Uncertainty Propagation in Neural Network Enabled Multi-Channel Optimisation

Chen Li, Schyler C. Sun, Saba Al-Rubaye, Antonios Tsourdous, Weisi Guo *IEEE Senior Member*

Abstract—**Multi-channel optimisation relies on accurate channel state information (CSI) estimation. Error distributions in CSI can propagate through optimisation algorithms to cause undesirable uncertainty in the solution space. The transformation of uncertainty distributions differs between classic heuristic and Neural Network (NN) algorithms. Here, we investigate how CSI uncertainty transforms from an additive Gaussian error in CSI into different power allocation distributions in a multi-channel system. We offer theoretical insight into the uncertainty propagation for both Water-filling (WF) power allocation in comparison to diverse NN algorithms. We use the Kullback–Leibler divergence to quantify uncertainty deviation from the trusted WF algorithm and offer some insight into the role of NN structure and activation functions on the uncertainty divergence, where we found that the activation function choice is more important than the size of the neural network.**

Index Terms—**machine learning; deep learning; XAI; wireless;**

I. Introduction

Multi-channel optimisation of radio resources is crucial to current 4G-5G systems and beyond. Traditional optimisation relies on heuristic algorithms which are often formulated as either convex optimisation (e.g. Lagrangian) or non-convex problems (e.g. Genetic algorithm, Mean-field games, Markov Decision Processes, reinforcement learning...etc.). As the scale of the complexity increases, neural networks (NN) [1]–[4] have been proposed to automate and accelerate the mapping between inputs (e.g. CSI, user demand) and output solutions (e.g. transmit power allocation) [5], [6]. One open challenge is the propagation of uncertainty from input to output via an optimisation algorithm. Classic uncertainty quantification (UQ) techniques such as Polynomial Chaos Expansion (PCE) cannot be readily applied due to the complex nature of the algorithms.

A. Uncertainty Propagation in Decision Modules

Here, we start by studying how uncertainty in CSI can propagate through a classic MIMO Water-filling (WF) power allocation algorithm (IEEE 802.xx series, OFDM systems) versus its contemporary NN accelerated versions [7], which offer equivalent accelerated real-time solutions [8], [9]. Other power allocation employ DRL, which adds a learning agent [10], [11]. The essential UQ problem is to quantify the distribution over the output $\boldsymbol{y}$ for: $\boldsymbol{y} = f(\boldsymbol{x} + n; \lambda)$, where inputs $\boldsymbol{x} \in \mathbb{R}^{n \times 1}$ (e.g. channel gains) have a noise n due to

mis-estimation, and map to an output $\boldsymbol{y} \in \mathbb{R}^{n \times 1}$ (e.g. power allocation) via a model $f(\cdot)$.

In classic WF, the Lagrangian optimisation produces an iterative solution; and in NNs an approximate non-linear mapping can achieve effective power allocation without iterative search for λ (WF level). Direct probability analysis or Bayesian inference can be used to understand the brittleness of classic and heuristic algorithms [12]. Other analytical methods include polynomial chaos expansion, which are more suited to dynamical systems [13]. Currently, black-box NNs cannot explain the essential mapping it performs. There are also legal requirements (e.g. GDPR) for AI to explain its reasoning. As such, there is the need to develop a range of **explainable AI (XAI)** solutions that attempt to quantify NN mappings [14], [15]. These XAI techniques range from visualising key hidden layer features to localised linear models (LIME) [16], [17].

B. Novelty

In this seminal initial results paper on robustness of NNs for wireless power control, we outline statistical results on how different NN architectures and activation functions transform CSI uncertainty into power allocation solutions. Our **novelties** in this paper include: (1) deriving a theoretical uncertainty transformation for WF power allocation, (2) provide the statistical results for uncertainty transformation for a range of NNs and measure their KL divergence from the theoretical distribution.

II. System Setup

A. WF Power Allocation

We consider a classical wireless parallel channel power allocation problem comprised of N channels with independent Rayleigh fading characteristics. WF power allocation is well-established and we will not detail it here. Suffice to say, under the Shannon capacity assumption and a total power budget, the solution form for power in channel $n \in N$ is of:

$$p(n) = f[h(n); \lambda] = \left(\frac{1}{\lambda} - \frac{N_0}{|h(n)|^2} \right)^+, \quad (1)$$

where N_0 is the AWGN power, $h(n)$ is the fading gain, and the parameter λ is the Lagrangian multiplier (WF level). We implement this WF algorithm for N parallel Rayleigh fading channels with an iterative search solution for parameter λ, so that to achieve power allocation $p(n)$ by $f(\boldsymbol{h}; \lambda)$.
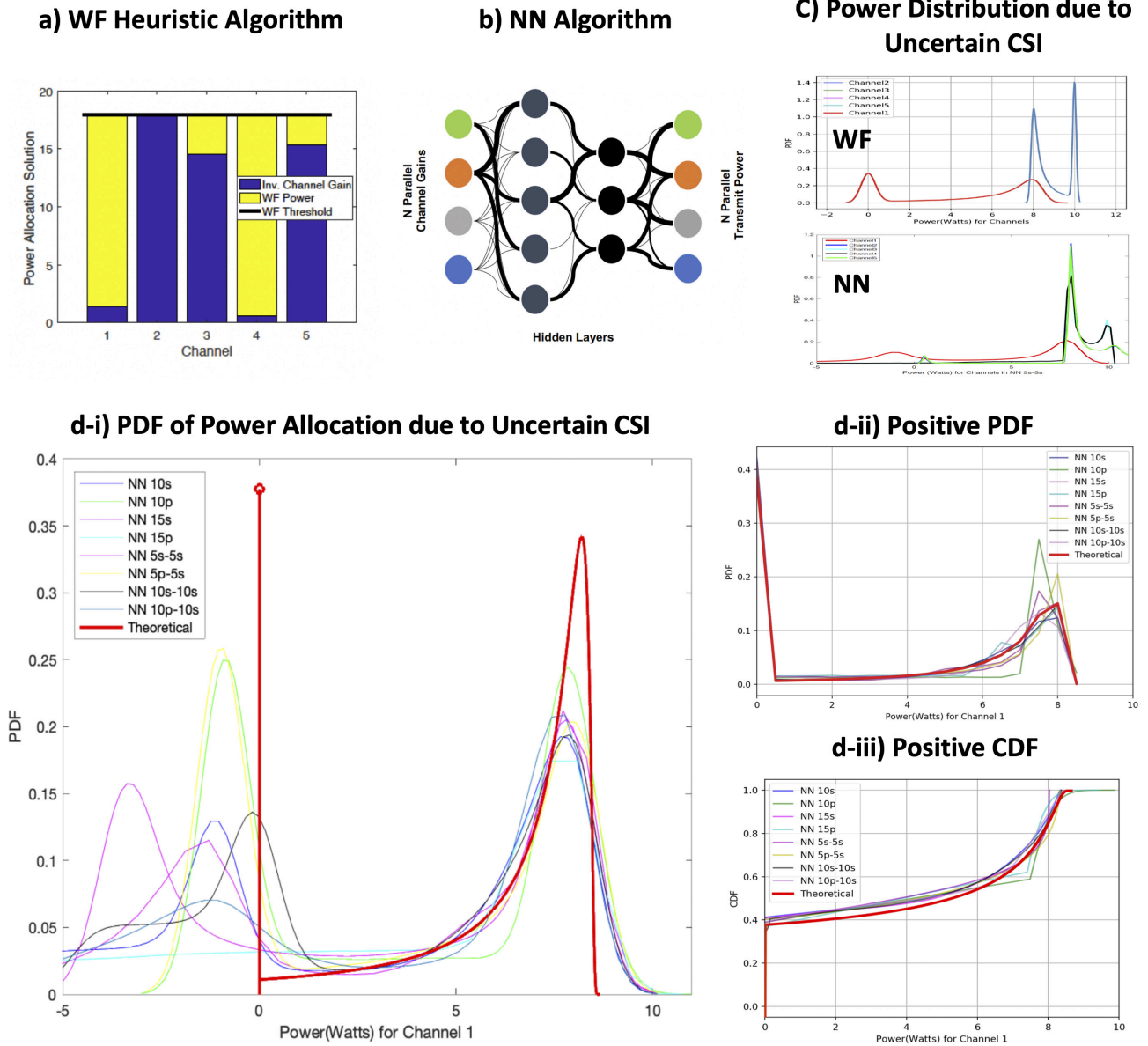
Fig. 1. Uncertainty Propagation in Parallel Channel Power Allocation: a) WF solution via iterative Lagrangian optimisation, b) NN implementation, c) uncertainty in power allocations, and d) PDF and CDF for uncertainty transformation from CSI into power solutions for WF and NNs.

B. NN Power Allocation

We reduce the WF heuristic search time by implementing a multi-layer NN with a number of neurons per layer. The input are the channel states and the output is the Lagrangian multiplier λ. The Lagrangian parameter in turn gives the power allocation output ($\lambda \rightarrow p(n)$). The implementation parameters are given in Table I. The training data is based on the conventional iterative WF solution, and the training results for different NN architectures is given in Fig. 2.

C. CSI Uncertainty and Theoretical Transformation

In both the WF and NN cases, we set up Gaussian distributed additive CSI uncertainty in just one of the multiple channels and examine its impact on the power allocation

TABLE I
SYSTEM SETUP

Parameter	Value
Channel	2.4GHz 3GPP Micro, Rayleigh
Wireless System	5 OFDMA channels, 40W budget
Iterative WF Monte-Carlo Loops	5e4
CSI Noise	Gaussian in Channel 1
Neurons per Layer	5-15
Activation Function	Sigmoid (s) or ReLU (p)
Training	Random division, $\leq$1e3 epochs

solution for all the channels. We treat WF as a trusted and reliable benchmark - familiar to engineers and used widely in society.
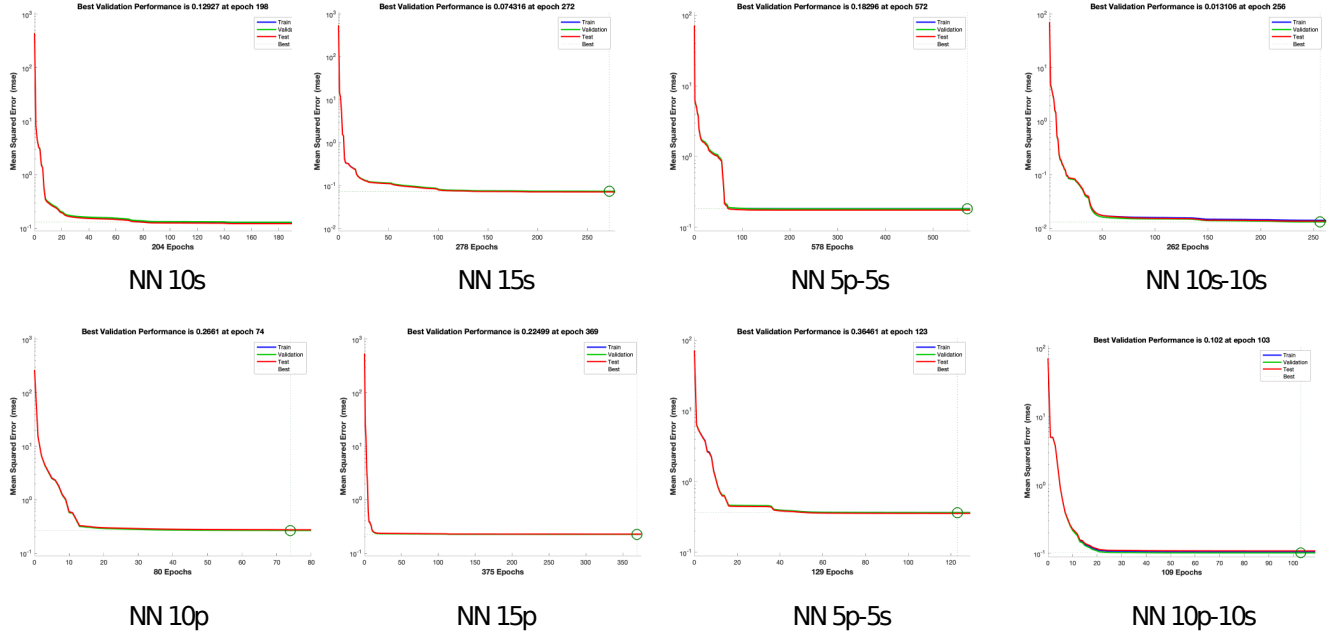
Fig. 2. NN training & testing statistics for different NN architectures.

The theoretical transformation via WF solution in Eq.(1) can be found as follows. We perturb one of the multiple channels with a Gaussian noise due to CSI mis-estimation: channel fading is $x \sim \mathcal{N}(\mu, \sigma^2)$, where μ is the accurate channel fading value and σ^2 is the input uncertainty variance due to CSI mis-estimation. We then find the uncertainty cascade for the perturbed channel under the split conditions of:

1) $x < \sqrt{N_0/\lambda}, y = 0$,
2) $x \geq \sqrt{N_0/\lambda}, y = (\frac{1}{\lambda} - \frac{N_0}{x^2})$.

Using the standard probability *Jacobian transformation* approach, we arrive at:

$$P(y = 0) = P(x_1 < \sqrt{N_0/\lambda}) \quad \text{Part 1}$$

$$f_Y(y) = \frac{(\frac{b}{a-y})^{3/2}}{2b\sigma\sqrt{2\pi}} \exp(-\frac{(\sqrt{\frac{b}{a-y}} - \mu)^2}{2\sigma^2}) \quad \text{Part 2} \quad (2)$$

where a, b are numerical values based on λ, N_0 for a given set of fading channels.

Given WF's theoretical uncertainty in Eq.(2) and that we trust its transparent nature, we now compare how the NN's uncertainty differs.

III. UNCERTAINTY PROPAGATION RESULTS

A. Power Allocation Uncertainty

We demonstrate that the PDF and CDF results for uncertainty transformation from CSI mis-estimation into power solutions' uncertainty for both WF and NNs in Fig. 1. We can observe that compared to the theoretical WF distribution, the NN solutions have two attributes. If no post-hoc $(\cdot)^+$ operator is used, then the NN produces negative power allocation solutions of diverse nature. If the operator is used (as is in

WF), then solutions are similar (see Fig. 1d-ii/-iii). In the other channels of Fig. 1c, we can see the power distribution from both WF and an example NN. It is clear that the impact on other channels can also be large and should be investigated in future.

TABLE II
UNCERTAINTY PROPAGATION DEVIATION FROM WF BASELINE

NN Module (Xz-Ys)	KL Div.	NN Module (Xz)	KL Div.
NN 10s-10s	0.009	NN 15s	0.013
NN 10p-10s	0.031	NN 15p	0.032
NN 5p-5s	0.028	NN 10s	0.01
NN 5s-5s	0.037	NN 10p	0.13

B. KL Divergence in Different NN Architectures

Using WF heuristic as a benchmark (novel theoretical bound derived in Eq. (2)), we examine the uncertainty transformations' KL divergence in Table II. The NN architectures experimented are labeled as Xz-Ys, where X stands for the number of neurons in the first layer, Y stands for the number of neurons in the second layer (if present), and z stands for the activation function in the first layer (s is sigmoid, y is ReLU), and the second layer activation is always sigmoid (s).

We show the results in descending architecture complexity from a multi-layer 10-10 to a single layer 10 structure. In general, the more sophisticated architectures offer a lower KL divergence (as expected), and we also see that the sigmoid function (s) offers a vastly superior performance in all cases (10z-10s, 15s, 10s) and a similar performance in the (5z-5s) case. We can conclude that indeed more sophisticated activation functions offer a more robust performance (by being closer to the original WF solution), whilst accelerating the algorithm speed by avoiding the search for λ.

IV. CONCLUSIONS & FUTURE WORK

Multi-channel optimisation relies on accurate channel state information (CSI) estimation. Error distributions in CSI can propagate through optimisation algorithms to cause undesirable uncertainty in the solution space. The transformation of uncertainty distributions differs between classic heuristic and Neural Network (NN) algorithms. Here, we examined uncertainty propagation in both classic heuristic Water-Filling (WF) power allocation and different Neural Network (NN) accelerated implementations. We derived a theoretical uncertainty distribution for WF and used KL divergence to measure the difference between different NN architectures against WF. Generally the activation function choice is more important than the size of the neural network, which will inform the design priority of future NNs.

Our future work will aim to quantify the reason for this finding via other supporting XAI features such as LIME [14].

REFERENCES

[1] R. Li, Z. Zhao, X. Zhou, G. Ding, Y. Chen, Z. Wang, and H. Zhang, "Intelligent 5G: When Cellular Networks Meet Artificial Intelligence," *IEEE Wireless Communications*, vol. 24, pp. 175–183, Oct. 2017.

[2] Q. Mao, F. Hu, and Q. Hao, "Deep learning for intelligent wireless networks: A comprehensive survey," *IEEE Communications Surveys Tutorials*, vol. 20, no. 4, pp. 2595–2621, Fourthquarter 2018.

[3] G. Li, J. Hoydis, E. Carvalho, A. Stimming, and Z. Qin, "Best Readings in Machine Learning in Communications," https://www.comsoc.org/publications/best-readings/machine-learning-communications, 2019.

[4] B. Ma, W. Guo, and J. Zhang, "A survey of online data-driven proactive 5g network optimisation using machine learning," *IEEE Access*, vol. 8, pp. 35 606–35 637, 2020.

[5] F. Liang, C. Shen, W. Yu, and F. Wu, "Towards optimal power control via ensembling deep neural networks," *IEEE Transactions on Communications*, pp. 1–1, 2019.

[6] N. Farsad and A. Goldsmith, "Neural network detection of data sequences in communication systems," *IEEE Transactions on Signal Processing*, vol. 66, no. 21, pp. 5663–5678, Nov 2018.

[7] L. Zhao, Y. Wang, and P. Chargé, "Efficient iterative water-filling power allocation method in MU-MIMO broadcast channels," in *Military Communications and Information Systems Conference*, Oct 2013, pp. 1–5.

[8] E. Altman, K. Avrachenkov, and A. Garnaev, "Closed form solutions for symmetric water filling games," in *IEEE INFOCOM 2008-The 27th Conference on Computer Communications*. IEEE, 2008, pp. 673–681.

[9] R. S. Prabhu and B. Daneshrad, "An energy-efficient water-filling algorithm for ofdm systems," in *2010 IEEE International Conference on Communications*. IEEE, 2010, pp. 1–5.

[10] Y. Lu, H. Lu, L. Cao, F. Wu, and D. Zhu, "Learning deterministic policy with target for power control in wireless networks," in *2018 IEEE Global Communications Conference (GLOBECOM)*, Dec 2018, pp. 1–7.

[11] X. Li, J. Fang, W. Cheng, and et. al., "Intelligent power control for spectrum sharing in cognitive radios: A deep reinforcement learning approach," *IEEE Access*, vol. 6, pp. 25 463–25 473, 2018.

[12] H. Owhadi, C. Scovel, and T. Sullivan, "On the brittleness of bayesian inference," *SIAM Review*, vol. 57, no. 4, p. 566–582, April 2015.

[13] M. Abbaszadeh, G. Moutsinas, P. J. Thomas, and W. Guo, "Uncertainty quantification in molecular signals using polynomial chaos expansion," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 4, no. 4, pp. 248–256, Dec 2018.

[14] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should i trust you?": Explaining the predictions of any classifier," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, 2016, pp. 1135–1144.

[15] W. Guo, "Explainable Artificial Intelligence (XAI) for 6G: Improving Trust between Human and Machines," *arXiv:1911.04542*, Nov 2019.

[16] S. C. et al., "Interpretability of deep learning models: A survey of results," in *IEEE SmartWorld*, Aug 2017, pp. 1–6.

[17] W. Guo, "Partially Explainable Big Data Driven Deep Reinforcement Learning for Green 5G UAV," in *IEEE International Conference on Communications (ICC)*. IEEE, 2020.