

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

Digital Global Systems, Inc.

Petitioner,

v.

DeepSig, Inc.,

Patent Owner.

IPR2024-01358

Patent 11,018,704

PETITION FOR *INTER PARTES* REVIEW OF

U.S. PATENT NO. 11,018,704

Mail Stop PATENT BOARD
Patent Trial and Appeal Board
U.S. Patent & Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
II.	GROUND FOR STANDING (37 C.F.R. §42.103(a))	2
III.	IDENTIFICATION OF CHALLENGE (37 C.F.R. §42.104(b))	2
A.	Citation of Prior Art	2
B.	Statutory Grounds for Challenge	3
C.	Level of Skill in the Art	3
IV.	TECHNOLOGY BACKGROUND	4
A.	Signals	4
B.	Nonlinear Signal Distortion	4
C.	Signal Processing and Predistortion.....	5
D.	Machine Learning and Deep Learning	6
E.	The '704 Patent's Alleged Innovations Were Well Known	7
F.	Prosecution History	9
G.	Prior Art	10
1.	Jüschke.....	10
2.	Holt	11
3.	Dzierwa.....	14
V.	Claim Construction	15
VI.	GROUND 1: Claims 1-3, 7-8, 9-11, 15-16, 17-19, and 23-24 were obvious over Jüschke.	15
A.	Detailed Description of Jüschke	15
B.	Claim 1 was obvious over Jüschke.	20
1.	Jüschke renders [1P] obvious	20
2.	Jüschke renders [1A] obvious	21
3.	Jüschke renders [1B.1] obvious.....	23
4.	Jüschke renders [1B.2] obvious.....	26
5.	Jüschke renders [1B.3] obvious.....	29
6.	Jüschke renders [1C] obvious.....	31
7.	Jüschke renders [1D] obvious	33

8.	Jüschke renders [1E] obvious.....	36
C.	Claim 2 was obvious over Jüschke.	37
1.	Jüschke renders [2A] obvious	37
2.	Jüschke renders [2B] obvious.....	39
D.	Claim 3 was obvious over Jüschke.	40
1.	Jüschke renders Claim 3 obvious	40
2.	Jüschke renders [3A] and [3C] obvious	43
3.	Jüschke renders [3B] and [3D] obvious	46
4.	Jüschke renders [3E] obvious.....	49
E.	Claim 7 was obvious over Jüschke.	51
1.	Jüschke renders Claim 7 obvious	51
F.	Claim 8 was obvious over Jüschke.	52
1.	Jüschke renders Claim 8 obvious	52
G.	Claim 9 was obvious over Jüschke.	53
1.	Jüschke renders [9P] obvious	53
2.	Jüschke renders [9A] obvious	53
3.	Jüschke renders [9B] obvious.....	54
4.	Jüschke renders [9C]-[9G] obvious.....	54
H.	Claim 10 was obvious over Jüschke.	55
1.	Jüschke renders [10A]-[10B] obvious.....	55
I.	Claim 11 was obvious over Jüschke.	55
1.	Jüschke renders [11A]-[11E] obvious.....	55
J.	Claim 15 was obvious over Jüschke.	55
1.	Jüschke renders Claim 15 obvious	55
K.	Claim 16 was obvious over Jüschke.	55
1.	Jüschke renders Claim 16 obvious	55
L.	Claim 17 was obvious over Jüschke.	55
1.	Jüschke renders [17P] obvious	55
2.	Jüschke renders [17A]-[17E] obvious.....	56
M.	Claim 18 was obvious over Jüschke.	56

1.	Jüschke renders [18A]-[18B] obvious.....	56
N.	Claim 19 was obvious over Jüschke.	56
1.	Jüschke renders [19A]-[19E] obvious.....	56
O.	Claim 23 was obvious over Jüschke.	56
1.	Jüschke renders Claim 23 obvious	56
P.	Claim 24 was obvious over Jüschke.	56
1.	Jüschke renders Claim 24 obvious	56
VII.	GROUND 2: Claims 1-24 were obvious over Jüschke in view of Holt.....	57
A.	Modifying Jüschke’s Training with Holt’s Loss Function	57
B.	Motivation for Modifying Jüschke’s Training.....	60
C.	Claims 1, 9, and 17 were obvious over Jüschke in view of Holt.....	64
1.	Jüschke in view of Holt renders [1E] obvious	64
2.	Jüschke in view of Holt renders similar limitations [9G] and [17E] obvious.....	67
D.	Claims 4, 12, and 20 were obvious over Jüschke in view of Holt.....	67
1.	Jüschke in view of Holt renders [4A] obvious	67
2.	Jüschke in view of Holt renders [4B] obvious	71
3.	Jüschke in view of Holt renders [4C] obvious	74
4.	Jüschke in view of Holt renders obvious similar claims 12 and 20	76
E.	Claims 5, 13, and 21 were obvious over Jüschke in view of Holt.....	77
1.	Jüschke in view of Holt renders Claim 5 obvious.....	77
2.	Jüschke in view of Holt renders obvious similar claims 13 and 21	78
F.	Claims 6, 14, and 22 were obvious over Jüschke in view of Holt.....	78
1.	Jüschke in view of Holt renders Claim 6 obvious.....	78
2.	Jüschke in view of Holt renders similar claims 14 and 22 obvious	80
VIII.	GROUND 3: Claims 2, 10, and 18 and dependents were obvious over Jüschke in view of Holt and Dzierwa.	81
A.	Modifying Jüschke’s Neural Network with Dzierwa’s Environment Data and Optimization Module	81
B.	Motivation for Modifying Jüschke’s Neural Network	83
C.	Claims 2, 10, and 18 were obvious over Jüschke in view of Holt and	

Dzierwa.	85
1. Jüschke in view of Holt and Dzierwa renders [2B] obvious.....	85
2. Jüschke in view of Holt and Dzierwa renders obvious similar limitations [10B] and [18B].....	86
IX. The Board Should Not Exercise Discretion to Deny Institution	87
A. <i>Fintiv</i> Favors Institution.....	87
B. The <i>Advanced Bionics</i> Factors Favor Institution.....	87
X. CONCLUSION	87
XI. MANDATORY NOTICES.....	88
A. Real Parties-in-Interest Under 37 C.F.R. §42.8(b)(1).....	88
B. Related Matters Under 37 C.F.R. §42.8(b)(2)	88
C. Lead and Back-Up Counsel Under 37 C.F.R. §42.8(b)(3)	88
D. Service Information Under 37 C.F.R. §42.8(b)(4)	88

EXHIBIT LIST	
Exhibit No.	Description
1001	U.S. Patent 11,018,704 to O'Shea et al. ("704 patent")
1002	U.S. Patent 11,018,704 File History
1003	Declaration of Zhi Ding, Ph. D.
1004	Curriculum Vitae of Zhi Ding, Ph. D.
1005	EP 2,538,553 A1 by Jüschke et al. ("Jüschke")
1006	U.S. Patent No. 10,552,738 to Holt et al. ("Holt")
1007	U.S. Patent No. 10,122,479 to Dzierwa ("Dzierwa")
1008	Cooper, V, and Morcom, W. <i>Band I television-transmitter design, with particular reference to the transmitters at Crystal Palace</i> , Proceedings of the IEE-Part B: Radio and Electronic Engineering, Paper No. 2075 R (1956), pp. 651-663 ("Cooper")
1009	Saleh, A. and Salz, J. <i>Adaptive linearization of power amplifiers in digital radio systems</i> , Bell System Technical Journal, Vol. 62, No. 4 (1983), pp. 1019-1033 ("Saleh")
1010	Rosenblatt, F. <i>The perceptron: a probabilistic model for information storage and organization in the brain</i> , Psychological Review, Vol. 65, No. 6, (1958), pp. 386-408 ("Rosenblatt")
1011	Sapna, S., et al. <i>Backpropagation learning algorithm based on Levenberg Marquardt Algorithm</i> . Comp Sci Inform Technol (CS and IT), (2012) pp. 393-398 ("Sapna")
1012	Van Trees, H., et al. <i>Detection, Estimation and Modulation Theory, Part I</i> , John Wiley & Sons, (2004), https://books.google.com/books?id=Xzp7VkuFqXYC (excerpted) ("Van Trees")
1013	Wang, N., et al. <i>Backhauling 5G small cells: A radio resource management perspective</i> , IEEE Wireless Communications Vol. 22, No. 5 (2015), pp. 41-49 ("Wang")
1014	Lanzani, C., et al. <i>Remote radio heads and the evolution towards 4G networks</i> , (2009), pp. 1-6, http://www.altera.com/literature/wp/wp-01096-rrh-4g.pdf ("Lanzani")
1015	Chen, S., et al. <i>Neural networks for nonlinear dynamic system modelling and identification</i> , IEEE Wireless Communications Vol. 56, No. 2 (1992), pp. 319-346 ("Chen")
1016	Ernst, O. <i>Stochastic Gradient Descent Learning and the Backpropagation Algorithm</i> , (2014) ("Ernst")
1017	Bottou, L. <i>Stochastic Gradient Learning in Neural Networks</i> ,

EXHIBIT LIST	
Exhibit No.	Description
	Proceedings of Neuro-Nîmes Vol. 91, No. 8 (1991), https://leon.bottou.org/publications/pdf/nimes-1991.pdf (“Bottou”).

I. INTRODUCTION

Petitioner Digital Global Systems, Inc. (“DGS” or “Petitioner”) requests *inter partes* review (“IPR”) of claims 1-24 of U.S. Patent No. 11,018,704 (“the ’704 patent”).

The ’704 patent purports to disclose and claim techniques to mitigate the problem of signal distortion by adding machine learning models such as neural networks for predistortion. This is nothing new.

Signal distortion is a longstanding problem in electronics and telecommunications. Nonlinear distortion, a phenomenon in components such as amplifiers, undesirably alters the shape of an electric signal by changing the magnitude/phase of its frequency components. Such unwanted modifications distort signal information and can cause signal interference. Predistortion is one method to counter radio signal distortion; it purposely distorts an electric signal to counteract the undesired effects of distortion from another source. It has been known and used for decades. *See* Ex-1008; Ex-1009.

Neural networks have been known since at least the 1950s. *See* Ex-1010. The use of neural networks for predistorting radio signals was disclosed years before the ’704 patent. As one example, European published patent application EP 2,538,553 (Jüschke) was published more than four years before the ’704 patent’s earliest

priority date and explicitly discloses a “a pre-distortion neural network processor.”
Ex-1005, Abstract.

As shown below, the claims of the ’704 patent are obvious over Jüschke and other prior art references.

II. GROUNDS FOR STANDING (37 C.F.R. §42.103(a))

The ’704 patent is available for IPR, and Petitioner is not barred or otherwise estopped from requesting IPR.

III. IDENTIFICATION OF CHALLENGE (37 C.F.R. §42.104(b))

The ’704 patent was filed March 2, 2020. It claims priority to U.S. Application Serial No. 15/955,485, filed April 17, 2018, which claims priority to U.S. Provisional Application No. 62/486,200, filed April 17, 2017. For purposes of this petition, the ’704 patent’s critical date is assumed to be April 17, 2017. All prior art is asserted pursuant to AIA statutes. *See* MPEP §21.59.01.

A. Citation of Prior Art

Each asserted reference was publicly available before the critical date of April 17, 2017.

- EP 2,538,553 A1 by Jüschke et al. (Ex-1005), published on December 26, 2012, is prior art at least under §102(a)(1) because it is a European patent publication published before the ’704 patent’s effective filing date.

- U.S. Patent No. 10,552,738 to Holt et al. (Ex-1006), filed December 15, 2016, is prior art at least under §102(a)(2) because it is a U.S. patent with an earlier effective filing date.
- U.S. Patent No. 10,122,479 to Dzierwa (Ex-1007), filed January 23, 2017, is prior art at least under §102(a)(2) because it is a U.S. patent with an earlier effective filing date.

B. Statutory Grounds for Challenge

Petitioner requests cancellation on the following statutory grounds:

GROUND 1: Claims 1-3, 7-8, 9-11, 15-16, 17-19, 23-24 are obvious over Jüschke under 35 U.S.C. §103.

GROUND 2: Claims 1-24 are obvious over Jüschke in view of Holt under 35 U.S.C. §103.

GROUND 3: Claims 2-6, 10-14, and 18-22 are obvious over Jüschke in view of Holt and Dzierwa under 35 U.S.C. §103.

C. Level of Skill in the Art

Persons of ordinary skill in the art (“POSITA”) would have a bachelor’s degree in electrical engineering, computer engineering, or an equivalent degree and at least two years of experience working with of radio communication systems and

machine learning. Additional work or research experience can substitute for less or different education, and vice-versa. Ex-1003, ¶¶33-34.

IV. TECHNOLOGY BACKGROUND

A. Signals

A signal in electronics and telecommunications is a time-varying electric wave that carries information. Examples of common electric signal applications include audio, video, speech, and image data carried on electromagnetic waves. *See* Ex-1003, ¶¶60-63.

B. Nonlinear Signal Distortion

Nonlinear signal distortion can occur when a signal passes through a nonlinear medium or electronic component. Nonlinear distortion commonly occurs when an amplifier fails to reproduce an output signal as an exact amplified replica of an input signal, leading to spurious frequency components that were not present in the original input signal. *See* Ex-1003, ¶¶64-69.

Generally speaking, at lower amplitudes, linear models accurately represent amplifier behavior. For example, an amplifier may receive an input 2V signal and output a 20V signal; this means the linear gain is 10. But linear models do not accurately represent the behavior of most amplifiers at higher amplitudes. For example, most audio amplifiers cannot linearly amplify a 50V signal to 500V. *See*

Ex-1003, ¶¶70-71.

Furthermore, radio transmitter amplifiers in telecommunications systems ideally are linear to accurately reproduce input signals as amplified output signals. Amplifiers with a nonlinear input/output splatter output signals onto adjacent radio frequencies, creating undesired signal interference on other radio channels. Ex-1003, ¶72.

C. Signal Processing and Predistortion

Signal processing is a subfield of electrical engineering that focuses on analyzing, modifying, and synthesizing signals. Predistortion is a signal processing technique, known for decades, that counteracts nonlinear signal distortion by pre-compensating for it. *See* Ex-1008. This improves amplifier linearity and minimizes distortion, splattering, and interference. *See* Ex-1003, ¶¶73-75.

As an example, an electronic circuit that improves linear performance, such as a linearizer, is paired with a nonlinear amplifier. Ex-1003, ¶75. The linearizer introduces an inverse distortion into the input signal to compensate for the amplifier's nonlinearity. Predistortion can be implemented both in analog and digitally. *See* Ex-1009.

D. Machine Learning and Deep Learning

Machine learning (“ML”) is a field of artificial intelligence (“AI”) that uses datasets and algorithms to generate desired outputs from inputted data. ML imitates human learning, gradually improving accuracy with more data and training. ML uses models for data-classification and decision-making. An example of an ML model in deep learning is a neural network that models a human brain. Ex-1003, ¶76.

Artificial neural network (“ANN”) models (or neural network models) consist of nodes or neurons interconnected in layered structures that relate inputs to outputs. They have existed since the 1950s. *See* Ex-1010. Neural networks are trained by iteratively updating neuron connectivities (i.e., parameters) so that inputs correctly map to desired output responses during training.

Neural networks also have hidden layers in between input and output layers. The more hidden layers, the “deeper” the neural network. Neural networks are nonlinear models owing to their use of so-called nonlinear activation functions. Given enough training data and neurons, a neural network can approximate a given function map between input and output data. Ex-1003, ¶77. Exemplary neural network models include convolutional neural networks (feed-forward, meaning information only flows forward) and recurrent neural networks (feedback loops to allow information to flow back in addition to feedforward layers). *See id.*, ¶78.

Neural networks are most commonly trained using backpropagation, an optimization technique to update deep learning model parameters before the output layer. Backpropagation computes the gradient of a loss function with respect to a neural network's parameter weights for input/output one layer at a time. Gradient descent, an iterative optimization process that aims to optimize an objective function optimum value (minimum or maximum), is typically used in backpropagation to adjust parameters. Common gradient descent types are Stochastic Gradient Descent and Batch Gradient Descent. Ex-1003, ¶79.

E. The '704 Patent's Alleged Innovations Were Well Known

The '704 patent purports to improve radio predistortion using “a nonlinear pre-distortion machine learning model” (hereinafter, “NPDMLM”). Ex-1001, Abstract. The NPDMLM “takes a transmit radio signal 104 as input and generates a pre-distorted signal 105 as output.” *Id.*, 5:60-61; *see also id.*, Figures 6A (“Analog RF in 601”), 7A (“Digital RF In 701”). Next, the “model provides pre-distorted signals 105 [to] the transmitting radio signal processing system 101 processes...to obtain a ready-to-transmit pre-distorted signal 107.” *Id.*, 10:7-10. The signal is then exposed to a distortion source (e.g., amplifier) and transmitted as a transmit output signal. Figure 1 of the '704 patent illustrates the overall system:

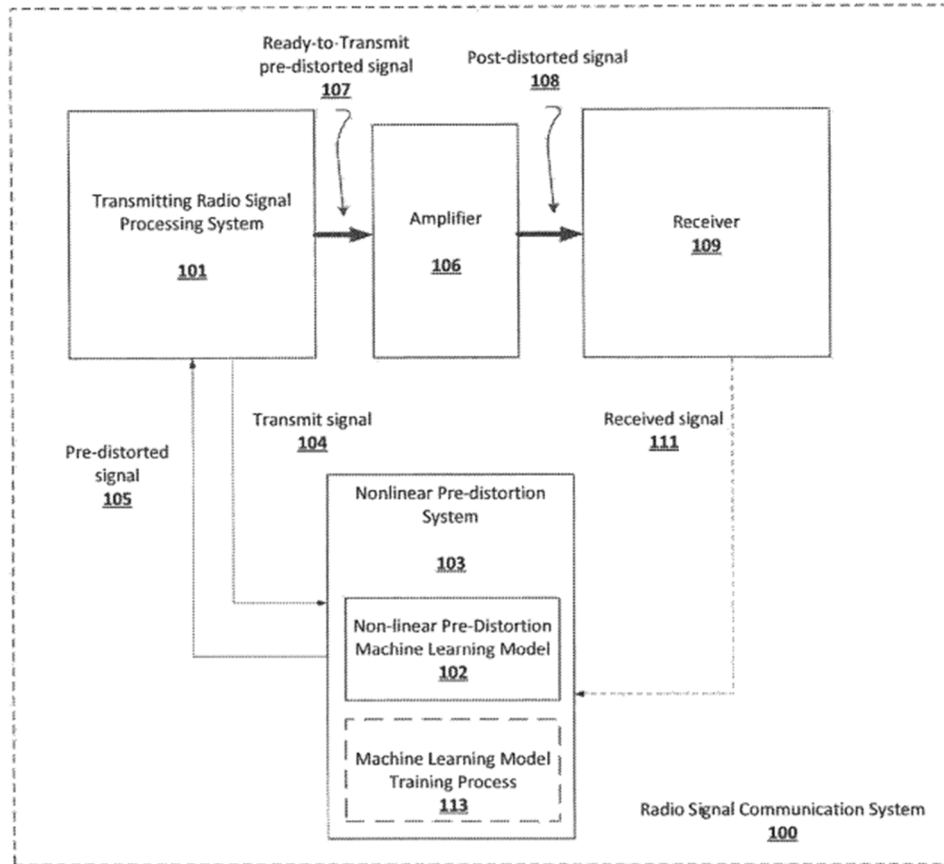


FIG. 1

Ex-1001, Figure 1; *see also id.*, 5:35-39 (noting that while Amplifier 106 is shown as separate, it can be part of Transmitting Radio Signal Processing System 101).

None of this was new. The NPDMLM is a typical nonlinear ML model, such as “a deep neural network, a convolutional neural network..., a recurrent neural network, or a combination of different types of neural networks.” *Id.*, 3:53-55. Nonlinear ML models were known for years before the ’704 patent’s earliest priority date, as was applying nonlinear ML models for predistortion. *See* Ex-1005, Abstract (“pre-distortion neural network processor...for mitigating impairments due to transmit noise” and “mitigating transmitter impairments of a transmit signal”); Ex-

1010; Ex-1003, ¶82. During prosecution, the Examiner found “generating a pre-distorted radio signal...using a nonlinear pre-distortion machine learning model” and “model parameters and at least one nonlinear function to correct radio signal distortion” disclosed by a single reference. Ex-1002, 63-64.

The '704 patent's model parameters are trained using a “distance metric [] computed between the candidate transmit radio signal and the received radio signal.” Ex-1001, 3:14-16. This distance metric, however, is simply a similarity value between input/output data used for training the NPDMLM. Ex-1003, ¶¶92-93.

The '704 patent also discloses “deployment scenarios.” Ex-1001, 2:38; *see id.*, Figures 6A, 6B, 7A, 7B, 9. These are nothing more than basic hardware configurations well within the knowledge of POSITA. Ex-1003, ¶¶84-91.

F. Prosecution History

During prosecution, the Examiner found the sole originally presented claim anticipated. Ex-1002, 63-64. To traverse prior art, the Applicant added amendments requiring that NPDMLM parameters be “associated with values corresponding to one or more deployment scenarios in [a] cellular communications system.” *Id.*, 44. The Applicant also added “cellular communications” and “base station” limitations, and required that the NPDMLM be “associated with the base station.” *Id.*

G. Prior Art

1. Jüschke

Jüschke was published more than four years before the '704 patent and addresses the same problem (transmit distortion such as amplifier nonidealities and transmit noise in radio signals) and discloses the same solution (predistortion using nonlinear ML models such as neural networks).

Jüschke teaches “concept[s] for mitigating impairments of a transmit signal due to transmitter imperfections” and “transmit noise” in wireless communications. Ex-1005, [0001], Abstract. Although transmit noise can be compensated via transmitter “calibration or pre-distortion” or “sophisticated compensation algorithms” at a receiver, these “fixed” techniques are a “bottle-neck” in complex mobile communication system architectures. Ex-1005, [0005]. Jüschke solves this with “a more flexible radio transmitter.” *Id.*, [0006].

Jüschke’s transmitter implements a “pre-distortion neural network processor” that pre-distorts input signals to generate pre-distorted signals, which are sent to an analog transmitter for transmission. Ex-1005, [0010]. The analog transmitter includes a distortion cause (e.g., an amplifier) that the pre-distortion neural network processor compensates for. *Id.* The “pre-distortion neural network processor may comprise a multi-layer feed forward neural network (MLNN).” *Id.*, [0013]; *see* Ex-1003, ¶100.

Jüschke trains its neural networks by analyzing input and output signals. “Sampled output values” from a transmitter and “ideal digital baseband IQ [input] samples” are sent to a “post distortion neural network 43” within a training processor. Ex-1005, [0048]. After training, weight and bias parameters are stored and transferred to the pre-distortion neural network processor. *Id.*, [0049]; see Ex-1003, ¶¶102-103.

2. Holt

Holt was filed four months before the '704 patent. It discloses training nonlinear ML models, including neural networks, implemented in wired and wireless communication systems.

Holt teaches techniques for “adaptive training of a channel coding model including an encoder model, a channel model..., and a decoder model.” Ex-1006, Abstract. The models can be neural networks or other “multi-layer non-linear model[s].” See *id.*, 4:62-63, 5:10-12, 5:27-29. “[T]he encoder model can be implemented by a transmitter device or component to encode communication data prior to transmission of the communication data via the communication channel, while the decoder model can be implemented by a receiver device or component to decode the communication data received via the communication channel.” *Id.*, 4:33-39. The channel model can be placed in between the encoder model and the decoder model, and can represent a “wireless channel (e.g., a radio frequency (RF) link)” and

“noise.” *Id.*, 5:18-19, 5:47; *see* Ex-1003, ¶105.

Holt’s models “can be trained on a loss function that describes a difference between input data input into the encoder model and output data received from the decoder model.” Ex-1006, Abstract. This “loss function can be backpropagated through the decoder model while modifying the decoder model, backpropagated through the channel model while the channel model is held constant, and then backpropagated through the encoder model while modifying the encoder model.” *Id.* Holt’s channel model can “emulate a communication channel” and account for “the addition of noise to an output of the encoder model.” *Id.*, 4:5-6, 5:47-48. Holt also teaches that an “actual communication channel” can be used. *Id.*, 18:1-2. Therefore, POSITA would have understood that Holt’s encoder and decoder models could be implemented with an actual communication channel. Ex-1003, ¶106.

Holt’s Figure 3 shows an exemplary channel coding model (200) with an encoder model (202), a channel model (208), a decoder model (212), a third set of outputs (214), and a first loss function (216):

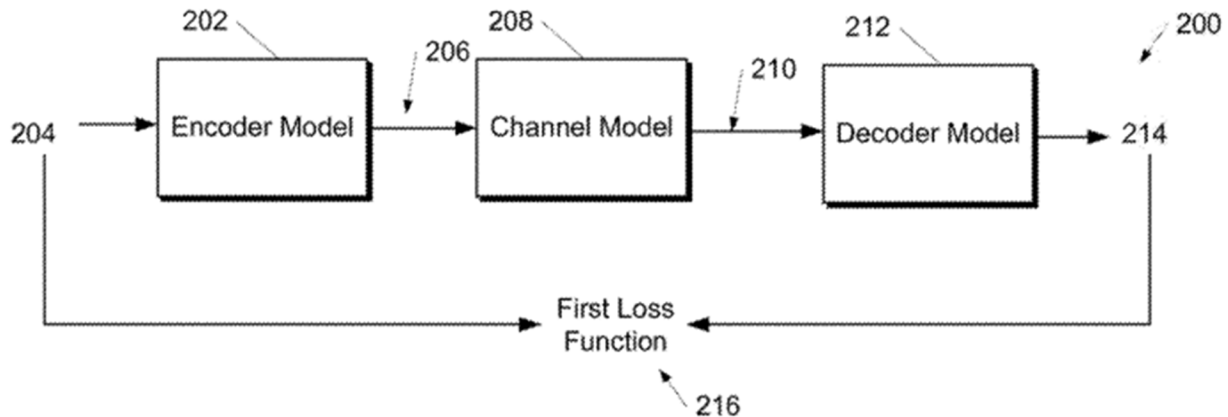
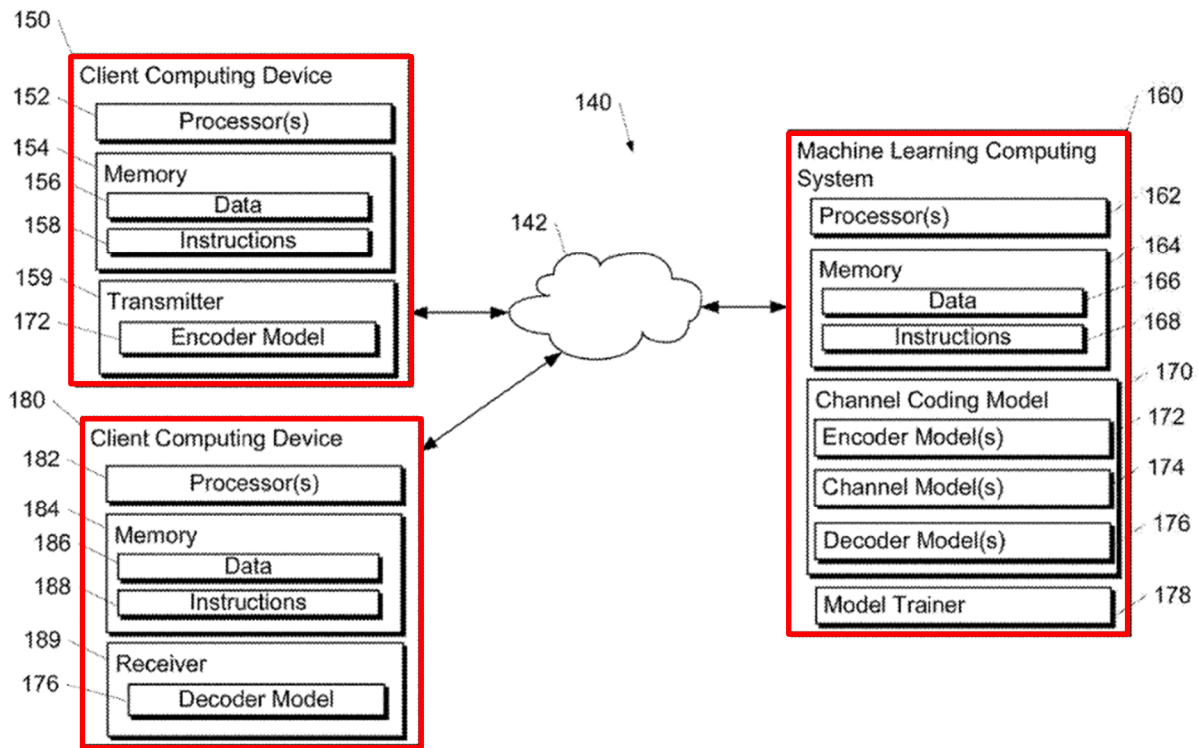


Figure 3

Ex-1006, Figure 3; *see also id.*, Figure 6 (“Actual Channel” 302 and “Fourth Loss Function” 318).

Holt’s system and models can be implemented within “any type of computing device,” such as “a mobile computing device (e.g., smartphone or tablet)” or “a network computing device such as a base station, router, beacon, other communication node.” Ex-1006, 9:34-39. As shown in Figure 2, Holt’s encoder, channel, and decoder model(s) are implemented on computing devices 150, 160, and 180:



Ex-1006, Figure 2 (annotations added).

3. Dzierwa

Dzierwa was filed about three months before the '704 patent and teaches radio signal detection and optimization using machine learning. Dzierwa “observe[s] and learn[s] an RF environment over time and identif[ies] temporal features” and generates “[a] knowledge map...based on learning data.” Ex-1007, 47:38-41. Detected signal events are “scrubbed against the knowledge map to determine if the real-time signal events are typical and expected for the environment.” *Id.*, 47:43-45. This enables detection of RF signals with “low power level[s]” and “narrowband signal[s] buried in a wideband signal.” *Id.*, 9:23-25.

V. Claim Construction

No claim term requires construction to resolve the patentability challenges here. Petitioner therefore interprets all claim terms in accordance with their plain and ordinary meaning.

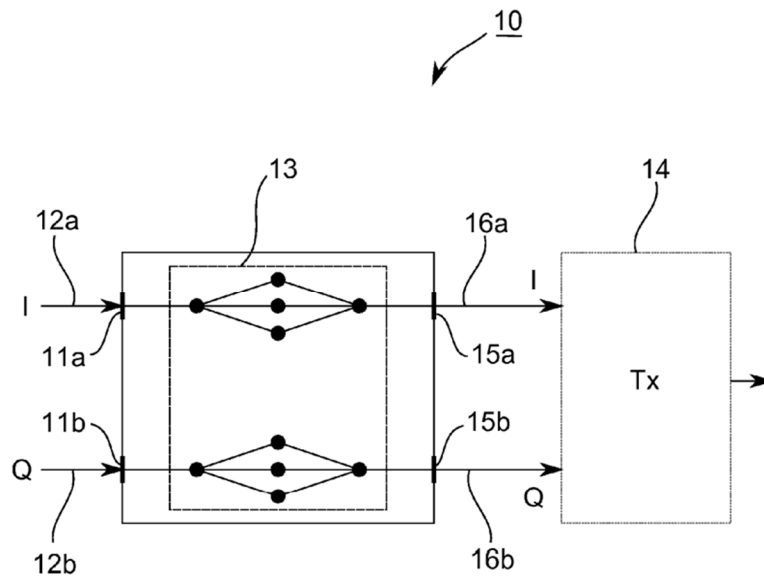
VI. GROUND 1: Claims 1-3, 7-8, 9-11, 15-16, 17-19, and 23-24 were obvious over Jüschke.

A. Detailed Description of Jüschke

Jüschke's predistortion teachings use neural networks to implement a "more flexible radio transmitter that can support as many standards and signals as possible." Ex-1005, [0006]. Jüschke discloses that "different parts of the system can be exchanged and adapted easily" and implemented in various deployments including "base stations," and throughout "mobile communication systems." *Id.*, [0006], [0059]; *see id.*, [0002] (disclosing exemplary wireless networks); *see also* Ex-1003, ¶¶112, 121-123. Given Jüschke's emphasis on a "flexible radio transmitter" (Ex-1005, [0006]), POSITA would have understood that Jüschke's teachings can be implemented in essentially any state-of-the-art transceiver system in wireless communications. Ex-1003, ¶123.

Figure 1 below illustrates apparatus 10. As shown, "input signals 12a, 12b are coupled to a pre-distortion neural network processor 13 for pre-distorting the input digital baseband transmit signal 12 in order to obtain a pre-distorted digital baseband

signal 16 with quadrature components 16a, 16b.” Ex-1005, [0039].¹ The pre-distorted signal (16, 16a, 16b) is then sent to analog transmitter device 14 where it is subjected to distortion (e.g., by an amplifier) and transmitted. *See id.*, [0050] (“analog amplifier device in the transmitter device 14”). See Figure 1 below:



¹ POSITA would have understood that “quadrature components” generally refer to the in-phase (I) and quadrature (Q) components in RF modulations. Ex-1003, ¶113; Ex-1005, [0039] (“in-phase and a quadrature component 16a, 16b”).

Ex-1005, Figure 1.²

Because POSITA would have understood that transmitted radio signals are analog signals, POSITA would have understood that any digital signals received by analog transmitter device 14, such as pre-distorted digital baseband signal 16, must be converted using a digital-to-analog converter before being transmitted as a radio signal. Therefore, POSITA would have understood that analog transmitter device 14 includes a digital-to-analog converter to transmit the pre-distorted digital signal 16. Ex-1003, ¶114.

Jüschke’s “(artificial) neural networks” mitigate “transmitter impairments of a transmit signal” such as “IQ [in-phase/quadrature-phase] imbalance due to an IQ up-conversion mixer in the transmit path of a transceiver, as well as nonlinearities due to a power amplifier (PA) in said transmit path.” Ex-1005, [0008]-[0009]; *see*

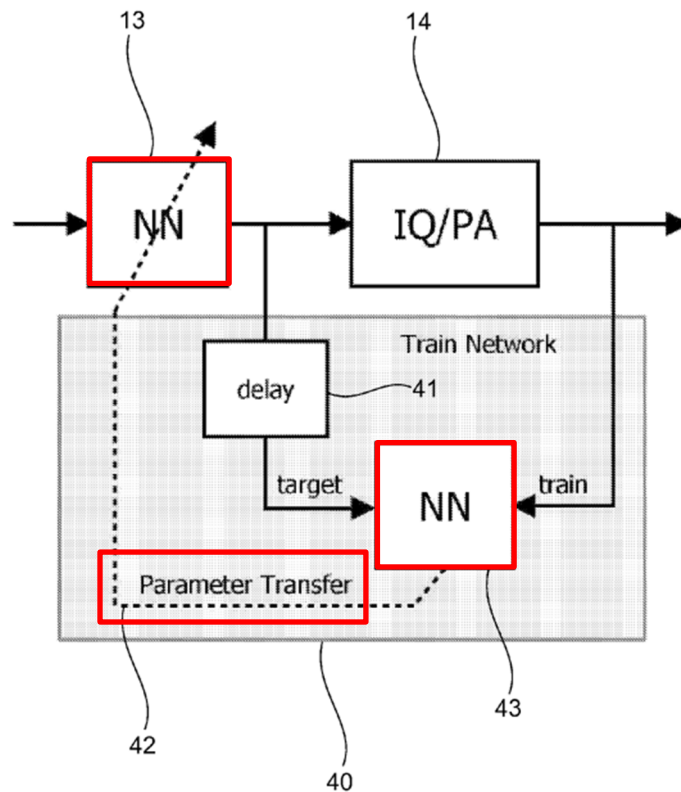
² Jüschke refers to item 13 as both a pre-distortional neural network processor and a pre-distortion neural network. *See* Ex. 1005, Abstract, [0039]-[0058]. POSITA would have understood that a pre-distortion neural network is implemented, through hardware, software, or a combination, on or by a pre-distortion neural network processor. Therefore, POSITA would have understood that both terms are accurate and often used interchangeably. Ex-1003, ¶102.

Ex-1003, ¶115. The pre-distortion neural network processor approximates an “inverse” transfer function to compensate for these impairments. Ex-1005, [0039]; *see also id.*, [0058].

For training its neural networks, Jüschke compares input and output signals in a feedback loop and updates the weight and bias parameters using a training processor. Ex-1003, ¶117. “Sampled output values of the non-ideal analog transmitter device 14 may be fed back as input samples for the post distortion neural network 43 [within training processor 40]. Undistorted and, hence, ideal digital baseband IQ samples 12a, 12b may be provided to the post distortion neural network 43 as target output values.” Ex-1005, [0048].

For training “weight (w_{kj}) and/or bias parameters (b_k),” Jüschke teaches that using “[g]radient-based training algorithms, like back-propagation, may be used,” while “focus[ing] on the well-known Levenberg-Marquardt algorithm.” Ex-1005, [0018]. POSITA would have understood that the Levenberg-Marquardt algorithm has been used in backpropagation training of neural networks. *See* Ex-1011; Ex-1003, ¶118. Jüschke clarifies that “training may take several epochs and may be terminated by MSE (Minimum Squared Error).” Ex-1005, [0048]. “After convergence of the post distortion neural network,” the trained weight and bias parameters are “transferred to the pre-distortion neural network processor 13.” *Id.*,

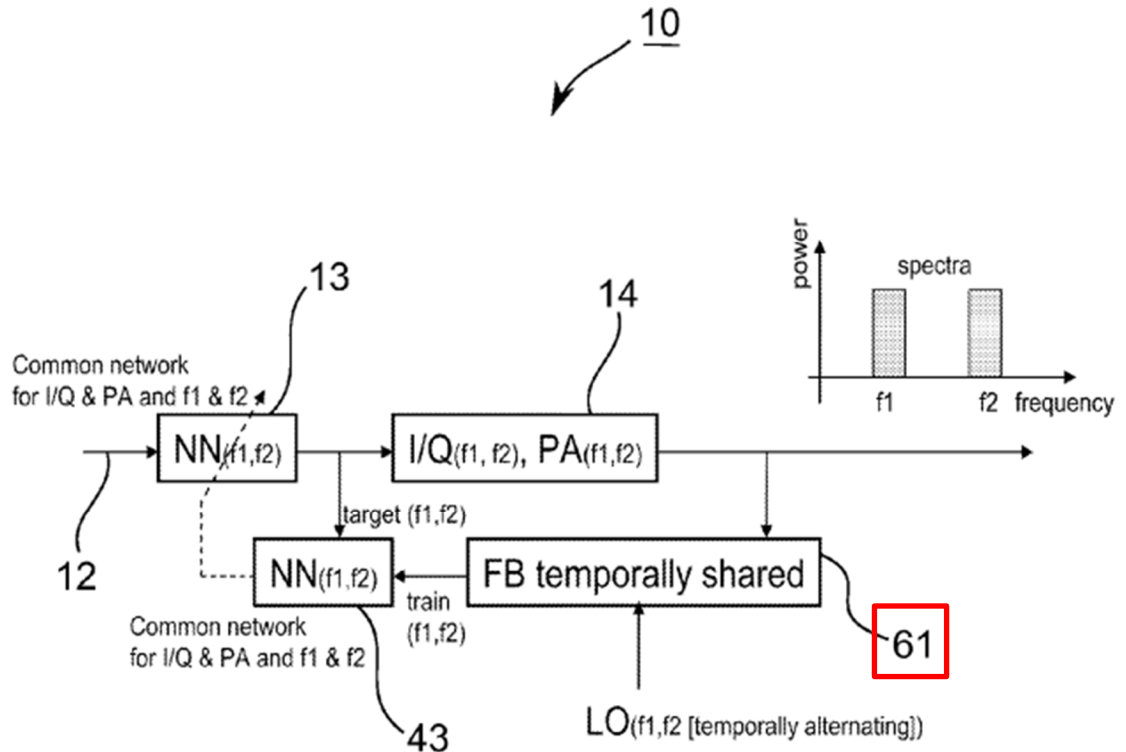
[0049]. Figure 4 below illustrates the parameters being trained and transferred:



Ex-1005, Figure 4 (annotations added).

Jüschke further teaches an adaptive “self learning/optimizing neural network” that “consider[s] and optimiz[es]...parameters.” Ex-1005, [0059]. “[I]nformation on the transmitter device’s 14 output signal quality may e.g. be provided to the neural networks 13, 43 by use of a feedback path 61.” *Id.*, [0060]. The “feedback path 61, preferably for mitigating TX noise caused by the transmitter device 14, ...provides an output signal of the transmitter device 14 to the post distortion neural network processing unit 43.” *Id.* The feedback signal “enabl[es] adaptive pre-distortion of the

baseband signal.” *Id.*; *see id.*, [0072] (“The neural network may learn from a feedback signal and adapt itself.”). Figure 6 below demonstrates feedback path 61 used for adaptive predistortion:



Ex-1005, Figure 6 (annotation added). *See* Ex-1003, ¶120.

B. Claim 1 was obvious over Jüschke.

1. Jüschke renders [1P] obvious

Jüschke renders obvious “[a] method for correcting distortion of radio signals in a cellular communications system” [1P].

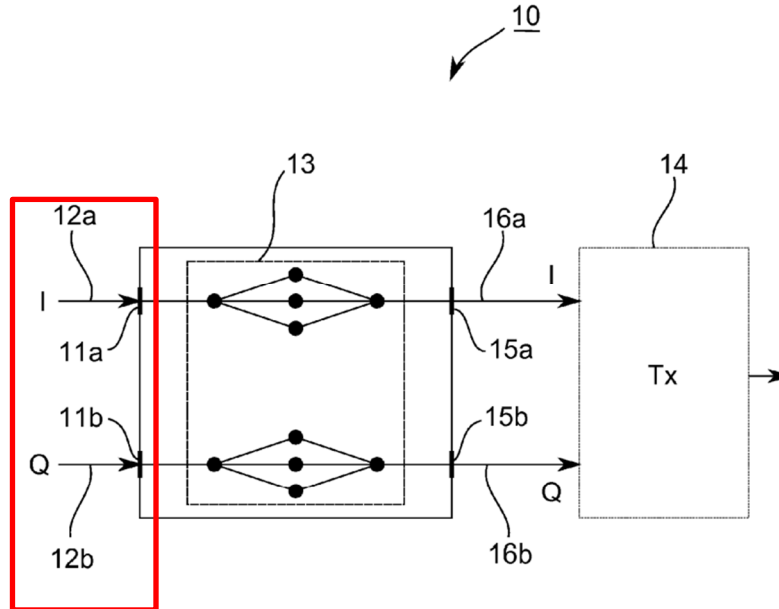
Jüschke’s teachings “generally relate to wireless communications and, more specifically, to a concept for mitigating impairments of a transmit signal due to

transmitter imperfections” and “transmit noise.” Ex-1005, [0001], Abstract. Jüschke discloses “apparatus (10) for mitigating transmitter impairments of a transmit signal” and a “a pre-distortion neural network processor.” *Id.*, Abstract. “[A]pparatus 10 may also be applied” for multiple “mobile communication systems, like e.g. GSM, WCDMA and/or LTE.” *Id.*, [0059]; *see also id.*, [0002].

2. Jüschke renders [1A] obvious

Jüschke renders obvious “*obtaining, at a base station in the cellular communications system, a transmit radio signal that is prepared for transmitting to one or more radio receivers associated with the cellular communications system*” [1A].

First, for *obtaining...a transmit radio signal that is prepared for transmitting to one or more radio receivers*, Jüschke discloses “apparatus (10) for mitigating transmitter impairments of a transmit signal, the apparatus comprising an input (11a; 11b) for [*obtaining*] an in-phase and a quadrature component (12a; 12b) of a digital baseband transmit signal (12).” Ex-1005, Abstract. As shown in Figure 1 below, “input digital baseband transmit signal (12; 12a; 12b)” is an input signal for apparatus 10 (*id.*):



Ex-1005, Figure 1 (annotation added). Given that input transmit signal 12 is shown flowing towards an analog transmitter (Tx 14), POSITA would have understood this to be a *transmit radio signal* being prepared for transmission to a radio receiver. This understanding would have been reinforced by the '704 patent's teaching of digital input signals for predistortion. *See* Ex-1001, Figures 7A, 7B; Ex-1003, ¶130.

Second, for the transmit radio signal being obtained *at a base station in the cellular communications system*, Jüschke teaches that “it is desirable to have a more flexible radio transmitter that can support as many standards and signals as possible,” Ex-1005, [0006], in “systems for mobile communications,” *id.*, [0002]. Therefore, POSITA would have understood that Jüschke's transmitter can be “flexibl[y]” implemented in scenarios and structures, including base stations such as “macro-cell base stations” and “small-cell base stations” and related transceiver

infrastructure. Ex-1005, [0006]. POSITA would have also understood that Jüschke's flexible transmitter can be implemented within a base station and that input transmit signal 12 can be obtained at a base station. *See* Ex-1003, ¶131.

Third, for the radio receivers being *associated with the cellular communications system*, Jüschke teaches that “apparatus 10” is applicable to “multiple generations of cellular systems” and multiple “mobile communication systems, like e.g. GSM, WCDMA and/or LTE,” which implement various radio receivers and transceivers in cellular communication. Ex-1005, [0002], [0059]. POSITA would have understood that a radio receiver was implemented in one of the exemplary cellular communication systems disclosed by Jüschke. *See* Ex-1003, ¶¶132-133.

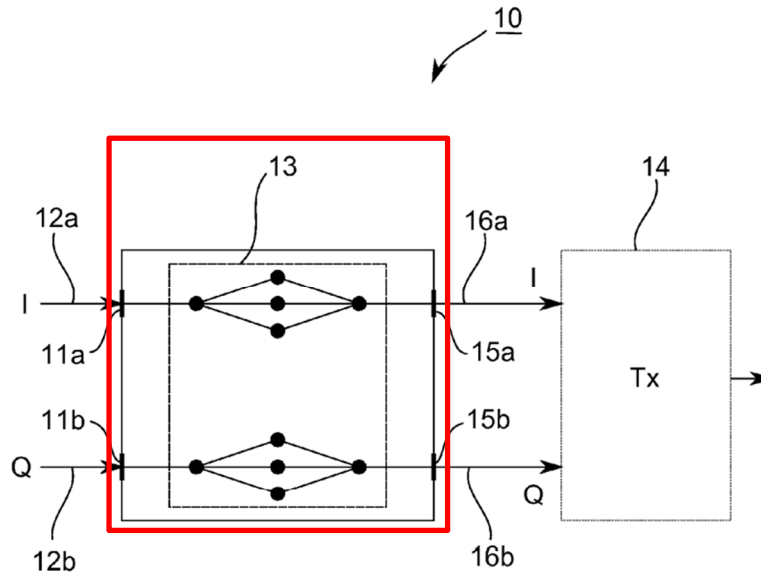
Therefore, Jüschke renders [1A] obvious.

3. Jüschke renders [1B.1] obvious

Jüschke renders obvious “*accessing a nonlinear pre-distortion machine learning model associated with the base station*” [1B.1].

Jüschke's “input signals 12a, 12b are coupled to a pre-distortion neural network processor 13 for pre-distorting the input digital baseband transmit signal 12.” Ex-1005, [0039]; *see id.*, Abstract. The neural network processor implements an artificial neural network, a type of ML model, that POSITA would have understood is *accessed* for pre-distorting signals. *See* Ex-1003, ¶136; Ex-1005,

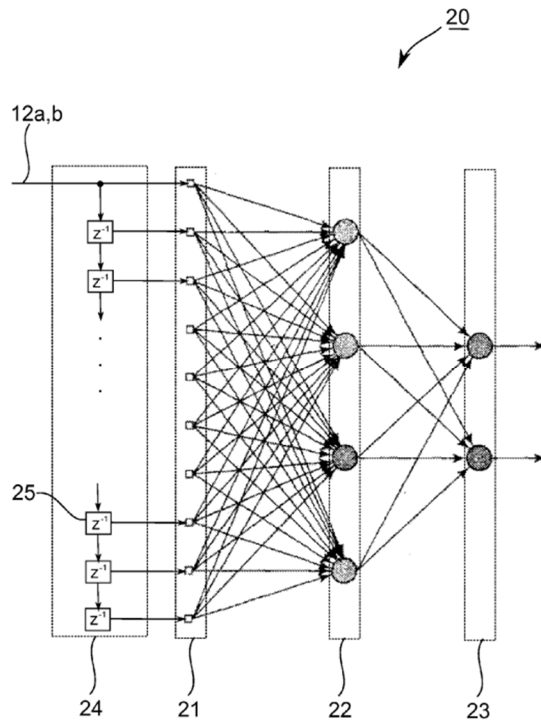
[0008], [0009]. Figure 1 shows apparatus 10 with pre-distortion neural network processor 13:



Ex-1005, Figure 1 (annotation added).

Jüschke teaches that its predistortion model “may comprise a multi-layer feed forward neural network (MLNN).” Ex-1005, [0013]. POSITA would have understood such an “MLNN” to be a deep neural network, which POSITA would have known is nonlinear. For example, Jüschke’s neural network processor and neural network teach a ***nonlinear pre-distortion machine learning model***. See Ex-1003, ¶137; Ex-1005, Figure 2, [0008]; *see also* Ex-1001, 4:52-53; Ex-1015, 319. Jüschke also teaches using neural network processor 13 and feed-forward neural network 20 for addressing non-linear distortion caused by amplifiers. Ex-1005, [0072]. This would have reinforced POSITA’s understanding that Jüschke teaches a

nonlinear model. Ex-1003, ¶138. Moreover, Figure 2 depicts neural network 20 with input layer 21, hidden layer 22, and output layer 23:



Ex-1005, Figure 2.

Furthermore, as explained for limitation [1A], Jüschke's teachings, including its flexible radio transmitter, can be implemented within a base station. Therefore, POSITA would have understood that Jüschke teaches that neural network 13, which is fed input 12 and generates output 16 (*accessing a nonlinear pre-distortion machine learning model*), can be *associated with the base station* by being implemented in a flexible transmitter that is itself implemented in a base station. See Ex-1003, ¶¶139-140. Jüschke thereby renders [1B.1] obvious.

4. Jüschke renders [1B.2] obvious

Jüschke renders obvious “*wherein the nonlinear pre-distortion machine learning model includes model parameters associated with values corresponding to one or more deployment scenarios in the cellular communications system*” [1B.2].

First, for *model parameters*, Jüschke’s pre-distortion neural network model includes “weight and/or bias parameters (w_{kj} , b_k) of neurons.” Ex-1005, [0017]; *see also* Ex-1005, [0013], [0014]. Neurons and parameters determine how data inputted into the model is transformed into desired outputs. Ex-1003, ¶142.

Second, for the parameters being *associated with values corresponding to one or more deployment scenarios in the cellular communications system*, Jüschke’s weight and bias parameters are implemented in deployments for GSM and LTE (*cellular communication systems*). For instance, “apparatus 10 may [] be applied to fragmented spectrum applications” for “mobile communication systems, like e.g. GSM, WCDMA and/or LTE.” Ex-1005, [0059]. “[T]he network parameters, i.e. weights w_{kj} , bias values b_k , etc., (addressing I/Q-imbalance and PA non-linearity in parallel for different frequency bands)” are “optimized” for the same. *Id.*

Jüschke’s models and parameters apply to configurations mirroring the *deployment scenarios* explicitly described in the ’704 patent. For instance, the ’704 patent describes one of its “particular deployment scenarios” as “an inline mode” of configurations. Ex-1001, 2:38-41. Figure 6A shows an “in-line mode [configuration]

within a tower-mounted amplifier, placing a nonlinear pre-distortion system 103 between an input 601 and an amplifier 603,” *id.*, 10:39-41, and Figure 7A illustrates an “in-line mode [configuration] within a remote radio unit” with pre-distortion system 103 between input 701 and amplifier 703 (*id.*, 10:47-48):

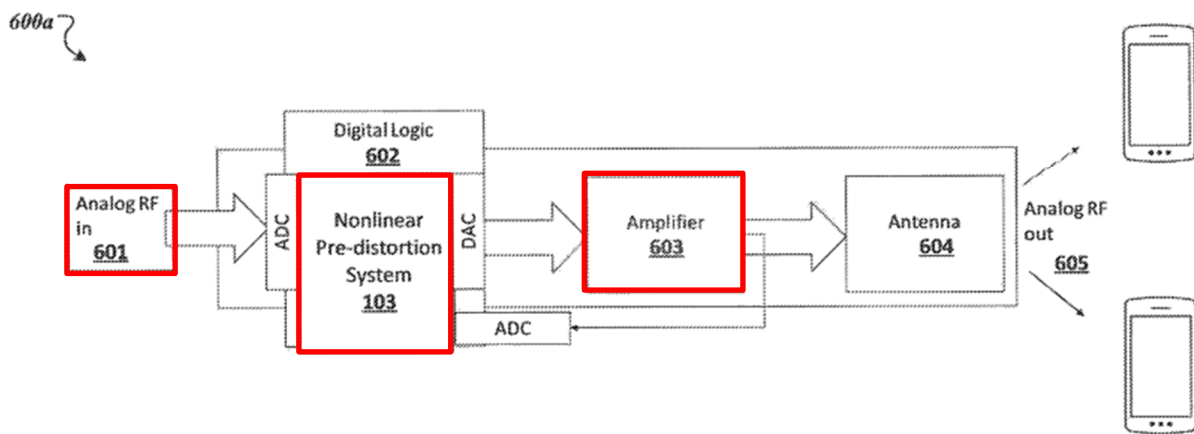


FIG. 6A

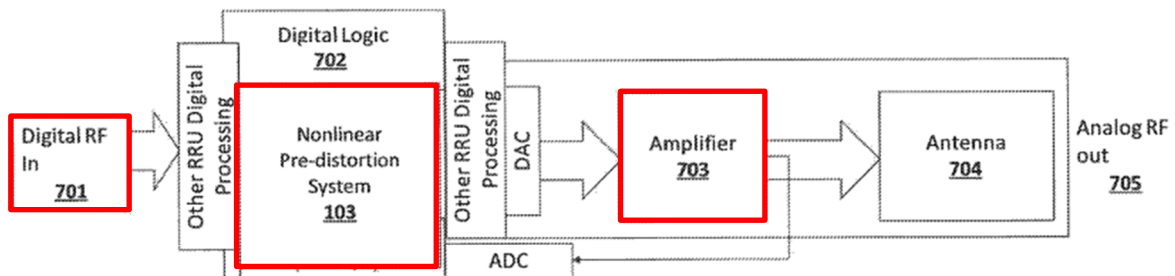
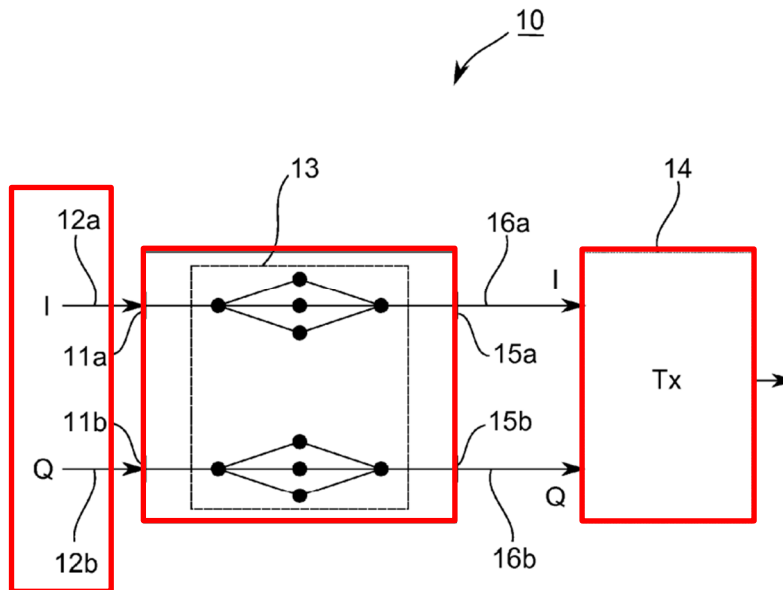


FIG. 7A

Ex-1001, Figures 6A, 7A (annotations added).

Jüschke likewise implements inline-mode deployment scenarios. Figure 1 shows pre-distortion neural network 13 (*compare to '704 patent pre-distortion system 103*) between input transmit signal 12 (*compare to '704 patent input 601/701*) and transmitter device 14 with a power amplifier (*compare to '704 patent amplifier 603/703*). Jüschke's weight and bias parameters (w_{kj} , b_k) are applied by the neural network to predistort transmission signals of this inline-mode deployment configuration:



Ex-1005, Figure 1 (annotations added); *see id.*, [0020] (“[A]t least one nonlinear analog power amplifier device...may be comprised by the analog transmitter device.”), [0050] (“analog amplifier device in the transmitter device 14”).

Moreover, as explained for limitation [1A], POSITA would have understood

that Jüschke’s “flexible radio transmitter” can be implemented in a variety of configurations for cellular communication systems, including base stations and related transceiver infrastructure. Ex-1005, [0006]. POSITA would have further understood these flexible configurations to encompass *one or more deployment scenarios in the cellular communications system* like those described by Figures 6A and 7A of the ’704 patent. Ex-1003, ¶146. And because Jüschke’s weight and bias parameters are implemented and trained for optimizing specific transmission environments (e.g., inline-mode deployment), POSITA would have understood that Jüschke’s *parameters correspond* to the particular configuration they are implemented for. *Id.* POSITA also would have understood that “block diagram of an apparatus 10” (Figure 1) for a flexible radio transmitter teaches a general transmitter hardware configuration that would fit multiple cellular deployment configurations. Ex-1005, [0038]; Ex-1003, ¶147.

Therefore, POSITA would have understood that Jüschke’s weight and bias parameters addressing I/Q and PA corrections teach *model parameters associated with values*, and Jüschke’s configurations, including in-line configurations, teach *one or more deployment scenarios in the cellular communications system*. Jüschke thereby renders [1B.2] obvious.

5. Jüschke renders [1B.3] obvious

Jüschke renders obvious “*and at least one nonlinear function to correct radio*

signal distortion” [1B.3].

As previously explained for limitation [1B.1], Jüschke teaches a neural network model (pre-distortion neural network 13), which POSITA would have understood is a non-linear machine learning model for mitigating transmitter impairments of a transmit signal. *See also* Ex-1001, 1:32-35 (“Neural networks are machine learning models that employ one or more layers of ***nonlinear units*** to predict an output for a received input.” (emphasis added)). Given that POSITA would have understood that Jüschke teaches a nonlinear neural network, POSITA would have also understood that pre-distortion neural network 13 must include at least a *nonlinear function*. *See* Ex-1003, ¶150.

Moreover, Jüschke’s neural network processor 13 is “adaptable or adapted to a transmitter transfer function of at least one analog transmitter device 14 arranged downstream...such that ***a transfer function of the pre-distortion neural network processor 13 approximates an inverse of said transmitter transfer function.***” Ex-1005, [0039] (emphasis added); *see id.*, Abstract, cl. 5. POSITA would have understood that an inverse of a nonlinear transmitter (transfer) function, including approximations of such an inverse, is a nonlinear function. *See* Ex-1003, ¶151.

In addition, the pre-distortion neural network processor may “compensate a nonlinear transmitter transfer function of at least one analog amplifier device in the

transmitter device 14.” Ex-1005, [0050]. “Neurons of the hidden layers 22, 22a and/or 22b each may have *nonlinear transfer or activation functions*.” *Id.*, [0053] (emphasis added). “Due to this nonlinear activation functions embodiments of the pre-distortion neural network processor 13 gain the ability to compensate arbitrary nonlinear power amplifiers.” *Id.* This would have further confirmed to POSITA that Jüschke’s pre-distortion neural network includes at least a nonlinear function for *correcting radio signal distortion*, such as compensating nonlinear power amplifiers. See Ex-1003, ¶152.

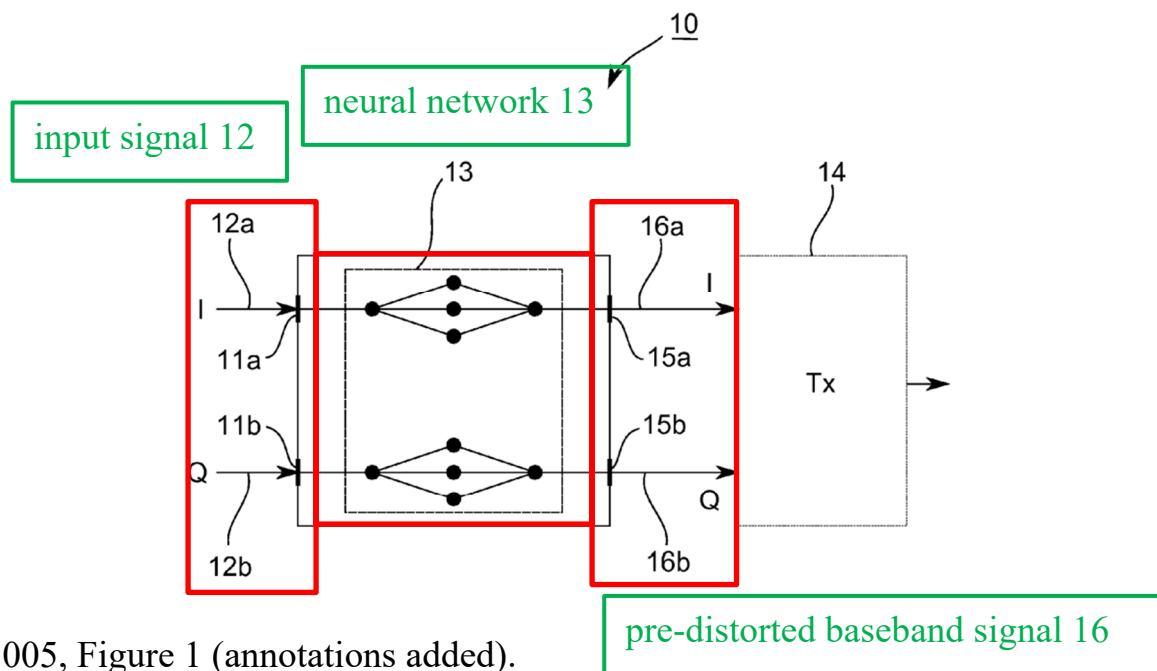
Therefore, POSITA would have understood that Jüschke teaches a neural network processor with *at least one nonlinear function to correct radio signal distortion* by virtue of being a neural network, by implementing inverse transfer functions, and by implementing nonlinear transfer or activation functions. Jüschke thereby renders [1B.3] obvious.

6. Jüschke renders [1C] obvious

Jüschke renders obvious “*generating a pre-distorted radio signal by processing the transmit radio signal using the nonlinear pre-distortion machine learning model*” [1C].

Jüschke’s “input signals 12a, 12b are coupled to a pre-distortion neural network processor 13 for pre-distorting the input digital baseband transmit signal 12 in order to obtain a pre-distorted digital baseband signal 16 with quadrature

components 16a, 16b.” Ex-1005, [0039]; *see id.*, Abstract. As shown for limitation [1A], POSITA would have understood that Jüschke’s input transmit signal 12 teaches a *transmit radio signal*. Figure 1 shows apparatus 10 with input signal components 12a and 12b being fed into neural network processor 13 (via inputs 11a and 11b) and pre-distorted output signal components 16a and 16b being outputted (via outputs 15a and 15b):



Ex-1005, Figure 1 (annotations added).

Therefore, POSITA would have understood that Jüschke’s pre-distorted baseband signal 16 teaches *generating a pre-distorted radio signal*, input signal 12 (into neural network 13) teaches *by processing the transmit radio signal*, and neural network processor 13 teaches *using the nonlinear pre-distortion machine learning model*. See Ex-1003, ¶¶155-156. Jüschke thereby renders [1C] obvious.

7. Jüschke renders [1D] obvious

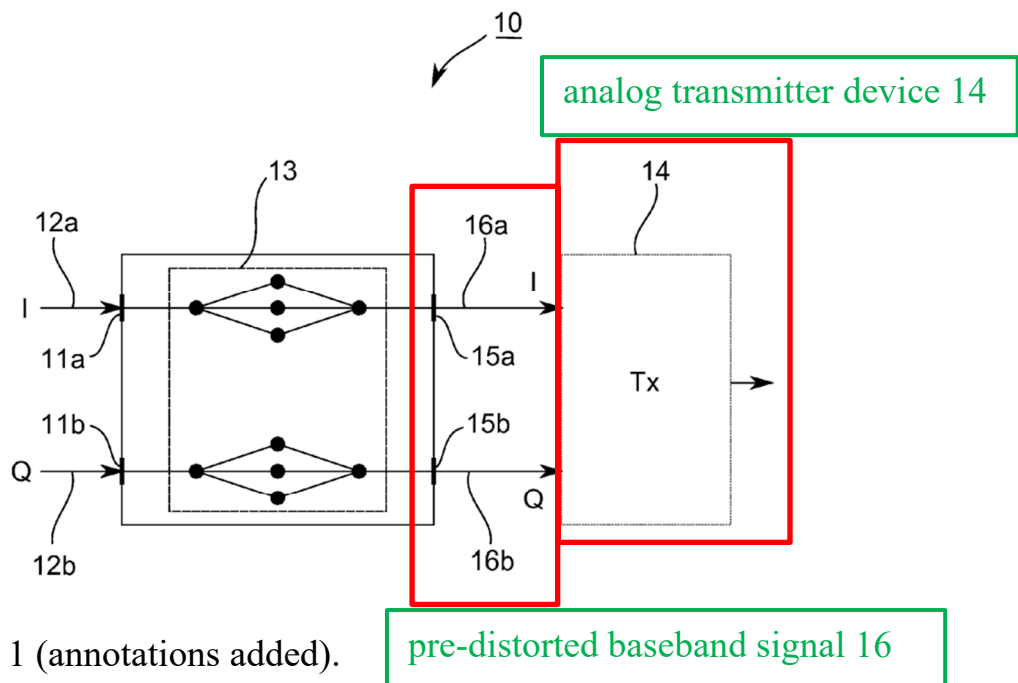
Jüschke renders obvious “*obtaining a transmit output radio signal by processing the pre-distorted radio signal through a transmitting radio signal processing stage of the base station*” [1D].

First, for *obtaining a transmit output radio signal by processing the pre-distorted radio signal*, Jüschke teaches an “analog transmitter device 14,” including a “transmitter transfer function,” that is “arranged downstream to the pre-distortion neural network processor 13.” Ex-1005, [0039]. Furthermore, the “in-phase and a quadrature component 16a, 16b of the pre-distorted digital baseband signal 16,” which are outputs of neural network processor 13, can “be coupled to an input of the at least one analog transmitter device 14.” Ex-1005, [0039]; *see id.*, Abstract. Hence, Jüschke’s pre-distorted digital signal 16 is fed into analog transmitter device 14, and after the pre-distorted signal is processed, the transmitter obtains and transmits an analog radio signal (*transmit output radio signal*).

POSITA would have understood that an analog transmitter, such as transmitter device 14, would have to first obtain and process a digital signal, such as pre-distorted digital baseband signal 16, before converting and transmitting the signal as an analog radio signal. Hence, POSITA would have understood that analog transmitter device 14, along with the analog radio signal it transmits corresponding to pre-distorted digital baseband signal 16, teaches *obtaining a transmit output radio*

signal by processing the pre-distorted radio signal. See Ex-1003, ¶159.

As shown in Figure 1 below, pre-distorted output signal components 16a, 16b are inputted into analog transmitter device 14. An output arrow originating from analog transmitter device 14 represents a *transmit output radio signal* that has been obtained and is being transmitted:



Ex-1005, Figure 1 (annotations added).

Second, for *processing the pre-distorted radio signal through a transmitting radio signal processing stage of the base station*, Jüschke's transmitter device 14, including "amplifier device" for increasing signal power for transmission, provides a *transmitting radio signal processing stage*. Ex-1005, [0020], [0050]. In addition, POSITA would have recognized that Jüschke's "analog" transmitter device 14 includes functionality and components for converting the received "pre-distorted

digital baseband signal 16” into an analog signal before ultimately obtaining and transmitting it as an output radio signal. *Id.*, [0039] (emphasis added). Because the analog transmitter receives a digital signal before converting and transmitting it as an analog radio signal, POSITA would have understood that the transmitter includes a digital-to-analog converter in addition to amplifier. Ex-1003, ¶161. Thus, POSITA would have understood that transmitter device 14 receives pre-distorted signal 16 and processes it, using an amplifier and a DAC module, to obtain an analog radio signal for transmission. *See id.*, ¶¶161-162.

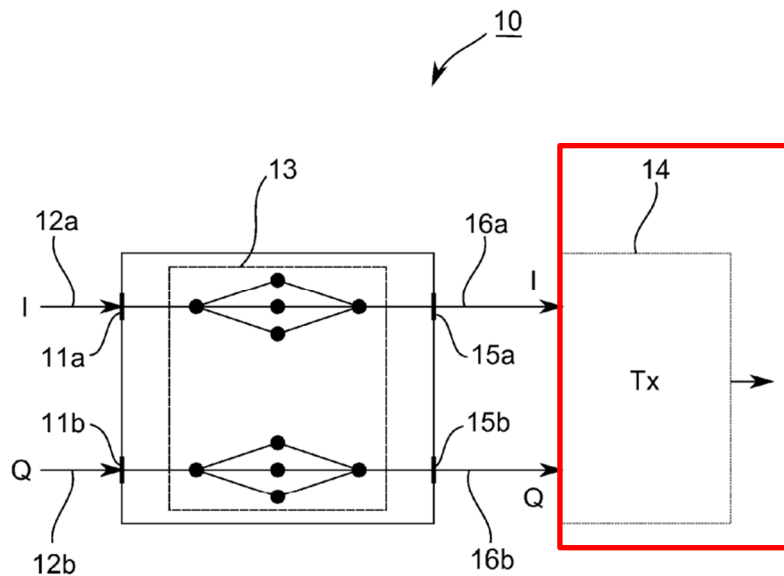
Third, as explained for limitation [1A], Jüschke’s flexible radio transmitter, including apparatus 10 and transmitter device 14, can be implemented within a base station. Ex-1005, [0006]. Therefore, POSITA would have understood that transmitter device 14, which implements the *transmitting radio signal processing stage*, can be implemented at a *base station*. *See* Ex-1003, ¶163.

Therefore, POSITA would have understood that Jüschke’s analog transmitter device 14 (and the radio signals it transmits) teaches *obtaining a transmit output radio signal*, that the transmitter’s amplifier and digital-to-analog conversion teach *processing the pre-distorted radio signal through a transmitting radio signal processing stage*, and that the transmitter can be implemented at a *base station*. Jüschke thereby renders [1D] obvious.

8. Jüschke renders [1E] obvious

Jüschke renders obvious “*transmitting the transmit output radio signal to the one or more radio receivers*” [1E].

Jüschke discloses an “analog transmitter device 14” along with a “transmitter transfer function.” Ex-1005, [0039]. As explained for limitation [1D], transmitter device 14 obtains an output radio signal (*transmit output radio signal*) for transmission before actually transmitting the output radio signal. As shown in Figure 1 below, an output arrow originating from analog transmitter device 14 represents a *transmit output radio signal* being transmitted:



Ex-1005, Figure 1 (annotation added).

Therefore, POSITA would have understood that a *transmit output radio signal* being transmitted would be received by a *radio receiver* or transceiver. See Ex-1003,

¶¶166-167.

Jüschke’s teachings are applicable to “mobile communication systems, like e.g. GSM, WCDMA and/or LTE” that implement numerous radio receivers and transceivers. Ex-1005, [0059]. Jüschke also states that its techniques “generally relate to wireless communications and, more specifically, to a concept for mitigating impairments of a transmit signal due to transmitter imperfections.” *Id.*, [0001]; *see id.*, [0004] (“Advanced receiver signal processing” and “thermal-noise at the receiver”). This would have confirmed POSITA’s understanding that Jüschke teaches output radio signals being transmitted by a transmitter device to and being received by a *radio receiver*. *See* Ex-1003, ¶168. Jüschke thereby renders [1E] obvious.

C. Claim 2 was obvious over Jüschke.

1. Jüschke renders [2A] obvious

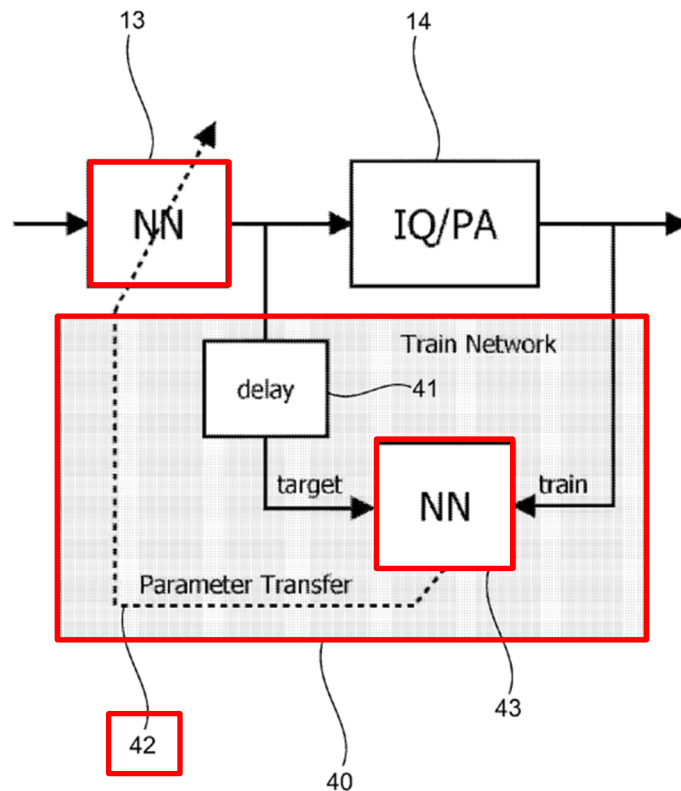
Jüschke renders obvious “[t]he method of claim 1, wherein the model parameters in the nonlinear pre-distortion machine learning model are loaded with values for a particular deployment scenario of the one or more deployment scenarios” [2A].

As explained for limitation [1B.2], Jüschke teaches weight and bias parameters (w_{kj} , b_k) for neural networks and addressing I/Q-imbalance and PA non-linearity in radio transmissions and cellular communication system deployment

scenarios such as, for example, in-line configurations.

Jüschke further teaches training the weight and bias parameters to fit unique deployment scenarios: “In order to fit the feed-forward pre-distortion neural network to a specific downstream hardware component, the apparatus may further comprise a training processor for training weight and/or bias parameters (w_{kj} , b_k) of neurons of the pre-distortion neural network processor.” Ex-1005, [0017]; *see id.*, [0048]. The training processor includes a “post distortion neural network processor having the same network structure as the pre-distortion neural network processor.” *Id.*, [0017].

After training, the weight and bias parameters are “transferred to the pre-distortion neural network processor 13...once (e.g. at the beginning of operation) or frequently in order to track time varying transfer function of the non-ideal transmitter devices 14.” *Id.*, [0058]; *see id.*, [0049]. As shown in Figure 4 below, dashed line 42 represents a parameter transfer from training processor 40 and post distortion neural network 43 to pre-distortion neural network processor 13:



Ex-1005, Figure 4 (annotations added).

POSITA therefore would have understood that Jüschke’s weight and bias parameters are trained and adjusted (*loaded with values*) to best fit the unique needs of a specific environment or transmission configuration (*particular deployment scenario*). See Ex-1003, ¶174. Jüschke thereby renders [2A] obvious.

2. Jüschke renders [2B] obvious

Jüschke renders obvious “*wherein generating the pre-distorted radio signal comprises generating the pre-distorted radio signal for the particular deployment scenario*” [2B].

As explained for limitation [2A], Jüschke teaches weight and bias parameters (w_{kj} , b_k) that are trained and adjusted to best fit particular deployment scenarios such as in-line configurations. After training, these parameters are transferred to the pre-distortion neural network processor.

Given that Jüschke's parameters are trained for specific environments and transmissions and then transferred to a pre-distortion neural network processor, the outputs of said pre-distortion neural network processor are thus also adjusted for said environments and transmissions—including output pre-distorted signals. See Ex-1005, [0039] (“[A] pre-distortion neural network processor 13 for pre-distorting the input digital baseband transmit signal 12 in order to obtain a pre-distorted digital baseband signal 16.”), Figure 1. Therefore, POSITA would have understood that Jüschke's pre-distorted baseband signal (*pre-distorted radio signal*) is generated for a specific environment or transmission (*particular deployment scenario*). See Ex-1003, ¶177. Jüschke thereby renders [2B] obvious.

D. Claim 3 was obvious over Jüschke.

1. Jüschke renders Claim 3 obvious

Jüschke renders obvious “[t]he method of claim 2, wherein the particular deployment scenario comprises *one of*: [3A] sending a radio transmission from the base station that includes the nonlinear pre-distortion machine learning model inside a tower mounted amplifier in an inline mode; [3B] sending a radio

transmission from the base station that includes the nonlinear pre-distortion machine learning model inside a tower mounted amplifier in a fault-tolerant subtractive mode; [3C] sending a radio transmission from the base station that includes the nonlinear pre-distortion machine learning model inside a remote radio unit in an inline mode; [3D] sending a radio transmission from the base station that includes the nonlinear pre-distortion machine learning model inside a remote radio unit in a fault-tolerant subtractive mode; or [3E] sending a radio transmission to a system that includes interference cancellation” [3] (emphasis added). As explained for limitation [2A], POSITA would have understood that Jüschke teaches weight and bias parameters (w_{kj} , b_k) that are trained and adjusted to best fit particular deployment scenarios such as in-line configurations.

Jüschke teaches the specific *deployment scenarios* of claim 3. Limitations [3A]-[3D] read on radio transmissions being sent from a base station with a NPDMLM being inside either a tower mounted amplifier [3A]/[3B] or a remote radio unit [3C]/[3D] in either an inline mode [3A]/[3C] or a fault-tolerant subtractive mode [3B]/[3D]. For ease of explanation, Petitioner will first address the *tower mounted amplifier* [3A]/[3B] and the *remote radio unit* [3C]/[3D] limitations.

As explained for limitation [1A], POSITA would have understood that Jüschke’s “flexible radio transmitter” can be implemented in a variety of

configurations for cellular communication systems including base stations, such as “macro-cell base stations” and “small-cell base stations.” Ex-1005, [0006]. POSITA further would have understood that a “macro-cell base station” is a cellular base station that sends and receives radio signals over a large coverage area and includes all necessary equipment such as amplifiers, including tower mounted amplifiers. *See* Ex-1003, ¶¶85, 122, 131, 181-182. Therefore, POSITA would have understood that Jüschke teaches its flexible radio transmitter, which includes a pre-distortion neural network (processor) (*nonlinear pre-distortion machine learning model*), can be implemented *inside tower mounted amplifiers* [3A]/[3B].

POSITA also would have understood that the “macro-cell base stations” that Jüschke discloses, Ex-1005, [0006], often consisted of a remote radio unit (“RRU”) (or remote radio head (“RRH”)). POSITA would have understood that an RRU is a transceiver component, typically collocated with a base station antenna, that directly connects to (and works together with) a baseband unit (“BBU”) or a radio control panel and processing unit to provide service to wireless devices (e.g., cellular phones) in a wireless network (e.g., cellular network). *See* Ex-1003, ¶¶86, 122, 131, 181, 183; Ex-1013, 42 (“baseband unit (BBU)-remote radio unit (RRU) structure of BS [base station] design to provide more flexible coverage and capacity”); Ex-1014. Therefore, POSITA also would have understood that Jüschke teaches its flexible

radio transmitter can be implemented in *remote radio units* [3C]/[3D] as well.

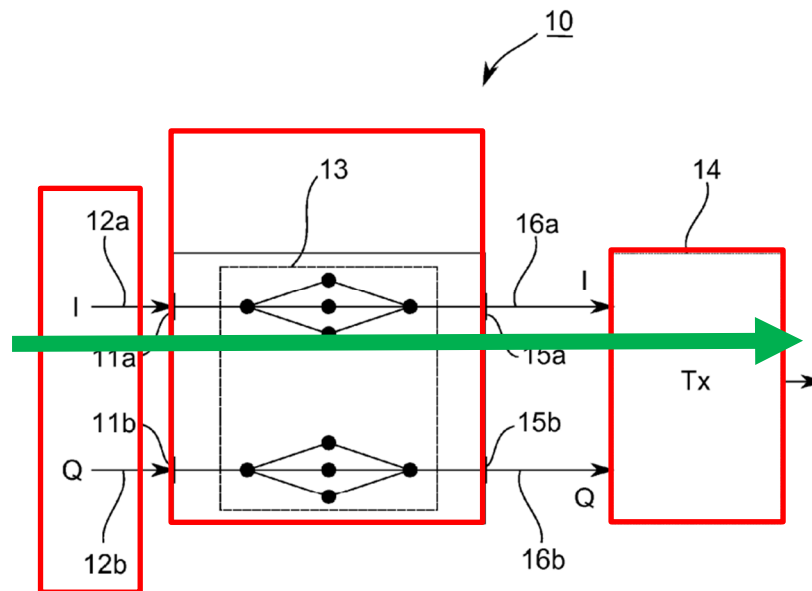
The remaining elements of limitations [3A]-[3E] are discussed below.

2. Jüschke renders [3A] and [3C] obvious

Jüschke renders obvious a *deployment scenario* for an *inline mode* [3A]/[3C].

See Section VI.B.4 ([1B.2]), *supra*.

Figure 1, shown below, illustrates apparatus 10 for a transmitter hardware configuration (*deployment scenario*) with a pre-distortion neural network 13 inline between an input transmit signal 12 and a power amplifier in a transmitter device 14 (*inline mode*). A green arrow demonstrates the components in tandem (i.e., in-line):



Ex-1005, Figure 1 (annotations added); *see id.*, [0050] (“analog amplifier device in the transmitter device 14”).

As explained for limitation [1B.2], POSITA would have understood that the

deployment configuration of apparatus 10 illustrated above was applicable to Jüschke's "flexible radio transmitter" and base station and remote radio unit hardware configuration *deployment scenarios*. Ex-1005, [0006]. POSITA would have further understood that *a radio transmission was sent* in such a *deployment scenario*, based on analog transmitter 14 transmitting analog radio signals. See Ex-1003, ¶187.

For comparison, the '704 patent discloses nearly identical inline-mode configurations. As shown below, "example configuration[s] for in-line mode," Ex-1001, 10:39, are illustrated in Figures 6A (tower-mounted amplifier [3A]) and 7A (remote radio unit [3C]) with a nonlinear pre-distortion system 103 placed in line between an input 601/701 and an amplifier 603/703. A green arrow demonstrates the components in tandem:

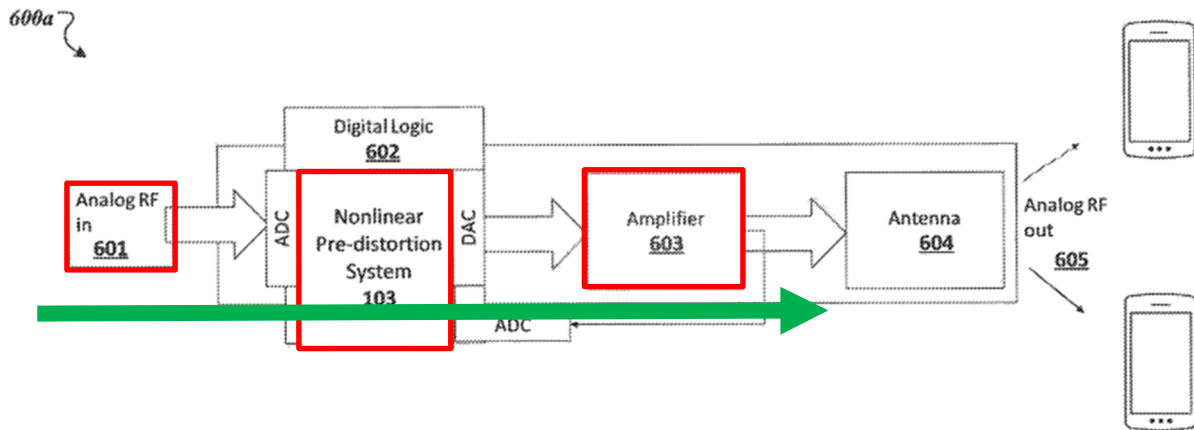


FIG. 6A

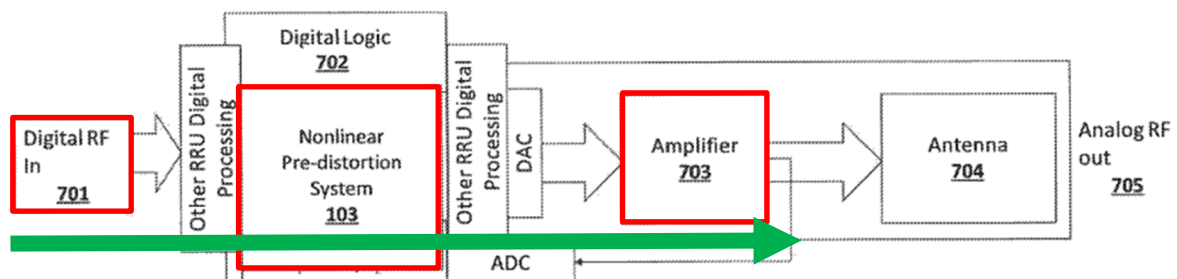


FIG. 7A

Ex-1001, Figures 6A, 7A (annotations added).

Therefore, POSITA would have understood that Jüschke teaches *inline mode deployment scenarios*, both *inside a tower-mounted amplifier* and *inside a remote radio unit* (Jüschke's flexible radio transmitter), via Jüschke's transmission

hardware configurations in which the input transmit signal 12, pre-distortion neural network processor 13, and transmitter device 14 are implemented in an inline configuration (*inline mode*). Jüschke thereby renders limitations [3A] and [3C] obvious. *See* Ex-1003, ¶189.

3. Jüschke renders [3B] and [3D] obvious

Jüschke also renders obvious a *deployment scenario* for a *fault-tolerant subtractive mode* [3B]/[3D]. The '704 patent generally describes “subtractive” predistortion as the practice of using a difference signal for predistortion. *See* Ex-1001, 7:51-53 (“[C]ompute a difference signal between the pre-distorted signal and the input signal (e.g. for subtractive pre-distortion).”).

For “the mitigation of IQ imbalance,” Jüschke teaches using a “piecewise linear transfer or activation function” or a “purely linear...activation function.” Ex-1005, [0016]; *see id.*, [0054]. POSITA would have understood that Jüschke teaches efficient signal correction by letting the nonlinear predistortion model compensate only a portion of an input. Ex-1003, ¶¶192-193. As Figures 5A and 5B illustrate, a nonlinear neural network with hidden layers (22, 22a, 22b) receives an original input signal 12 connected to direct path 51 from input layer 21 and forms a total compensated signal 16 at output layer 23. *See* Ex. 1005, [0054] (“bypassing the intermediate hidden layers 22, 22a or 22b”). POSITA would have understood that the ML networks in Figures 5A and 5B compensate for the difference between

output 16 and input 12 that bypassed hidden layers 22 (see Ex-1003, ¶¶193-195):

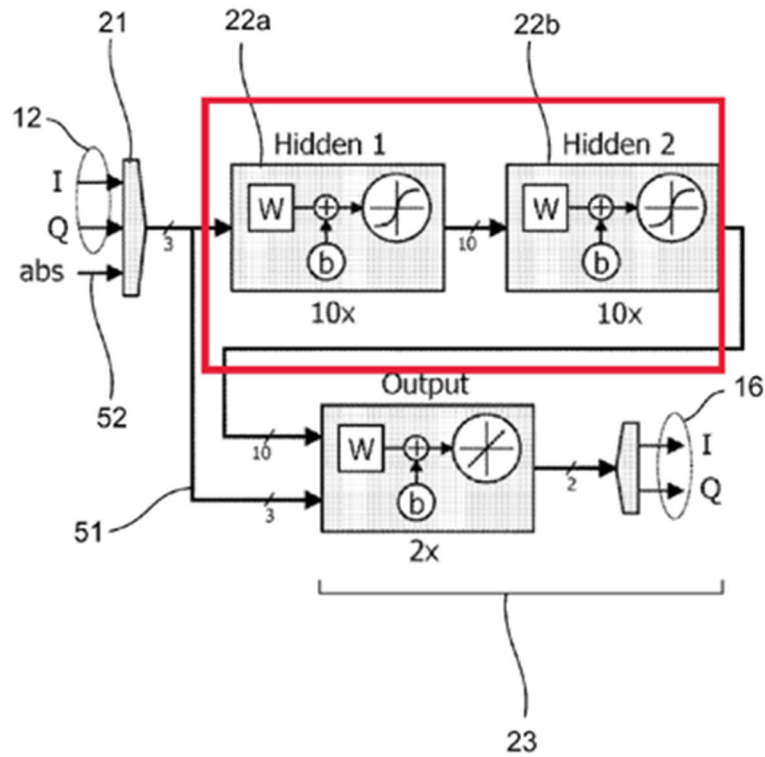


Fig. 5a

Ex-1005, Figure 5A (annotation added).

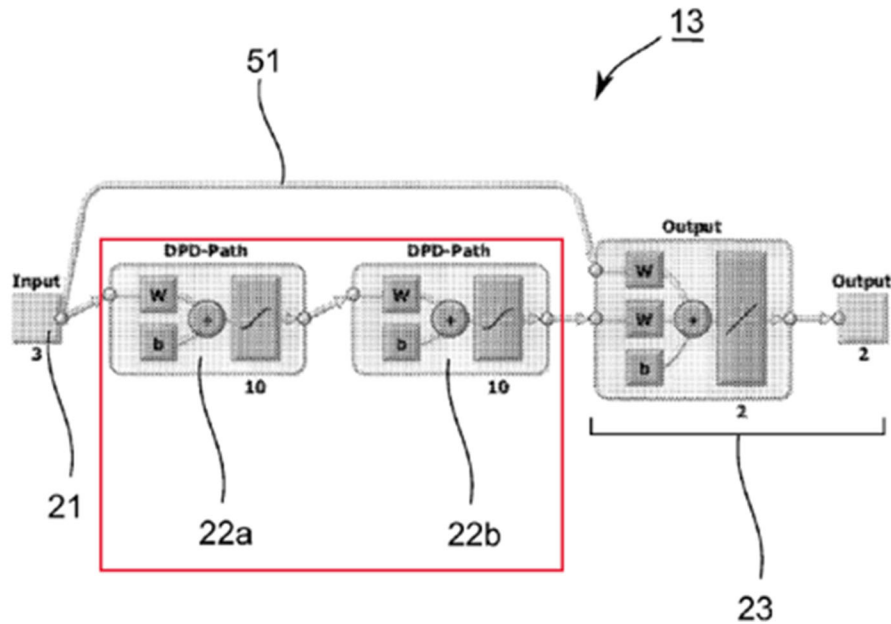


Fig. 5b

Ex-1005, Figure 5B (annotation added).

POSITA would have understood that Figures 5A and 5B, and their “nonlinear part (for PA linearization) mainly handled by the hidden layers, which relaxes the requirements for the hidden layer size,” teach a subtractive mode because the nonlinear part of the model and the signal generated by bypass signal 51 (linear part) are added together (output layer 23) to form the final predistortion signal. Ex 1005, [0054]; *see id.*, [0024]. POSITA would have understood that the nonlinear part is only responsible for the *difference* between the final predistortion signal and the bypass, hence a subtractive mode. Ex-1003, ¶196. POSITA also would have understood that Jüschke’s topology with hidden layers that adjust size requirements

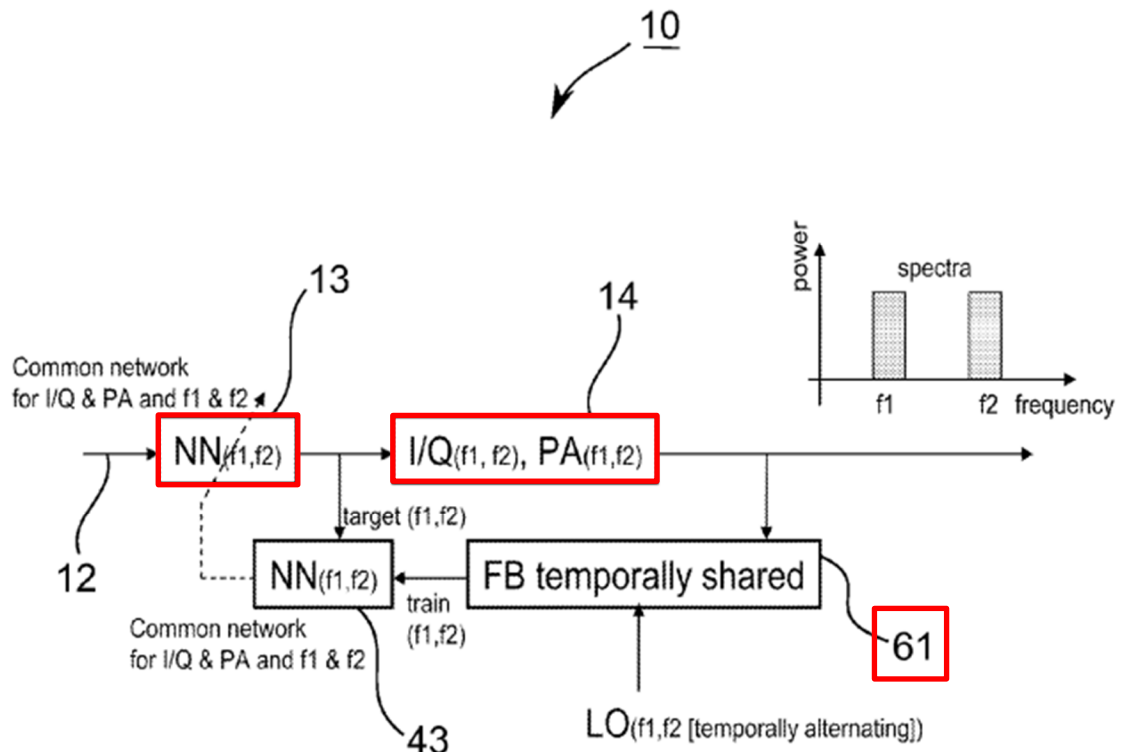
and a bypassing path that simplifies the predistortion model teaches improvements to system resilience and fault tolerance. *Id.*, ¶197. Therefore, POSITA would have understood that Jüschke teaches a *deployment scenario for a fault-tolerant subtractive mode* in view of Figures 5A and 5B and paragraphs [0024] and [0054].

Therefore, POSITA would have understood that Jüschke teaches network hidden layers with adjustable requirements (*subtractive mode*) and improved system resilience (*fault-tolerant*), both *inside a tower-mounted amplifier* and *inside a remote radio unit* (Jüschke’s flexible radio transmitter), Jüschke thereby renders limitations [3B] and [3D] obvious.

4. Jüschke renders [3E] obvious

In addition, Jüschke renders obvious *sending a radio transmission to a system that includes interference cancellation* [3E].

Jüschke teaches “a feedback path 61, preferably for mitigating TX noise caused by the transmitter device 14...being adapted for temporally alternately generating feedback on the different frequency bands.” Ex-1005, [0060]. Figure 6 below illustrates feedback path 61 along with pre-distortion neural network 13 and transmitter device 14:



Ex-1005, Figure 6 (annotations added). *See id.*, [0065], [0011].

POSITA would have understood that feedback loop 61 can provide information about interference that may be sensed at the transmit antenna receiving the PA output, thereby allowing the pre-distortion network to compensate for the same. Because the pre-distortion network would be able to effectively cancel the interference, POSITA would have understood that Jüschke's feedback path 61 teaches a system with *interference cancellation*. *See* Ex-1003, ¶201.

Further, given that Jüschke's teachings "address[] I/Q-imbalance and PA non-linearity in parallel for different frequency bands," Ex-1005, [0059], in addition to

“TX noise,” *id.*, [0004], POSITA would have understood that Jüschke’s predistortion techniques include interference cancellation and the feedback path 61 can be used by receivers and transceivers that receive the *radio transmissions* of limitation [3E]. *See id.*, Abstract (“mitigating impairments due to transmit noise”); Ex-1003, ¶202. *See also* Ex-1001, 2:66-67 (“Correcting distortion of radio signals may include correcting interference of radio signals.”).

Therefore, Jüschke renders the *deployment scenario* of limitation [3E] obvious, in addition to limitations [3A]-[3D] explained *supra*.

E. Claim 7 was obvious over Jüschke.

1. Jüschke renders Claim 7 obvious

Jüschke renders obvious “[t]he method of claim 1, wherein the values associated with the model parameters are optimized to minimize the distortion or interference of radio signals associated with the base station in the cellular communications system” [7].

As previously explained, Jüschke teaches techniques for mitigating transmit signal impairments due to imperfections in cellular communication systems and base stations. *See* Section VI.B.1 ([1P]); Ex-1005, [0001] (“mitigating impairments of a transmit signal due to transmitter imperfections”). Jüschke also teaches the use of weight and bias parameters for addressing I/Q-imbalance and PA non-idealities or non-linearity in cellular communication systems including base station

configurations. *See* Section VI.B.4 ([1B.2]), *supra*. Furthermore, the weight and bias parameters are trained and updated for deployment scenarios including transmissions from base stations in cellular communication systems and are specifically fitted for certain hardware. *See* Ex-1005, [0048] (“training weight (w_{kj}) and/or bias parameters (b_k),”), [0017] (“[F]it the feed-forward pre-distortion neural network to a specific downstream hardware component.”).

Given that POSITA would have understood that Jüschke teaches that its weight and bias parameters (*values associated with the model parameters*) are constantly updated via training to be fitted for specific hardware and used for addressing I/Q-imbalance and PA non-linearity in cellular communication systems (*optimized to minimize the distortion or interference of radio signals*) including deployments from a base station, Jüschke renders claim 7 obvious. *See* Ex-1003, ¶206.

F. Claim 8 was obvious over Jüschke.

1. Jüschke renders Claim 8 obvious

Jüschke renders obvious “[t]he method of claim 1, wherein the nonlinear pre-distortion machine learning model is one of a deep neural network, a convolutional neural network, a recurrent neural network, or a combination of different types of neural networks” [8].

As explained for limitation [1B.1], Jüschke teaches neural networks, pre-

distortion neural network 13, and feed-forward neural network 20. *See also* Ex-1005, [0041] (“Although not limited thereto, the pre-distortion neural network processor 13 may comprise a feed-forward neural network having a plurality of layers...”). POSITA would have understood that artificial neural networks include deep neural networks, convolutional neural networks, and recurrent neural networks. Therefore, Jüschke renders claim 8 obvious. *See* Ex-1003, ¶208.

G. Claim 9 was obvious over Jüschke.

1. Jüschke renders [9P] obvious

For limitation [9P], the analysis for [1P] applies. *See* Section VI.B.1, *supra*. POSITA would have understood that Jüschke’s teachings apply to apparatus claims. *See* Ex-1003, ¶209; Ex-1005, [Abstract] (“Provided is apparatus (10) for mitigating transmitter impairments of a transmit signal.”).

2. Jüschke renders [9A] obvious

Jüschke renders obvious “*one or more processors associated with a base station in the cellular communications system*” [9A]. As explained for limitation [1A], Jüschke’s teachings can be implemented within a base station of a cellular communications system including processors. *See* Ex-1005, Abstract (“pre-distortion neural network processor (13) for pre-distorting the input digital baseband transmit signal (12; 12a; 12b)”). POSITA would have understood that Jüschke’s “flexible radio transmitter,” including network processor 13, can be implemented in

“base stations.” *Id.*, [0006], Figure 1; Ex-1003, ¶212. Therefore, Jüschke renders [9A] obvious.

3. Jüschke renders [9B] obvious

Jüschke renders obvious “*one or more machine-readable media storing instructions that, when executed, are configured to cause the one or more processors to perform operations comprising*” [9B]. As explained for limitation [9A], processors are implemented in Jüschke’s teachings, including network processor 13. “[E]mbodiments are also intended to cover program storage devices, e.g., digital data storage media, which are machine or computer readable and encode machine-executable or computer-executable programs of instructions, wherein said instructions perform some or all of the steps of said above-described methods.” Ex-1005, [0071]. And “[f]unctions of various elements...may be provided through the use of dedicated hardware, as e.g. a processor, as well as hardware capable of executing software in association with appropriate software.” *Id.*, [0075]. Therefore, Jüschke renders [9B] obvious. Ex-1003, ¶¶215-216.

4. Jüschke renders [9C]-[9G] obvious

Claim limitations [9C]-[9G] are nearly identical to [1A]-[1E] and rendered obvious by Jüschke for the reasons discussed in Sections VI.B.2-8, *supra*.

H. Claim 10 was obvious over Jüschke.

1. Jüschke renders [10A]-[10B] obvious

Claim limitations [10A]-[10B] are nearly identical to [2A]-[2B] and rendered obvious by Jüschke for the reasons discussed in Sections VI.C.1-2, *supra*.

I. Claim 11 was obvious over Jüschke.

1. Jüschke renders [11A]-[11E] obvious

Claim limitations [11A]-[11E] are nearly identical to [3A]-[3E] and rendered obvious by Jüschke for the reasons discussed in Sections VI.D.1-4, *supra*.

J. Claim 15 was obvious over Jüschke.

1. Jüschke renders Claim 15 obvious

Claim 15 is nearly identical to claim 7 and rendered obvious by Jüschke for the reasons discussed in Section VI.E.1, *supra*.

K. Claim 16 was obvious over Jüschke.

1. Jüschke renders Claim 16 obvious

Claim 16 is nearly identical to claim 8 and rendered obvious by Jüschke for the reasons discussed in Section VI.F.1, *supra*.

L. Claim 17 was obvious over Jüschke.

1. Jüschke renders [17P] obvious

Claim limitation [17P] is nearly identical to [1P], [9A]-[9B] and rendered obvious by Jüschke for the reasons discussed in Sections VI.B.1, VI.G.2-3 *supra*.

2. Jüschke renders [17A]-[17E] obvious

Claim limitations [17A]-[17E] are nearly identical to [1A]-[1E] and rendered obvious by Jüschke for the reasons discussed in Sections VI.B.2-8, *supra*.

M. Claim 18 was obvious over Jüschke.

1. Jüschke renders [18A]-[18B] obvious

Claim limitations [18A]-[18B] are nearly identical to [2A]-[2B] and rendered obvious by Jüschke for the reasons discussed in Sections VI.C.1-2, *supra*.

N. Claim 19 was obvious over Jüschke.

1. Jüschke renders [19A]-[19E] obvious

Claim limitations [19A]-[19E] are nearly identical to [3A]-[3E] and rendered obvious by Jüschke for the reasons discussed in Sections VI.D.1-4, *supra*.

O. Claim 23 was obvious over Jüschke.

1. Jüschke renders Claim 23 obvious

Claim 23 is nearly identical to claim 7 and rendered obvious by Jüschke for the reasons discussed in Section VI.E.1, *supra*.

P. Claim 24 was obvious over Jüschke.

1. Jüschke renders Claim 24 obvious

Claim 24 is nearly identical to claim 8 and rendered obvious by Jüschke for the reasons discussed in Section VI.F.1, *supra*.

VII. GROUND 2: Claims 1-24 were obvious over Jüschke in view of Holt.

As shown for Ground 1 in Section VI, *supra*, Jüschke renders obvious claims/limitations [1P]-[1E], [2A]-[2B], [3A]-[3E], 7, 8 [9P]-[9G], [10A]-[10B], [11A]-[11E], 15, 16, [17P]-[17E], [18A]-[18B], [19A]-[19E], 23, and 24.

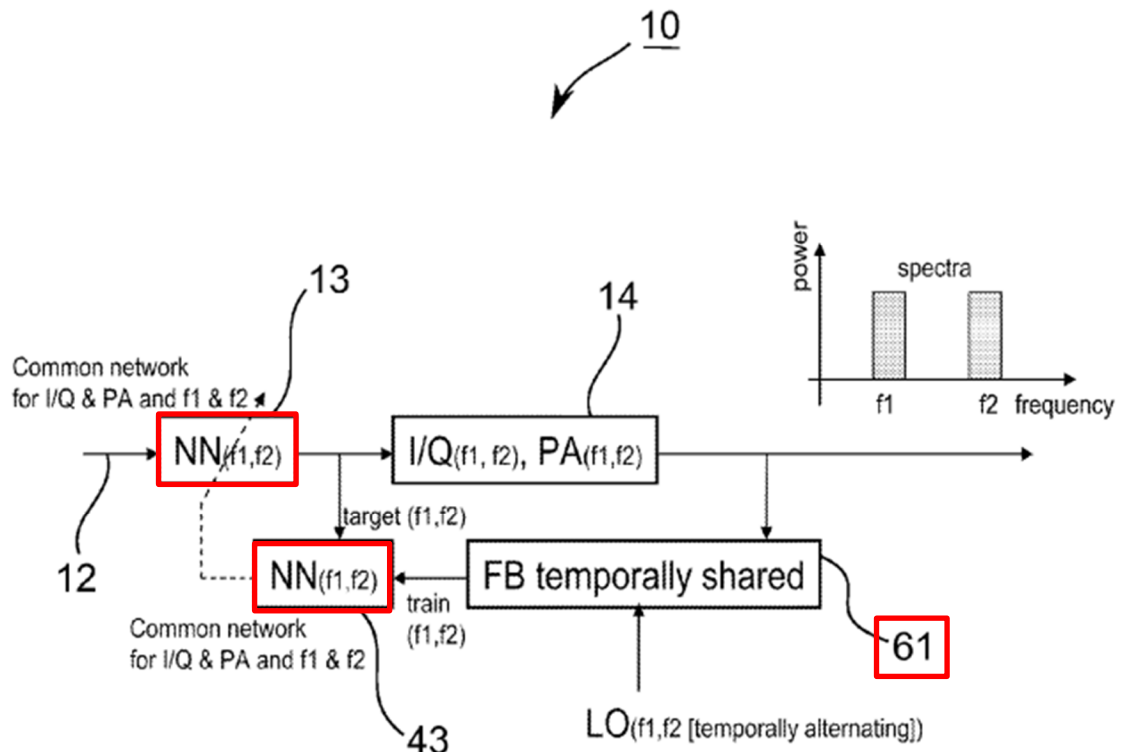
To the extent Jüschke does not render obvious [1E] and similar limitations [9G] and [17E], they would have been rendered obvious by Jüschke in view of Holt.

Jüschke in view of Holt also renders obvious claims 4-6 and similar claims 14-16 and 20-22.

A. Modifying Jüschke's Training with Holt's Loss Function

Combining Jüschke with Holt implements Holt's loss function, used for training neural networks in radio communications, into Jüschke's sampling techniques and feedback loop, which are also used for training radio neural networks. This would have been a simple substitution for POSITA.

As previously discussed, Jüschke trains its models using “[s]ampled [transmitter] output values” and “ideal [input] samples,” Ex-1005, [0048], and implements a “feedback signal for enabling adaptive pre-distortion.” *Id.*, [0060]. Figure 6 of Jüschke demonstrates training and adaptive pre-distortion, with pre-distortion neural network 13, post-distortion neural network 43, and feedback signal 61 marked with red boxes:



Ex-1005, Figure 6 (annotations added).

Jüschke explicitly discloses “[g]radient-based training algorithms, like back-propagation,” Ex-1005, [0018], and “Hessian-based algorithms” such as Levenberg-Marquardt, *id.*, [0049], which has also been used in backpropagation training of neural networks, Ex-1011, 397. Therefore, POSITA would have understood that Jüschke teaches backpropagation for training. Ex-1003, ¶222.

Holt’s models similarly “can be trained on a loss function that describes a difference between input data input into the encoder model and output data received from the decoder model” using backpropagation. Ex-1006, Abstract. Figure 3 below exemplifies a first loss function 216 from an encoder model input 204 and decoder

model output 214 with channel model 208 in between. Holt teaches that channel model 208 can “emulate a communication channel” (*id.*, 4:5-6):

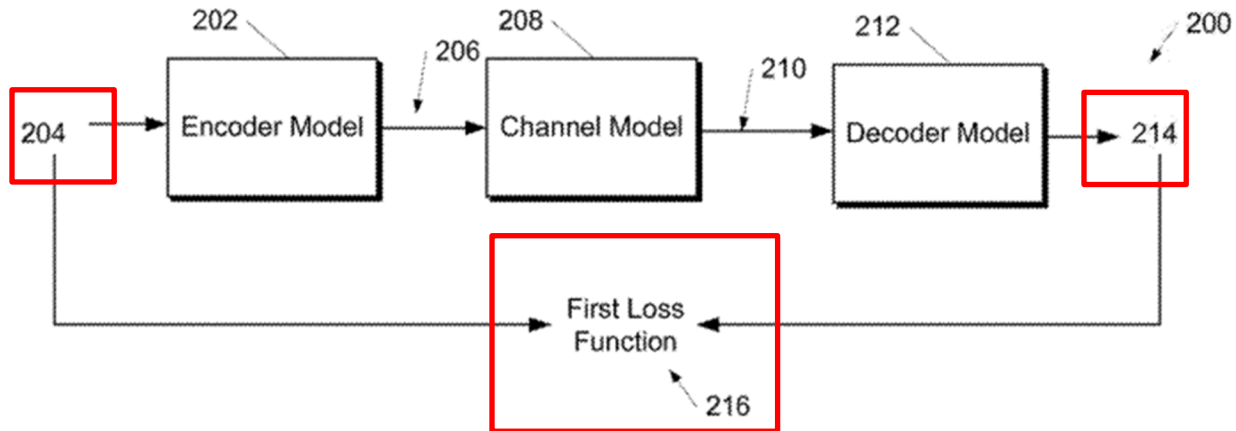


Figure 3

Ex-1006, Figure 3 (annotations added).

In addition, POSITA would have understood that Holt teaches using an “actual communication channel” instead of a channel model. Ex-1006, 18:1-2. Therefore, POSITA would have understood that Holt’s encoder and decoder models could be implemented with an actual communication channel for loss functions. *See* Ex-1003, ¶¶223-224. Figure 6 illustrates fourth loss function 318 from an encoder model input 316 and decoder model output 312 with an “Actual [Communication] Channel” 302 in between:

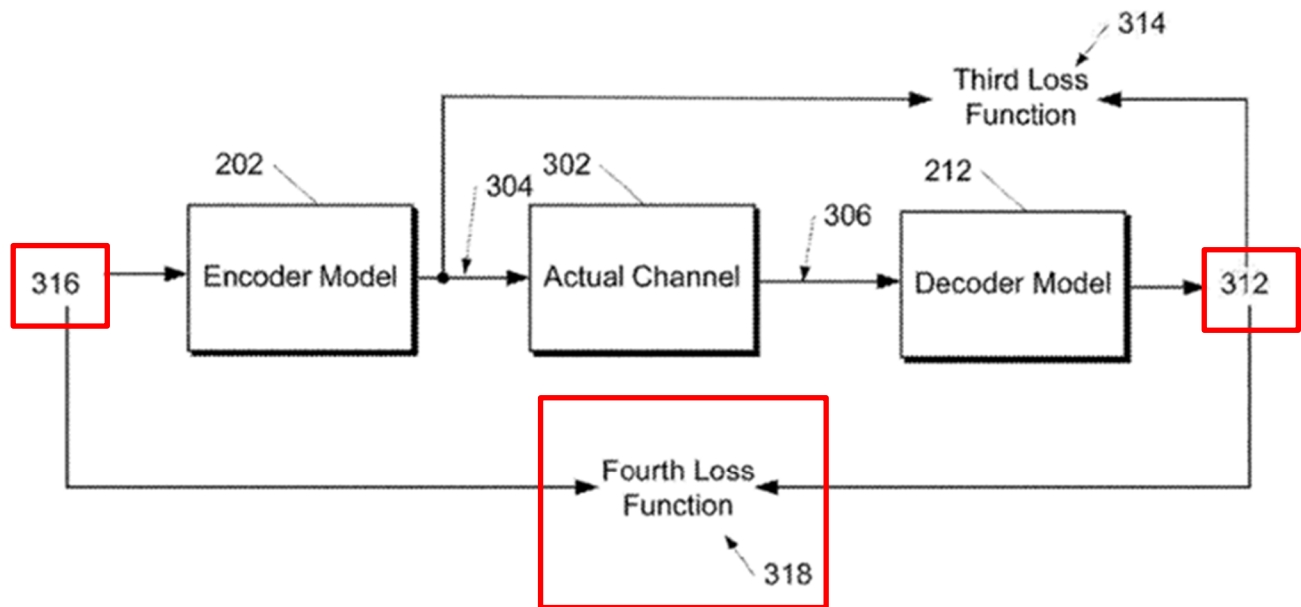


Figure 6

Ex-1006, Figure 6 (annotations added).

Therefore, POSITA would have easily been able to combine Jüschke with Holt’s loss function—which also uses input/output samples, feedback loops, and backpropagation—for training, because doing so would have been an exchange of known training techniques. Ex-1003, ¶225. POSITA also would have recognized that Jüschke and Holt implement their teachings within similar (if not identical) devices. *Compare* Ex-1005, [0006] (“macro-cell base stations” and “small-cell base stations”) to Ex-1006, 9:34-38 (“any type of computing device, including...a network computing device such as a base station”).

B. Motivation for Modifying Jüschke’s Training

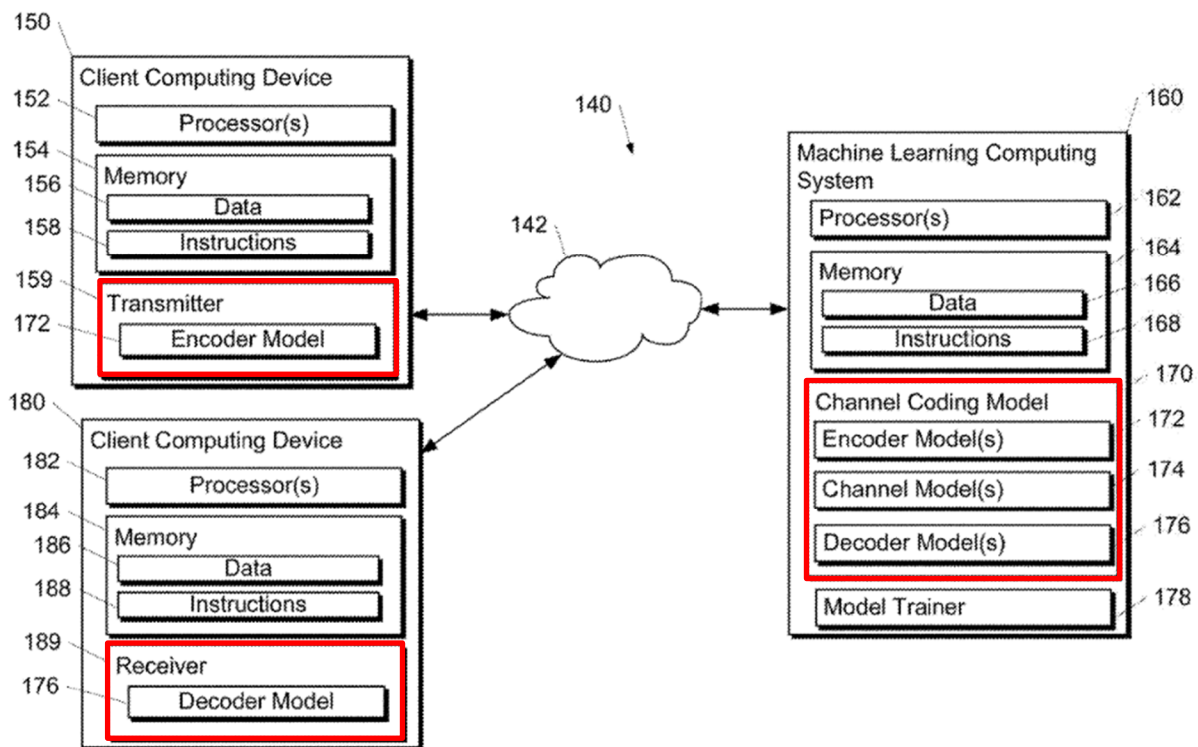
POSITA would have been motivated to modify Jüschke’s training techniques

using Holt's loss function because incorporating Holt's loss function would have improved the accuracy of Jüschke's training model.

Jüschke and Holt both use ML and neural networks in wireless communications systems to improve signal processing performance. *Compare* Ex-1005, Abstract ("mitigating impairments due to transmit noise") to Ex-1006, 5:45-48 ("Providing a channel model between an encoder model and a decoder model provides an ability to model the addition of noise to an output of the encoder model."). Because Jüschke and Holt are in similar (if not identical) fields and solve similar problems, POSITA would have been familiar with both and would have recognized that their elements could be interchanged and combined for improved model training. Ex-1003, ¶227; *see KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 401 (2007).

For example, POSITA would have been motivated to implement Holt's sampling and feedback training techniques into Jüschke because doing so would have improved similar devices (ML models in radio transceivers). Holt's loss functions, based on its encoder and decoder models communicating over a communication channel, include derivation from input signals and output signals actually recovered by a receiver. *See* Ex-1006, Figure 6 (fourth loss function 318 from an encoder model input 316 and decoder model output 312 with an actual

communication 302). POSITA would have understood that Holt uses actual receiver data as output data for training, and that this would have improved training by sampling signals actually received and recovered by a receiver. *See* Ex-1003, ¶1228. POSITA would have understood that such samples provide more accurate training data that accounts for the effects of transmission and reception. *See id.* Figure 2 below shows Holt’s ML models implemented in devices communicating across a network:



Ex-1006, Figure 2 (annotations added).

Therefore, POSITA would have understood implementing Holt into Jüschke to be “mere application of a known [training] technique” to Jüschke’s self-contained

feedback system, a “similar device” “ready for the [training data] improvement” taught by Holt. *See* Ex-1003, ¶229; *KSR*, 550 U.S. at 401.

Combining Jüschke’s and Holt’s sampling and feedback training techniques also would have been a “combination of familiar elements according to known methods” and a “simple substitution of one known element for another.” *KSR*, 550 U.S. at 401. Both references teach training models using input/output sampling. Thus, exchanging the training techniques of one for the other, including the source of input/output data, would have been both a combination of known methods and a simple substitution that “does no more than yield predictable results.” *Id.*; *see* Ex-1003, ¶230.

POSITA would have had a reasonable expectation of success in combining Holt’s training techniques with Jüschke for more accurate output data and better ML training. POSITA would have understood that Holt’s received and recovered output signal samples would have been more accurate replacement than Jüschke’s output signal sampling. *See, e.g.*, Ex-1005, Figure 4 (analog transmitter output in a feedback loop for “train[ing]” an “NN”). POSITA would have known that Holt’s “encoder model can be implemented by a transmitter device...while the decoder model can be implemented by a receiver device.” Ex-1006, 4:33-37. POSITA also would have known that Jüschke’s teachings are embodied in a “flexible radio transmitter that

can support as many standards and signals as possible.” Ex-1005, [0006]; *see id.* [0072] (“neural network may be part of the transmitter loop”). Therefore, POSITA would have had a reasonable expectation of success for improving output data sampling and model training in such systems by adding Holt’s receiver data into Jüschke’s transmission output feedback loop. *See* Ex-1003, ¶231.

In short, POSITA would have been motivated to combine Jüschke with Holt for multiple reasons, including similar (if not identical) fields, solving similar problems, known improvements for similar devices, familiar elements and known methods, and simple substitution. *See KSR*, 550 U.S. at 416-417.

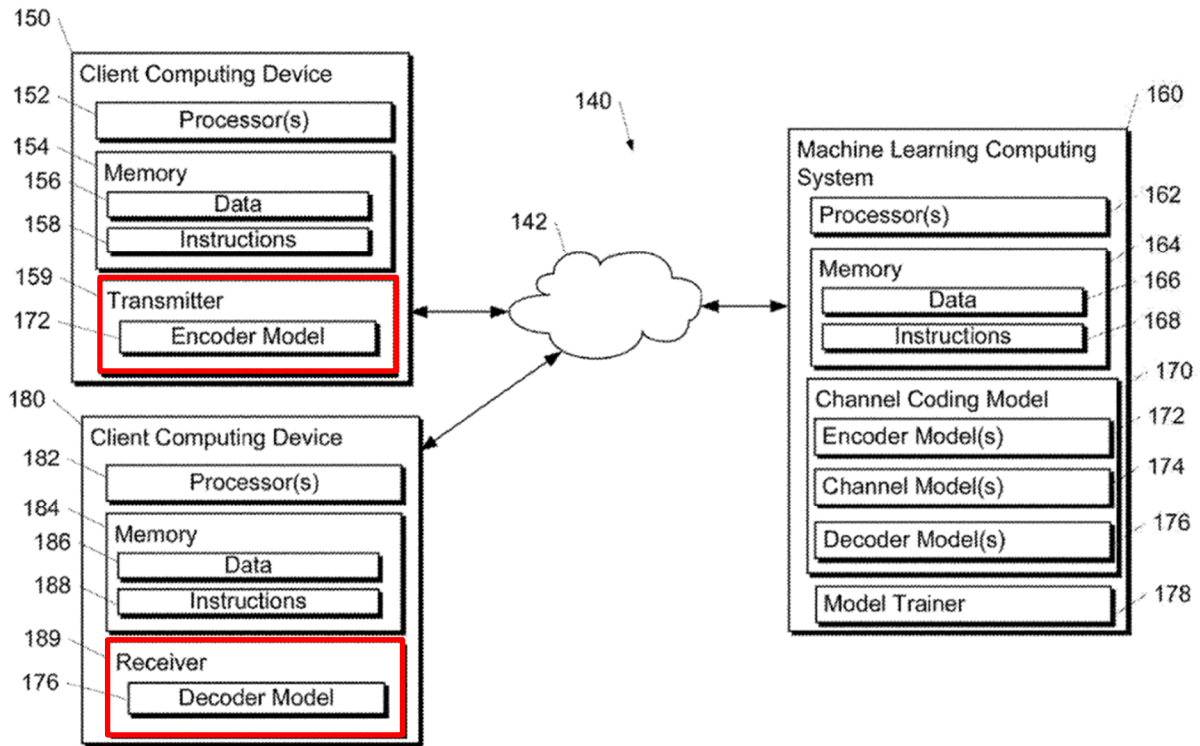
C. Claims 1, 9, and 17 were obvious over Jüschke in view of Holt.

1. Jüschke in view of Holt renders [1E] obvious

As shown for Ground 1, “*transmitting the transmit output radio signal to the one or more radio receivers*” [1E] was rendered obvious by Jüschke. *See* Section VI.B.8, *supra*. Jüschke in view of Holt also renders [1E] obvious.

Holt teaches communication data corresponding to a *transmit output radio signal* being *transmitted* by a transmitter (with an encoder model) and received by a *receiver* (with a decoder model). Indeed, “transmitter 159...encode[s] communication data prior to transmission of the communication data via a communication channel,” and “receiver 189...decode[s] communication data received from communication channel.” Ex-1006, 14:49-53, 15:16-19.

Figure 2 below illustrates transmitter 159 and receiver 189 and associated ML models in communication over network 142:



Ex-1006, Figure 2 (annotations added).

Figure 13 illustrates a method for transmitting and receiving data over a communication channel:

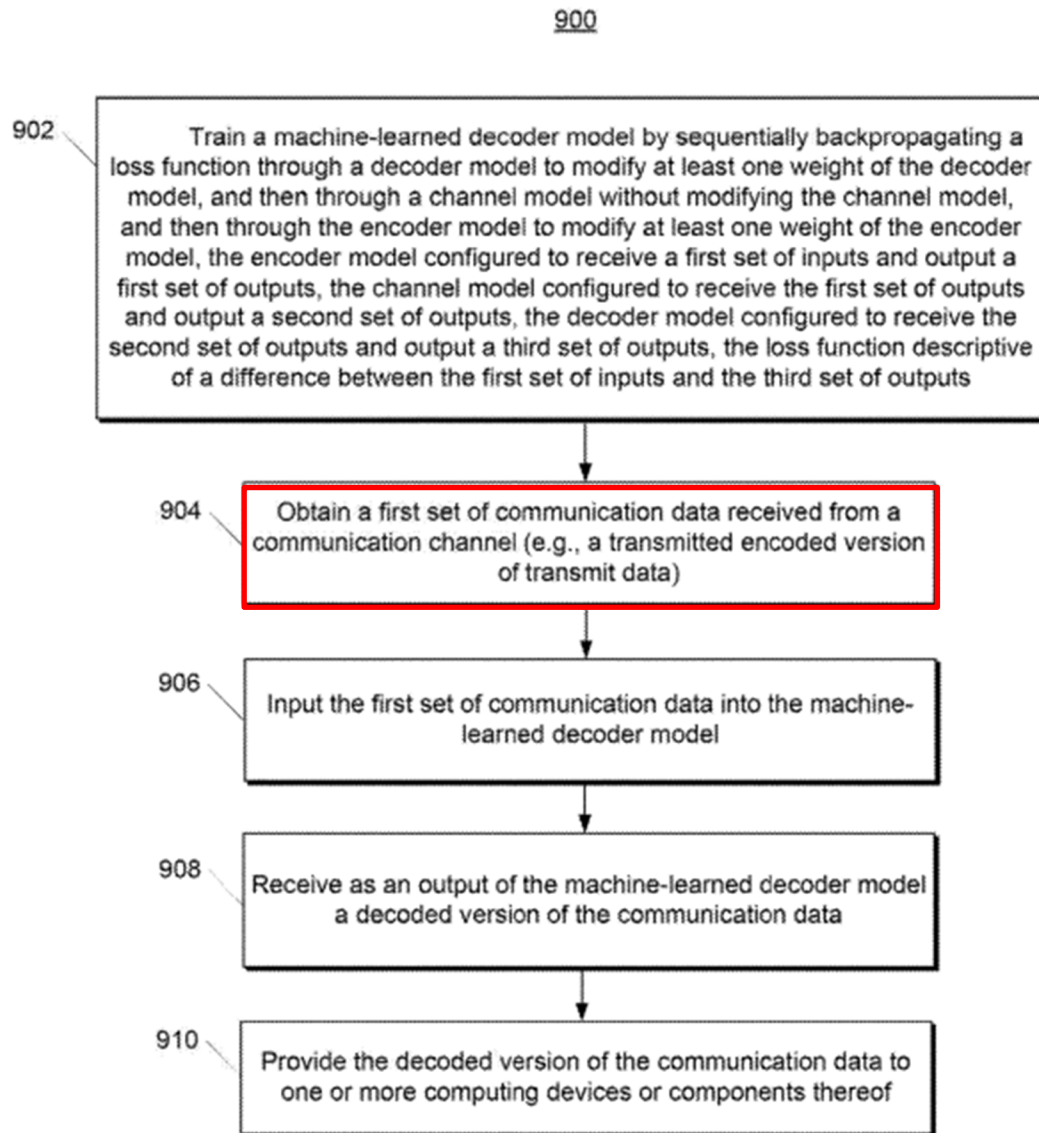


Figure 13

Ex-1006, Figure 13 (annotation added); *see id.*, 22:12-14.

POSITA would have understood that Holt's transmitter/encoder model teaches *transmitting the transmit output radio signal*, and that Holt's transmission of said data to a receiver/decoder model teaches *one or more radio receivers*. *See*

Ex-1003, ¶¶236-237. And given that Jüschke and Holt both implement ML techniques for improving telecommunications, POSITA would have been motivated to implement Holt's techniques into Jüschke because doing so would have been a simple substitution of known elements and would have provided known improvements. *See* Sections VII.A-B, *supra*.

Therefore, Jüschke in view of Holt renders [1E] obvious.

2. Jüschke in view of Holt renders similar limitations [9G] and [17E] obvious

For the reasons stated above, Jüschke in view of Holt also renders obvious [9G] and [17E]. *See* Sections VII.A, VI.G.4, VI.L.2, *supra*.

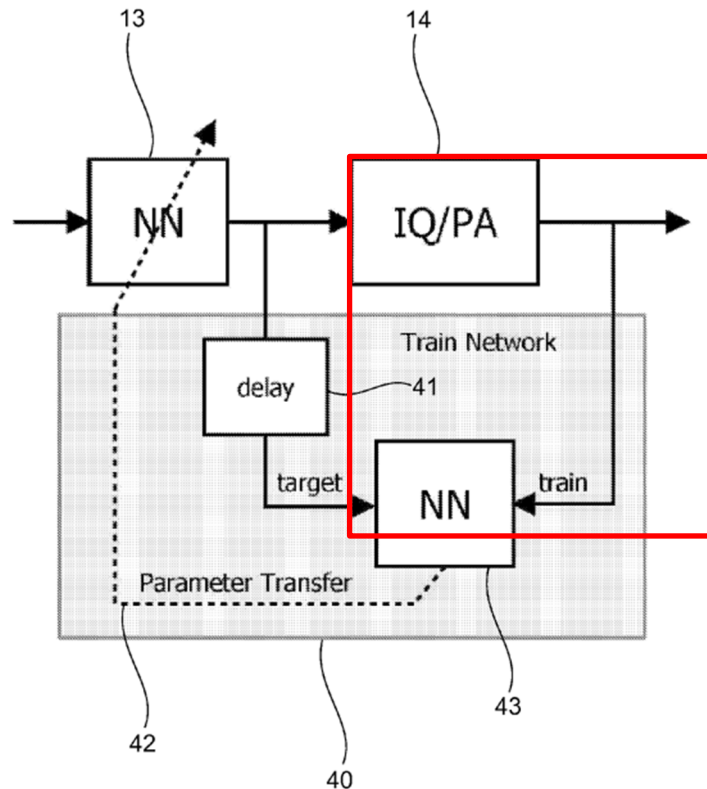
D. Claims 4, 12, and 20 were obvious over Jüschke in view of Holt.

1. Jüschke in view of Holt renders [4A] obvious

Jüschke in view of Holt renders obvious “[t]he method of claim 2, further comprising: obtaining a received radio signal that corresponds to a representation of the transmit output radio signal received at the one or more radio receivers” [4A].

As explained for limitation [1E], Jüschke teaches transmitting a *transmit output radio signal* to a *radio receiver*. Jüschke further teaches a feedback loop of the signal to be transmitted. Ex-1005, [0072] (“[T]he neural network may be part of the transmitter loop including transmit and feedback path.”). Figure 4 shows an output arrow originating from analog transmitter device 14, representing a *transmit*

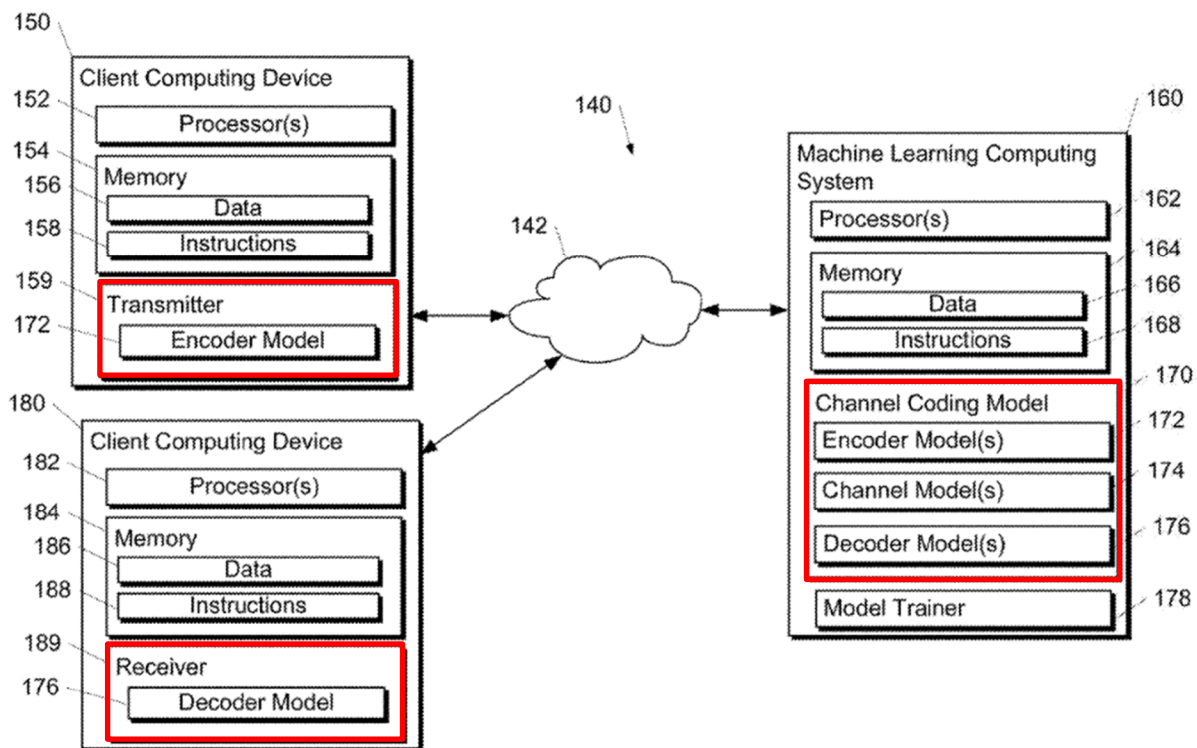
output radio signal, and a feedback loop corresponding to the *transmit output radio signal* being received by a *radio receiver*:



Ex-1005, Figure 4 (annotation added); *see also id.*, Figures 1, 6.

POSITA would have understood that a *transmit output radio signal* being from transmitter device 14 would be the same signal received by a *radio receiver*. *See* Section VI.B.8, *supra*. Therefore, POSITA would have understood that feedback from a *transmit output radio signal* transmitted to a *radio receiver* would correspond to a representation of the *transmit output radio signal* received at a *radio receiver*. *See* Ex-1003, ¶243.

In addition, as explained for limitation [1E], Holt teaches communication data (*transmit output radio signal*) received by a receiver. See Section VII.C.1, *supra*. This data is further processed for machine learning and training. For example, as shown in Figure 2, a client computing device 150 with transmitter 159 transmits a signal over network 142. This signal can be received and processed at computing system 160 with ML models 172, 174, and 176 and model trainer 178. The signal can also be received and processed at client computing device 180 with receiver 189 and decoder model 176:



Ex-1006, Figure 2 (annotations added); *see also id.*, 4:33-37 (“[T]he encoder model can be implemented by a transmitter device...while the decoder model can be

implemented by a receiver device.”), 7:30-34 (“[A] trained decoder model can be implemented by a receiver device or component to decode communication data (e.g., a transmitted encoded version of transmit data) *after receipt of the communication data from a communication channel*” (emphasis added).).

POSITA would have understood that Holt’s disclosure of transmitted and received communication data received by a decoder model implemented in a receiver device teaches a *received radio signal that corresponds to a representation of the transmit output radio signal received at the one or more radio receivers. See Ex-1003, ¶245.*

POSITA also would have been motivated to combine Holt’s techniques for machine learning and training with Jüschke at least because doing so would have been a simple substitution of known elements. *See Ex-1003, ¶246.* Additionally, POSITA would have understood that using transmit signals actually received and recovered at a receiver as taught in Holt would have provided a known improvement to Jüschke’s training techniques by using more accurate training data that accounts for the effects of transmission and reception. *Id.* Adding Holt’s received and recovered transmission data into Jüschke’s feedback loop would have introduced more accurate training data into Jüschke’s neural network training techniques, and therefore POSITA would have had a reasonable expectation of success

implementing and improving one form of output data sampling (Jüschke) for another that samples more accurate output data (Holt). *See id.*; Section VII.B, *supra*.

Therefore, POSITA would have understood *obtaining a received radio signal that corresponds to a representation of the transmit output radio signal received at the one or more radio receivers* [4A] to have been taught by Jüschke in view of Holt and would have been motivated to combine them. Jüschke in view of Holt thereby renders [4A] obvious.

2. Jüschke in view of Holt renders [4B] obvious

Jüschke in view of Holt renders obvious “*determining a distance metric between the transmit radio signal and the received radio signal*” [4B].

As explained for limitation [4A], Jüschke teaches transmitting an output radio signal to a radio receiver and implementing a feedback loop for training a pre-distortion neural network model. For the *transmit radio signal*, “[u]ndistorted and, hence, ideal digital baseband IQ samples 12a, 12b [undistorted input values] may be provided to the post distortion neural network 43 as target output values.” Ex-1005, [0048]. For the *received radio signal*, “[s]ampled output values of the non-ideal analog transmitter device 14 may be fed back as input samples for the post distortion neural network 43.” *Id.*, [0048]; *see id.*, Figure 4, 6; *see also* limitation [1A] (Jüschke’s input digital baseband transmit signal 12 teaching a *transmit radio signal*).

Jüschke's neural network 43 is trained using these input/output samples until its response is "the inverse of the transfer or distortion function of the analog transmitter device 14." *Id.*, [0048]. POSITA would have understood that training the neural network for an (input/output) inverse of the transmitter accounts for the difference (or distance) between the sampled input and output values. *See* Ex-1003, ¶249. Therefore, Jüschke teaches determining a *distance metric* by using the difference (or distance) between undistorted input values and sampled output values for training its post-distortion network and pre-distortion neural network. *See id.*

Holt's loss functions also use difference metrics between signals sent and received over a communication channel. For instance, Holt teaches calculating a loss function by comparing the inputs of an encoder model (implemented by a transmitter device) and outputs of a decoder model (implemented by a receiver device) sent over a communication channel. Ex-1006, Abstract (referencing "a loss function that describes a difference between input data...and output data"). POSITA would have understood that a loss function describes the difference between an ML model's predicted output values and the true values, and that a distance metric quantifies this difference. Ex-1003, ¶250; *see also* Ex-1001, 7:34-38. Hence, POSITA would have understood that a loss function calculates the distance, and thereby a *distance metric*, between an ML model's predicted values at the decoder and the true values at the

encoder input. Ex-1003, ¶250. Therefore, POSITA would have understood that Holt's loss function teaches *determining a distance metric between the transmit radio signal and the received radio signal. Id.*

POSITA also would have been motivated to implement Holt's loss functions with transmitters and receivers in Jüschke because it would have provided known advantages. *See* Section VII.B, *supra*. For example, Jüschke's training would have been improved by using Holt's loss functions, which compare the difference between the input signals and *a representation of the transmit output radio signal* actually received at a radio receiver. *See* Ex-1003, ¶252. POSITA also would have understood that implementing Holt's loss function into Jüschke's feedback sampling would have been an exchange of one form of output data sampling for training neural networks (Jüschke's) for another more accurate version (Holt's) of sampling output data actually received and recovered at a receiver. *Id.* Therefore, POSITA would have understood that implementing Holt's loss functions, in lieu of or in addition to Jüschke's transmission input/output feedback loop, also would have been a combination of known elements, methods, and techniques to predictably improve (with a reasonable expectation of success) Jüschke's training and output data sampling. *Id.*

Therefore, POSITA would have understood *determining a distance metric*

between the transmit radio signal and the received radio signal [4B] to have been taught by Jüschke in view of Holt. Jüschke in view of Holt thereby renders [4B] obvious.

3. Jüschke in view of Holt renders [4C] obvious

Jüschke renders obvious “*updating the values of the model parameters for the particular deployment scenario based at least on the distance metric*” [4C] in view of Holt.

As explained for limitation [4B], Jüschke trains its pre-distortion neural network model using undistorted inputs and sampled outputs. Jüschke’s training processor “train[s] weight (w_{kj} ,) and/or bias parameters (b_k),” Ex-1005, [0048], and “the trained weight (w_{kj} ,) and/or bias parameters (b_k) of the post distortion neural network 43 may be transferred to the pre-distortion neural network processor 13.” *Id.*, [0049]. *See id.*, [0072]. Jüschke’s parameters are used “to fit the feed-forward pre-distortion neural network to a specific downstream hardware component.” *Id.*, [0017]. POSITA would have understood that Jüschke’s training and optimizing its parameters teaches *updating the values of the model parameters*. POSITA also would have understood that fitting parameters to a specific hardware component teaches training the parameters for a certain configuration, and therefore *updating the parameters for the particular deployment scenario*. *See* Ex-1003, ¶255.

Moreover, “a pair of the transmitter device’s input-output signals is presented

to the post distortion neural network processor during each training iteration.” Ex-1005, [0018]. POSITA would have understood that Jüschke uses input and output signals, and therefore a difference or *distance metric* between them, for training its models. See Ex-1003, ¶256. Therefore, by training and optimizing its model parameters using inputs and outputs, Jüschke alone renders [4C] obvious.

In addition, as explained for limitation [4B], Holt’s loss functions are used to train and update model parameters. As an example, “***fourth loss function 318 can be backpropagated*** through the decoder model 212 while modifying the decoder model 212 to train the decoder model (e.g., by ***modifying one or more weights associated with the decoder model 212***) to generate outputs that attempt to reconstruct the data provided to the encoder model 202.” Ex-1006, 18:42-48. POSITA would have therefore understood that Holt’s training and updating ML model parameters teach *updating the values of the model parameters for the particular deployment scenario*, and that Holt’s loss functions teach *based at least on the distance metric*. See Ex-1003, ¶257.

POSITA would have been motivated to implement Holt’s loss function and backpropagation techniques for updating ML models into Jüschke because doing so would have provided known advantages. See Section VII.B, *supra*. For example, POSITA would have understood that updating parameters using Holt’s loss

functions based on actually received transmissions would have provided a known accuracy improvement to Jüschke's training techniques by accounting for the actual effects of transmission and reception into Jüschke's output data sampling for neural network training. *See* Ex-1003, ¶258. Updating parameters using Holt's loss functions in lieu of or in addition to Jüschke's transmission feedback loop also would have been a combination of known elements, methods, and techniques to achieve predictable results (with a reasonable expectation of success) because doing so would have been implementing one form of output data sampling (using more accurate sampling data actually received and recovered at a receiver) for another form of output data sampling using a transmission output feedback. *See id.*

Therefore, POSITA would have understood *updating the values of the model parameters for the particular deployment scenario based at least on the distance metric* [4C] to have been taught by Jüschke in view of Holt and would have been motivated to combine them. Jüschke in view of Holt thereby renders [4C] obvious.

4. Jüschke in view of Holt renders obvious similar claims 12 and 20

Limitations [12A]-[12C] and [20A]-[20C] are nearly identical to [4A]-[4C] and rendered obvious by Jüschke in view of Holt for the reasons discussed above.

E. Claims 5, 13, and 21 were obvious over Jüschke in view of Holt.

1. Jüschke in view of Holt renders Claim 5 obvious

Jüschke in view of Holt renders obvious “[t]he method of claim 4, wherein the distance metric is selected from one of: a mean squared error, an absolute error, or another quantitative measurement of an error value computed between the transmit radio signal and the received radio signal” [5].

As explained for limitation [4B], Jüschke determines the difference between input samples and output values. Jüschke further teaches using “MSE (Minimum Squared Error)” when training its pre-distortion neural network. Ex-1005, [0048]. MSE is a measure of estimator quality (the minimum error), i.e., a distance metric, based on *mean squared error*. Ex-1003, ¶261 In addition, Jüschke uses the Levenberg-Marquardt algorithm, an algorithm for nonlinear least squares optimization, for training. *Id.*, [0049]. POSITA would have understood that least squares optimization implements a *mean squared error*. See Ex-1003, ¶261.

In addition to Jüschke, as explained for limitation [4B], Holt determines loss functions to calculate the distance between transmitted inputs and received outputs. See Ex-1006, 18:40-42. POSITA would have understood that mean squared error is a common example of a loss function. See Ex Ex-1003, ¶262; Ex-1012, 55-57 (describing cost functions, equivalent to loss functions, including mean-square error and absolute error); Ex-1017, 2 (describing cost functions and loss functions and

backpropagation minimizing “Mean Squared Error”). Therefore, POSITA would have understood that Holt’s loss function in view of Jüschke teaches a *mean squared error* and a *mean absolute value error*. Moreover, POSITA would have understood that Holt’s fourth loss function 318, by describing a mathematical difference between inputs 316 and outputs 312, teaches a *quantitative measurement of an error value computed between the transmit radio signal and the received radio signal*. See Ex-1003, ¶262.

Therefore, POSITA would have understood that Jüschke in view of Holt teaches claim 5, and would have been motivated to use Holt’s loss function with Jüschke for the reasons explained for limitation [4B]. Jüschke in view of Holt thereby renders claim 5 obvious.

2. Jüschke in view of Holt renders obvious similar claims 13 and 21

Claims 13 and 21 are nearly identical to claim 5 and rendered obvious by Jüschke in view of Holt for the reasons discussed above.

F. Claims 6, 14, and 22 were obvious over Jüschke in view of Holt.

1. Jüschke in view of Holt renders Claim 6 obvious

Jüschke in view of Holt renders obvious “[t]he method of claim 4, wherein updating the values of the model parameters includes using optimizers based on Stochastic Gradient Descent that iteratively minimize the distance metric based on

an exact or approximate gradient of the distance metric with regard to the model parameters” [6].

As explained for limitation [4B], Jüschke determines the difference between input samples and output values, and Holt determines loss functions to calculate the distance between transmitted inputs and received outputs. Jüschke also teaches the use of “[g]radient-based training algorithms, like back-propagation” and the Levenberg-Marquardt algorithm. Ex-1005, [0018]. POSITA would have understood that the Stochastic Gradient Descent (“SGD”) is the preferred common gradient descent type used in backpropagation for neural networks for optimizing and learning model parameters, forming the basis of neural networks. Ex-1003, ¶265; *see* Ex-1016, 1 (SGD being a “more economical” method for minimizing a neural network error function.); Ex-1017, 2 (SGD being “more reliable, and less prone to reach bad local minima than standard gradient descent” for neural networks.). POSITA also would have understood that the Levenberg-Marquardt algorithm taught in Jüschke uses gradient descent. *See* Ex-1005, [0018] (“Hessian-based algorithms allow the neural network to learn more subtle features of a complicated mapping.”). In addition, because both Jüschke and Holt teach backpropagation to train models based on loss functions, POSITA would have understood that a Jüschke/Holt combination would implement SGD without issue. *See* Section

VII.D.3 ([4C]), *supra*. Hence, POSITA would have understood that SGD, being the preferred gradient descent type used in backpropagation for training neural networks, to have been taught by Jüschke's disclosure of "[g]radient-based training algorithms, like back-propagation" for neural networks. Ex-1005, [0018].

Therefore, POSITA would have understood that Jüschke in view of Holt teaches claim 6, and would have been motivated to use Holt's loss function with Jüschke for the reasons explained for limitation [4B]. Jüschke in view of Holt thereby renders claim 6 obvious.

2. Jüschke in view of Holt renders similar claims 14 and 22 obvious

Claims 14 and 22 are nearly identical to claim 6 and rendered obvious by Jüschke in view of Holt for the reasons discussed above.

VIII. GROUND 3: Claims 2, 10, and 18 and dependents were obvious over Jüschke in view of Holt and Dzierwa.

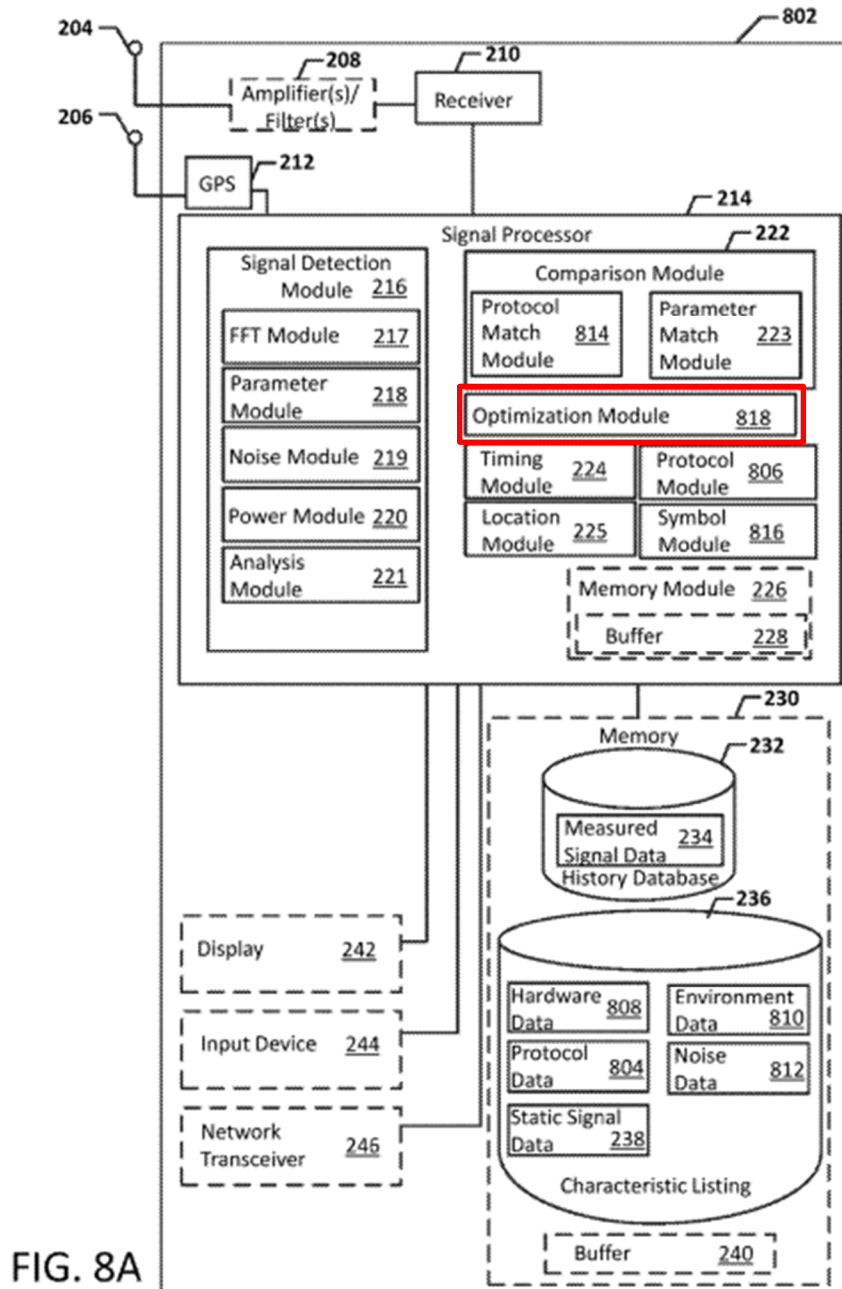
As shown for Ground 2, *supra*, Jüschke in view of Holt renders claims 1-24 obvious. *See* Section VII, *supra*.

To the extent that Jüschke, alone or in combination with Holt, does not render obvious [2B] and similar limitations [10B] and [18B], they and their respective dependent claims (3-6, 11-14, 19-22) would have been rendered obvious by Jüschke in view of Holt and Dzierwa.

A. Modifying Jüschke’s Neural Network with Dzierwa’s Environment Data and Optimization Module

The combined neural network predistortion system implements Dzierwa’s environment data into Jüschke’s data sampling for its pre-distortion neural network. This would have been a simple substitution within the skill and knowledge of POSITA.

As previously explained, Dzierwa implements machine learning for “observ[ing] and learn[ing] an RF environment,” and generating “[a] knowledge map.” Ex-1007, 47:38-41. Dzierwa further teaches an “optimization module” for “noise figure parameters, antenna hardware parameters, and environmental parameters correlated with an identified signal of interest.” *Id.*, 27:52-57. The optimization module can also “enable the signal processor 214 to provide signal optimization parameters.” Figure 8A below illustrates optimization module 818:



Ex-1007, Figure 8A (annotation added).

Dzierwa also teaches “processor 214 may output the tracking information” used to make signal transmission predictions in a wireless environment. Ex-1007,

33:20. Dzierwa uses “deltas or differentials from the baseline, as well as actual data, to provide onsite sensing, measurement, and analysis for a given environment and spectrum” for “apparatus or device unit[s].” *Id.* 37:45-49.

B. Motivation for Modifying Jüschke’s Neural Network

POSITA would have been motivated to modify Jüschke’s neural network using Dzierwa’s environment data and optimization module. This motivation also applies to an already-modified version of Jüschke implementing Holt’s loss function in Jüschke’s training techniques. *See* Section VII.B, *supra*.

Jüschke and Dzierwa (as well as Holt) use ML in wireless environments to improve signal processing performance. *Compare* Ex-1005, Abstract (“mitigating impairments due to transmit noise”) *to* Ex-1007, 27:52-54 (“[O]ptimization module 818 may gather information from the characteristic listing, such as noise.”) *and id.*, 27:33-34 (“communication protocol (e.g., LTE , CDMA , etc.) associated with a signal of interest”). Because Jüschke and Dzierwa are in identical fields and solve similar problems, POSITA also would have been familiar with both and would have recognized that their elements could be interchanged and combined to optimize the predistortion functionality of Jüschke’s networks. *See* Ex-1003, ¶276.

POSITA also would have recognized that combining Dzierwa’s environment data into Jüschke’s neural network and data sampling would have been a combination of familiar elements according to known methods and elements and a

simple substitution. POSITA would have understood that Dzierwa’s “knowledge map...based on the *learning data*” and “*RF environment* from the knowledge map,” Ex-1007, 9:19-21 (emphasis added), could be implemented in Jüschke’s input/output data sampling and neural network. Therefore, POSITA would have understood that implementing Dzierwa’s knowledge map and RF environment data would have predictably modified Jüschke’s data sampling. *See* Ex-1003, ¶277.

Moreover, POSITA would have been motivated to implement Dzierwa’s optimization module and RF environment data into Jüschke because they would have provided a known improvement for similar devices (ML models in radio transceivers). POSITA would have understood that Dzierwa’s optimization module gathers “noise figure parameters, antenna hardware parameters, and environmental parameters...to optimize a signal of interest.” Ex-1007, 27:53-60. POSITA would have understood that these environmental optimizations—as well as “deltas or differentials” and “actual data” for analyzing a “given environment” (*id.*, 37:45-48)—would have provided a predictable improvement to Jüschke’s “flexible radio transmitter.” Ex-1005, [0006]. Therefore, POSITA would have been sufficiently motivated to implement Dzierwa’s optimization module and environment data in Jüschke’s transmitter and neural networks because doing so would have predictably improved performance with more complete environment data for similar ML model

devices, including any versions of Jüschke already modified by Holt. *See* Ex-1003, ¶278. Given that the combination of Jüschke and Holt does not rely on Holt for teaching claims 2 or similar claims 10 and 18, POSITA would have understood the same combination would be improved by the teachings of Dzierwa. *Id.*

C. Claims 2, 10, and 18 were obvious over Jüschke in view of Holt and Dzierwa.

1. Jüschke in view of Holt and Dzierwa renders [2B] obvious

As shown for Grounds 1 and 2, [2B] would have been obvious over Jüschke in view of Holt. *See* Sections VI.C.2, VI.H.1, VI.M.1, and VII, *supra*. To the extent that Jüschke, on its own (Ground 1) or in combination with Holt (Ground 2), is found to not render obvious [2B], it would have been rendered obvious in view of Dzierwa.

Dzierwa teaches machine learning techniques “to observe and learn an RF environment,” and generate “[a] knowledge map...based on learning data,” Ex-1007, 47:38-41, that can be used with an “optimization module.” *Id.*, 27:52. POSITA would have understood that Dzierwa’s RF environment data and optimization module, designed to optimize “noise figure parameters, antenna hardware parameters, and environmental parameters,” *id.*, 27:53-55, could be implemented in Jüschke’s neural network to optimize predistortion *for a particular deployment scenario*.

POSITA would have been sufficiently motivated to combine Jüschke and

Dzierwa because they are both in wireless communications and machine learning and signal processing. Hence, POSITA would have been familiar with both and combining them would have been a simple substitution of known elements. POSITA also would have understood that combining them would have provided predictable improvements, such as more thorough transmitter environment data and optimization values. *See* Section VIII.B, *supra*. Therefore, POSITA would have understood that Jüschke in view of Holt and Dzierwa teaches [2B], and would have been motivated to combine them. Jüschke in view of Holt and Dzierwa thereby renders claim 2 obvious.

2. Jüschke in view of Holt and Dzierwa renders obvious similar limitations [10B] and [18B]

Claim limitations [10B] and [18B] are nearly identical to [2B] and rendered obvious by Jüschke in view of Holt and Dzierwa for the reasons discussed above. *See also* Section VII (Ground 2); Sections VI.C.2 ([2B]), VI.H.1 ([10B]), and VI.M.1 ([18B]), *supra*.

IX. SECONDARY CONSIDERATIONS OF NON-OBVIOUSNESS

Dr. Ding performed an internet search on the Patent Owner and the '704 patent. Ex-1003, ¶287. Dr. Ding found no publicly-available information indicating that the Petitioned Claims are non-obvious. *Id.* Dr. Ding is unaware of any nexus supporting an alleged point of novelty in the Petitioned Claims. *Id.*, ¶¶287-288.

X. The Board Should Not Exercise Discretion to Deny Institution

A. *Fintiv* Favors Institution

Factors 1-5 (favors institution): The '704 patent is not subject to any litigation.

Factor 6 (favors institution): Petitioner presents strong grounds for invalidity.

B. The *Advanced Bionics* Factors Favor Institution

None of the references relied on were previously applied, nor are they cumulative to those used during prosecution.

XI. CONCLUSION

Petitioner requests the Board institute *inter partes* review and cancel each challenged claim.

XII. MANDATORY NOTICES

A. Real Parties-in-Interest Under 37 C.F.R. §42.8(b)(1)

The real party-in-interest is Digital Global Systems, Inc.

B. Related Matters Under 37 C.F.R. §42.8(b)(2)

To the best of Petitioner's knowledge, the '704 patent is not, and has not been, involved in any district court litigation.

C. Lead and Back-Up Counsel Under 37 C.F.R. §42.8(b)(3)

Lead Counsel	First Back-up Counsel
Brian Eutermoser, Reg. No. 64,058 KING & SPALDING LLP 1401 Lawrence Street Suite 1900 Denver, CO 80202 (720) 535-2300 beutermoser@kslaw.com	Chris Campbell, Reg. No. 37,291 KING & SPALDING LLP 1700 Pennsylvania Avenue, NW Suite 900 Washington, DC 20006 (202) 626-5578 ccampbell@kslaw.com

Second Back-up Counsel
John Roehrick, Reg. No. 77,252 KING & SPALDING LLP 633 West Fifth Street Suite 1600 Los Angeles, CA 90071 (213) 443-4355 jroehrick@kslaw.com

D. Service Information Under 37 C.F.R. §42.8(b)(4)

Please address all correspondence to counsel at the addresses above. Petitioner consents to service by email at the above email addresses.

Petition for *Inter Partes* Review of
U.S. Patent No. 11,018,704

Respectfully submitted,

KING & SPALDING LLP

By: /s/ Brian Eutermoser

Brian Eutermoser, Reg. No. 64,058

Chris Campbell, Reg. No. 37,291

John Roehrick, Reg. No. 77,252

Counsel for Petitioner

CERTIFICATION OF WORD COUNT

The undersigned certifies pursuant to 37 C.F.R. §42.24 that the foregoing Petition for *Inter Partes* Review, excluding any table of contents, mandatory notices under 37 C.F.R. §42.8, certificates of service or word count, or appendix of exhibits, contains 13,722 words according to the word-processing program used to prepare this document (Microsoft Word).

Dated: August 28, 2024

Respectfully submitted,

By: /s/ Brian Eutermoser
KING & SPALDING LLP
1401 Lawrence Street
Suite 1900
Denver, CO 80202
(720) 535-2300

Counsel for Petitioner

CERTIFICATE OF SERVICE

Pursuant to 37 C.F.R. §§42.6(e) and 42.105(b), the undersigned certifies that on August 28, 2024, a complete and entire copy of this Petition for *Inter Partes* Review and Exhibits were provided overnight mail to counsel for Patent Owner and Patent Owner by serving copies to the address below for the '704 patent:

Ayan Roy-Chowdhury
Fish & Richardson P.C.
P.O. BOX 1022
Minneapolis, MN 55440-1022

DeepSig, Inc.
1201 Wilson Blvd., Floor 27
Arlington, VA 22209

DeepSig, Inc.
950 N Glebe Road, Suite 910
Arlington, VA 22203

Respectfully submitted,

By: /s/ Brian Eutermoser
KING & SPALDING LLP
1401 Lawrence Street
Suite 1900
Denver, CO 80202
(720) 535-2300

Counsel for Petitioner

Appendix 1

Listing of Claims

Claim 1	Limitation
[1P]	A method for correcting distortion of radio signals in a cellular communications system:
[1A]	obtaining, at a base station in the cellular communications system, a transmit radio signal that is prepared for transmitting to one or more radio receivers associated with the cellular communications system;
[1B.1]	accessing a nonlinear pre-distortion machine learning model associated with the base station,
[1B.2]	wherein the nonlinear pre-distortion machine learning model includes model parameters associated with values corresponding to one or more deployment scenarios in the cellular communications system,
[1B.3]	and at least one nonlinear function to correct radio signal distortion;
[1C]	generating a pre-distorted radio signal by processing the transmit radio signal using the nonlinear pre-distortion machine learning model;
[1D]	obtaining a transmit output radio signal by processing the pre-distorted radio signal through a transmitting radio signal processing stage of the base station; and
[1E]	transmitting the transmit output radio signal to the one or more radio receivers.

Claim 2	Limitation
[2A]	The method of claim 1, wherein the model parameters in the nonlinear pre-distortion machine learning model are loaded with values for a particular deployment scenario of the one or more deployment scenarios, and
[2B]	wherein generating the pre-distorted radio signal comprises generating the pre-distorted radio signal for the particular deployment scenario.

Claim 3	Limitation
[3A]	The method of claim 2, wherein the particular deployment scenario comprises one of: sending a radio transmission from the base station that includes the nonlinear pre-distortion machine learning model inside a tower mounted amplifier in an inline mode;
[3B]	sending a radio transmission from the base station that includes the nonlinear pre-distortion machine learning model inside a tower mounted amplifier in a fault-tolerant subtractive mode;
[3C]	sending a radio transmission from the base station that includes the nonlinear pre-distortion machine learning model inside a remote radio unit in an inline mode;
[3D]	sending a radio transmission from the base station that includes the nonlinear pre-distortion machine learning model inside a remote radio unit in a fault-tolerant subtractive mode; or
[3E]	sending a radio transmission to a system that includes interference cancellation.

Claim 4	Limitation
[4A]	The method of claim 2, further comprising: obtaining a received radio signal that corresponds to a representation of the transmit output radio signal received at the one or more radio receivers;
[4B]	determining a distance metric between the transmit radio signal and the received radio signal; and
[4C]	updating the values of the model parameters for the particular deployment scenario based at least on the distance metric.

5. The method of claim 4, wherein the distance metric is selected from one of:
a mean squared error, an absolute error, or another quantitative measurement of an
error value computed between the transmit radio signal and the received radio signal.

6. The method of claim 4, wherein updating the values of the model parameters includes using optimizers based on Stochastic Gradient Descent that iteratively minimize the distance metric based on an exact or approximate gradient of the distance metric with regard to the model parameters.

7. The method of claim 1, wherein the values associated with the model parameters are optimized to minimize the distortion or interference of radio signals associated with the base station in the cellular communications system.

8. The method of claim 1, wherein the nonlinear pre-distortion machine learning model is one of a deep neural network, a convolutional neural network, a recurrent neural network, or a combination of different types of neural networks.

Claim 9	Limitation
[9P]	An apparatus for correcting distortion of radio signals in a cellular communications system, the apparatus comprising:
[9A]	one or more processors associated with a base station in the cellular communications system; and
[9B]	one or more machine-readable media storing instructions that, when executed, are configured to cause the one or more processors to perform operations comprising:
[9C]	obtaining, at the base station, a transmit radio signal that is prepared for transmitting to one or more radio receivers associated with the cellular communications system;
[9D.1]	accessing a nonlinear pre-distortion machine learning model associated with the base station,
[9D.2]	wherein the nonlinear pre-distortion machine learning model includes model parameters associated with values corresponding to one or more deployment scenarios in the cellular communications system,
[9D.3]	and at least one nonlinear function to correct radio signal distortion;
[9E]	generating a pre-distorted radio signal by processing the transmit radio signal using the nonlinear pre-distortion machine learning model;
[9F]	obtaining a transmit output radio signal by processing the pre-distorted radio signal through a transmitting radio signal processing stage of the base station; and
[9G]	transmitting the transmit output radio signal to the one or more radio receivers.

Claim 10	Limitation
[10A]	The apparatus of claim 9, wherein the model parameters in the nonlinear pre-distortion machine learning model are loaded with values for a particular deployment scenario of the one or more deployment scenarios, and
[10B]	wherein generating the pre-distorted radio signal comprises generating the pre-distorted radio signal for the particular deployment scenario.

Claim 11	Limitation
[11A]	The apparatus of claim 10, wherein the particular deployment scenario comprises one of: sending a radio transmission from the base station that includes the nonlinear pre-distortion machine learning model inside a tower mounted amplifier in an inline mode;
[11B]	sending a radio transmission from the base station that includes the nonlinear pre-distortion machine learning model inside a tower mounted amplifier in a fault-tolerant subtractive mode;
[11C]	sending a radio transmission from the base station that includes the nonlinear pre-distortion machine learning model inside a remote radio unit in an inline mode;
[11D]	sending a radio transmission from the base station that includes the nonlinear pre-distortion machine learning model inside a remote radio unit in a fault-tolerant subtractive mode; or
[11E]	sending a radio transmission to a system that includes interference cancellation.

Claim 12	Limitation
[12A]	The apparatus of claim 10, wherein the operations further comprise: obtaining a received radio signal that corresponds to a representation of the transmit output radio signal received at the one or more radio receivers;
[12B]	determining a distance metric between the transmit radio signal and the received radio signal; and
[12C]	updating the values of the model parameters for the particular deployment scenario based at least on the distance metric.

13. The apparatus of claim 12, wherein the distance metric is selected from one of: a mean squared error, an absolute error, or another quantitative measurement

of an error value computed between the transmit radio signal and the received radio signal.

14. The apparatus of claim 12, wherein updating the values of the model parameters includes using optimizers based on Stochastic Gradient Descent that iteratively minimize the distance metric based on an exact or approximate gradient of the distance metric with regard to the model parameters.

15. The apparatus of claim 9, wherein the values associated with the model parameters are optimized to minimize the distortion or interference of radio signals associated with the base station in the cellular communications system.

16. The apparatus of claim 9, wherein the nonlinear pre-distortion machine learning model is one of a deep neural network, a convolutional neural network, a recurrent neural network, or a combination of different types of neural networks.

Claim 17	Limitation
[17P]	One or more non-transitory machine-readable media storing instructions that, when executed, are configured to cause one or more processors to perform operations comprising:
[17A]	obtaining, at a base station in a cellular communications system, a transmit radio signal that is prepared for transmitting to one or more radio receivers associated with the cellular communications system;
[17B.1]	accessing a nonlinear pre-distortion machine learning model associated with the base station,
[17B.2]	wherein the nonlinear pre-distortion machine learning model includes model parameters associated with values corresponding to one or more deployment scenarios in the cellular communications system,
[17B.3]	and at least one nonlinear function to correct radio signal distortion;
[17C]	generating a pre-distorted radio signal by processing the transmit radio signal using the nonlinear pre-distortion machine learning model;
[17D]	obtaining a transmit output radio signal by processing the pre-distorted radio signal through a transmitting radio signal processing stage of the base station; and
[17E]	transmitting the transmit output radio signal to the one or more radio receivers.

Claim 18	Limitation
[18A]	The one or more non-transitory machine-readable media of claim 17, wherein the model parameters in the nonlinear pre-distortion machine learning model are loaded with values for a particular deployment scenario of the one or more deployment scenarios, and
[18B]	wherein generating the pre-distorted radio signal comprises generating the pre-distorted radio signal for the particular deployment scenario.

Claim 19	Limitation
[19A]	The one or more non-transitory machine-readable media of claim 18, wherein the particular deployment scenario comprises one of: sending a radio transmission from the base station that includes the nonlinear pre-distortion machine learning model inside a tower mounted amplifier in an inline mode;
[19B]	sending a radio transmission from the base station that includes the nonlinear pre-distortion machine learning model inside a tower mounted amplifier in a fault-tolerant subtractive mode;
[19C]	sending a radio transmission from the base station that includes the nonlinear pre-distortion machine learning model inside a remote radio unit in an inline mode;
[19D]	sending a radio transmission from the base station that includes the nonlinear pre-distortion machine learning model inside a remote radio unit in a fault-tolerant subtractive mode; or
[19E]	sending a radio transmission to a system that includes interference cancellation.

Claim 20	Limitation
[20A]	The one or more non-transitory machine-readable media of claim 18, wherein the operations further comprise: obtaining a received radio signal that corresponds to a representation of the transmit output radio signal received at the one or more radio receivers;
[20B]	determining a distance metric between the transmit radio signal and the received radio signal; and
[20C]	updating the values of the model parameters for the particular deployment scenario based at least on the distance metric.

21. The one or more non-transitory machine-readable media of claim 20, wherein the distance metric is selected from one of: a mean squared error, an

absolute error, or another quantitative measurement of an error value computed between the transmit radio signal and the received radio signal.

22. The one or more non-transitory machine-readable media of claim 20, wherein updating the values of the model parameters includes using optimizers based on Stochastic Gradient Descent that iteratively minimize the distance metric based on an exact or approximate gradient of the distance metric with regard to the model parameters.

23. The one or more non-transitory machine-readable media of claim 17, wherein the values associated with the model parameters are optimized to minimize the distortion or interference of radio signals associated with the base station in the cellular communications system.

24. The one or more non-transitory machine-readable media of claim 17, wherein the nonlinear pre-distortion machine learning model is one of a deep neural network, a convolutional neural network, a recurrent neural network, or a combination of different types of neural networks.