U}^Y\right)$ and $q\left(\mathbf{U}^X\right)$ are free-form variational distributions and $q(\mathbf{X})$, $q(\mathbf{Z})$ are factorised according to (3.12).

$$\log p(\mathbf{Y}) = \log \int_{\mathbf{X},\mathbf{Z}} p(\mathbf{Y} \mid \mathbf{X}) p(\mathbf{X} \mid \mathbf{Z}) p(\mathbf{Z}), \quad p(\mathbf{Z}) = \mathcal{N}(\mathbf{Z} \mid \mathbf{0}, I) \tag{3.7}$$

$$\mathcal{F}_v \leq \log p(\mathbf{Y}), \quad \mathcal{F}_v = \int_{\mathbf{X},\mathbf{Z},\mathbf{F}^Y,\mathbf{F}^X} \mathcal{Q} \log \frac{p\left(\mathbf{Y}, \mathbf{F}^Y, \mathbf{F}^X, \mathbf{X}, \mathbf{Z}\right)}{\mathcal{Q}} \tag{3.8}$$

$$\begin{gathered} p\left(\mathbf{Y}, \mathbf{F}^Y, \mathbf{F}^X, \mathbf{X}, \mathbf{Z}\right) = \\ p\left(\mathbf{Y} \mid \mathbf{F}^Y\right) p\left(\mathbf{F}^Y \mid \mathbf{X}\right) p\left(\mathbf{X} \mid \mathbf{F}^X\right) p\left(\mathbf{F}^X \mid \mathbf{Z}\right) p(\mathbf{Z}) \end{gathered} \tag{3.9}$$

$$\begin{aligned} p\left(\mathbf{Y}, \mathbf{F}^Y, \mathbf{F}^X, \mathbf{X}, \mathbf{Z}, \mathbf{U}^Y, \mathbf{U}^X, \tilde{\mathbf{X}}, \tilde{\mathbf{Z}}\right) = {} & p\left(\mathbf{Y} \mid \mathbf{F}^Y\right) p\left(\mathbf{F}^Y \mid \mathbf{U}^Y, \mathbf{X}\right) p\left(\mathbf{U}^Y \mid \tilde{\mathbf{X}}\right) \\ & \cdot p\left(\mathbf{X} \mid \mathbf{F}^X\right) p\left(\mathbf{F}^X \mid \mathbf{U}^X, \mathbf{Z}\right) p\left(\mathbf{U}^X \mid \tilde{\mathbf{X}}\right) p(\mathbf{Z}) \end{aligned} \tag{3.10}$$

$$\mathcal{Q} = p\left(\mathbf{F}^Y \mid \mathbf{U}^Y, \mathbf{X}\right) q\left(\mathbf{U}^Y\right) q(\mathbf{X}) p\left(\mathbf{F}^X \mid \mathbf{U}^X, \mathbf{Z}\right) q\left(\mathbf{U}^X\right) q(\mathbf{Z}) \tag{3.11}$$

$$q(\mathbf{X}) = \prod_{q=1}^{Q} \mathcal{N}\left(\boldsymbol{\mu}_q^X, \mathbf{s}_q^X\right), q(\mathbf{Z}) = \prod_{q=1}^{Q_z} \mathcal{N}\left(\boldsymbol{\mu}_q^Z, \mathbf{S}_q^Z\right) \tag{3.12}$$

With few additional transforms, variational lower bound can be written in analytical (3.13) and computationally feasible (3.14,3.15) forms. There, $\mathcal{H}$ denotes the entropy with respect to a distribution, KL – the Kullback-Leibler divergence [166] and $\langle\cdot\rangle$ stands for expectations. Terms $\mathbf{g}_Y$ and $\mathbf{r}_X$ are associated with D-GP model's vertical span (nodes per layer) and depth (number of layers).

$$\mathcal{F}_v = \int \mathcal{Q} \log \frac{p\left(\mathbf{Y} \mid \mathbf{F}^Y\right) p\left(\mathbf{U}^Y\right) p\left(\mathbf{X} \mid \mathbf{F}^X\right) p\left(\mathbf{U}^X\right) p(\mathbf{Z})}{\mathcal{Q}'}, \quad \mathcal{Q}' = q\left(\mathbf{U}^Y\right) q(\mathbf{X}) q\left(\mathbf{U}^X\right) q(\mathbf{Z}) \tag{3.13}$$

$$\mathcal{F}_v = \mathbf{g}_Y + \mathbf{r}_X + \mathcal{H}_{q(\mathbf{X})} - \mathrm{KL}(q(\mathbf{Z}) \| p(\mathbf{Z})) \tag{3.14}$$

$$\begin{aligned} \mathbf{g}_Y &= g\left(\mathbf{Y}, \mathbf{F}^Y, \mathbf{U}^Y, \mathbf{X}\right) = \left\langle \log p\left(\mathbf{Y} \mid \mathbf{F}^Y\right) + \log \frac{p\left(\mathbf{U}^Y\right)}{q\left(\mathbf{U}^Y\right)} \right\rangle_{p(\mathbf{F}^Y \mid \mathbf{U}^Y, \mathbf{X}) q(\mathbf{U}^Y) q(\mathbf{X})} \\ \mathbf{r}_X &= r\left(\mathbf{X}, \mathbf{F}^X, \mathbf{U}^X, \mathbf{Z}\right) = \left\langle \log p\left(\mathbf{X} \mid \mathbf{F}^X\right) + \log \frac{p\left(\mathbf{U}^X\right)}{q\left(\mathbf{U}^X\right)} \right\rangle_{p(\mathbf{F}^X \mid \mathbf{U}^X, \mathbf{Z}) q(\mathbf{U}^X) q(\mathbf{X}) q(\mathbf{Z})} \end{aligned} \tag{3.15}$$

If user wants to modify the model's hierarchy, variational bound must be updated accordingly. For example, a deeper model (more layers) will only require additional $\mathbf{r}_X$ terms. Resulting in a sum $\sum_{h=1}^{H-1} \mathbf{r}_{X_h}$, where $\mathbf{r}_{X_h} = r\left(X_h, \mathbf{F}^{X_h}, \mathbf{U}^{X_h}, \mathbf{X}_{h+1}\right), \mathbf{X}_H = \mathbf{Z}$. Vertical expansion (equivalent to increasing a number of neurons in DNN), under independence assumption $\log p\left(\mathbf{X}_h \mid \mathbf{X}_{h+1}\right) = \sum_{m=1}^{M_h} \log p\left(\mathbf{X}_h^{(m)} \mid \mathbf{X}_{h+1}\right)$ would transform $\mathbf{r}_{X_h}$ term into $\sum_{m=1}^{M_h} \mathbf{r}_{X_h}^{(m)}$. Combined expansion can be written as (3.16) and it's visualisation is provided in figure 3.2 (c). Further analytical study of D-GPs can found on studies of Dunlop et. al. [167].

$$\mathcal{F}_v = \sum_{m=1}^{M_Y} \mathbf{g}_Y^{(m)} + \sum_{h=1}^{H-1} \sum_{m=1}^{M_h} \mathbf{r}_{X_h}^{(m)} + \sum_{h=1}^{H-1} \mathcal{H}_{q(\mathbf{x}_h)} - \mathrm{KL}(q(\mathbf{Z}) \| p(\mathbf{Z})) \tag{3.16}$$

As a tool for engineers and researchers, D-GPs (and GPs) found successful applications in different areas: surrogate modelling of physical systems [168], inverse reinforcement learning (RL) [169], image classification [170], source localisation [171], time series prediction for hardware degradation [172], survival analysis with competing risks [173], simultaneous localisation and mapping (SLAM) [174], astrophysics and astronomy [175], signal processing [176]–[180], system identification and control [181]–[185] and many others.
These works serve as evidence to a notion of universal modelling property of D-GPs. Supporting our hypothesis on joint modelling of RF hardware impairments with wireless channel.

4. Overview of programming frameworks

Given the complexity of the problem, developers may encounter issues during implementation of proposed GP models. Even though knowledgeable users may already know how to work with probabilistic programming packages, we believe that it would be beneficial to overview various software options for practitioners who may not have specific knowledge or are new to the field. Thus, we select open source projects which are being continuously updated and have an supporting documentation (with examples). It is worth pointing out that usage of common open source tools among researchers helps to avoid unnecessary repetition of existing works and encourages reproducibility of simulation (experimental) results. Despite multiple benefits of "paper + code" paradigm only a relatively small portion of existing papers adheres to it, judging by trend in figure 4.2.

Below is a list of frameworks[1], which either solely focus on GP models or contain their implementations as a part of larger probabilistic modelling tool set. Each entry has a link to package's homepage and references to a respective introductory papers. Additional description is provided when package has a specific functionality.

1. General purpose libraries:
    - PYRO and NumPyro — [187], [188]
    - TensorFlow Probability — [189]
    - Turing.jl — [190]
    - Edward2 (Edward is no longer being actively developed) — [191]–[193]
    - PyMC3 and PyMC4 — [194], [195] — PyMC4 is in active development.
    - BayesDB with CGPM — [196]–[198]
    - MXFusion — [199]
    - CrossCat — [200]
    - Gen.jl — [201]
    - Stan with PyStan and Stan.jl interfaces — [202], [203]
2. GP specific libraries:
    - GPy — [136]
    - GPflow — [204], [205]
    - GPyTorch — [139]
    - GaussianProcesses.jl — [206]
    - Stheno.jl and Stheno — [207], [208]
    - Neural Tangents — [209] — High-level DNN API for specifying complex, hierarchical, DNN of both finite and infinite width. Neural Tangents allows researchers to define, train, and evaluate infinite networks as easily as finite ones. It supports construction of a neural network model with the usual building blocks like convolutions, pooling, residual connections, non-linearities and obtain not only the finite model, but also the kernel function of the respective GP.

[1] The interested reader may find an older, but still useful overview by Erickson et. al. [186].

- AugmentedGaussianProcesses.jl — [210] — GP package based on data augmentation, sparsity and natural gradients. Contains a collection of models for different gaussian and non-gaussian likelihoods, which are transformed via data augmentation into conditionally conjugate likelihood allowing for extremely fast inference via block coordinate updates.
- CandleGP — Port of GPflow features to PyTorch. Structure and syntaxes are kept close to the former.
- AutoGP — [211] — An inference framework for GP models that explores three complementary directions: scalable and statistically efficient variational inference, flexible kernels and objective functions for hyperparameter learning alternative to the marginal likelihood.
- JuliaGaussianProcesses — a new promising set of Julia libraries. In active development.

3. Supporting libraries:
    - JAX — [212] — Composable transformations of python and numpy programs. Supports automatic differentiation, vectorisation, just-in-time (JIT) compilation to graphics processing unit (GPU) and tensor processing unit (TPU). Usable if one want to implement GP models "from scratch". For example: kalman-jax by Wilkinson et. al. [213], [214].
    - GPyOpt — GPy extension for global optimization with different acquisition functions and physical experiments. Supports large data sets via sparse GP models.
    - BoTorch and Ax — [215], [216] — The first provides a modular and easily extensible interface for composing bayesian optimisation primitives. The second is a platform for designing and optimizing computational experiments.
    - Emukit — [217] — Adaptable toolkit for enriching decision making under uncertainty. Supports surrogate models when data is obtained from multiple information sources that have different fidelity and cost, bayesian optimisation, bayesian quadratures, active learning and analysis of the inputs influence on the outputs of a given system.
    - numpy-ml — General purpose "ml" package with minimal amount of python dependencies, that includes basic GP modelling tools.
    - BayesianOptimization — Bayesian optimisation package with minimal amount of python dependencies.
    - PyProb — [218] — A probabilistic programming system for simulators and high-performance computing (HPC). The main focus is on coupling existing simulation code bases with probabilistic inference with minimal intervention. In active development.
    - Trieste — A Bayesian optimization toolbox built on TensorFlow, created by Secondmind Labs (former PROWLER.io).
    - ZhuSuan — [219] — probabilistic programming library for Bayesian deep learning. ZhuSuan provides deep learning style primitives and algorithms for building probabilistic models and applying Bayesian inference. Supported inference algorithms include customisable methods of: variational inference, importance sampling and Hamiltonian Monte Carlo.

We would also like to point out an interesting trend[2] in figure 4.1, indicating a continual increase in popularity of the PyTorch framework [220].

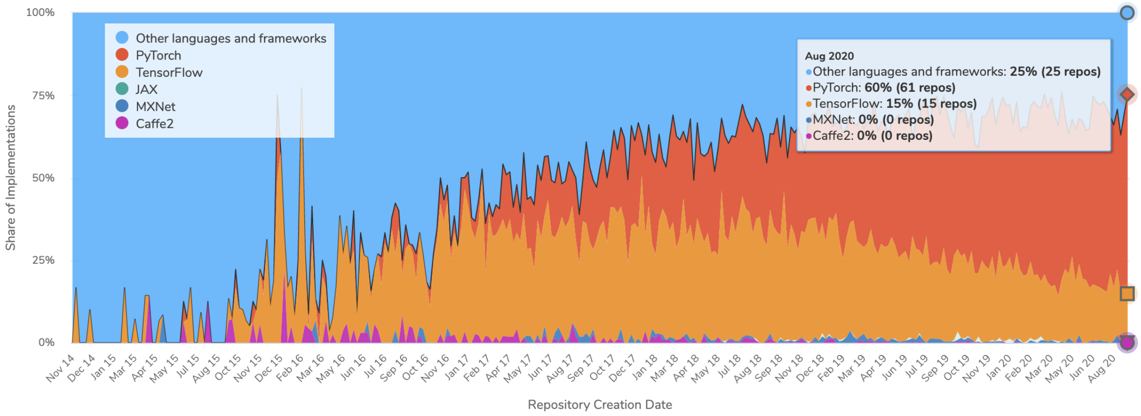

Figure 4.1: Paper implementations grouped by framework, from November 2014 to August 2020, weekly scale.

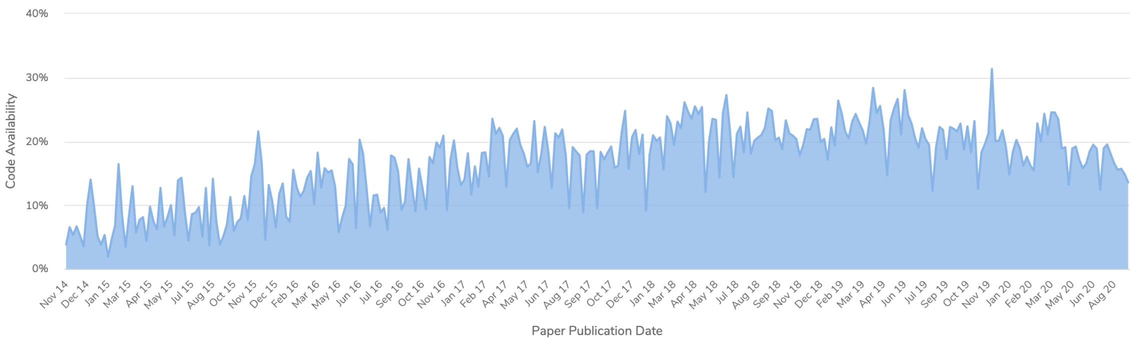

Figure 4.2: Percentage of papers with code support, from November 2014 to August 2020, weekly scale.

[2] Information is taken from – paperswithcode.com. Code repositories for some papers might have been added later, which explains PyTorch and TensorFlow entries before their respective release dates.

5. Conclusions and future research

In this report we've introduced a problem of combined modelling of wireless channels and hardware RF impairments. Because of growing amount of beyond 5G scenarios and complexity of RF hardware, we foresee continual growth of interest in "joint channel" simulators.
Motivated by lack of proper solution and complexity of the problem, we've given an overview of most challenging impairments and existing simulators. To address their disadvantages and formulated simulator requirements we've proposed use of bayesian non-parametric tools. Specifically deep GPs.
Deep GP regression is novel data-driven modelling technique, continuously researched by "ml" community. As all generative models they are able to produce new samples for link simulations. Uncertainty quantification helps with model transparency and interpretation. Lastly, universal functional approximation property of deep GPs together with automated kernel selection make it possible to incorporate any new information into the model. Like new types of hardware impairments or channel scenarios, without changes in general modelling framework.
Our future research includes experimental evaluations of GPs methods on non-stationary wireless channel scenarios and selected RF hardware impairments. Specifically we want to analyse how relevant automatic kernel selection can be in the above-mentioned context. Is domain knowledge is enough to construct deep GP model? Would it require additional automation similar to NAS? These and other questions will help to design a proof-of-concept "full channel" simulation. In a long term, we see a possibility of expanding proposed framework into an open-source software product, like QuaDRiGa or NYUSIM.

6. References

[1] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus and M. Fallgren, 'Scenarios for 5G mobile and wireless communications: The vision of the METIS project,' *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35,

[2] A. Osseiran, V. Braun, T. Hidekazu, P. Marsch, H. Schotten, H. Tullberg, M. A. Uusitalo and M. Schellman, 'The Foundation of the Mobile and Wireless Communications System for 2020 and Beyond: Challenges, Enablers and Technology Solutions,' in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, pp. 1–5.

[3] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang and D. J. Love, 'Prospective Multiple Antenna Technologies for Beyond 5G,' *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1637–1660,

[4] K. David and H. Berndt, '6G Vision and Requirements: Is There Any Need for Beyond 5G?' *IEEE Vehicular Technology Magazine*, vol. 13, no. 3, pp. 72–80,

[5] E. Bjornson, L. Van der Perre, S. Buzzi and E. G. Larsson, 'Massive MIMO in Sub-6 GHz and mmWave: Physical, Practical, and Use-Case Differences,' *IEEE Wireless Communications*, vol. 26, no. 2, pp. 100–108,

[6] E. Björnson, Ö. Özdogan and E. G. Larsson, 'Reconfigurable Intelligent Surfaces: Three Myths and Two Critical Questions,' version 1,

[7] F. Sohrabi and W. Yu, 'Hybrid Digital and Analog Beamforming Design for Large-Scale Antenna Arrays,' *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 501–513,

[8] W. Wang, S. L. Capitaneanu, D. Marinca and E.-S. Lohan, 'Comparative Analysis of Channel Models for Industrial IoT Wireless Communication,' *IEEE Access*, vol. 7, pp. 91 627–91 640,

[9] P. Ferrand, M. Amara, S. Valentin and M. Guillaud, 'Trends and challenges in wireless channel modeling for evolving radio access,' *IEEE Communications Magazine*, vol. 54, no. 7, pp. 93–99,

[10] P. Testolina, M. Lecci, M. Polese, M. Giordani and M. Zorzi, 'Scalable and Accurate Modeling of the Millimeter Wave Channel,' in *2020 International Conference on Computing, Networking and Communications (ICNC)*, pp. 969–974.

[11] J. Huang, C.-X. Wang, L. Bai, J. Sun, Y. Yang, J. Li, O. Tirkkonen and M.-T. Zhou, 'A Big Data Enabled Channel Model for 5G Wireless Communication Systems,' *IEEE Transactions on Big Data*, vol. 6, no. 2, pp. 211–222,

[12] S. Jaeckel, L. Raschkowski, F. Burkhardt and L. Thiele, 'Efficient Sum-of-Sinusoids-Based Spatial Consistency for the 3GPP New-Radio Channel Model,' in *2018 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–7.

[13] S. Jaeckel, L. Raschkowski, S. Wu, L. Thiele and W. Keusgen, 'An Explicit Ground Reflection Model for mm-Wave Channels,' in *2017 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pp. 1–5.

[14] T. S. Rappaport, G. R. MacCartney, M. K. Samimi and S. Sun, 'Wideband Millimeter-Wave Propagation Measurements and Channel Models for Future Wireless Communication System Design,' *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3029–3056,

[15] V. Saxena, J. E. Gonzalez, I. Stoica, H. Tullberg and J. Jaldén, 'Constrained Thompson Sampling for Wireless Link Optimization,'

[16] S. Park, R. C. Daniels and R. W. Heath, 'Optimizing the Target Error Rate for Link Adaptation,' in *2015 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6.

[17] Ö. T. Demir and E. Björnson, 'Channel Estimation in Massive MIMO Under Hardware Non-Linearities: Bayesian Methods Versus Deep Learning,' *IEEE Open Journal of the Communications Society*, vol. 1, pp. 109–124,

[18] U. Gustavsson, C. Sanchéz-Perez, T. Eriksson, F. Athley, G. Durisi, P. Landin, K. Hausmair, C. Fager and L. Svensson, 'On the impact of hardware impairments on massive MIMO,' in *2014 IEEE Globecom Workshops (GC Wkshps)*, pp. 294–300.

[19] Q. Zhang, T. Q. S. Quek and S. Jin, 'Scaling Analysis for Massive MIMO Systems With Hardware Impairments in Rician Fading,' *IEEE Transactions on Wireless Communications*, vol. 17, no. 7, pp. 4536–4549,

[20] T. Wirth, L. Thiele, M. Kurras, M. Mehlhose and T. Haustein, 'Massive MIMO proof-of-concept: Emulations and hardware-field trials at 3.5 GHz,' in *2016 50th Asilomar Conference on Signals, Systems and Computers*, pp. 1793–1798.

[21] H. Madsen, *Time Series Analysis*. Chapman and Hall/CRC.

[22] L. Ljung, *System Identification: Theory for the User*, 2 edition. Upper Saddle River, NJ: Prentice Hall, 640 pp.

[23] A. Papazafeiropoulos, T. Ratnarajah, P. Kourtessis and S. Chatzinotas, 'Nuts and Bolts of a Realistic Stochastic Geometric Analysis of mmWave HetNets: Hardware Impairments and Channel Aging,' *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 5657–5671,

[24] E. Björnson, J. Hoydis, M. Kountouris and M. Debbah, 'Massive MIMO Systems with Non-Ideal Hardware: Energy Efficiency, Estimation, and Capacity Limits,' version 3, *IEEE Trans. Inform. Theory*, vol. 60, no. 11, pp. 7112–7139,

[25] C. Mollén, U. Gustavsson, T. Eriksson and E. G. Larsson, 'Impact of Spatial Filtering on Distortion from Low-Noise Amplifiers in Massive MIMO Base Stations,' version 3,

[26] X. Jiang, F. Kaltenberger and L. Deneire, 'How accurately should we calibrate a Massive MIMO TDD system?' In *2016 IEEE International Conference on Communications Workshops (ICC)*, pp. 706–711.

[27] X. Zhang, M. Matthaiou, E. Björnson, M. Coldrey and M. Debbah, 'On the MIMO capacity with residual transceiver hardware impairments,' in *2014 IEEE International Conference on Communications (ICC)*, pp. 5299–5305.

[28] S. Ali, W. Saad, N. Rajatheva, K. Chang, D. Steinbach, B. Sliwa, C. Wietfeld, K. Mei, H. Shiri, H.-J. Zepernick, T. M. C. Chu, I. Ahmad, J. Huusko, J. Suutala, S. Bhadauria, V. Bhatia, R. Mitra, S. Amuru, R. Abbas, B. Shao, M. Capobianco, G. Yu, M. Claes, T. Karvonen, M. Chen, M. Girnyk and H. Malik, '6G White Paper on Machine Learning in Wireless Communication Networks,' version 1,

[29] U. Challita, H. Ryden and H. Tullberg, 'When Machine Learning Meets Wireless Cellular Networks: Deployment, Challenges, and Applications,' *IEEE Communications Magazine*, vol. 58, no. 6, pp. 12–18,

[30] W. Saad, M. Bennis and M. Chen, 'A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems,' *IEEE Network*, vol. 34, no. 3, pp. 134–142,

[31] N. H. Mahmood, S. Böcker, A. Munari, F. Clazzer, I. Moerman, K. Mikhaylov, O. Lopez, O.-S. Park, E. Mercier, H. Bartz, R. Jäntti, R. Pragada, Y. Ma, E. Annanperä, C. Wietfeld, M. Andraud, G. Liva, Y. Chen, E. Garro, F. Burkhardt, H. Alves, C.-F. Liu, Y. Sadi, J.-B. Dore, E. Kim, J. Shin, G.-Y. Park, S.-K. Kim, C. Yoon, K. Anwar and P. Seppänen, 'White Paper on Critical and Massive Machine Type Communication Towards 6G,' version 2,

[32] M. Levi and H. Messer, 'Sufficient conditions for array calibration using sources of mixed tapes,' in *International Conference on Acoustics, Speech, and Signal Processing*, 2943–2946 vol.5.

[33] B. D. Steinberg, 'Self-calibration of large phased-array antennas for radar,' *International Journal of Imaging Systems and Technology*, vol. 4, no. 4, pp. 275–284,

[34] M. Koerber and D. Fuhrmann, 'Array calibration by Fourier series parameterization: Scaled principal components method,' in *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, 340–343 vol.4.

[35] H.-S. Lui, H. T. Hui and M. S. Leong, 'A Note on the Mutual-Coupling Problems in Transmitting and Receiving Antenna Arrays,' *IEEE Antennas and Propagation Magazine*, vol. 51, no. 5, pp. 171–176,

[36] X. Artiga, B. Devillers and J. Perruisseau-Carrier, 'Mutual coupling effects in multi-user massive MIMO base stations,' in *Proceedings of the 2012 IEEE International Symposium on Antennas and Propagation*, pp. 1–2.

[37] O. Raeesi, Y. Zou, A. Tölli and M. Valkama, 'Closed-form analysis of channel non-reciprocity due to transceiver and antenna coupling mismatches in multi-user massive MIMO network,' in *2014 IEEE Globecom Workshops (GC Wkshps)*, pp. 333–339.

[38] C.-M. Chen, V. Volski, L. Van der Perre, G. A. E. Vandenbosch and S. Pollin, 'Finite Large Antenna Arrays for Massive MIMO: Characterization and System Impact,' *IEEE Trans. Antennas Propagat.*, vol. 65, no. 12, pp. 6712–6720,

[39] F. Kaltenberger, H. Jiang, M. Guillaud and R. Knopp, 'Relative channel reciprocity calibration in MIMO/TDD systems,' in *2010 Future Network Mobile Summit*, pp. 1–10.

[40] E. G. Larsson, O. Edfors, F. Tufvesson and T. L. Marzetta, 'Massive MIMO for next generation wireless systems,' *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195,

[41] D. Rainish, A. Freedman and D. Goberman, 'Calibration Techniques for an Antenna Array,' pat. 20 190 058 530.

[42] I. Şeker, 'Calibration methods for phased array radars,' in *Radar Sensor Technology XVII*, vol. 8714, International Society for Optics and Photonics, 87140W.

[43] S. Andersson, U. Forssen, F. B. Ovesjo and S. O. Petersson, 'Antenna array calibration,' U.S. Patent 6157343A.

[44] A. Bourdoux, B. Come and N. Khaled, 'Non-reciprocal transceivers in OFDM/SDMA systems: Impact and mitigation,' in *Radio and Wireless Conference, 2003. RAWCON '03. Proceedings*, pp. 183–186.

[45] M. Guillaud, D. Slock and R. Knopp, 'A practical method for wireless channel reciprocity exploitation through relative calibration,' in *Proceedings of the Eighth International Symposium on Signal Processing and Its Applications, 2005.*, vol. 1, Sydney, Australia: IEEE, pp. 403–406.

[46] F. Kaltenberger and M. Guillaud, 'Exploitation of reciprocity in measured mimo channels,'

[47] N. Fistas and A. Manikas, 'A new general global array calibration method,' in *Proceedings of ICASSP '94. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. iv, IV/73–IV/76 vol.4.

[48] G. Efstathopoulos and A. Manikas, 'A blind array calibration algorithm using a moving source,' in *2008 5th IEEE Sensor Array and Multichannel Signal Processing Workshop*, pp. 455–458.

[49] A. M. Mengot and A. Manikas, 'Global calibration of CDMA-based arrays,' in *2006 14th European Signal Processing Conference*, pp. 1–5.

[50] A. Manikas and N. Fistas, 'Modelling and estimation of mutual coupling between array elements,' in *Proceedings of ICASSP '94. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. iv, IV/553–IV/556 vol.4.

[51] B. Friedlander, 'Array calibration in the presence of linear manifold distortion,' in *2017 51st Asilomar Conference on Signals, Systems, and Computers*, pp. 1199–1203.

[52] C. Fulton and W. Chappell, 'Calibration techniques for digital phased arrays,' in *2009 IEEE International Conference on Microwaves, Communications, Antennas and Electronics Systems*, pp. 1–10.

[53] J. Vieira, F. Rusek and F. Tufvesson, 'A receive/transmit calibration technique based on mutual coupling for massive MIMO base stations,' in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 1–6.

[54] J. Vieira, F. Rusek, O. Edfors, S. Malkowsky, L. Liu and F. Tufvesson, 'Reciprocity Calibration for Massive MIMO: Proposal, Modeling and Validation,' version 3,

[55] T. Moon, J. Gaun and H. Hassanieh, 'Online Millimeter Wave Phased Array Calibration Based on Channel Estimation,' in *2019 IEEE 37th VLSI Test Symposium (VTS)*, pp. 1–6.

[56] C. Shan, X. Chen, H. Yin, W. Wang, G. Wei and Y. Zhang, 'Diagnosis of Calibration State for Massive Antenna Array via Deep Learning,' *IEEE Wireless Communications Letters*, vol. 8, no. 5, pp. 1431–1434,

[57] H. Bertrand, D. Grenier and S. Roy, 'Experimental antenna array calibration with artificial neural networks,' *Signal Processing*, vol. 88, no. 5, pp. 1152–1164,

[58] J. Schlee, 'Absolute timing and Tx power calibration of the Tx path in a distributed system,' U.S. Patent 8731005B2.

[59] X. Jiang, M. Čirkić, F. Kaltenberger, E. G. Larsson, L. Deneire and R. Knopp, 'MIMO-TDD reciprocity under hardware imbalances: Experimental results,' in *2015 IEEE International Conference on Communications (ICC)*, pp. 4949–4953.

[60] K. Gopala and D. Slock, 'Antenna Array Absolute Self-Calibration and Application to Separate Tx/Rx Array Full Duplex MIMO,' in *2017 IEEE Globecom Workshops (GC Wkshps)*, Singapore: IEEE, pp. 1–6.

[61] X. Jiang, A. Decurninge, K. Gopala, F. Kaltenberger, M. Guillaud, D. Slock and L. Deneire, 'A Framework for Over-the-Air Reciprocity Calibration for TDD Massive MIMO Systems,' *IEEE Transactions on Wireless Communications*, vol. 17, no. 9, pp. 5975–5990,

[62] S. Das, F. Tariq, M. Rahman, F. Frederiksen, E. De Carvalho and R. Prasad, 'Impact of Nonlinear Power Amplifier on Link Adaptation Algorithm of OFDM Systems,' in *2007 IEEE 66th Vehicular Technology Conference*, pp. 1303–1307.

[63] Y. Zou, P. Zetterberg, U. Gustavsson, T. Svensson, A. Zaidi, T. Kadur, W. Rave and G. Fettweis, 'Impact of Major RF Impairments on mm-Wave Communications Using OFDM Waveforms,' in *2016 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–7.

[64] S. Jacobsson, U. Gustavsson, G. Durisi and C. Studer, 'Massive MU-MIMO-OFDM Uplink with Hardware Impairments: Modeling and Analysis,' in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, pp. 1829–1835.

[65] W. Sneijers, 'Doherty Architectures in UHF White Paper,' p. 14,

[66] R. Pengelly, C. Fager and M. Ozen, 'Doherty's Legacy: A History of the Doherty Power Amplifier from 1936 to the Present Day,' *IEEE Microwave Magazine*, vol. 17, no. 2, pp. 41–58,

[67] D. S. Kozlov, A. P. Shitvov, A. G. Schuchinsky and M. B. Steer, 'Passive Intermodulation of Analog and Digital Signals on Transmission Lines With Distributed Nonlinearities: Modelling and Characterization,' *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 5, pp. 1383–1395,

[68] M. Gutsche, 'When connectors made entirely of non-ferromagnetic materials are used, they do not have to become the weak link in a communications system, causing harmful and hard-to-find intermod interference.,' *San Diego*, p. 4,

[69] W. Damm and A. Global, 'PIM: Components, Material, Handling & Testing,' p. 4,

[70] D. Weinstein, 'Passive Intermodulation Distortion in Connectors, Cable and Cable Assemblies,'

[71] 'Identifying Sources of External PIM,'

[72] S. Chen, 'Analog Dialogue 50-03, March 2017,' p. 5,

[73] T. S. D. Costa, 'Characterization of Passive Intermodulation Distortion in MultiBand FDD Radio Systems,' 86 pp.

[74] A. Mohammadi and F. M. Ghannouchi, *RF Transceiver Design for MIMO Wireless Communications*, ser. Lecture Notes in Electrical Engineering. Berlin Heidelberg: Springer-Verlag.

[75] J. C. Cahuana, 'Digital Predistortion for the Linearization of Power Amplifiers,' Chalmers University of Technology.

[76] Y. Guo, C. Yu and A. Zhu, 'Power Adaptive Digital Predistortion for Wideband RF Power Amplifiers With Dynamic Power Transmission,' *IEEE Trans. Microwave Theory Techn.*, vol. 63, no. 11, pp. 3595–3607,

[77] G. Yang, H. Wang, L. Li and F. Liu, 'One-step model extraction method for direct learning digital predistortion,' *Electronics Letters*, vol. 50, no. 16, pp. 1148–1150,

[78] D. Schreurs, M. O'Droma, A. A. Goacher and M. Gadringer, Eds., *RF Power Amplifier Behavioral Modeling*, Illustrated Edition. Cambridge, UK ; New York: Cambridge University Press, 288 pp.

[79] F. M. Ghannouchi and O. Hammi, 'Behavioral modeling and predistortion,' *IEEE Microwave Magazine*, vol. 10, no. 7, pp. 52–64,

[80] J. Xu, 'Practical Digital Pre-Distortion Techniques for PA Linearization in 3GPP LTE.'

[81] (). 'Digital Pre-Distortion (DPD) Concept,' [Online]. Available: http://rfmw.em.keysight.com/wireless/helpfiles/n7614/Content/Main/Digital%20Pre-Distortion%20(DPD)%20Concept.htm.

[82] (). 'Dual-Band Digital Pre-Distortion (DPD) Concept,' [Online]. Available: http://rfmw.em.keysight.com/wireless/helpfiles/n7614/Content/Main/Dual%20Band%20DPD%20Concept.htm.

[83] T. Ota, T. Kawasaki, S. Kimura, K. Tamanoi, T. Maniwa and M. Yoshida, 'A reduced complexity digital predistorter with a single common feedback loop for concurrent multi-band power amplifiers,' in *2017 IEEE Asia Pacific Microwave Conference (APMC)*, pp. 596–599.

[84] T. Ota, T. Kawasaki, S. Kimura, K. Tamanoi, H. Ishikawa, M. Shimizu and T. Maniwa, 'A Novel Multi-Band Look-Up Table Based Digital Predistorter with a Single Common Feedback Loop,' in *2018 Asia-Pacific Microwave Conference (APMC)*, pp. 551–553.

[85] P. Pratt, 'Wideband digital predistortion,' U.S. Patent 10224970B2.

[86] X. Hu, T. Liu, Z. Liu, W. Wang and F. M. Ghannouchi, 'A Novel Single Feedback Architecture With Time-Interleaved Sampling for Multi-Band DPD,' *IEEE Communications Letters*, vol. 23, no. 6, pp. 1033–1036,

[87] F. Mkadem, A. Islam and S. Boumaiza, 'Multi-Band Complexity-Reduced Generalized-Memory-Polynomial Power-Amplifier Digital Predistortion,' *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 6, pp. 1763–1774,

[88] C. Yu, L. Guan, E. Zhu and A. Zhu, 'Band-Limited Volterra Series-Based Digital Predistortion for Wideband RF Power Amplifiers,' *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 12, pp. 4198–4208,

[89] P. L. Gilabert and G. Montoro, 'Look-Up Table Implementation of a Slow Envelope Dependent Digital Predistorter for Envelope Tracking Power Amplifiers,' *IEEE Microwave and Wireless Components Letters*, vol. 22, no. 2, pp. 97–99,

[90] O. Tanovic, A. Megretski, Y. Li, V. Stojanovic and M. Osqui, 'Equivalent Baseband Models and Corresponding Digital Predistortion for Compensating Dynamic Passband Nonlinearities in Phase-Amplitude Modulation-Demodulation Schemes,' *IEEE Transactions on Signal Processing*, vol. 66, no. 22, pp. 5972–5987,

[91] K. Hausmair, P. Landin, U. Gustavsson, C. Fager and T. Eriksson, 'Digital Predistortion for Multi-Antenna Transmitters Affected by Antenna Crosstalk,' *IEEE Transactions on Microwave Theory and Techniques*,

[92] X. Wang, Y. Li, C. Yu, W. Hong and A. Zhu, 'Digital Predistortion of 5G Massive MIMO Wireless Transmitters Based on Indirect Identification of Power Amplifier Behavior With OTA Tests,' *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 1, pp. 316–328,

[93] T. Ackermann, J. Potschka, T. Maiwald, A. Hagelauer, G. Fischer and R. Weigel, 'A Robust Digital Predistortion Algorithm for 5G MIMO: Modeling a MIMO Scenario With Two Nonlinear MIMO Transmitters Including a Cross-Coupling Effect,' *IEEE Microwave Magazine*, vol. 21, no. 7, pp. 54–62,

[94] S. Amin, P. N. Landin, P. Händel and D. Rönnow, 'Behavioral Modeling and Linearization of Crosstalk and Memory Effects in RF MIMO Transmitters,' *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 4, pp. 810–823,

[95] S. A. Bassam, M. Helaoui and F. M. Ghannouchi, 'Crossover Digital Predistorter for the Compensation of Crosstalk and Nonlinearity in MIMO Transmitters,' *IEEE Transactions on Microwave Theory and Techniques*, vol. 57, no. 5, pp. 1119–1128,

[96] G. M. van de Ven and A. S. Tolias, 'Three scenarios for continual learning,' version 1,

[97] S. Habu, Y. Yamao and H. Suzuki, 'Unified Feedback Beamforming Digital Predistorter,' in *2019 49th European Microwave Conference (EuMC)*, pp. 904–907.

[98] J. Jing and C. Yu, 'Multibeam Digital Predistortion for Millimeter-Wave Analog Beamforming Transmitters,' *IEEE Microwave and Wireless Components Letters*, vol. 30, no. 2, pp. 209–212,

[99] M. B. Salman and G. M. Guvensen, 'On the Effects of PA Nonlinearities for Hybrid Beamforming Based Wideband Massive MIMO Systems,' in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, pp. 1–7.

[100] J. Cavers, 'Amplifier linearization using a digital predistorter with fast adaptation and low memory requirements,' *IEEE Transactions on Vehicular Technology*, vol. 39, no. 4, pp. 374–382,

[101] J. Schoukens, J. Suykens and L. Ljung, 'Wiener-Hammerstein Benchmark,' p. 4,

[102] Z. Yu, 'A Generalized Digital Predistortion Model Based on Artificial Neural Networks,' in *2018 Asia-Pacific Microwave Conference (APMC)*, pp. 935–937.

[103] M. Tanio, N. Ishii and N. Kamiya, 'Efficient Digital Predistortion Using Sparse Neural Network,' *IEEE Access*, vol. 8, pp. 117 841–117 852,

[104] C. Tarver, A. Balatsoukas-Stimming and J. R. Cavallaro, 'Design and Implementation of a Neural Network Based Predistorter for Enhanced Mobile Broadband,' in *2019 IEEE International Workshop on Signal Processing Systems (SiPS)*, pp. 296–301.

[105] Y. Wu, U. Gustavsson, A. G. i Amat and H. Wymeersch, 'Residual Neural Networks for Digital Predistortion,'

[106] K. He, X. Zhang, S. Ren and J. Sun, 'Deep Residual Learning for Image Recognition,' version 1,

[107] D. Wang, M. Aziz, M. Helaoui and F. M. Ghannouchi, 'Augmented Real-Valued Time-Delay Neural Network for Compensation of Distortions and Impairments in Wireless Transmitters,' *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 1, pp. 242–254,

[108] R. Hongyo, Y. Egashira, T. M. Hone and K. Yamaguchi, 'Deep Neural Network-Based Digital Predistorter for Doherty Power Amplifiers,' *IEEE Microwave and Wireless Components Letters*, vol. 29, no. 2, pp. 146–148,

[109] P. Jaraut, M. Rawat and F. M. Ghannouchi, 'Composite Neural Network Digital Predistortion Model for Joint Mitigation of Crosstalk, I/Q Imbalance, Nonlinearity in MIMO Transmitters,' *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 11, pp. 5011–5020,

[110] D. Duvenaud, O. Rippel, R. P. Adams and Z. Ghahramani, 'Avoiding pathologies in very deep networks,' version 3,

[111] T. Elsken, J. H. Metzen and F. Hutter, 'Neural Architecture Search: A Survey,' *Journal of Machine Learning Research*, vol. 20, no. 55, pp. 1–21,

[112] S. Jaeckel, L. Raschkowski, K. Börner and L. Thiele, 'QuaDRiGa: A 3-D Multi-Cell Channel Model With Time Evolution for Enabling Virtual Field Trials,' *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 6, pp. 3242–3256,

[113] S. Sun, G. R. MacCartney Jr. and T. S. Rappaport, 'A Novel Millimeter-Wave Channel Simulator and Applications for 5G Wireless Communications,'

[114] S. Ju, O. Kanhere, Y. Xing and T. S. Rappaport, 'A Millimeter-Wave Channel Simulator NYUSIM with Spatial Consistency and Human Blockage,'

[115] C. Slezak, M. Zhang, M. Mezzavilla and S. Rangan, 'Understanding End-to-End Effects of Channel Dynamics in Millimeter Wave 5G New Radio,' in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5.

[116] C. Herranz, M. Zhang, M. Mezzavilla, D. Martin-Sacristán, S. Rangan and J. F. Monserrat, 'A 3GPP NR Compliant Beam Management Framework to Simulate End-to-End mmWave Networks,' in *Proceedings of the 21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, ser. MSWIM '18, New York, NY, USA: Association for Computing Machinery, pp. 119–125.

[117] B. Winkel and A. Jessner, 'Spectrum management and compatibility studies with Python,'

[118] D. Hilbert, 'Mathematical problems,' *Bull. Amer. Math. Soc.*, vol. 8, no. 10, pp. 437–479,

[119] A. N. Kolmogorov, 'On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition,' in *American Mathematical Society Translations: Series 2*. Providence, Rhode Island: American Mathematical Society, vol. 28, pp. 55–59.

[120] J. Schmidt-Hieber, 'The Kolmogorov-Arnold representation theorem revisited,' version 1,

[121] J. Schwarz, D. Altman, A. Dudzik, O. Vinyals, Y. W. Teh and R. Pascanu, 'Towards a natural benchmark for continual learning,' p. 13,

[122] S. Kapoor, T. Karaletsos and T. D. Bui, 'Variational Auto-Regressive Gaussian Processes for Continual Learning,' version 1,

[123] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, ser. Adaptive Computation and Machine Learning. Cambridge, Mass: MIT Press, 248 pp.

[124] D. Duvenaud, H. Nickisch and C. E. Rasmussen, 'Additive Gaussian Processes,' version 1,

[125] D. Duvenaud, 'Automatic model construction with Gaussian processes,' Thesis, University of Cambridge.

[126] E. Schulz, M. Speekenbrink and A. Krause, 'A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions,' *Journal of Mathematical Psychology*, vol. 85, pp. 1–16,

[127] S. Roberts, M. Osborne, M. Ebden, S. Reece, N. Gibson and S. Aigrain, 'Gaussian processes for time-series modelling,' *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1984, p. 20 110 550,

[128] A. G. d. G. Matthews, M. Rowland, J. Hron, R. E. Turner and Z. Ghahramani, 'Gaussian Process Behaviour in Wide Deep Neural Networks,' version 2,

[129] A. Damianou, 'Deep Gaussian Processes and Variational Propagation of Uncertainty,' Thesis, University of Sheffield.

[130] K. Vafa, 'Training and Inference for Deep Gaussian Processes.'

[131] R. Frigola, 'Bayesian Time Series Learning with Gaussian Processes,' Thesis, University of Cambridge.

[132] D. J. C. MacKay, 'Bayesian methods for adaptive models,' phd, California Institute of Technology.

[133] ——, *Information Theory, Inference and Learning Algorithms*, Sixth Printing 2007 edition. Cambridge, UK ; New York: Cambridge University Press, 640 pp.

[134] C. E. Rasmussen and Z. Ghahramani, 'Occam\textquotesingle s Razor,' in *Advances in Neural Information Processing Systems 13*, T. K. Leen, T. G. Dietterich and V. Tresp, Eds., MIT Press, pp. 294–300.

[135] D. J. Rezende, S. Mohamed and D. Wierstra, 'Stochastic Backpropagation and Approximate Inference in Deep Generative Models,' version 3,

[136] Z. Dai, A. Damianou, J. Hensman and N. Lawrence, 'Gaussian Process Models with Parallelization and GPU acceleration,' version 1,

[137] C. E. Rasmussen and H. Nickisch, 'Gaussian Processes for Machine Learning (GPML) Toolbox,' *Journal of Machine Learning Research*, vol. 11, no. 100, pp. 3011–3015,

[138] E. Snelson and Z. Ghahramani, 'Sparse Gaussian Processes using Pseudo-inputs,' in *Advances in Neural Information Processing Systems 18*, Y. Weiss, B. Schölkopf and J. C. Platt, Eds., MIT Press, pp. 1257–1264.

[139] J. Gardner, G. Pleiss, K. Q. Weinberger, D. Bindel and A. G. Wilson, 'GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration,' in *Advances in Neural Information Processing Systems 31*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett, Eds., Curran Associates, Inc., pp. 7576–7586.

[140] M. Titsias, 'Variational Learning of Inducing Variables in Sparse Gaussian Processes,' in *Artificial Intelligence and Statistics*, PMLR, pp. 567–574.

[141] J. Hensman, N. Fusi and N. D. Lawrence, 'Gaussian Processes for Big Data,' version 1,

[142] J. Hensman, A. Matthews and Z. Ghahramani, 'Scalable Variational Gaussian Process Classification,' in *Artificial Intelligence and Statistics*, PMLR, pp. 351–360.

[143] J. Hensman, A. G. Matthews, M. Filippone and Z. Ghahramani, 'MCMC for Variationally Sparse Gaussian Processes,' in *Advances in Neural Information Processing Systems 28*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama and R. Garnett, Eds., Curran Associates, Inc., pp. 1648–1656.

[144] H. Salimbeni and M. Deisenroth, 'Doubly Stochastic Variational Inference for Deep Gaussian Processes,' in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, Eds., Curran Associates, Inc., pp. 4588–4599.

[145] M. van der Wilk, C. E. Rasmussen and J. Hensman, 'Convolutional Gaussian Processes,' in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, Eds., Curran Associates, Inc., pp. 2849–2858.

[146] K. Vafa, 'Training Deep Gaussian Processes with Sampling,' p. 5,

[147] A. Naish-guzman and S. Holden, 'The Generalized FITC Approximation,' in *Advances in Neural Information Processing Systems 20*, J. C. Platt, D. Koller, Y. Singer and S. T. Roweis, Eds., Curran Associates, Inc., pp. 1057–1064.

[148] J. Shi, M. K. Titsias and A. Mnih, 'Sparse Orthogonal Variational Inference for Gaussian Processes,'

[149] D. Duvenaud, J. Lloyd, R. Grosse, J. Tenenbaum and G. Zoubin, 'Structure Discovery in Nonparametric Regression through Compositional Kernel Search,' in *International Conference on Machine Learning*, PMLR, pp. 1166–1174.

[150] G. Schwarz, 'Estimating the Dimension of a Model,' *Ann. Statist.*, vol. 6, no. 2, pp. 461–464,

[151] C. Steinruecken, E. Smith, D. Janz, J. Lloyd and Z. Ghahramani, 'The Automatic Statistician,' in *Automated Machine Learning: Methods, Systems, Challenges*, ser. The Springer Series on Challenges in Machine Learning, F. Hutter, L. Kotthoff and J. Vanschoren, Eds., Cham: Springer International Publishing, pp. 161–173.

[152] A. G. Wilson, Z. Hu, R. Salakhutdinov and E. P. Xing, 'Deep Kernel Learning,' version 1,

[153] ——, 'Stochastic Variational Deep Kernel Learning,' version 2,

[154] S. Sun, G. Zhang, C. Wang, W. Zeng, J. Li and R. Grosse, 'Differentiable Compositional Kernel Learning for Gaussian Processes,' version 3,

[155] H. Kim and Y. W. Teh, 'Scaling up the Automatic Statistician: Scalable Structure Discovery using Gaussian Processes,' version 2,

[156] J. Lee, Y. Bahri, R. Novak, S. S. Schoenholz, J. Pennington and J. Sohl-Dickstein, 'Deep Neural Networks as Gaussian Processes,' version 3,

[157] R. M. Neal, *Bayesian Learning for Neural Networks*, ser. Lecture Notes in Statistics. New York: Springer-Verlag.

[158] G. Yang, 'Wide Feedforward or Recurrent Neural Networks of Any Architecture are Gaussian Processes,' in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d\ textquotesingle Alché-Buc, E. Fox and R. Garnett, Eds., Curran Associates, Inc., pp. 9951–9960.

[159] ——, 'Tensor Programs I: Wide Feedforward or Recurrent Neural Networks of Any Architecture are Gaussian Processes,' version 2,

[160] ——, 'Tensor Programs II: Neural Tangent Kernel for Any Architecture,' version 3,

[161] J. Hron, Y. Bahri, R. Novak, J. Pennington and J. Sohl-Dickstein, 'Exact posterior distributions of wide Bayesian neural networks,' version 1,

[162] R. Novak, L. Xiao, J. Lee, Y. Bahri, G. Yang, J. Hron, D. A. Abolafia, J. Pennington and J. Sohl-Dickstein, 'Bayesian Deep Convolutional Networks with Many Channels are Gaussian Processes,' version 4,

[163] A. Damianou and N. Lawrence, 'Deep Gaussian Processes,' in *Artificial Intelligence and Statistics*, pp. 207–215.

[164] N. Lawrence, 'Probabilistic Non-linear Principal Component Analysis with Gaussian Process Latent Variable Models,' *Journal of Machine Learning Research*, vol. 6, pp. 1783–1816, Nov.

[165] M. Titsias and N. D. Lawrence, 'Bayesian Gaussian Process Latent Variable Model,' in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, JMLR Workshop and Conference Proceedings, pp. 844–851.

[166] S. Kullback and R. A. Leibler, 'On Information and Sufficiency,' *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86,

[167] M. M. Dunlop, M. A. Girolami, A. M. Stuart and A. L. Teckentrup, 'How Deep Are Deep Gaussian Processes?' *Journal of Machine Learning Research*, vol. 19, no. 54, pp. 1–46,

[168] M. I. Radaideh and T. Kozlowski, 'Surrogate modeling of advanced computer simulations using deep Gaussian processes,' *Reliability Engineering & System Safety*, vol. 195, p. 106 731,

[169] M. Jin, A. Damianou, P. Abbeel and C. Spanos, 'Inverse Reinforcement Learning via Deep Gaussian Process,' version 4,

[170] K. Blomqvist, S. Kaski and M. Heinonen, 'Deep Convolutional Gaussian Processes,' in *Machine Learning and Knowledge Discovery in Databases*, U. Brefeld, E. Fromont, A. Hotho, A. Knobbe, M. Maathuis and C. Robardet, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, pp. 582–597.

[171] Y.-J. Park, P. M. Tagade and H.-L. Choi, 'Deep Gaussian Process-Based Bayesian Inference for Contaminant Source Localization,' *IEEE Access*, vol. 6, pp. 49 432–49 449,

[172] P. Tagade, K. S. Hariharan, S. Ramachandran, A. Khandelwal, A. Naha, S. M. Kolake and S. H. Han, 'Deep Gaussian process regression for lithium-ion battery health prognosis and degradation mode diagnosis,' *Journal of Power Sources*, vol. 445, p. 227 281,

[173] 'Deep Multi-task Gaussian Processes for Survival Analysis with Competing Risks,' in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, Eds., Curran Associates, Inc., pp. 2329–2337.

[174] X. Wang, X. Wang, S. Mao, J. Zhang, S. C. Periaswamy and J. Patton, 'Indoor Radio Map Construction and Localization with Deep Gaussian Processes,' *IEEE Internet of Things Journal*, pp. 1–1,

[175] D. Foreman-Mackey, E. Agol, S. Ambikasaran and R. Angus, 'Fast and scalable Gaussian process modeling with applications to astronomical time series,'

[176] F. Perez-Cruz, S. Van Vaerenbergh, J. J. Murillo-Fuentes, M. Lazaro-Gredilla and I. Santamaria, 'Gaussian Processes for Nonlinear Signal Processing: An Overview of Recent Advances,' *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 40–50,

[177] A. Solin, 'Machine Learning with Signal Processing - Part I: Signal Processing Tooling,' presented at the ICML 2020.

[178] ——, 'Machine Learning with Signal Processing - Part II: Stochastic Differential Equations,' presented at the ICML 2020.

[179] ——, 'Machine Learning with Signal Processing - Part III: Three Views into Gaussian Processes,' presented at the ICML 2020.

[180] ——, 'Machine Learning with Signal Processing - Part IV: Application Examples,' presented at the ICML 2020.

[181] C. L. C. Mattos, Z. Dai, A. Damianou, G. A. Barreto and N. D. Lawrence, 'Deep recurrent Gaussian processes for outlier-robust system identification,' *Journal of Process Control*, DYCOPS-CAB 2016, vol. 60, pp. 82–94,

[182] S. Särkkä, 'The Use of Gaussian Processes in System Identification,' version 1,

[183] S. Eleftheriadis, T. Nicholson, M. Deisenroth and J. Hensman, 'Identification of Gaussian Process State Space Models,' in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, Eds., Curran Associates, Inc., pp. 5309–5319.

[184] R. Calandra, J. Peters, C. E. Rasmussen and M. P. Deisenroth, 'Manifold Gaussian Processes for regression,' in *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 3338–3345.

[185] L. Ljung, 'Approaches to identification of nonlinear systems,' in *Proceedings of the 29th Chinese Control Conference*, pp. 1–5.

[186] C. B. Erickson, B. E. Ankenman and S. M. Sanchez, 'Comparison of Gaussian process modeling software,' version 1,

[187] E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. Szerlip, P. Horsfall and N. D. Goodman, 'Pyro: Deep Universal Probabilistic Programming,' version 1,

[188] D. Phan, N. Pradhan and M. Jankowiak, 'Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro,' version 1,

[189] J. V. Dillon, I. Langmore, D. Tran, E. Brevdo, S. Vasudevan, D. Moore, B. Patton, A. Alemi, M. Hoffman and R. A. Saurous, 'TensorFlow Distributions,' version 1,

[190] H. Ge, K. Xu and Z. Ghahramani, 'Turing: A Language for Flexible Probabilistic Inference,' in *International Conference on Artificial Intelligence and Statistics*, PMLR, pp. 1682–1690.

[191] D. Tran, A. Kucukelbir, A. B. Dieng, M. Rudolph, D. Liang and D. M. Blei, 'Edward: A library for probabilistic modeling, inference, and criticism,'

[192] D. Tran, M. Hoffman, D. Moore, C. Suter, S. Vasudevan, A. Radul, M. Johnson and R. A. Saurous, 'Simple, Distributed, and Accelerated Probabilistic Programming,'

[193] D. Tran, M. W. Dusenberry, M. van der Wilk and D. Hafner, 'Bayesian Layers: A Module for Neural Network Uncertainty,'

[194] J. Salvatier, T. Wiecki and C. Fonnesbeck, 'Probabilistic Programming in Python using PyMC,' version 1,

[195] M. Kochurov, C. Carroll, T. Wiecki and J. Lao, 'PyMC4: Exploiting Coroutines for Implementing a Probabilistic Programming Framework,'

[196] V. Mansinghka, R. Tibbetts, J. Baxter, P. Shafto and B. Eaves, 'BayesDB: A probabilistic programming system for querying the probable implications of data,'

[197] F. Saad and V. Mansinghka, 'Probabilistic Data Analysis with Probabilistic Programming,'

[198] F. Saad, L. Casarsa and V. Mansinghka, 'Probabilistic Search for Structured Data via Probabilistic Programming and Nonparametric Bayes,' version 1,

[199] Z. Dai, E. Meissner and N. D. Lawrence, 'MXFusion: A Modular Deep Probabilistic Programming Library,' p. 5,

[200] V. Mansinghka, P. Shafto, E. Jonas, C. Petschulat, M. Gasner and J. B. Tenenbaum, 'CrossCat: A Fully Bayesian Nonparametric Method for Analyzing Heterogeneous, High Dimensional Data,' *Journal of Machine Learning Research*, vol. 17, no. 138, pp. 1–49,

[201] M. F. Cusumano-Towner, F. A. Saad, A. K. Lew and V. K. Mansinghka, 'Gen: A general-purpose probabilistic programming system with programmable inference,' in *Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation*, ser. PLDI 2019, New York, NY, USA: Association for Computing Machinery, pp. 221–236.

[202] A. Gelman, D. Lee and J. Guo, 'Stan: A Probabilistic Programming Language for Bayesian Inference and Optimization,' *Journal of Educational and Behavioral Statistics*, vol. 40, no. 5, pp. 530–543,

[203] A. Vehtari, A. Gelman and J. Gabry, 'Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC,' *Stat Comput*, vol. 27, no. 5, pp. 1413–1432,

[204] A. G. d. G. Matthews, M. van der Wilk, T. Nickson, K. Fujii, A. Boukouvalas, P. Le{\ 'o}n-Villagr{\ 'a}, Z. Ghahramani and J. Hensman, 'GPflow: A Gaussian Process Library using TensorFlow,' *Journal of Machine Learning Research*, vol. 18, no. 40, pp. 1–6,

[205] N. Knudde, J. van der Herten, T. Dhaene and I. Couckuyt, 'GPflowOpt: A Bayesian Optimization Library using TensorFlow,' version 1,

[206] J. Fairbrother, C. Nemeth, M. Rischard, J. Brea and T. Pinder, 'GaussianProcesses.jl: A Nonparametric Bayes package for the Julia Language,' version 2,

[207] W. Tebbutt, W. Bruinsma and R. E. Turner, 'Probabilistic Programming with Gaussian Processes in Stheno.jl,' presented at the JuliaCon 2019.

[208] ——, 'Gaussian Process Probabilistic Programming,' presented at the JuliaCon 2019.

[209] R. Novak, L. Xiao, J. Hron, J. Lee, A. A. Alemi, J. Sohl-Dickstein and S. S. Schoenholz, 'Neural Tangents: Fast and Easy Infinite Neural Networks in Python,' presented at the Eighth International Conference on Learning Representations.

[210] T. Galy-Fajou, F. Wenzel, C. Donner and M. Opper, 'Multi-Class Gaussian Process Classification Made Conjugate: Efficient Inference via Data Augmentation,' version 1,

[211] K. Krauth, E. V. Bonilla, K. Cutajar and M. Filippone, 'AutoGP: Exploring the Capabilities and Limitations of Gaussian Process Models,' version 3,

[212] R. Frostig, M. Johnson and C. Leary, 'Compiling machine learning programs via high-level tracing.'

[213] W. J. Wilkinson, P. E. Chang, M. R. Andersen and A. Solin, 'State Space Expectation Propagation: Efficient Inference Schemes for Temporal Gaussian Processes,' version 1,

[214] P. E. Chang, W. J. Wilkinson, M. E. Khan and A. Solin, 'Fast Variational Learning in State-Space Gaussian Process Models,' version 2,

[215] M. Balandat, B. Karrer, D. R. Jiang, S. Daulton, B. Letham, A. G. Wilson and E. Bakshy, 'BoTorch: Programmable Bayesian Optimization in PyTorch,' version 2,

[216] S. Daulton, M. Balandat and E. Bakshy, 'Differentiable Expected Hypervolume Improvement for Parallel Multi-Objective Bayesian Optimization,' version 2,

[217] A. Klein, Z. Dai, F. Hutter, N. Lawrence and J. Gonzalez, 'Meta-Surrogate Benchmarking for Hyperparameter Optimization,' version 2,

[218] A. G. Baydin, L. Shao, W. Bhimji, L. Heinrich, L. Meadows, J. Liu, A. Munk, S. Naderiparizi, B. Gram-Hansen, G. Louppe, M. Ma, X. Zhao, P. Torr, V. Lee, K. Cranmer, Prabhat and F. Wood, 'Etalumis: Bringing Probabilistic Programming to Scientific Simulators at Scale,' version 2, *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 1–24,

[219] J. Shi, J. Chen, J. Zhu, S. Sun, Y. Luo, Y. Gu and Y. Zhou, 'ZhuSuan: A Library for Bayesian Deep Learning,' version 1,

[220] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai and S. Chintala, 'PyTorch: An Imperative Style, High-Performance Deep Learning Library,' version 1,

www.windmill-itn.eu