

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

Digital Global Systems, Inc.

Petitioner,

v.

DeepSig, Inc.,

Patent Owner

IPR2025-00049

Patent **10,581,469**

PETITION FOR *INTER PARTES* REVIEW OF

U.S. PATENT NO. 10,581,469

DECLARATION OF ZHI DING, PH. D.

TABLE OF CONTENTS

I.	Introduction And Qualifications	1
A.	Qualifications and Experience.....	2
B.	Materials Considered.....	8
II.	Person of Ordinary Skill in the Art	10
III.	Statement Of Legal Principles	12
A.	Claim Construction.....	12
B.	Legal Standards for Prior Art	14
C.	Anticipation	16
D.	Obviousness.....	17
IV.	Technology Background	22
A.	Signals	22
B.	Nonlinear Signal Distortion.....	24
C.	Signal Processing and Predistortion	30
D.	Machine Learning and Deep Learning	31
V.	The '469 Patent	33
A.	The '469 Patent's Prosecution History.....	47
VI.	Overview of the Prior Art References.....	48
A.	Jüschke	48
B.	Holt.....	54
C.	Dzierwa.....	58
VII.	Opinions Regarding the '469 patent's Invalidity	59
VIII.	GROUND 1: Claims 1-8, 10, 12-15, and 17-24 were obvious over Jüschke in view of Holt.	60
A.	Description of Jüschke	60
B.	The Combined Neural Network Predistortion System: Modifying Jüschke's Training with Holt's Loss Function.....	69

Declaration of Zhi Ding, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 10,581,469

C.	Motivation for Modifying Jüschke’s Training	74
D.	Jüschke in view of Holt teaches claim 1	78
1.	Jüschke in view of Holt teaches [1P]	78
2.	Jüschke in view of Holt teaches [1A]	79
3.	Jüschke in view of Holt teaches [1B].....	81
4.	Jüschke teaches [1C]	89
5.	Jüschke teaches [1D].....	92
6.	Jüschke in view of Holt teaches [1E]	98
E.	Jüschke in view of Holt teaches claim 2	111
F.	Jüschke in view of Holt teaches claim 3	119
G.	Jüschke in view of Holt teaches claim 4	121
H.	Jüschke in view of Holt teaches claim 5	122
1.	Jüschke in view of Holt teaches [5A]/[5C]	123
I.	Jüschke in view of Holt teaches claim 6	130
J.	Jüschke in view of Holt teaches claim 7	132
K.	Jüschke in view of Holt teaches claim 8	135
1.	Jüschke in view of Holt teaches claim [8P]	138
2.	Jüschke in view of Holt teaches claim [8A].....	138
3.	Jüschke in view of Holt teaches claim [8B.1]	139
4.	Jüschke in view of Holt teaches claim [8B2].....	140
5.	Jüschke in view of Holt teaches claim [8C]	141
6.	Jüschke in view of Holt teaches claim [8D].....	142
7.	Jüschke in view of Holt teaches claim [8E.1]	143
8.	Jüschke in view of Holt teaches claim [8E.2]	145
9.	Jüschke in view of Holt teaches claim [8F]	146
L.	Jüschke in view of Holt teaches claim 10	147

Declaration of Zhi Ding, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 10,581,469

M.	Jüschke in view of Holt teaches claim 12	148
N.	Jüschke in view of Holt teaches claim 13	150
O.	Jüschke in view of Holt teaches claim 14	152
P.	Jüschke in view of Holt teaches claim 15	153
Q.	Jüschke in view of Holt teaches claim 17	154
R.	Jüschke in view of Holt teaches claim 18	156
S.	Jüschke in view of Holt teaches claim 19	157
T.	Jüschke in view of Holt teaches claim 20	159
U.	Jüschke in view of Holt teaches claims 21-22.....	159
V.	Jüschke in view of Holt teaches claim 21	161
	1. Jüschke in view of Holt teaches [21A]	161
	2. Jüschke in view of Holt teaches [21B].....	162
W.	Jüschke in view of Holt teaches claims 23-24.....	163
IX.	GROUND 2: Claims 2, 22, and 24 were obvious over Jüschke in view of Holt and Dzierwa.	166
A.	The Combined Neural Network Predistortion System: Modifying Jüschke’s Neural Network with Dzierwa’s Environment Data and Optimization.	166
B.	Motivation for Modifying Jüschke’s Neural Network	169
C.	Jüschke in view of Holt and Dzierwa teaches claim 2	171
	1. Jüschke in view of Holt and Dzierwa teaches claim 2.....	172
D.	Jüschke in view of Holt and Dzierwa teaches claim 3	173
E.	Jüschke in view of Holt and Dzierwa teaches claim 4	173
F.	Jüschke in view of Holt and Dzierwa teaches claim 5	173
G.	Jüschke in view of Holt and Dzierwa teaches claims 22 and 24.....	174
X.	No Secondary Considerations of Non-Obviousness	175
XI.	Conclusion.....	176

Declaration of Zhi Ding, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 10,581,469

I, Zhi Ding, Ph.D., declare as follows:

I. INTRODUCTION AND QUALIFICATIONS

1. I, Zhi Ding, submit this declaration to state my opinions on the matters described below.

2. I have been retained as an independent expert in this proceeding before the United States Patent and Trademark Office by DGS Global Systems, Inc.

3. I understand this proceeding involves U.S. Patent No. 10,581,469 (“the ’469 patent”), and I have been asked to provide my opinions as to the patentability or unpatentability of the claims of the ’469 patent as would have been understood by one of ordinary skill in the art. Specifically, I have been asked to consider the validity of claims 1-8, 10, 12-15, and 17-24 of the ’469 patent (the “Challenged Claims”) in view of prior art, anticipation and obviousness considerations, and understanding of a person of ordinary skill in the art (“POSITA”) as it relates to the ’469 patent. I have personal knowledge of the facts and opinions set forth in this declaration, and believe them to be true. If called upon to do so, I would testify competently thereto.

4. This declaration sets forth my opinions, which I have formed based on my study of the evidence; my education, training, research, knowledge, and personal and professional experience; and my understanding as an expert in the field.

5. I am being compensated for my time at the rate of \$820 per hour. This compensation is not contingent upon the nature of my findings, the presentation of my findings in testimony, or the outcome of this proceeding. I have no financial interest in the outcome of this matter or on any potential litigation between Petitioner and Patent Owner.

A. Qualifications and Experience

6. I presently hold the title of Distinguished Professor in the Department of Electrical and Computer Engineering at the University of California, Davis. Before my appointment on July 1, 2020, I held the position of professor for the prior 23 years and have served as a faculty member at several United States universities for over 33 years. I am also a private technical consultant on various technologies related to information systems. In total, I have more than three decades of research experience on a wide range of topics related to data communications and signal processing.

7. I earned my Bachelor of Science degree in 1982 in wireless engineering from the Nanjing Institute of Technology (later renamed as Southeast University) in Nanjing, China. I earned my Master of Science degree in 1987 in electrical engineering from the University of Toronto in Toronto, Canada. I earned my Ph.D. in 1990 in electrical engineering from Cornell University in Ithaca, New York.

Declaration of Zhi Ding, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 10,581,469

8. My responsibilities as a Professor at University of California, Davis, include classroom instruction on various topics of communication systems and signal analysis, as well as mentoring undergraduate students and supervising graduate students in their research and development efforts on various topics related to digital communications. I have directly supervised such research and development works ranging from signal detection to wireless networking. As the chief academic advisor, I have also directly supervised the completion of over 20 Masters' theses and over 30 Ph.D. dissertations on various topics related to digital communications. I have served full time as a faculty member at three major research universities in the United States over the past 30 years, including Auburn University from 1990 to 1998, University of Iowa from 1999 to 2000, and University of California, Davis, from 2000 to present.

9. Since 1990, I have been selected as the principal investigator of multiple highly competitive federal and local research grants, including more than twenty major research projects supported by the National Science Foundation and two research projects funded by the U.S. Army Research Office. These competitive research projects focused on developing more efficient and effective digital communication transceivers, networks, and signal processing tools. I have also participated in several large-scale projects supported by the Defense Advanced

Declaration of Zhi Ding, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 10,581,469

Research Projects Agency (DARPA) with teams of researchers. I have applied for, and received support from, other federal, state, and industry sponsors.

10. I have published over 250 peer-reviewed research articles in premier international journals, in addition to over 250 refereed technical articles at top international conferences on communications and information technologies. According to Google Scholar, my published works have been cited by over 20,000 times by peers. I also authored two books on communications technologies. My most recent book, coauthored with B.P. Lathi, is entitled, “Modern Digital and Analog Communication Systems,” 5th edition, and was published by the Oxford University Press in 2018. The 4th edition of this book (published in 2009) had been widely adopted as an introductory textbook to communication systems.

11. In addition to the over 500 published technical papers that have been cited over 20,000 times according to Google Scholar, I am also co-inventor of 4 issued U.S. patents on communication technologies.

12. I am a member of the Institute of Electrical and Electronics Engineers (IEEE) and was elevated to the grade of Fellow in January 2003 for contributions made in signal processing for communication. The IEEE is the world's largest professional society of engineers, with over 400,000 members in more than 160 countries. The IEEE has led the development of many standards for modern digital

Declaration of Zhi Ding, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 10,581,469

communications and networking, most notably, the IEEE 802 series of network standards. The IEEE Grade of Fellow is conferred by the Boards of Directors upon a person with an extraordinary record of accomplishments in any of the IEEE fields of interest. The total number selected in any one year does not exceed one-tenth of one percent of the total voting Institute membership.

13. I have served the IEEE in the following capacities:

14. Chief Information Officer of the IEEE Communications Society from Jan. 2018 to present.

15. Chief Marketing Officer of the IEEE Communications Society from Jan. 2020 to present.

16. General Chair of the 2016 IEEE International Conference on Acoustics, Speech, and Signal Processing, the flagship conference of the IEEE Signal Processing Society.

17. Chair of the Steering Committee for the IEEE Transactions on Wireless Communications from 2008 to 2010.

18. Distinguished Lecturer of the IEEE Communications Society from January 2008 to December 2009.

19. Technical Program Chair of the 2006 IEEE Globecom, one of two flagship annual IEEE Communication Society conferences.

Declaration of Zhi Ding, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 10,581,469

20. Distinguished Lecturer of the IEEE Circuits and Systems Society from 2004 to 2005.

21. Associate Editor of the IEEE Transactions on Signal Processing from 1994 to 1997 and from 2001 to 2004.

22. Member of the IEEE Statistical Signal and Array Processing for Communications Technical Committee from 1993 to 1998.

23. Member of the IEEE Signal Processing for Communications Technical Committee from 1998 to 2004.

24. In 2012, I received the Wireless Communications Technical Committee Recognition Award from the IEEE Communications Society, an award given to a person with a high degree of visibility and contribution in the field of "Wireless and Mobile Communications Theory, Systems, and Networks." I received the 2020 Education Award from the IEEE Communications Society. According to the Society, this award "recognizes distinguished and significant contributions to education within the Society's technical scope."

25. I have also served as a technical consultant for the telecommunication industry. For example, in 1995 I consulted for Analog Devices, Inc., on the development of the first generation DOCSIS cable modem systems. I have also consulted for other companies, including Nortel Networks and NEC US

Declaration of Zhi Ding, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 10,581,469

Laboratories. I worked as a visiting faculty research fellow at NASA Glenn Research Center in 1992 and at U.S. Air Force Wright Laboratory in 1993. I have served on multiple review panels of the National Science Foundation to evaluate competitive research proposals in the field of communication. I have also reviewed a large number of research proposals at the request of the National Science and Engineering Research Council (NSERC) of Canada as an expert panelist from 2010 to 2013, and also at the request of the Research Grant Council (RGC) of Hong Kong as an external reviewer.

26. I have served as an expert witness or consulting expert on a number of matters related to intellectual property, mostly in the arena of telecommunications, including cellular communications, Wi-Fi technologies, Bluetooth, and optical communications. For example, since 2007, I have been engaged to work on various litigations involving cellular, WiFi, Bluetooth, and optical communication networks. Many of my expert works related to the standard essentiality of patents with respect to the 3GPP 2G/3G/4G/5G technical specifications, the IEEE 802.11 standards, and the Bluetooth standards. I have provided testimonies at court trials or hearings in multiple jurisdictions, both in the United States and internationally. I have submitted declarations to PTAB as expert on a number of inter partes reviews of challenged patents.

27. Further experience and a complete list of my publications are presented in my curriculum vitae. *See* Ex-1004, *infra*.

B. Materials Considered

28. In reaching the conclusions described in this declaration, I have relied on the documents and materials cited herein as well as those identified in this declaration, including the '469 patent (Ex. 1001), the file history of the '469 patent (Ex. 1002), the prior art references, and information discussed and any other references specifically identified in this declaration, including the materials identified in the chart below. I have also relied on my education, training, research, knowledge, and personal and professional experience in the relevant technologies and systems that were already in use prior to, and within the timeframe of the earliest priority date of the claimed subject matter in the '469 patent, which I understand is April 17, 2017.

29. I also reviewed portions of the Manual of Patent Examining procedure, including Sections 2131, 2141-2145. Those sections explain the legal requirements of anticipation, obviousness, and examining procedures relevant to obviousness.

30. Provided below are the Exhibit numbers for documents cited in this declaration:

Declaration of Zhi Ding, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 10,581,469

Exhibit No.	Description
1001	U.S. Patent 10,581,469 to O'Shea et al. (the '469 patent)
1002	U.S. Patent 10,581,469 File History
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> , Proceedings of Neuro-Nîmes Vol. 91, No. 8 (1991),

Exhibit No.	Description
	https://leon.bottou.org/publications/pdf/nimes-1991.pdf (“Bottou”).

II. PERSON OF ORDINARY SKILL IN THE ART

31. I understand that an assessment of claims of the ’469 patent should be undertaken from the perspective of a person of ordinary skill in the art as of the earliest claimed priority date, which I understand is April 17, 2017. I have also been advised that to determine the appropriate level of a person having ordinary skill in the art, the following factors may be considered: (1) the types of problems encountered by those working in the field and prior art solutions thereto; (2) the sophistication of the technology in question, and the rapidity with which innovations occur in the field; (3) the educational level of active workers in the field; and (4) the educational level of the inventor.

32. The ’469 patent relates to implementing predistortion signal processing techniques using nonlinear machine learning models. In particular, the ’469 patent describes “[m]ethods, systems, and apparatus, including computer programs encoded on computer storage media, for correcting distortion of radio signals.” ’469 patent, Abstract. The ’469 patent further describes obtaining a transmit radio signal, processing the transmit radio signal using a nonlinear pre-distortion machine learning model to generate a pre-distorted radio signal, processing the pre-distorted

radio signal through the transmitting radio signal processing system to obtain a transmit output radio signal, and then transmitting the output radio signal is then transmitted to one or more radio receivers. *Id.* The '469 patent further describes the nonlinear pre-distortion machine learning model as “includ[ing] model parameters and at least one nonlinear function to correct radio signal distortion or interference.” *Id.*

33. A POSITA is someone who understands radio communication systems (including RF amplification, distortions, precorrection, physical-layer applications, software and hardware, components such as resistors, capacitors, inductors, diodes, transistors, processors, and circuitry such as storage units, transmitters, receivers, and amplifiers), signal processing and machine learning (including deep learning and neural networks). In view of the technology described in the '469 patent, the POSITA would have had a bachelor's degree in electrical engineering, computer engineering, computer science, or an equivalent degree and at least two years of experience working in a relevant industry at the time of the '469 patent's filing. Additional work or research experience in the field of signal precorrection and machine learning at the time of the '469 patent's filing would have been a suitable substitute for less or different education, and vice-versa.

34. My opinions regarding the level of ordinary skill in the art are based on, among other things, my over 35 years of experience in electrical and computer engineering, radio communication systems, and machine learning including neural networks, my understanding of the basic qualifications that would be relevant to an engineer or scientist tasked with investigating methods and systems in the relevant area, and my familiarity with the backgrounds of colleagues, co-workers, and employees, both past and present.

35. Although my qualifications and experience exceed those of the hypothetical person having ordinary skill in the art defined above, my analysis and opinions regarding the '469 patent have been based on the perspective of a person of ordinary skill in the art as of April 17, 2017.

III. STATEMENT OF LEGAL PRINCIPLES

36. I am not a lawyer. My understanding of legal principles comes from counsel. I have applied the following legal principles in arriving at the opinions set forth in this declaration.

A. Claim Construction

37. I understand that under the legal principles, claim terms are generally given their ordinary and customary meaning, which is the meaning that the term would have to a person of ordinary skill in the art in question at the time of the

invention, i.e., as of the effective filing date of the patent application. I further understand that the person of ordinary skill in the art is deemed to read the claim term not only in the context of the particular claim in which a claim term appears, but in the context of the entire patent, including the specification.

38. I am informed by counsel that the patent specification, under the legal principles, has been described as the single best guide to the meaning of a claim term, and is thus highly relevant to the interpretation of claim terms. And I understand for terms that do not have a customary meaning within the art, the specification usually supplies the best context of understanding the meaning of those terms.

39. I am further informed by counsel that other claims of the patent in question, both asserted and unasserted, can be valuable sources of information as to the meaning of a claim term. Because the claim terms are normally used consistently throughout the patent, the usage of a term in one claim can often illuminate the meaning of the same term in other claims. Differences among claims can also be a useful guide in understanding the meaning of particular claim terms.

40. I understand that the prosecution history can further inform the meaning of the claim language by demonstrating how the inventors understood the invention and whether the inventors limited the invention in the course of prosecution, making

the claim scope narrower than it otherwise would be. Extrinsic evidence may also be consulted in construing the claim terms, such as my expert testimony.

41. I have been informed by counsel that, in *Inter Partes* Review (IPR) proceedings, a claim of a patent shall be construed using the same claim construction standard that would be used to construe the claim in a civil action under 35 U.S.C. §282(b) (the “Phillips” claim construction standard), including construing the claim in accordance with the ordinary and customary meaning of such claim as understood by one of ordinary skill in the art and the prosecution history pertaining to the patent.

42. I have applied the “Phillips” claim construction standard for purposes of interpreting the claims in this proceeding, to the extent they require an explicit construction. The description of the legal principles set forth above thus provides my understanding of the “Phillips” standard as provided to me by counsel.

B. Legal Standards for Prior Art

43. I understand that a patent or other publication must first qualify as prior art before it can be used to invalidate a patent claim.

44. I understand that a U.S. or foreign patent qualifies as prior art to a patent if the date of issuance of the U.S. or foreign patent is prior to the invention of the patent. I further understand that a printed publication, such as an article published in

Declaration of Zhi Ding, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 10,581,469

a magazine or trade publication or a patent application, qualifies as prior art to a patent if the date of publication is prior to the invention of the patent.

45. I understand that, regardless of the date of invention of the patent, a U.S. or foreign patent qualifies as prior art to a patent if the date of issuance of the U.S. or foreign patent is more than one year before the earliest effective filing date of the patent. I further understand that a printed publication, such as an article published in a magazine or trade publication or a patent application, constitutes prior art to a patent if the publication occurs more than one year before the earliest effective U.S. filing date of the patent, again regardless of the date of invention of the patent.

46. I understand that a U.S. patent or published U.S. application qualifies as prior art to a patent if the application for that patent was filed in the United States before the invention of the patent. My understanding is that, for such prior art references, a patentee may attempt to show that the claimed invention (if the corresponding application was filed before March 16, 2013) was conceived prior to the filing in the United States of the purported prior art U.S. patent or application. To do so, it is my understanding that patentee must prove with corroborating evidence that the named inventors conceived of the complete claimed invention before the prior art, and were diligent in reducing the claimed inventions to practice.

47. I understand that to qualify as prior art, a reference must contain an enabling disclosure that allows one of ordinary skill to practice the claims without undue experimentation.

48. I understand that documents and materials that qualify as prior art can be used to invalidate a patent claim as anticipated or as obvious. I understand that documents and materials qualify as prior art if they were available to the public or those interested in the technology before the priority date of the '469 patent.

C. Anticipation

49. I understand that to be granted a patent, the invention must not have been previously disclosed by a reference that is available as prior art. I am not a lawyer, but I have been informed by counsel, that in order to anticipate a patent, the reference must disclose each and every element required by the patent's claims.

50. As part of anticipation, I have been further informed that a prior art reference may anticipate a reference even if it does not disclose all elements of the claim, if those claim elements are inherent. I have been informed that a prior art reference inherently teaches an element if the inherent characteristic necessarily flows from the teachings of the prior art. I further understand that inherency can be demonstrated using extrinsic evidence that makes clear that the missing descriptive matter is necessarily present in what the reference describes.

D. Obviousness

51. I am not a lawyer, and my opinions are limited to my understanding of the relevant field around the time that the inventor conceived the technology behind the '469 patent. Namely, my knowledge of the art at that time, and what that art would have taught and motivated a POSITA at that time. Nevertheless, I have an understanding of obviousness as explained to me by counsel.

52. I have been informed that a claim may be invalid under 35 U.S.C. §103(a) if the subject matter described by the claim as a whole would have been obvious to a hypothetical person of ordinary skill in the art in view of a prior art reference or in view of a combination of references at the time the claimed invention was made. I have been informed that obviousness is determined from the perspective of a hypothetical POSITA and that the asserted claims of the patent should be read from the point of view of such a person at the time the claimed invention was made.

53. I have been informed that a hypothetical POSITA is assumed to know and to have all relevant prior art in the field of endeavor covered by the claims of the patent.

54. I have been informed that there are two criteria for determining whether prior art is analogous and thus can be considered prior art: (1) whether the art is from the same field of endeavor, regardless of the problem addressed; and (2) if the

reference is not within the field of the patentee's endeavor, whether the reference still is reasonably pertinent to the particular problem with which the patentee is involved. I also have been informed that the field of endeavor of a patent is not limited to the specific point of novelty, the narrowest possible conception of the field, or the particular focus within a given field. I also have been informed that a reference is reasonably pertinent if, even though it may be in a different field from that of the patentee's endeavor, it is one that, because of the matter with which it deals, logically would have commanded itself to a patentee's attention in considering his problem.

55. I have been informed that an analysis of whether a claimed invention would have been obvious should be considered in light of the scope and content of the prior art, the differences (if any) between the prior art and the claimed invention, and the level of ordinary skill in the pertinent art involved. I have been informed as well that a prior art reference should be viewed as a whole.

56. I have been informed that in considering whether an invention for a claimed combination would have been obvious, I may assess whether there are apparent reasons to combine known elements in the prior art in the manner claimed in view of interrelated teachings of multiple prior art references, the effects of demands known to the design community or present in the market place, and/or the

Declaration of Zhi Ding, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 10,581,469

background knowledge possessed by a person having ordinary skill in the art. I have been informed that other principles may be relied on in evaluating whether a claimed invention would have been obvious, and that these principles include the following:

- a combination of familiar elements according to known methods is likely to be obvious when it does no more than yield predictable results;
- when a device or technology is available in one field of endeavor,
- design incentives and other market forces can prompt variations of it, either in the same field or in a different one, so that if a person of ordinary skill can implement a predictable variation, the variation is likely obvious;
- if a technique has been used to improve one device, and a person of ordinary skill in the art would recognize that it would improve similar devices in the same way, using the technique is obvious unless its actual application is beyond his or her skill;
- an explicit or implicit teaching, suggestion, or motivation to combine two prior art references to form the claimed combination may demonstrate obviousness, but proof of obviousness does not depend on or require showing a teaching, suggestion, or motivation to combine;

Declaration of Zhi Ding, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 10,581,469

- market demand, rather than scientific literature, can drive design trends and may show obviousness;
- in determining whether the subject matter of a patent claim would have been obvious, neither the particular motivation nor the avowed purpose of the named inventor controls;
- one of the ways in which a patent's subject can be proved obvious is by noting that there existed at the time of invention a known problem for which there was an obvious solution encompassed by the patent's claims;
- any need or problem known in the field of endeavor at the time of invention and addressed by the patent can provide a reason for combining the elements in the manner claimed; "common sense" teaches that familiar items may have obvious uses beyond their primary purposes, and in many cases a person of ordinary skill will be able to fit the teachings of multiple patents together like pieces of a puzzle;
- a person of ordinary skill in the art is also a person of ordinary creativity and is not an automaton;
- a patent claim can be proved obvious by showing that the claimed combination of elements was "obvious to try," particularly when there

is a design need or market pressure to solve a problem and there are a finite number of identified, predictable solutions such that a person of ordinary skill in the art would have had good reason to pursue the known options within his or her technical grasp; and

- one should be cautious of using hindsight in evaluating whether a claimed invention would have been obvious.

57. I also have been informed that, in making a determination as to whether or not the claimed invention would have been obvious to a person of ordinary skill, the Board may consider certain objective factors if they are present, such as: commercial success of products practicing the claimed invention; long-felt but unsolved need; teaching away; unexpected results; copying; and praise by others in the field. These factors are generally referred to as “secondary considerations” or “objective indicia” of nonobviousness. I have been informed, however, that for such objective evidence to be relevant to the obviousness of a claim, there must be a causal relationship (called a “nexus”) between the claim and the evidence and that this nexus must be based on what is claimed and novel in the claim rather than something in the prior art. I have also been informed that even when they are present, secondary considerations may be unable to overcome primary evidence of

obviousness (e.g., motivation to combine with predictable results) that is sufficiently strong.

58. In sum, my understanding is that prior art teachings are properly combined where a POSITA at the time of the alleged invention having the understanding and knowledge reflected in the prior art and motivated by the general problem facing the inventor would have been led to make the combination of elements recited in the claims. Under this analysis, the prior art references themselves, or any need or problem known in the field of endeavor at the time of the invention, can provide a reason for combining the elements of multiple prior art references in the claimed manner.

59. I have been informed and understand that the obviousness analysis requires a comparison of the properly construed claim language to the prior art on a limitation-by-limitation basis.

IV. TECHNOLOGY BACKGROUND

A. Signals

60. Generally speaking, a signal is waveform that varies with time and/or space and is transmitted over a given medium such as coaxial cables or optic fibers. The type and format of a signal depend on the medium of transmission used in signaling. For example, a signal may be electric, optic, audio, or visual.

61. Signals may be periodic or non-periodic. Periodic signals show patterns that repeat over time. Non-periodic or aperiodic signals, on the other hand, do not. An example of a periodic signal is a musical note sound wave. It is a pressure wave (i.e., sound) consisting of sinusoidal components whose periods or frequencies define a particular note (e.g., B-flat). An example of a non-periodic signal is a radio signal propagated by an AM radio station for a talk show that is irregular in nature (i.e., spoken words constantly change and are not continuously repeated).

62. In radio frequency (RF) telecommunications, electric signals are functions in an electromagnetic wave form used for carrying information. Electric signals themselves are essentially the movement of electrons through a conductor medium, such as a copper wire. Oftentimes, electric signals are referred to as voltage signals (voltage over time) or current signals (current or ampere over time). Electric signals are commonly used to represent and transmit data for use in audio, video, speech, and images (e.g., audio signals and video signals) as well as specialized applications such as sonar and radar.

63. There are two main types of electric signals: analog and digital. Analog electric signals are continuous variations of voltage over time or current over time and are typically bound to a certain range (e.g., +5V to -5V). Analog signals use voltage, current, and/or frequency (and variations in the same) to represent

information, such as, for example, an electric current moving through a power cable. Analog signals typically are continuous in both time and values. Digital signals, on the other hand, can only take on (two or more) discrete values (or step values) for representing data as a sequence of values (e.g., 0/1, off/on, low/high, etc.). Digital signals are used in all digital electronics such as computer equipment and digital data transmitters. Binary digital signals are typically represented as voltage over time with a range including a positive value (e.g., +1.8V, +3.3V, or +5V) and a zero or negative value (e.g., 0V or -1.8V, -3.3V, -5V).

B. Nonlinear Signal Distortion

64. Signal distortion in electronics and telecommunications generally refers to any alterations to the original shape or other characteristics in an electric signal. The primary reasons for signal distortion in electric signals are environmental parameters that directly affect a signal,¹ the path complexity between a transmitter

¹ Examples of environmental parameters that cause signal distortion are flaws in electronic components and devices such as non-linear performance in components such as amplifiers and imperfect frequency response characteristics of electronic devices. Signal processing errors are another example of environmental signal distortion.

endpoint and receiver endpoint for a transmitted signal, and the properties and characteristics of the medium through which a signal travels through. Although there are exceptions in which controlled distortion is applied deliberately (e.g., noise reduction systems, musical effects with electronic instruments), signal distortion is generally undesirable and a phenomenon that electrical engineers strive to minimize or effectively mitigate.

65. Interference and noise are concepts related to unwanted signal distortion. Interference refers to electric signal disruptions that affect reception or processing. Interference is generally caused by external sources, such as multiple transmitters operating on a shared or an adjacent frequency band, electromagnetic radiation from electronic devices and components, and crosstalk or “splattering” from nearby communication channels. Noises refer to random signals introduced in circuits or channels during transmission, amplification, or processing often caused by thermal effects and electromagnetic interference. Noises can manifest as additive, multiplicative, or as random fluctuations in electric signal amplitude, phase, or frequency and interferes with transmission and processing. In general, distortion, interference, and noise all affect—most often undesirably—electric signals.

66. As stated above, distortion is generally an undesired effect. However, there are exceptions when controlled distortion is deliberately applied. For example,

noise reduction systems used in audio engineering deliberately distort audio signals to emphasize or strengthen specific characteristics of an electric signal that are subject to unwanted noise, such as noise introduced during transmission through a communication channel. The audio signal is then transmitted through the noisy environment, such as a noisy communication channel, and symmetrically distorted inversely after passing through. The end result is a received signal with a reduction in undesired noise. One such example is the Dolby sound system.

67. Distortion in electronics and telecommunications can be classified as linear and nonlinear. Linear distortion is a change in electric signal that can be modeled by a linear system, and is typically caused by multipath channels, signal reflections, and related phenomena in the electric signal propagation path in passive components such as coaxial cables and optical fibers. Nonlinear distortion, on the other hand, is caused by nonlinearities in the input/output relationship of active devices such as vacuum tubes, transistors, and amplifiers. Nonlinear distortions often increase the frequency (spectrum) range of the original signal.

68. An electrical component or system can be described mathematically by an ideal transfer function F below where output function $y(t)$ is a function of input signal $x(t)$:

$$y(t) = F(x(t))$$

Assuming a perfect gain constant A and perfect delay variable T under ideal conditions, an output function $y(t)$ is considered to be distortionless if :

$$y(t) = A \times x(t - T)$$

Such ideal distortionless output $y(t)$ provides a perfect replica of the input $x(t)$ which may be larger or smaller based on the gain A (like a larger or smaller replica of a photography) and may be generated with a fixed delay T because of the processing time needed by the distortionless system/device. A transfer function such as this can describe an ideal component such as a perfect linear amplifier. Perfect gain and delay, however, are merely ideal approximations and do not accurately describe how actual electrical components behave when considering imperfection. Distortion occurs when transfer function F deviates from the ideal form above. For example, time-varying gain constant A (amplitude) or imperfect delay variable T (or phase) will result in linear distortion. A change in the delay or gain that varies with the frequency components in input $x(t)$, will result in nonlinear distortion.

69. Focusing specifically on nonlinear signal distortion, nonlinear distortion occurs when a signal passes through a nonlinear medium or electrical component such as an amplifier. When the signal passes through such non-ideal devices/systems, the output signal is no longer a distortionless replica of the original input. In other words, using a nonlinear (imperfect) amplifier as an example, the

output signal is not an exact amplified replica (same shape) as the original input signal. This undesirably alters the variation of the original electrical input signal and hence the information therein. Nonlinear distortions often generate new frequency (spectral) components that were previously absent in the original input signal.

70. At small signal and amplification levels, linear system models are generally accurate in representing amplifier behavior. For example, an audio amplifier may receive an input 2V signal and output a 20V signal of the same shape. This means that the linear gain for the amplifier is 10. Assuming the same example amplifier, a 3V input signal would be amplified as an approximately 30V output signal and so on. However, linear models do not accurately characterize the behavior of most amplifiers at higher signal levels. For example, most audio amplifiers cannot linearly amplify a 50V signal to 500V.

71. Nonlinear distortion creates a variety of undesired effects. For instance, a non-uniformity of signal amplification due to nonlinear distortion (e.g., a 50 V input signal only being amplified to a 400 V output for a gain of 8 instead of a 5 V input being amplified to a 50 V output for the desired gain of 10) is a loss in amplification and leads to variation of the signal shapes (i.e., distorted signals). Nonlinear distortion also can cause spectrum distortion by generating spurious frequency components as a form of interference. Barring techniques to compensate

or mitigate distortion, this means an inaccurate information would be interpreted by the signal receiver. This can be particularly problematic for devices such as wireless computing and/or communication mobile devices.

72. Nonlinear distortion also introduces a problem called intermodulation. Intermodulation means the output signal has additional frequency components that were absent in the original input signal. These new frequency components now appear within the signal bandwidth of the output signal as a form of interference that drastically alters the information in the original signal to be communicated to the receiver. Intermodulation also has the undesired effect of expanding the bandwidth of a signal (i.e., expanding the upper and lower frequency limits or increasing the overall range). Because of this, amplifiers exhibiting significant nonlinear signal distortions behavior, in addition to altering the information of their own input signals, may cause their respective output signals to splatter onto adjacent radio frequency bands, thereby causing undesired interference on other radio signals in these adjacent user channels. For these reasons, radio frequency (RF) transmitter amplifiers for most telecommunications systems are desired to be linear such that they accurately amplify input signals for transmission over a large distance without distortion and without causing interference to other RF signals in adjacent user channels.

C. Signal Processing and Predistortion

73. Signal processing is a subfield of electrical engineering that focuses on analyzing, modifying, and synthesizing signals that convey information. Various forms of signal processing include analog signal processing for analog signals (i.e., most 20th-century radio, telephone, and television systems), and digital signal processing for digitized discrete-time sampled signals (i.e., signals consisting of signal samples; applies to most modern computer systems). Signal processing tools may be linear or nonlinear, time-invariant or time-varying, deterministic or statistical. Examples of signal processing applications include audio signal processing (e.g., signals representing audible information such as speech and music), image signal processing (e.g., signals representing visual information as used in digital cameras, computers, and imaging systems), video signal processing (e.g., interpreting image and sound information within electric signals by televisions, video players, etc.), and wireless communications (e.g., waveform generation, demodulation, filtering, and equalization).

74. Predistortion is a signal processing technique used to counteract unwanted nonlinear signal distortions before they take place. Simply put, predistortion improves the performance of electric devices/systems such as RF transmitter amplifiers to improve data integrity (e.g., maintain original input signal

information), power efficiency, and achieve requisite output power level (e.g., power amplification for sufficient geographical coverage) while minimizing signal splattering and interference caused by intermodulation.

75. Predistortion has been used for many decades as demonstrated by the work of Cooper and Morcom published in 1956. *See* Ex-1008. As an example, an electronic circuit that improves linear performance, such as a linearizer, is paired with a nonlinear component such as an amplifier. A linearizer introduces inverse distortion into the input of an amplifier to compensate for nonlinearity and reduce distortion. In addition to minimizing interference, predistortion gets more usable signal power out of an amplifier without having to use a larger, less efficient, and more expensive replacement. Predistortion can be implemented both in analog and digitally. One such design can be found in the work of Saleh and Salz. *See* Ex-1009.

D. Machine Learning and Deep Learning

76. Machine learning (“ML”) is a field of artificial intelligence (“AI”) that focuses on using datasets to train learning models and/or algorithms to generate desired outputs from inputted data. ML enables AI to imitate learning the way humans learn, by gradually improving learning accuracy based on training through more data and iteration cycles and thereby “learning.” ML examples include data classifiers or ML models used for classification and decision making. For example,

a decision tree is a well-known and relatively simple ML model that implements a tree-like structure as a predictive model to draw inference about observations. A more advanced and popular subset of machine learning is *deep learning* that implements neural networks that model a human brain.

77. Artificial neural network (ANN) models or neural network models have been developed and used in computers at least since the 1950s as shown in the work of Rosenblatt. *See* Ex-1010. Neural networks imitate the structure of the human brain and consist of interconnected nodes or neurons in a layered structure that relates inputs to desired outputs. Neural networks are trained by iteratively learning the neuron connectivities (i.e., parameters) so that the inputs correctly map to desired output responses during training. In addition to having input and output layers, neural networks also have hidden layers in between. The “deeper” neural networks boast more hidden layers. Neural networks are nonlinear models owing to the use of the so-called nonlinear activation functions. When given enough training data and sufficient numbers and layers of neurons, a neural network can be trained to successfully approximate a given function map between input and output data.

78. An exemplary neural network model used for deep learning is a convolutional neural network (“CNN”). CNNs are feed-forward models, meaning that information only flows forward from the input, and each layer receives the

previous layer's output data in one cycle. A recurrent neural network ("RNN"), another well-known neural network architecture, has feedback loops to allow information to flow back in addition to feedforward layers.

79. Neural networks are most commonly trained using backpropagation, a technique that uses an optimization algorithm to update model parameters before the output layer. Backpropagation computes the gradient (i.e., slope) 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 (i.e., to minimize a loss function) during training is typically used in backpropagation to adjust the model parameters. A common gradient descent type is Stochastic Gradient Descent, a variant that uses a single random training sample (or small batch) to approximate the gradient and update the parameters. Batch Gradient Descent, another common variant, uses the full training set.

V. THE '469 PATENT

80. The '469 patent purports to improve radio predistortion for correcting distortion and interference in radio signals using "a nonlinear pre-distortion machine learning model." '469 patent, Abstract (hereinafter, "NPDMLM"). The radio signal processing unit NPDMLM "includes at least one nonlinear function to correct one or more of radio signal distortion or interference." '469 patent, 3:7-9.

81. The NPDMLM receives a “transmit radio signal corresponding to an output of a transmitting radio signal processing system” as an input that has not yet been distorted or pre-distorted. ’469 patent, Abstract; *see id.*, Figures 6A (“Analog RF in 601”), 7A (“Digital RF In 701”) The NPDMLM “takes a transmit radio signal 104 as input and generates a pre-distorted signal 105 as output.” ’469 patent, 5:60-61. By processing the input radio signal, the NPDMLM “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 pre-distorted output signal from the NPDMLM is then exposed to a distortion source, such as an amplifier, and transmitted as a transmit output signal. “In some implementations, an amplifier 106 amplifies the signal to generate a post-distorted signal 108 that is received by a receiver 109. In other implementations, the receiver 109 may amplify and retransmit the signal.” *Id.*, 10:10-14. In addition, “the amplifier can be tightly integrated (e.g., within a handset) or loosely coupled (e.g., a tower mounted amplifier) within the transmitting radio signal processing system 101.” *Id.*, 5:36-39. Figure 1 of the ’469 patent, shown below, illustrates the overall system:

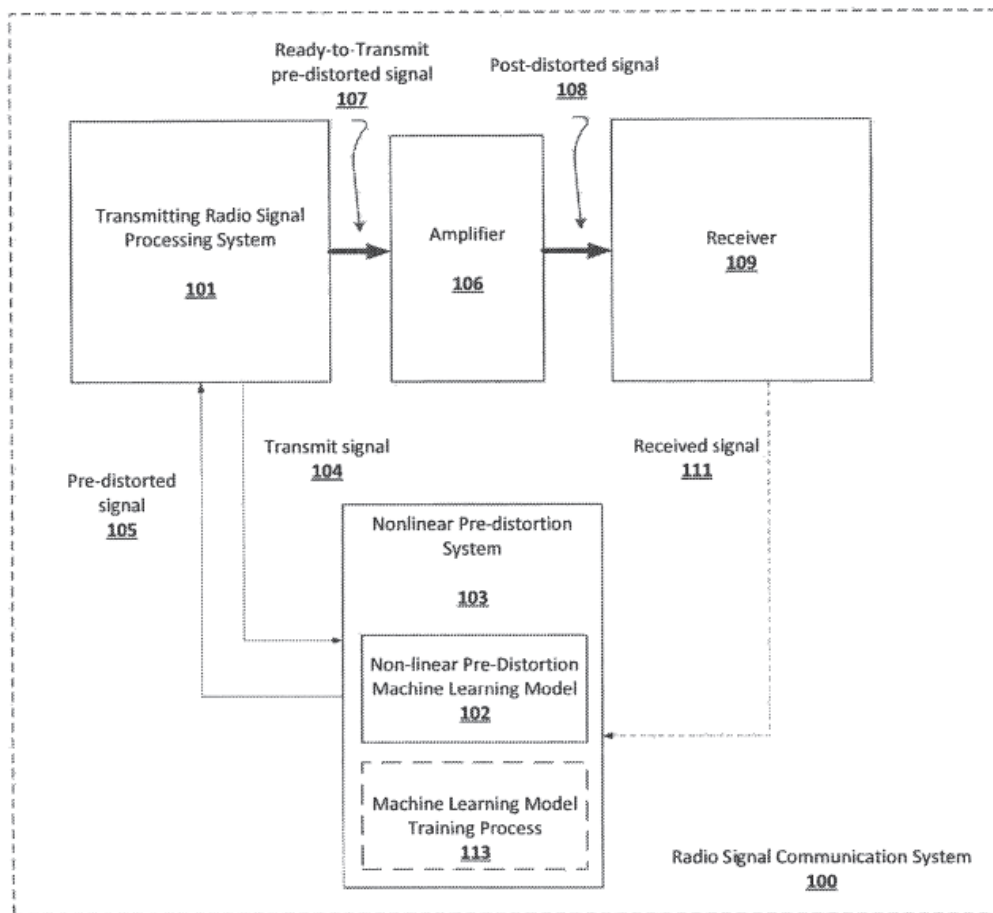


FIG. 1

'469 patent, Figure 1.

82. Everything previously described for the '469 patent was known in the art well before the critical date. The NPDMLM is a nonlinear neural network 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.*, 3:52-54. Nonlinear ML models, including artificial neural networks, were known and used years before the '469 patent. *See* Ex-1010. Adding ML models such as neural networks to predistortion was also not a new or inventive concept at the time of the

critical date. As will be discussed in more detail further below, prior art reference Jüschke explicitly discloses this concept. Indeed, the very first page of Jüschke discloses a “a pre-distortion neural network processor” used “for mitigating impairments due to transmit noise” and “for mitigating transmitter impairments of a transmit signal.” Ex-1005, Abstract.

83. The '469 patent also implements concepts for the NPDMLM parameters being trained using a “distance metric [] computed between the candidate transmit radio signal and the received radio signal.” '469 patent, 3:12-14. Distance metrics are numerical values used to calculate a similarity between data points in ML supervised and unsupervised learning and quantify how well an ML model output signal matches a desired output. Distance metrics help ML algorithms quantify the accuracy, and improve the performance, of learning tasks such as classification, clustering, and information retrieval.

84. The use of distance metrics in the '469 patent is consistent with the general understanding of distance metrics as used in ML. For example, “[d]etermining updated values for the model parameters may include using Stochastic Gradient Descent style optimizers that iteratively minimize the distance metric based on an exact or approximate gradient of the distance metric with regard to the model parameters.” '469 patent, 3:55-59; *see id.*, 7:36-38 (“...one of many

other potential forms of nonlinear activation function that works optimally to minimize a distance metric or loss function.”).² In other words, the ’469 patent determines a distance metric between expected data signal and the ML generated data signal (for example, a distance metric between the transmitted radio signal and a recovered radio signal at the receiver), and uses this distance metric in feedback to update the NPDMLM parameters (i.e., train the model). Using a distance metric (or loss function) to train ML models as described in iterative feedback was known in the art before the critical date.

85. Although not a concept claimed in the independent claims, the ’469 patent also implements a concept called “deployment scenarios.” These are general hardware configurations that the ’469 patent can be implemented in such as radio transmissions being sent from a cellular base station with the NPDMLM

² A loss function describes the difference between an ML model’s predicted output values and the true output values; a distance metric is used to quantify this difference. Common examples of loss functions are mean squared error (MSE), cross-entropy loss, and hinge loss. In my opinion, a POSITA would have understood loss functions and distance metrics as I have described.

implemented within a tower amplifier. The '469 patent briefly describes the deployment scenarios as follows:

The particular deployment scenarios may be one or: [i] sending a radio transmission from a cellular base station that includes the nonlinear pre-distortion machine learning model inside a tower mounted amplifier in an inline mode; [ii] sending a radio transmission from a cellular base station that includes the nonlinear pre-distortion machine learning model inside a tower mounted amplifier in a fault-tolerant subtractive mode; [iii] sending a radio transmission from a cellular base station that includes the nonlinear pre-distortion machine learning model inside a remote radio unit in an inline mode; [iv] sending a radio transmission from a cellular base station that includes the nonlinear pre-distortion machine learning model inside a remote radio unit in a fault-tolerant subtractive mode; [v] sending a radio transmission to a mobile handheld device that includes the nonlinear pre-distortion machine learning model, which obtains model parameter updates off a hardware chip and provides pre-distorted signals on the hardware chip or off the hardware chip; [vi] sending a radio transmission to a mobile handheld device that includes the nonlinear pre-distortion machine learning model, which obtains model parameter updates on-chip and provides pre-distorted signals on a hardware chip; [vii] sending a radio transmission to a backhaul or satellite relay radio system and using feedback from one or more amplifiers to determine model parameters of the nonlinear pre-distortion machine learning model; or [viii] sending radio transmission to a system that includes interference cancellation.

Ex-1005, 2:38-65.

86. A “tower mounted amplifier” ([i] and [ii]) refers to, as its name suggests, an amplifier installed in a tower (or a mast) or tower-like structure commonly used in radio communication systems such as cellular network communications. In my opinion, a POSITA would have understood that a tower mounted amplifier is an amplifier that is typically mounted closely to the antenna of base stations (such as macro-cell base stations, a type of cellular base stations cellular base station that sends and receives radio signals over a large coverage area) used in radio communication systems. Tower mounted amplifiers boost the sensitivity of base stations to improve overall network performance. By boosting the sensitivity of a base stations (e.g., ability of a base station receiver to detect radio signals), a tower mounted amplifier helps to improve the uplink budget of the base station for better capacity (e.g., total traffic a base station can handle such as number of users and peak load) and coverage (e.g., connectivity provided by the base station in terms of signal strength and distance). Tower mounted amplifiers are also commonly referred to as mast head amplifiers, in reference to mobile masts (also referred to as cellular towers). In my opinion, a POSITA would have been familiar with all of these concepts in radio communication systems and supporting infrastructure.

87. A “remote radio unit” (“RRU”) ([iii] and [iv]), also known as a remote radio head (“RRH”), refers to a radio frequency (RF) transceiver component, typically collocated with an antenna of an extended base station, that has a direct high-speed connection to an operator radio control panel or a baseband unit via a fast broadband connection such as fiber optic. It transmits RF signals to, and receives RF signals from, wireless devices (e.g., cellular phones) in a wireless network (e.g., cellular network such as a Long Term Evolution 4G network). Remote radio units are connected to and work together with another component called a baseband unit (“BBU”)³ of the radio control panel. Remote radio units are responsible for generating and extracting radio frequency signals which they can amplify, transmit, receive, and filter. In my opinion, a POSITA would have been familiar with remote radio units as part of a POSITA’s general knowledge of radio communication systems and supporting infrastructure. As an example, the work of Wang, Hossain, and Bhargava discusses a brief history of certain base station infrastructure including

³ A baseband unit (BBU) is a component in base stations that process baseband signals, signals carrying the original frequencies of transmission signals before modulation or alteration. Baseband units are typically located in a base-station room under a base station tower.

a “the separated baseband unit (BBU)-remote radio unit (RRU) structure of BS design to provide more flexible coverage and capacity solutions with improved processing resource utilization.” Ex-1013, 42. In addition, the work of Lanzani, Kardaras, Boppana discusses remote radio units (remote radio heads in this example) as part of a cellular base station infrastructure. *See* Ex-1014. A snippet from Lanzani is provided below as an illustrative example. Therefore, it is my opinion that a POSITA also would have been familiar with the concept, and the role played by, remote radio units (or remote radio heads) as part of radio communication systems and supporting infrastructure.

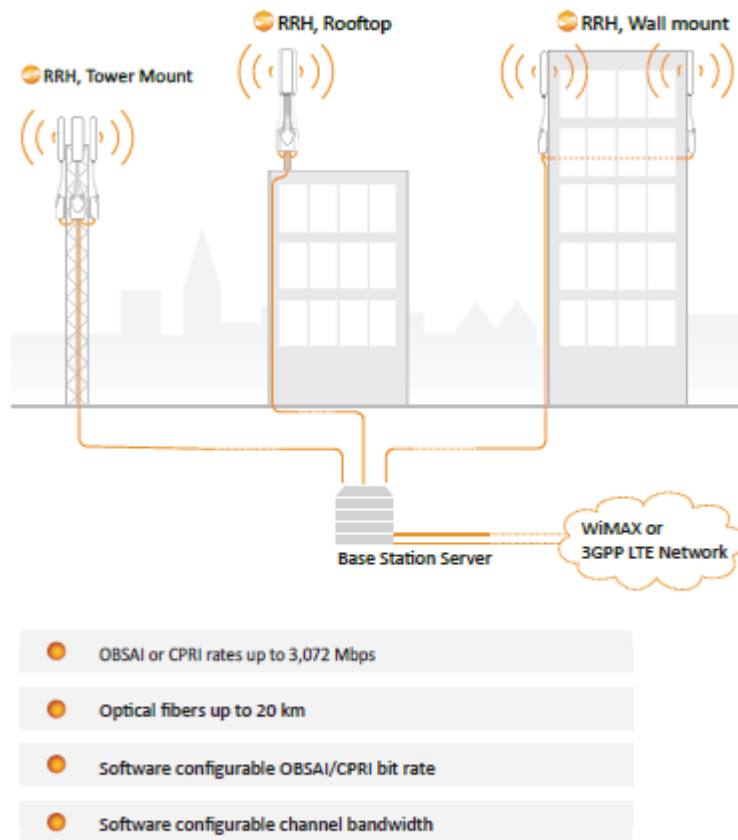


Figure 1: Distributed Wireless Base Station system.

Ex-1014, Figure 1.

88. A brief look at the two “inline mode[s]” ([i] and [iii] above) demonstrates that the ’469 patent’s deployment scenarios are technically basic hardware configurations. For configuration deployment [i], “FIG. 6A illustrates an example configuration for in-line mode within a tower-mounted amplifier, placing a nonlinear pre-distortion system 103 between an input 601 and an amplifier 603.”

'469 patent, 10:38-41. FIG. 6A with red box annotations below highlights analog input 601, nonlinear pre-distortion system 103, and amplifier 603. As illustrated below, "in-line mode" simply means input signal and components are in a tandem configuration. FIG. 6A is shown below:

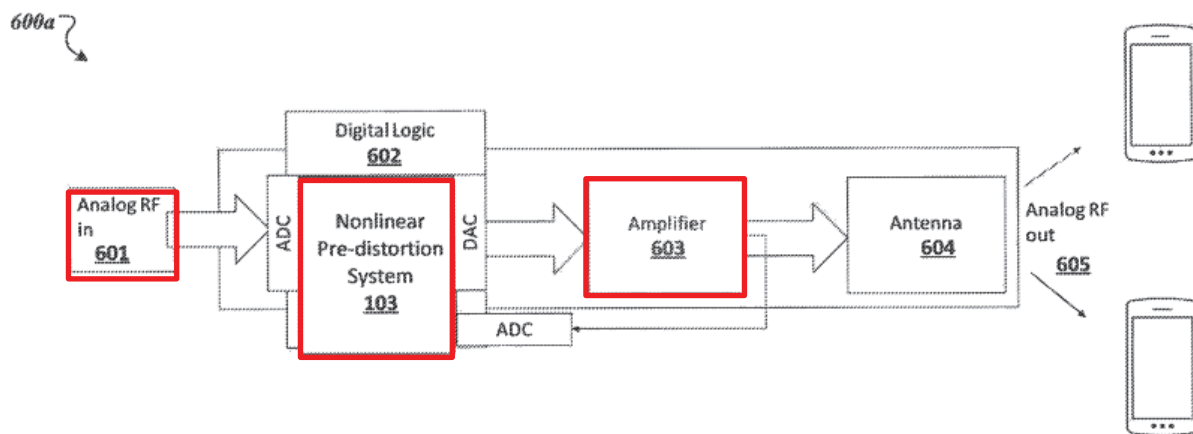


FIG. 6A

'469 patent, Figure 6A, (annotations added).

89. Configuration deployment [iii] is similar to configuration deployment [i] but for the NPDMLM now being implemented in a remote radio unit (e.g., remote radio unit for a cellular base station). "FIG. 7A illustrates a configuration for in-line mode within a remote radio unit, with the nonlinear pre-distortion system 103 configured between the input 701 and the amplifier 703." '469 patent, 10:47-50.

FIG. 7A with red box annotations below highlights digital input 701, nonlinear pre-distortion system 103, and amplifier 703. Again, the input signal and components are in a linear configuration. FIG. 7A is shown below:

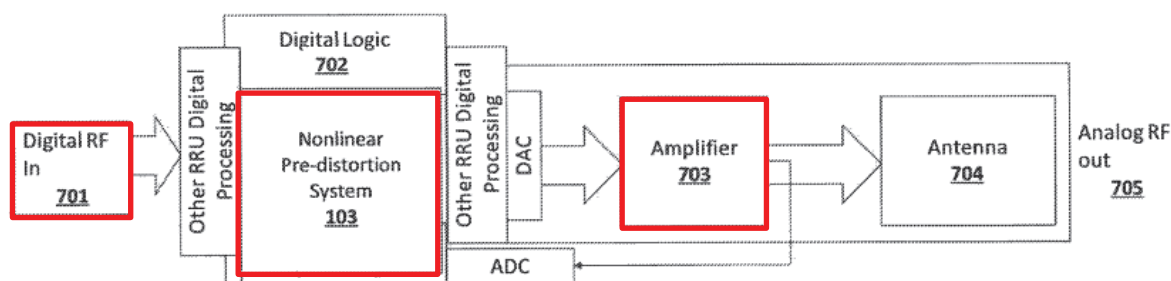


FIG. 7A

'469 patent, Figure 7A (annotations added).

90. A look at the two “fault-tolerant subtractive mode[s]” ([*ii*] and [*iv*] above) further demonstrates that the '469 patent's deployment scenarios are basic hardware configurations. For configuration deployment [*ii*], “FIG. 6B illustrates an example configuration for fault-tolerant subtractive mode within the tower-mounted amplifier, with the nonlinear pre-distortion system 103 configured between an RF splitter 606 and an analog subtract 607.” '469 patent, 10:42-45. FIG. 6B with red box annotations below highlights analog input 601, to RF (radio frequency) splitter 606, to analog subtract 607, with nonlinear pre-distortion system 103 in a feedback

loop. This configuration is used for “comput[ing] a difference signal between the pre-distorted signal and the input signal (e.g. for subtractive pre-distortion).” *Id.*, 7:51-53. In short, the inline configuration previously discussed for [i] has now been modified to implement a feedback system for subtractive predistortion. FIG. 6B is shown below:

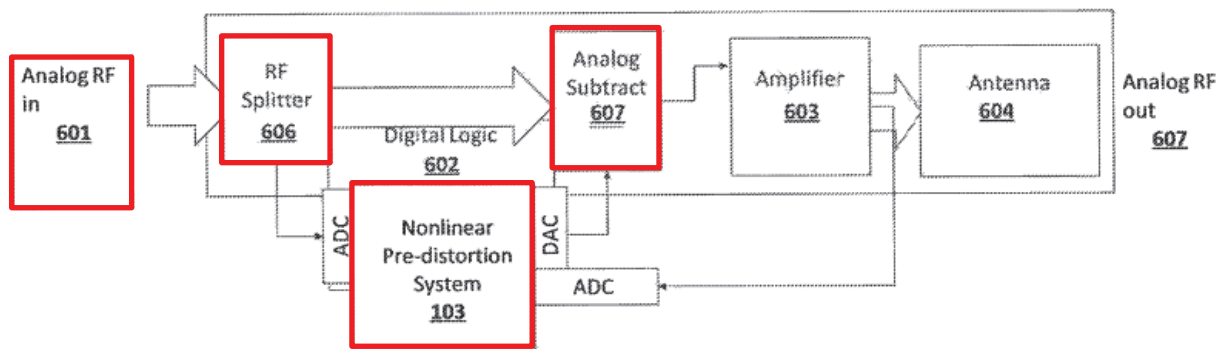


FIG. 6B

'469 patent, Figure 6B (annotations added).

91. Configuration deployment [iv] is similar to configuration deployment [ii] but for the NPDMLM again being implemented at a remote radio unit. “FIG. 7B illustrates an example configuration for fault-tolerant subtractive mode within the remote radio unit with the machine learning model 102 and the machine learning model training process 113 separated.” '469 patent, 10:50-54. FIG. 7B is shown below with red box annotations highlighting digital input 701 to analog subtract 707,

with ML model training process 113 and nonlinear pre-distortion ML model 102 in a feedback loop:

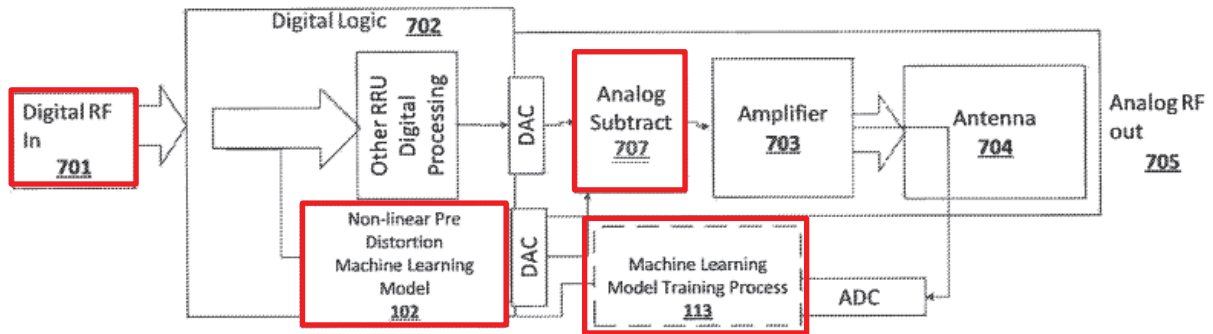


FIG. 7B

'469 patent, Figure 7B (annotations added).

92. Another configuration deployment [vii] is for a “system that includes interference cancellation.” “This configuration uses an equivalent network which modifies a received radio signal 912 and passes it through a distortion/interference removal network 905 in order to remove artifacts resulting from an interference source which may be sensed at another antenna 911.” '469 patent, 11:25-30. The system uses a distance metric “to minimize or remove any dependent signal content from the interference signal 911 present within the received signal 912.” *Id.*, 11:31-33. FIG. 9 is shown below with red box annotations highlighting digital input 903,

digital output 912, nonlinear pre-distortion system 103, removal network 905, and
monitor antenna 911:

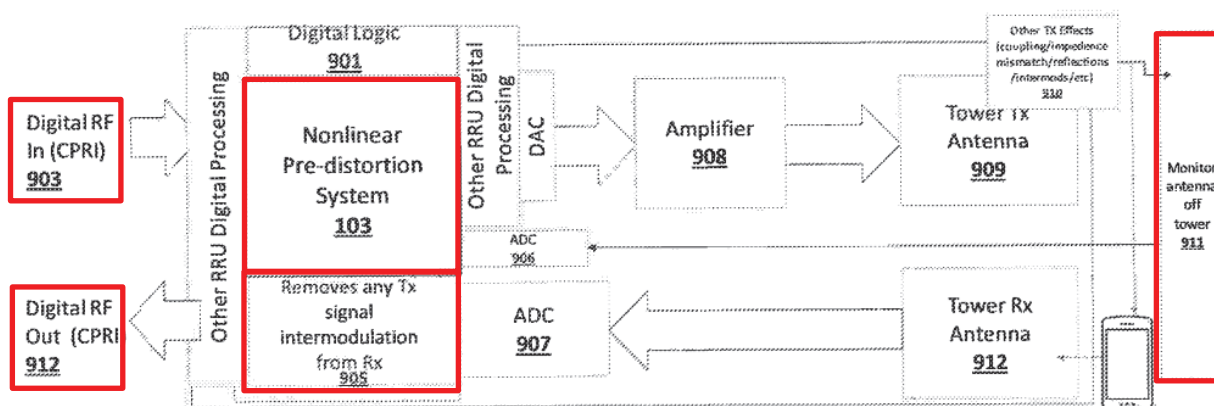


FIG. 9

'469 patent, Figure 9 (annotations added).

A. The '469 Patent's Prosecution History

93. The Examiner found independent claims 1, 21, and 23 anticipated. Ex-1002, 52. To distinguish from the prior art, the Applicant added amendments directed to the NPDMLM's model parameters as follows: "updating the model parameters based on a distance metric computed between the transmit radio signal and a received signal that is obtained from the one or more radio receivers, the received signal corresponding to the transmit output radio signal that was transmitted to the one or more radio receivers." *Id.*, 43-49.

VI. OVERVIEW OF THE PRIOR ART REFERENCES

94. As shown below, the references (Jüschke, Holt, Dzierwa) all implement machine learning in wireless communications and signal processing. Hence, a POSITA would have understood that their teachings are related to highly related to each other and many, if not all, of their techniques were interchangeable. A POSITA also would have understood that machine learning, and in particular neural networks, were highly adaptable to solving numerous challenges in signal processing such as distortion, noise, and environmental optimization.

A. Jüschke

95. Jüschke was published more than four years before the effective filing date of the '469 patent. Jüschke focuses on the same problem as the '469 patent (transmit distortion such as power amplifier PA nonidealities and transmit TX noise in radio signals) and the same solution (predistortion using nonlinear ML models such as neural networks).

96. Jüschke teaches “concept[s] for mitigating impairments due to transmit noise” and “impairments of a transmit signal due to transmitter imperfections” in “wireless communications.” Ex. 1005, Abstract, [0001]. Examples of “prominent non-idealities,” i.e., transmission distortion, noise, and interference, that Jüschke mitigates include “carrier-frequency and sampling-rate offset, phase-noise, IQ-

imbalance (I = In-phase signal component, Q = Quadrature signal component), and Power Amplifier (PA) nonlinearities.” Ex. 1005, [0004]. Jüschke refers to “the combination of all those impairments at a transmitter...as transmit noise (TX noise).” *Id.* “[T]ransmit noise may be interpreted as a deviation of a physical, i.e. real-world RF transmit signal from an ideal transmit signal constellation due to the aforementioned transmitter non-idealities.” *Id.*

97. Carrier frequency offset is the mismatching error between the frequency of a received signal and the frequency of a local oscillator (LO)⁴ at a receiver. Sampling-rate offset refers to the deviation of sampling clock rate from the pre-determined sampling rate. Phase noise refers to random phase fluctuations in a signal’s frequency components. IQ imbalance is a form circuitry error in direct conversion radio receivers⁵ in which the in-phase signal component (I or real) and quadrature signal component (Q or imaginary) have mismatched phase (degrees)

⁴ A local oscillator is an electrical oscillator component used with a mixer to change the frequency of a signal.

⁵ Direct conversion receivers directly translate received radio frequency signals (RF or passband) from a carrier frequency (frequency of a wave that is modified by an information-carrying signal) to a digital baseband signal using a mixing stage.

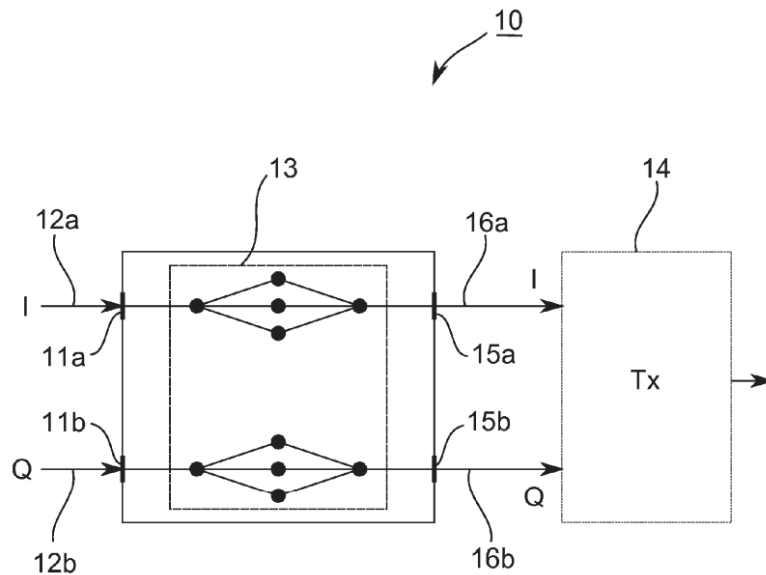
and/or amplitude. It is commonly caused by nonideal frequency mixers. Power amplifier (PA) nonlinearity refers to nonlinear output signal distortions from nonlinear (imperfect) amplifiers. Jüschke goes on to explain that “the mixer after the digital to analog converter, which is responsible for mixing the baseband signal with the carrier signal, may cause so called IQ imbalance imperfections.” Ex. 1005, [0004]. Moreover, “non-linear behavior of the PA characteristic close to the saturation point adds impairments to the transmit signal...result[ing] in sideband emission and affect[ing] spectrum emission requirements.” *Id.*

98. Jüschke further explains that although “TX noise [at the transmitter] may be at least partially compensated by means of calibration or pre-distortion, or at the receiver using sophisticated compensation algorithms,” that “such classical algorithms are quite fixed and, hence, not really flexible and a bottle-neck, in particular for increasingly complex mobile communication system architectures.” Ex. 1005, [0005]. “Therefore, it is desirable to have a more flexible radio transmitter that can support as many standards and signals as possible.” *Id.*, [0006]. “This is not only true for macro-cell base stations but also for small-cell base stations.” *Id.*

99. As part of its systems and techniques for “mitigating impairments due to transmit noise” and “transmitter imperfections,” Jüschke’s flexible radio transmitter implements a “pre-distortion neural network processor” that pre-distorts

an “input digital baseband transmit signal” and generates “a pre-distorted digital baseband signal.” Ex. 1005, Abstract, [0001]. The pre-distortion neural network processor sends the pre-distorted signal to an analog transmitter device for transmission. *Id.*, [0010]. The analog transmitter device includes a cause of distortion (e.g., an amplifier) that the pre-distortion neural network processor compensates for using a transfer function that “approximates an inverse of [the analog] transmitter [device’s] transfer function.” *Id.* The “pre-distortion neural network processor may comprise a multi-layer feed forward neural network (MLNN), also referred to as multi-layer perceptron (MLP).” *Id.*, [0013].

100. Figure 1 below illustrates Jüschke apparatus 10 and input digital baseband transmit signal (12), pre-distortion neural network processor (13), pre-distorted digital baseband signal (16), and analog transmitter device (14) arranged downstream from pre-distortion neural network processor (13):



Ex-1005, Figure 1.

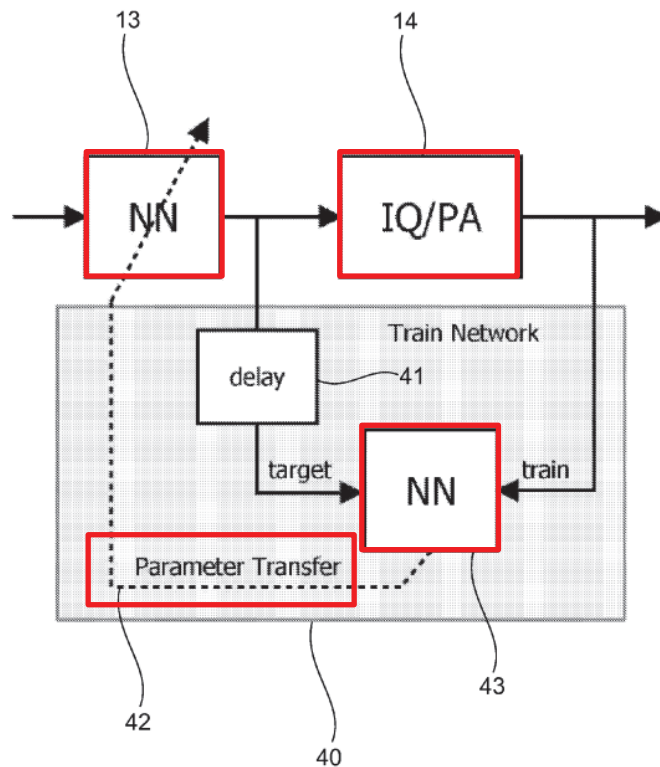
101. Furthermore, Jüschke teaches training its neural networks at a transmitter by analyzing undistorted input and sampled output signals. *See* Ex. 1005, [0047] (“Fig. 4 illustrates how...the pre-distortion neural network processor 13 may be trained at a transmitter.”). Jüschke implements a “a training processor 40 for training weight (w_k) and/or bias parameters (b_k) of neurons of the pre-distortion neural network processor 13,” which “may comprise a post distortion neural network 43 having the same network structure as the pre-distortion neural

network 13.” *Id.*, [0048].⁶ “Sampled output values of the non-ideal analog transmitter device 14” and “ideal digital baseband IQ samples 12a, 12b may be provided to the post distortion neural network 43 as target output values.” *Id.* After training, the weight and bias parameters are stored and “transferred to the pre-distortion neural network processor 13.” *Id.*, [0049].

102. Figure 4, shown below with red boxes added, illustrates Jüschke’s pre-distortion neural network processor (13) (with “NN” referring to a neural network), analog transmitter device (14) (with “IQ” referring to in-phase/quadrature

⁶ I note that Jüschke refers to item 13 as both a pre-distortion neural network processor 13 (*see* Ex. 1005, Abstract, [0039]-[0042], [0045]-[0058], [0068], [0070]) and a pre-distortion neural network 13. *See* Ex. 1005, [0048]-[0049], [0058]. In my opinion, a POSITA would have understood that a pre-distortion neural network 13 is implemented, either through hardware, software, or a combination, on or by a pre-distortion neural network 13 processor. Therefore, a POSITA would have understood that both pre-distortion neural network processor 13 and pre-distortion neural network 13 are accurate and can be referred to interchangeably in many circumstances except when the physical attributes of the processor are relevant (e.g., physical installation).

components of a radio signal and “PA” referring to a power amplifier), post distortion neural network (43), and the trained parameters being transferred (42) from the post distortion neural network (43) to the pre-distortion neural network processor (13):



Ex-1005, Figure 4 (annotations added).

B. Holt

103. Holt was filed four months before the effective filing date of the '469 patent. It focuses on training nonlinear machine learning models, including neural networks, implemented in wired and wireless communication systems.

104. Holt teaches “systems and methods that enable adaptive training of a channel coding model including an encoder model, a channel model positioned structurally after the encoder model, and a decoder model positioned structurally after the channel model.” Ex-1006, Abstract. The models can be “neural network[s]...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, which can be in between the encoder model and decoder model, can represent a “wireless channel (e.g., a radio frequency (RF) link).” *Id.*, 5:18-19.

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.* The channel model, which can be placed in between the encoder model

and the decoder model, can “emulate a communication channel, for example, by training the channel model on example data that has been transmitted via the communication channel.” *Id.*, 4:5-8. Holt’s channel model also accounts for “the addition of noise to an output of the encoder model.” *Id.*, 5:47-48. In addition, in Holt, an “actual communication channel” can also be used instead of a channel model. *Id.*, 17:65-18:2 (“[D]ecoder model 212 can be further trained...to learn how to decode data coming out of the communication channel 302 (e.g., the actual communication channel 302.”). Therefore, in my opinion, a POSITA would have understood that Holt’s encoder and decoder models could be implemented in conjunction with either channel model or actual communication channel.

106. Holt Figure 3 is shown below illustrating a channel coding model (200) exemplifying a first set of inputs (204), an encoder model (202), a first set of outputs (206), a channel model (208), a second set of outputs (210), a decoder model (212), a third set of outputs (214), and a first loss function (216) for the first set of inputs (204) and the third set of outputs (214):

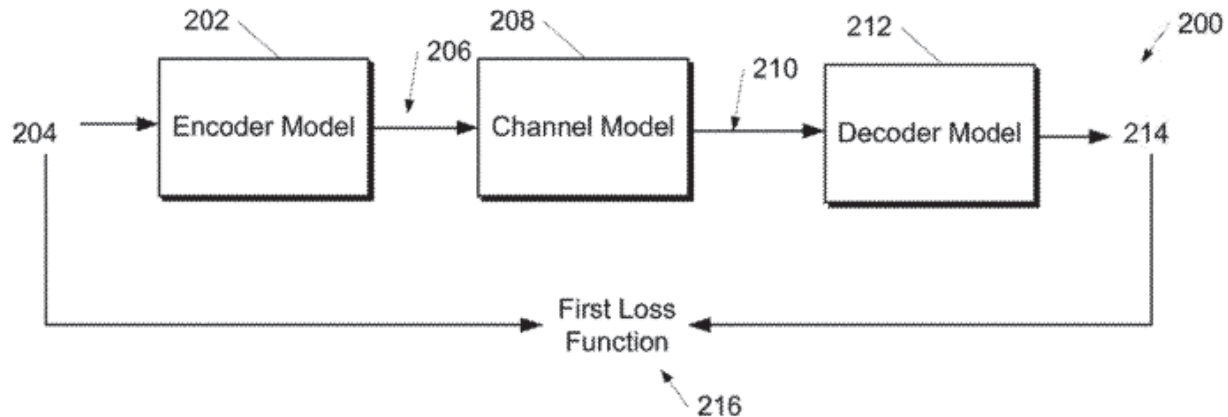
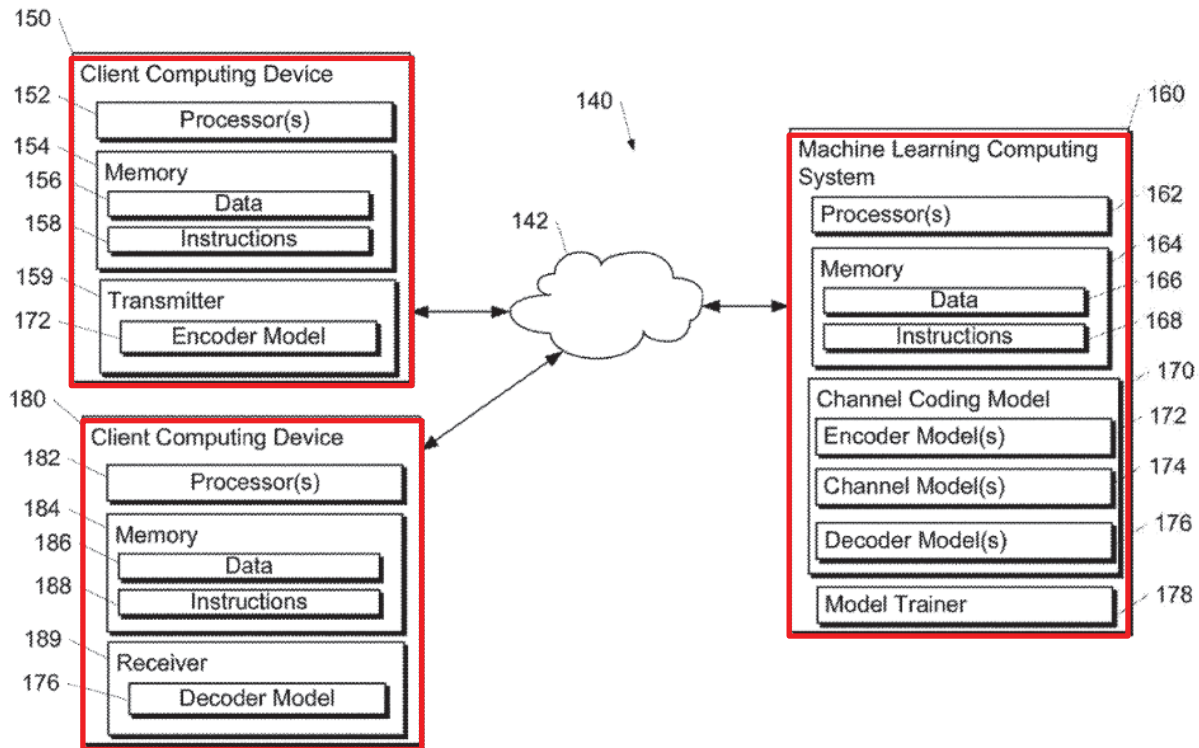


Figure 3

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

107. In addition, Holt’s system and models can be implemented within one or more of “any type of computing device, including a personal computer (e.g., desktop or laptop), a mobile computing device (e.g., smartphone or tablet), an embedded computing device, a server computing device, a network computing device such as a base station, router, beacon, other communication node, or other forms of computing devices.” Ex-1006, 9:34-40. As shown in Figure 2 below, Holt’s channel coding model(s), including encoder model, channel model, and decoder model, are implemented on computing devices 150, 160, and 180:



Ex-1006, Figure 2 (annotations added).

C. Dzierwa

108. Dzierwa was filed about three months before the effective filing date of the '469 patent and teaches radio signal detection and optimization using machine learning techniques. Dzierwa uses machine learning “to observe and learn an RF environment over time and identify 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” or “not typical.” *Id.*, 47:43-46. This enables detection of RF signals with “low power level[s]” and “narrowband

signal[s] buried in a wideband signal, and which cannot be identified otherwise.”

Id., 9:24-26. Dzierwa also references numerous patents and publications within its body.

VII. OPINIONS REGARDING THE '469 PATENT'S INVALIDITY

109. I have reviewed and analyzed the prior art references and materials cited herein. In my opinion the claims of the '469 patent are rendered obvious based on the following prior art:

- **GROUND 1:** Claims 1-7, 8, 10, 12-15, 17-20, 21-22, and 23-24 are obvious over Jüschke in view of Holt under 35 U.S.C. §103.
- **GROUND 2:** Claims 2-5, 22, and 24 are obvious over Jüschke in view of Holt and Dzierwa under 35 U.S.C. §103.

VIII. GROUND 1: CLAIMS 1-8, 10, 12-15, AND 17-24 WERE OBVIOUS OVER JÜSCHKE IN VIEW OF HOLT.

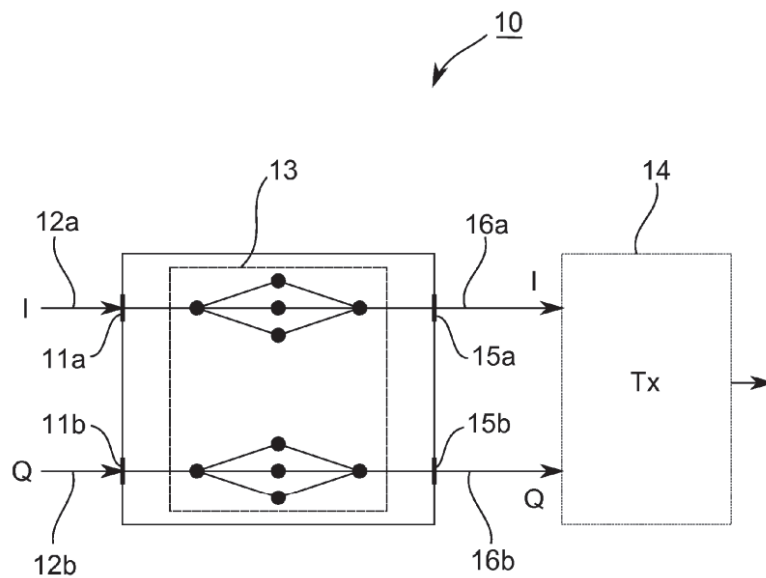
110. As I will show below, in my opinion, Jüschke in view of Holt teaches claims 1-8, 10, 12-15, and 17-24. *See* Sections VIII.D-W, *infra*.

A. Description of Jüschke

111. Jüschke’s predistortion techniques implement neural networks “for mitigating impairments due to transmit noise” and “for mitigating transmitter impairments of a transmit signal,” Ex-1005, [Abstract], using a “more flexible radio transmitter that can support as many standards and signals as possible.” *Id.*, [0006]. For instance, Jüschke’s transmitter can be implemented in systems and devices for “mobile communications” such as “2nd generation cellular communication systems, like e.g. GSM (Global System for Mobile Communications),” an example of a 2G wireless network, and “3rd and 4th generation mobile communication systems, like e.g. UMTS (Universal Mobile Telecommunication System) and LTE (Long-Term Evolution),” examples of 3G and 4G wireless networks. *Id.* [0002]; *see id.*, [0059] (“mobile communication systems, like e.g. GSM [cellular 2G], WCDMA [Wideband Code Division Multiple Access; cellular 3G] and/or LTE [cellular 4G]”).

112. Figure 1 below illustrates Jüschke apparatus 10. As shown, “an in-phase component 12a and a quadrature component 12b of a digital baseband transmit signal 12...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]. I note that a POSITA would have understood that the “quadrature components” generally refer to both the in-phase component and the quadrature component in RF modulations. The pre-distorted signal (16, 16a, 16b) is then sent to analog transmitter device 14 for transmission where it is subjected to distortion (e.g., by an amplifier or a power amplifier PA) and transmitted as an analog radio signal. *See id.*, [0050] (“analog amplifier device in the transmitter device 14”). See Figure 1 below:



Ex-1005, Figure 1.

113. Because a POSITA would have understood that transmitted radio signals are analog signals, a POSITA would have further understood that any digital

signals received by analog transmitter device 14, such as pre-distorted digital baseband signal 16, would need to be converted using a digital-to-analog converter component prior to being transmitted as a radio signal. Therefore, a POSITA would have understood that analog transmitter device 14 has a digital-to-analog converter in order to transmit the pre-distorted digital baseband signal 16. *See* Ex-1005, [0011] (“The at least one analog transmitter device may comprise an IQ up-conversion mixer...[s]uch an IQ up-conversion mixer up-converts IQ baseband signals transmitted from an IQ Digital-to-Analog Converter (DAC) to Radio Frequency (RF) IQ signals.”)

114. Jüschke’s “(artificial) neural networks (NNs) have high flexibility and advantages regarding solving (non-) deterministic and complex problems,” Ex-1005, [0008], and can be used to mitigate “transmitter impairments of a transmit signal” such as “IQ 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.” *Id.*, [0009]. For IQ imbalance, “as few as six nodes or neurons, i.e. three nodes or neurons (one input node, one hidden neuron, one output node) for each quadrature component 12a, 12b” can be used. *Id.*, [0046]. PA linearization, on the other hand, “is more complex than IQ imbalance compensation due to the nonlinear behavior of the power amplifier.” *Id.*, [0050].

115. Jüschke's pre-distortion neural network processor approximates an "inverse" of the distortive transmitter transfer function to compensate for the nonideal analog transmitter impairments. Ex-1005, [0039]-[0040]. "The pre-distortion neural network processor 13 is [] adaptable or adapted to a transmitter transfer function of at least one analog transmitter device 14 arranged downstream to the pre-distortion neural network processor 13 such that a transfer function of the pre-distortion neural network processor 13 approximates an inverse of said transmitter transfer function." *Id.*; *see also id.*, [0058] ("[P]ost distortion neural network 43 approximates an inverse of the transmitter transfer function, including a power amplifier's AM-AM and/or AM-PM distortion functions, as well as an IQ up-conversion mixer's transfer function.").

116. For training the pre-distortion neural network 13, Jüschke compares input and output signals in a feedback loop and updates the weight and bias parameters of a post-distortion neural network 43 having the same network structure as the pre-distortion neural network 13 using a training processor. "The training processor 40 may comprise a post distortion neural network 43 having the same network structure as the pre-distortion neural network 13 with respect to layers, nodes and neurons." Ex-1005, [0048]. "Sampled output values of the non-ideal analog transmitter device 14 may be fed back as input samples for the post distortion

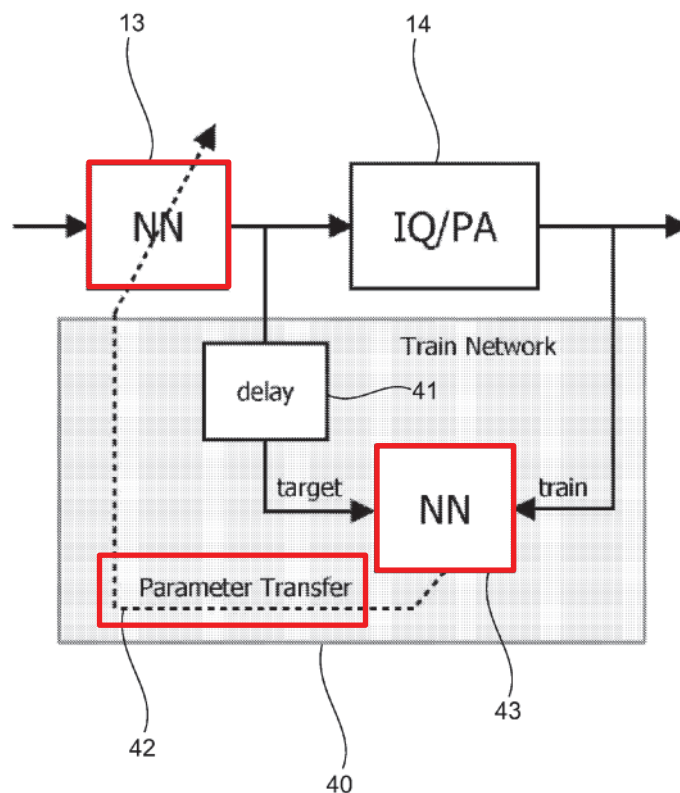
neural network 43 [within training processor 40]. Undistorted and, hence, ideal digital baseband IQ samples 12a, 12b may be provided to the post distortion neural network 43 as target output values...as training data.” *Id.*

117. For training its “weight (w_{kj}) and/or bias parameters (b_k),” Jüschke teaches the use of “the Levenberg-Marquardt training algorithm, such that a transfer function of the post distortion neural network approximates an inverse of the transmitter transfer function when the training has ended.” *Id.*, [0049]. Jüschke also teaches that “[g]radient-based training algorithms, like back-propagation, may be used.” Ex-1005, [0018]. Jüschke clarifies that “training may take several epochs and may be terminated by MSE (Minimum Squared Error).” *Id.*, [0048]. “After convergence of the post distortion neural network 43,” its learned weight and bias parameters through training are “transferred to the pre-distortion neural network processor 13.” *Id.*, [0049]. In my opinion, a POSITA would have understood that the Levenberg-Marquardt algorithm has been used in backpropagation training of neural networks as demonstrated, e.g., by the work of Sapna, Tamilarasi, and Kumar. *See* Ex-1011, 394-397. I note that Sapna was published a little over a year after Jüschke was originally filed as an EP patent application.

118. Jüschke further teaches training techniques that are flexible and can “be repeated at subsequent time instants in order to adapt the pre-distortion neural

network processor 13 to time varying transmit transfer functions of the non-ideal analog transmitter device 14.” *Id.* For example, the pre-distortion neural network can “be used for online signal processing while the second [post distortion] neural network 43 is needed for live adaptation of the pre-distortion neural network 13.”

Id. Figure 4 below illustrates the parameters being trained and transferred:

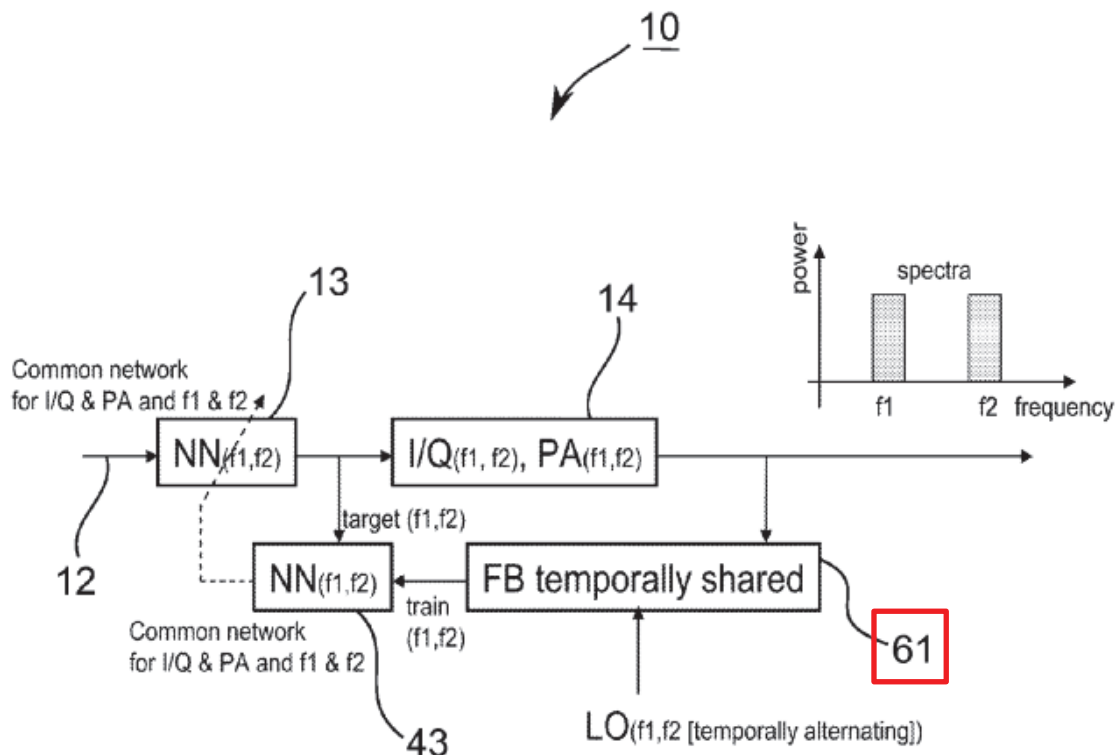


Ex-1005, Figure 4 (annotations added).

119. Furthermore, Jüschke teaches adaptive “self learning/optimizing neural network” that “consider[s] and optimiz[es]...the 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).” 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. The post distortion neural network processing unit 43 uses the feedback signal for enabling adaptive pre-distortion of the baseband signal, the pre-distortion being reciprocal to the distortion caused by the transmitter device's 14 non-ideal components, thus allowing for linearization of the amplification, for example." *Id.*; *see also id.*, [0028] ("Information of the signal quality can e.g. be provided to the neural network by use of a feedback path."), [0072] ("The neural network may learn from a feedback signal and adapt itself to achieve required results."). Figure 6 below

demonstrates feedback path 61 used by Jüschke for adaptive predistortion:



Ex-1005, Figure 6 (annotation added).

120. As previously discussed, Jüschke's teaches a "more flexible radio transmitter that can support as many standards and signals as possible." Ex-1005, [0006]. Jüschke explains that "a high flexibility within the transceiver system has the advantage, that different parts of the system can be exchanged and adapted easily, like the flexibility to use the digital part with different analog components without reconfiguration." *Id.* Being taught as highly flexible, in my opinion, a POSITA would have understood that Jüschke's flexible radio transmitter can be implemented

in a variety of deployments scenarios for “mobile communication systems, like e.g. GSM [2G cellular], WCDMA [3G cellular] and/or LTE [4G cellular].” *Id.*, [0059]; *see id.* [0002].

121. For instance, Jüschke teaches that its flexible transmitter can be implemented in a base station for wireless communications, such as “macro-cell base stations” and “small-cell base stations” used in cellular communication systems. Ex-1005, [0006]. In my opinion, a POSITA would have understood that a macro-cell base station is a cellular base station that sends and receives radio signals through large towers and antennas. In my opinion, another flexibility in a transmitter configuration may involve the deploying the transmitter antennas collocated with the signal processing and predistortion unit in centralized way or deploying the transmitter antennas (and the associated power amplifiers) in a remote radio unit (also known as remote radio head) that is directly connected to the signal processing and predistortion unit in a distributed way. Also in my opinion, a POSITA would have understood that a small-cell or micro-cell base station is a low-power cellular base station covering a limited area such as malls, hotels, office buildings, or transportation hubs.

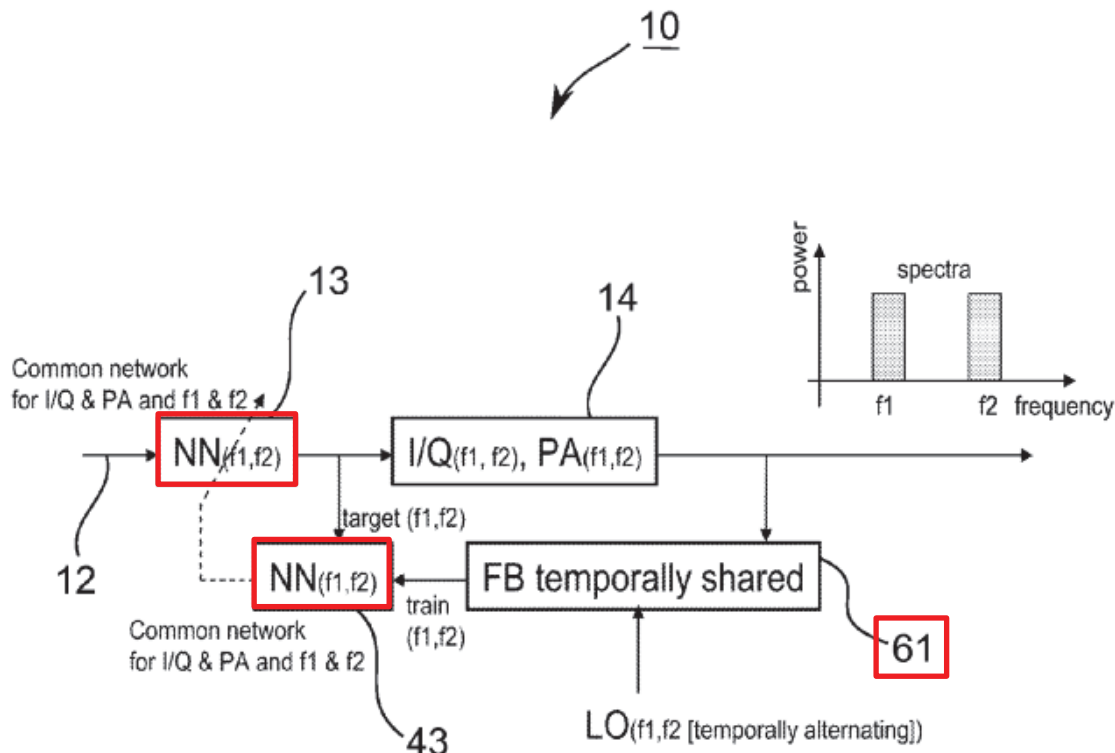
122. Moreover, given Jüschke’s emphasis on a “more flexible radio transmitter” with “a high flexibility within the transceiver system,” it is my opinion

that a POSITA also would have understood that Jüschke's teachings can be implemented in essentially any state-of-the-art transceiver system in a wireless communication system. Ex-1005, [0006].

B. The Combined Neural Network Predistortion System: Modifying Jüschke's Training with Holt's Loss Function

123. The combined neural network predistortion system comprises implementing Holt's loss function, used for training neural network models in radio communications, into Jüschke's sampling techniques and feedback loop, also used for training radio neural network models. In my opinion, this would have been a simple substitution well within the skill and knowledge of a POSITA.

124. As I previously discussed, Jüschke trains its ML models using "[s]ampled [transmitter] output values" and "ideal [input] samples." Ex-1005, [0048]. Jüschke also implements a "feedback signal for enabling adaptive pre-distortion of the baseband signal, the pre-distortion being reciprocal to the distortion caused by the transmitter device's 14 non-ideal components." *Id.*, [0060]. Figure 6 of Jüschke is illustrated below demonstrating training and adaptive pre-distortion, with pre-distortion distortion neural network 13, post distortion neural network 43, and feedback signal 61 marked with red boxes:



Ex-1005, Figure 6 (annotations added).

125. Given that Jüschke implements input/output samples, feedback loops, and backpropagation for training, in my opinion, a POSITA would have understood that combining it with Holt’s loss function would have been a simple substitution of known training techniques used in radio communication neural networks.

126. First, Jüschke explained that the training processor 40 may be adapted to train the weight and/or bias parameters of the post distortion neural network 43. Jüschke taught that “during each training iteration,” “[g]radient-based training algorithms, like back-propagation, may be used.” Ex-1005, [0018]. Second, Jüschke

further suggested that to improve efficiency “due to the fact that the gradient vanishes at the solution[,] Hessian-based algorithms allow the neural network to learn more subtle features of a complicated mapping.” *Id.* “To benefit from the advantages of Hessian based training,” Jüschke “focus[es] on the well-known Levenberg-Marquardt algorithm.” *Id.*; *see id.*, [0049] (“The training processor 40 may be adapted to train the weight (w_{kj}) and/or bias parameters (b_k) of the post distortion neural network 43 using the Levenberg-Marquardt training algorithm.”). A POSITA would have known that the Levenberg-Marquardt algorithm has been used in backpropagation training of neural networks. *See* Ex-1011, 397. Therefore, in my opinion, a POSITA would have understood that Jüschke teaches backpropagation like Holt amongst its training techniques.

127. Furthermore, 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” using “backpropagat[ion].” 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):

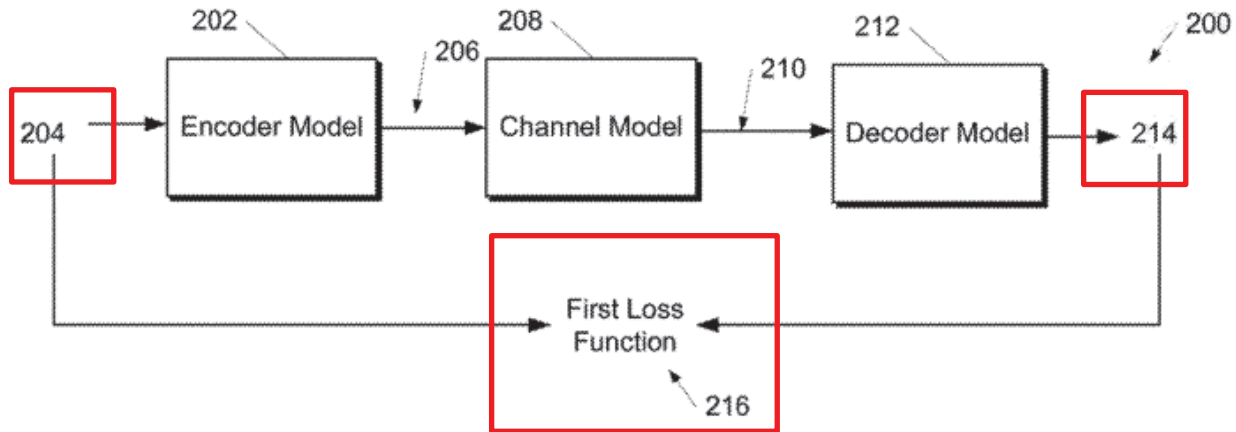


Figure 3

Ex-1006, Figure 3 (annotations added).

128. In addition, Holt teaches that an “actual communication channel” can be used as well. Ex-1006, 18:1-2. Therefore, as I previously explained, a POSITA would have understood that Holt’s encoder and decoder models were not limited to a channel model and could be implemented with an actual communication channel for loss functions. As shown below, 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:

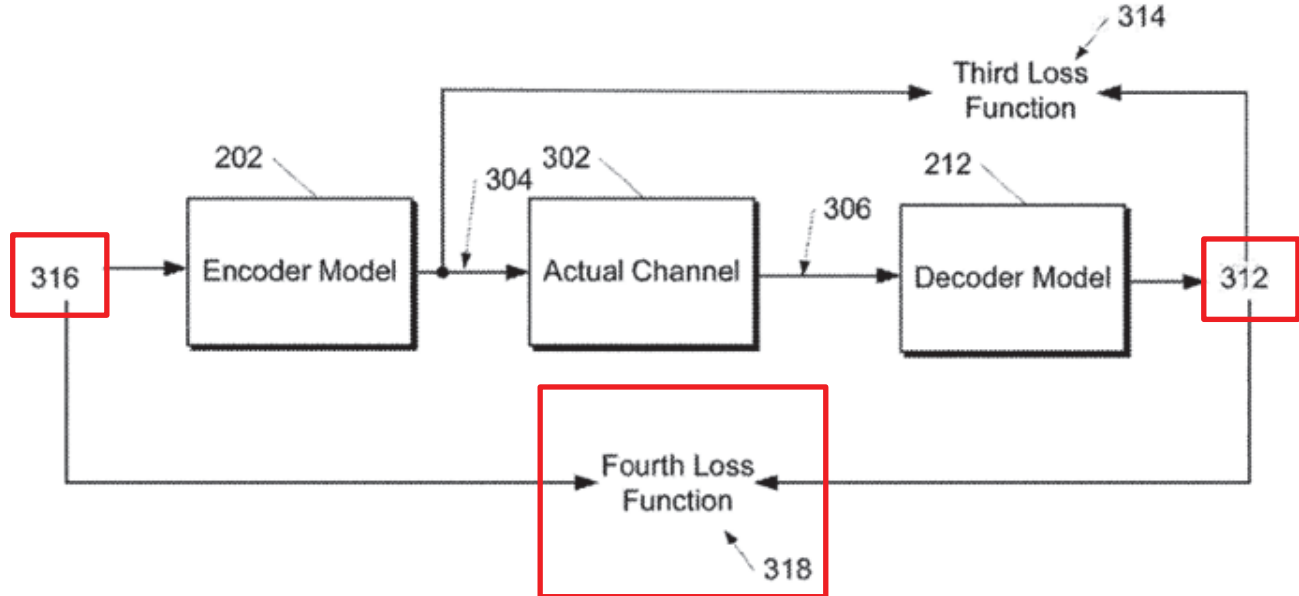


Figure 6

Ex-1006, Figure 6 (annotations added).

129. Therefore, in my opinion, a POSITA would have easily been able to combine Jüschke with Holt’s loss function and training configuration for training Jüschke’s model and parameters. Jüschke, like Holt, uses input/output samples, feedback loops, and backpropagation to train its respective models. Hence, substituting Holt’s loss function into Jüschke would have been an exchange of known model training techniques in wireless communication systems. In addition, a POSITA would have recognized that Jüschke and Holt implement their teachings within 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”).

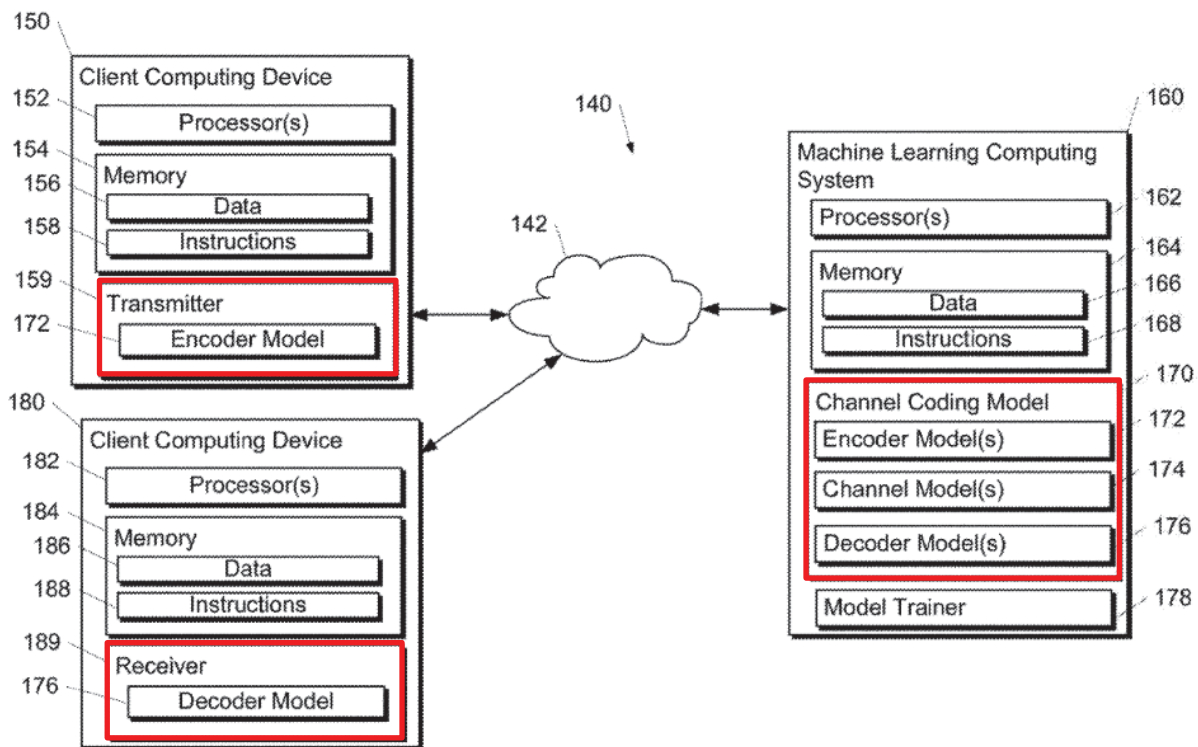
C. Motivation for Modifying Jüschke's Training

130. In my opinion, a POSITA would have been motivated to modify Jüschke's training techniques using Holt's loss function.

131. Jüschke and Holt already use ML and neural networks in wireless communications systems, such as cellular, to improve signal processing performance (e.g., reduce distortion, noise, etc.). *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."). Therefore, in my opinion, because Jüschke and Holt are in identical fields and solve similar problems, a POSITA would have been familiar with both and would have recognized that their elements could be interchanged and combined for telecommunications model training.

132. In my opinion, a POSITA would have been motivated to implement Holt's sampling and feedback training techniques into Jüschke because it would have provided a known improvement. Holt's loss functions, based on its encoder and decoder models communicating over a communication channel, include derivation from input signals inputted into the encoder model and output signals actually received by a decoder model implemented at a receiver after an actual

communication channel 302. *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). A POSITA would have understood that Holt uses actual receiver data as output data for training, and this would have improved training by sampling the signals actually received and recovered by a receiver. Therefore, in my opinion, a POSITA would have recognized implementing Holt into Jüschke as a known improvement. Figure 2 below shows Holt's ML models implemented in devices communicating across a network:



Ex-1006, Figure 2 (annotations added).

133. Therefore, in my opinion, a POSITA would have understood implementing Holt into Jüschke would have been the mere application of a known training technique to Jüschke's self-contained feedback system for training its ML models. A POSITA would have understood that Jüschke's and Holt's ML models were similar devices and that Jüschke's feedback system and ML models were ready for the training data improvement by sampling output data actually received by a receiver as taught by Holt.

134. Furthermore, in my opinion, a POSITA also would have recognized that combining Jüschke's and Holt's sampling and feedback training techniques would have been a combination of familiar elements according to known methods and a simple substitution of one known element for another. Both references teach training models using input/output sampling in feedback. Therefore, exchanging the training techniques of one into the other, such as where input/output data comes from and how it is gathered, would have been a combination of known methods well within the knowledge and skill of a POSITA. In addition, exchanging Jüschke's and Holt's training techniques would have been a simple substitution of known elements/steps for a POSITA. A POSITA would have understood that implementing Holt's training techniques into Jüschke would have predictably modified one training technique for another using another form of input/output sampling.

135. In my opinion, a POSITA also 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. As previously stated, both references focus on ML and neural networks for improving telecommunications. A 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. And as previously discussed, a POSITA also would have known that Jüschke’s teaches 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, a POSITA would have understood that adding Holt’s receiver feedback to Jüschke’s model training data would have had a reasonable expectation of anticipated success for improving output data collection and training quality in such systems.

136. Hence, as I have demonstrated, a POSITA would have been motivated to combine Jüschke with Holt for numerous reasons, including similar (if not identical) fields of work, solving similar problems, known improvement for similar devices, familiar elements and known methods, and simple substitution.

D. Jüschke in view of Holt teaches claim 1

137. In my opinion, Jüschke in view of Holt teaches claim 1. For ease of explanation, I break claim 1 into separate limitations and organize them under sections [1P]-[1E.2], *infra*.

1. Jüschke in view of Holt teaches [1P]

138. In my opinion, Jüschke teaches “[a] *method for correcting distortion of radio signals*” [1P].

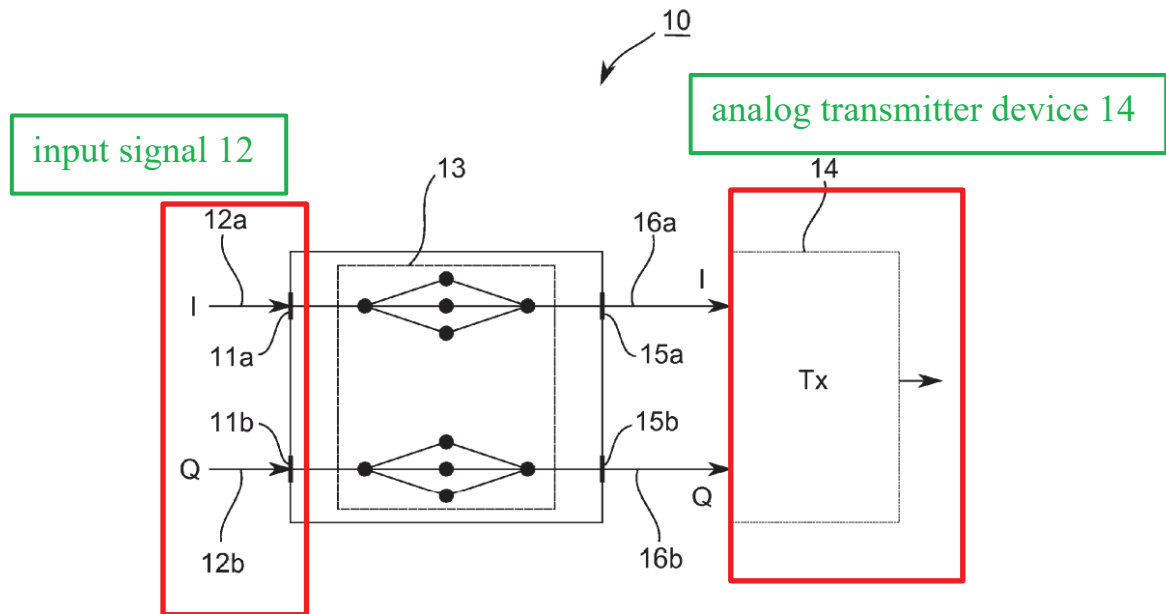
139. Jüschke’s teachings “generally relate to wireless communications and, more specifically, to a concept for mitigating impairments of a transmit signal due to transmitter imperfections” and “concept[s] for mitigating impairments due to transmit noise.” Ex-1005, [0001], Abstract. Indeed, Jüschke discloses “apparatus (10) for mitigating transmitter impairments of a transmit signal” and a “a pre-distortion neural network processor.” *Id.*, Abstract. Moreover, “apparatus 10 may also be applied” for multiple “mobile communication systems, like e.g. GSM, WCDMA and/or LTE.” *Id.*, [0059]. *See also id.*, [0002] (“2nd generation cellular communication systems, like e.g. GSM (Global System for Mobile Communications)” and “3rd and 4th generation mobile communication systems, like e.g. UMTS (Universal Mobile Telecommunication System) and LTE (Long-Term Evolution)”).

140. In my opinion, a POSITA would have understood that Jüschke's teachings apply to cellular communication systems (e.g., GSM and LTE) and implement a pre-distortion neural network processor used for pre-distorting "a digital baseband transmit signal (12)." Ex-1005, [Abstract]. A POSITA would have understood this was for mitigating transmitter impairments of a transmit signal such as, for example, "compensat[ing], i.e. lineariz[ing], arbitrary nonlinear power amplifiers," a well-known source of nonlinear distortion. Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches limitation [1P].

2. Jüschke in view of Holt teaches [1A]

141. In my opinion, Jüschke teaches "*obtaining a transmit radio signal corresponding to an output of a transmitting radio signal processing system*" [1A].

142. 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 (annotations added).

143. In my opinion, because input transmit signal 12 is illustrated as flowing towards an analog transmitter device for transmitting radio signals (Tx 14), a POSITA would have understood this to be a digital *transmit radio signal* at baseband (I and Q) being prepared for transmission by analog transmitter device 14 to a radio receiver. A POSITA would have understood that the *transmit radio signal* as taught by the '469 patent includes digital input signals as disclosed by its examples. See '469 patent, Figures 7A (“Digital RF In 701”), 7B (“Digital RF In 701”). Therefore, in my opinion, a POSITA would have understood that input transmit signal 12 of Jüschke teaches a *transmit radio signal*.

144. As shown, a POSITA would have understood that Jüschke teaches *obtaining a transmit radio signal* by the disclosure of apparatus 10 and inputs 11a and 11b. A POSITA also would have understood that Jüschke teaches a *transmit radio signal corresponding to an output of a transmitting radio signal processing system* based on Jüschke's analog transmitter device 14 transmitting radio signals based on input transmit signal 12. Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches limitation [1A].

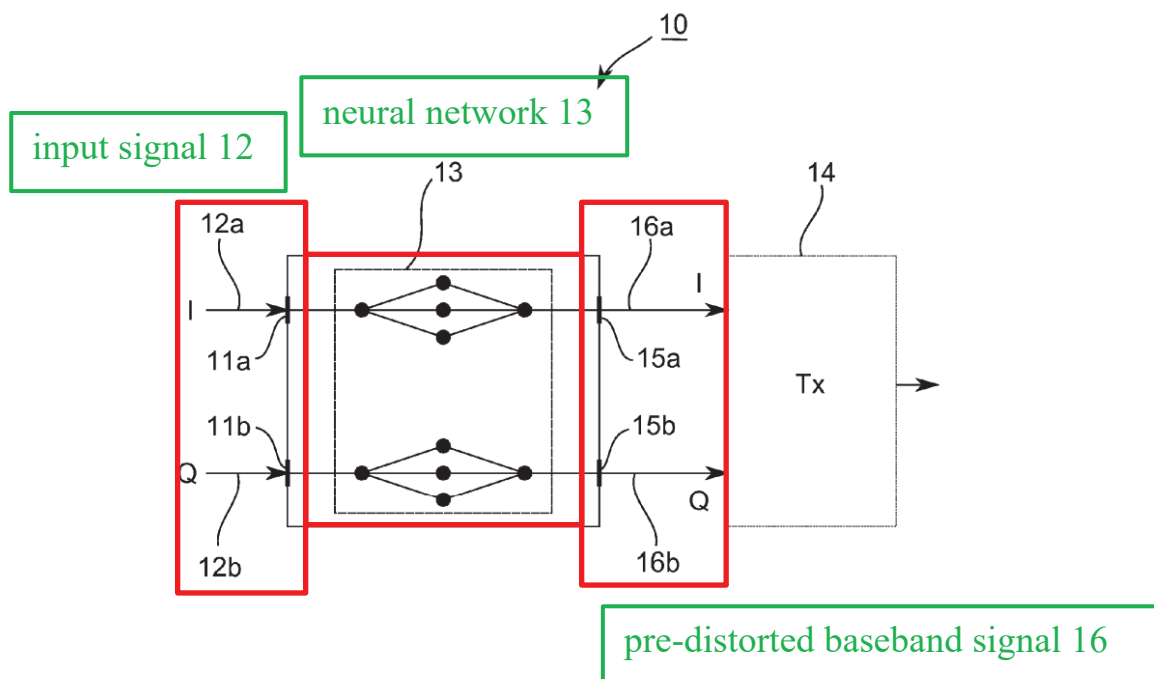
3. Jüschke in view of Holt teaches [1B]

145. In my opinion, Jüschke in view of Holt teaches “*generating a pre-distorted radio signal by processing the transmit radio signal using a nonlinear pre-distortion machine learning model, wherein the nonlinear pre-distortion machine learning model includes model parameters and at least one nonlinear function to correct radio signal distortion*” [1B]. For ease of explanation, I break [1B] into two separate limitations and organize them under sections [1B.1]-[1B.2], *infra*.

a) Jüschke in view of Holt teaches [1B.1]

146. In my opinion, Jüschke in view of Holt teaches “*generating a pre-distorted radio signal by processing the transmit radio signal using a nonlinear pre-distortion machine learning model*” [1B.1].

147. First, for *generating a pre-distorted radio signal by processing the transmit radio signal*, 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 I previously explained for limitation [1A], a POSITA would have understood that Jüschke’s input transmit signal 12 teaches a *transmit radio signal*. Figure 1, shown below, illustrates apparatus 10 with input signal components 12a, 12b being fed into neural network processor 13 (via inputs 11a, 11b) and pre-distorted output signal components 16a, 16b being outputted (via outputs 15a, 15b):

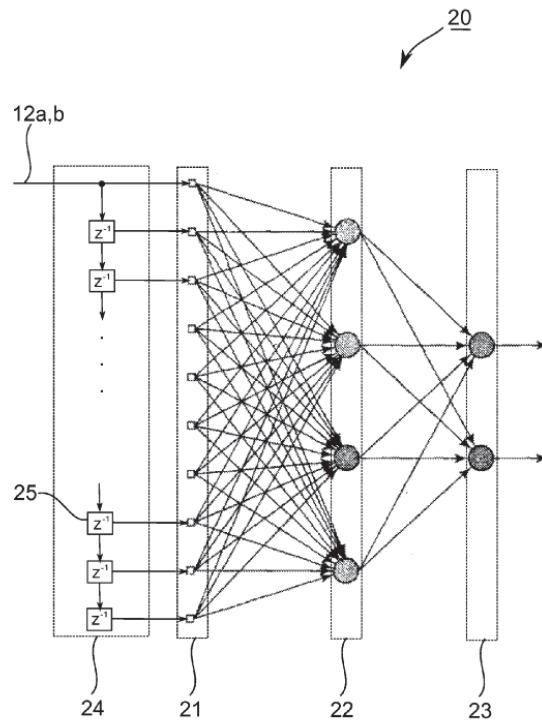


Ex-1005, Figure 1 (annotations added).

148. Second, for *using a nonlinear pre-distortion machine learning model*, Jüschke further teaches that its “pre-distortion neural network processor may comprise a multi-layer feed forward neural network (MLNN), also referred to as multi-layer perceptron (MLP).” Ex-1005, [0013]. For the MLNN, Jüschke teaches that “the pre-distortion neural network processor gain[s] the ability to split the linearization problem into a dominant linear part (output layer) and a **nonlinear** part (hidden layer(s)).” *Id.*, [0024] (emphasis added). In my opinion, a POSITA would have understood such an “MLNN” to be nonlinear owing to its nonlinear hidden layers. *Compare to* ’469 patent, 3:51-54 (“The nonlinear pre-distortion machine learning model may be a deep neural network, a convolutional neural network, a

convolutional neural network, a recurrent neural network, or a combination of different types of neural networks”). Since such a (deep or MLNN) neural network is nonlinear, a POSITA would have understood that Jüschke’s neural network processor 13 and neural network 20 teach a ***nonlinear pre-distortion machine learning model***. See Ex-1005, Figure 2, [0008] (“([A]rtificial) neural networks (NNs) have high flexibility and advantages regarding solving (non-) deterministic and complex problems without the necessity to describe them algorithmically.”). As shown by the work of Chen and Billings, neural networks “provide[] an ideal means of modelling complicated nonlinear systems.” Ex-1015, 319.

149. Jüschke also discloses using neural network processor 13 and feed-forward neural network 20 for addressing non-linear distortion caused by amplifiers. Ex-1005, [0072] (“neural networks according to embodiments, the parameters (addressing I/Q-imbalance, PA non-linearity”). This disclosure would have further reinforced a POSITA’s understanding that Jüschke teaches a nonlinear model. Figure 2 is illustrated below depicting multilayer feed-forward neural network 20 with input layer 21, hidden layer 22 (with hidden neurons), and output layer 23 (with output neurons) receiving input signal components 12a, 12b:



Ex-1005, Figure 2.

150. Therefore, in my opinion, a POSITA would have understood that Jüschke teaches limitation [1B.1] via pre-distorted baseband signal 16 (*generating a pre-distorted radio signal*), input signal 12 (*by processing the transmit radio signal*), and neural network processor 13 (*using a nonlinear pre-distortion machine learning model*). Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches limitation [1B.1].

b) Jüschke in view of Holt teaches [1B.2]

151. In my opinion, Jüschke in view of Holt teaches “*wherein the nonlinear pre-distortion machine learning model includes model parameters and at least one nonlinear function to correct radio signal distortion*” [1B.2].

152. 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] (“A function of hidden neurons is to intervene between the input layer and the output layer in some useful manner.”), [0014] (“A neuron is an information-processing unit that is fundamental to the operation of the neural network.”). Neurons and parameters determine how data inputted into the model is transformed into desired outputs.

153. Second, for *at least one nonlinear function*, as I previously explained for limitation [1B.1], Jüschke discloses a neural network model (pre-distortion neural network 13) that comprises nonlinear hidden layer 22, which a POSITA would have understood is a non-linear machine learning model implemented for mitigating transmitter impairments of a transmit signal. *See also* ’469 patent, 1:31-33 (“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)). A POSITA would have understood that a feedforward neural network is linear if it is

formed strictly with linear functions. Thus, a POSITA would have understood a feedforward neural network is nonlinear only if it is formed with at least one nonlinear function. Since Jüschke teaches a nonlinear predistortion system implementing “a multi-layer feed forward neural network” or deep neural network, Ex-1005, [0013], a POSITA would have understood that Jüschke’s neural network processor 13 and neural network 20 teach a ***nonlinear pre-distortion machine learning model***, *see id.*, Figure 2. It is my opinion that a POSITA would have further understood that pre-distortion neural network 13 as a machine learning model necessarily includes at least a *nonlinear function*.

154. 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. (“the pre-distortion neural network processor (13) is adapted to compensate for a nonlinear transmitter transfer function of at least one analog amplifier device (14)”). An inverse function of a nonlinear function is also nonlinear. In my opinion, a POSITA would have understood that an inverse of a nonlinear transmitter (transfer) function, including a close approximation of such an inverse, is a nonlinear function.

155. Third, for *to correct radio signal distortion*, 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 with a hyperbolic tangent characteristic, in particular sigmod 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.* In my opinion, a POSITA would have further understood from these disclosures that Jüschke’s pre-distortion neural network includes at least a nonlinear function for *correcting radio signal distortion*, such as compensating nonlinear power amplifiers.

156. As shown, a POSITA would have understood that Jüschke’s weight and bias parameters addressing I/Q and PA corrections teach the *model parameters*. And a 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 implementing a neural network, by implementing inverse transfer functions, and by implementing nonlinear transfer or activation functions. Therefore, in my opinion, a

POSITA would have understood that Jüschke in view of Holt teaches limitation [1B.2].

4. Jüschke teaches [1C]

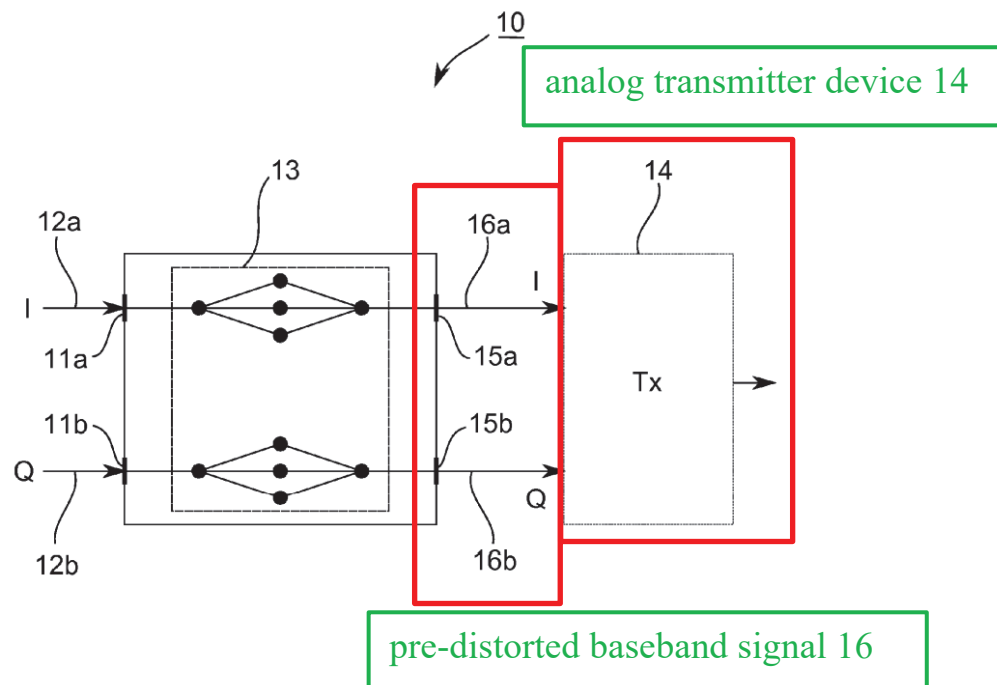
157. In my opinion, Jüschke in Holt teaches “*obtaining a transmit output radio signal by processing the pre-distorted radio signal through the transmitting radio signal processing system*” [1C].

158. 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*).

159. A POSITA would have understood that an RF transmitter, such as transmitter device 14, would have to first obtain and process a digital information-bearing data signal, such as pre-distorted digital baseband signal 16, before

converting and transmitting the signal as an analog radio signal. In my opinion, a 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*.

160. 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).

161. Second, for *processing the pre-distorted radio signal through the transmitting radio signal processing system*, Jüschke's transmitter device 14, including "amplifier device" for increasing signal power for transmission, provides a *transmitting radio signal processing system*. Ex-1005, [0020], [0050]. In addition, a 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, a POSITA would have understood that the transmitter includes a digital-to-analog converter in addition to amplifier. Thus, a 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.

162. Furthermore, Jüschke's analog transmitter device can include "an IQ up-conversion mixer" in the transmit path. Ex. 1005, [0040]. The mixer is applied to signals received by the analog transmitter (e.g., pre-distorted signals) after the digital to analog converter and before the final output radio signal is obtained and transmitted. Ex. 1005, [0004]. In addition, the transmitter device is modeled by a

transmitter transfer function. This model mathematically represents all processing (including interference) applied to a signal for transmission. In my opinion, a POSITA would have understood this disclosed IQ up-conversion mixer and transmitter transfer function reinforce Jüschke's teachings of processing a pre-distorted signal at *a transmitting radio signal processing system*.

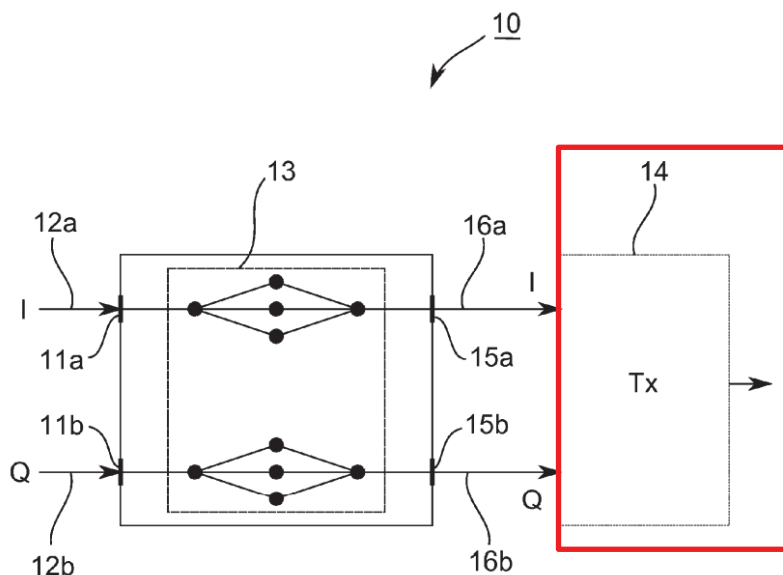
163. As shown, a POSITA would have understood that Jüschke's analog transmitter device 14, and the radio signals it transmits, teaches *obtaining a transmit output radio signal*. A POSITA also would have understood that analog transmitter device 14, via its amplifier and digital-to-analog conversion, teaches *processing the pre-distorted radio signal through the transmitting radio signal processing system*. Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches limitation [1C].

5. Jüschke teaches [1D]

164. In my opinion, Jüschke teaches "*transmitting the transmit output radio signal to one or more radio receivers*" [1D].

165. Jüschke discloses an "analog transmitter device 14" along with a "transmitter transfer function." Ex-1005, [0039]. As I previously explained for limitation [1C], 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).

166. In my opinion, a POSITA therefore would have understood that a *transmit output radio signal* being transmitted by analog transmitter device 14 would be received by a *radio receiver* or transceiver.

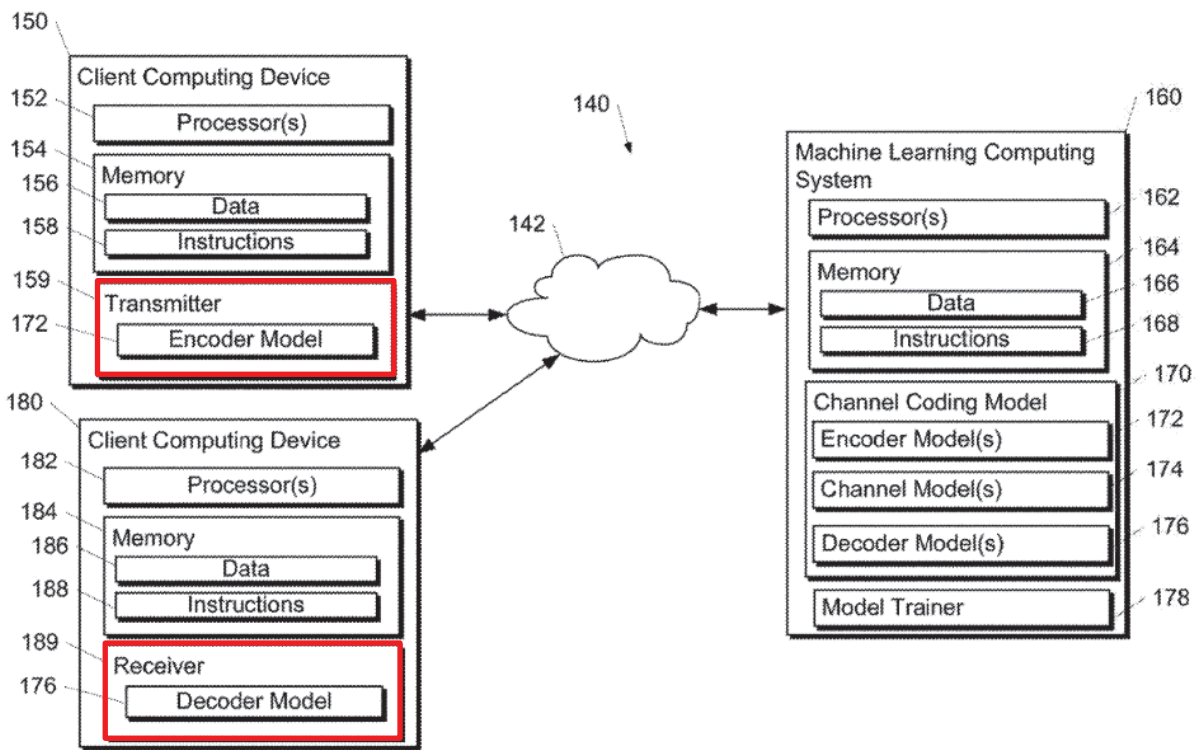
167. Moreover, Jüschke's teachings are applicable to "mobile communication systems, like e.g. GSM, WCDMA and/or LTE" that implement numerous radio receivers and transceivers. Ex-1005, [0059]. Jüschke also states that its techniques "generally relate to wireless communications and, more specifically, to a concept for mitigating impairments of a transmit signal due to transmitter imperfections." *Id.*, [0001]. Jüschke also discloses "[a]dvanced receiver signal

processing” in the context of “real world-communication systems.” *Id.*, [0004]. “RF transmitters suffer from a number of impairments that degrade the quality of the received signal beyond the impact of fading or thermal-noise at the receiver.” *Id.* In my opinion, this would have confirmed a POSITA’s understanding that Jüschke teaches output radio signals being transmitted by a transmitter device to and being received by a *radio receiver*. Therefore, in my opinion, a POSITA would have understood that Jüschke teaches limitation [1D].

168. In addition, Holt teaches communication data corresponding to a *transmit output radio signal* being *transmitted* by a transmitter and received by a *receiver*. “[T]ransmitter 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. The “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, such as would be present in numerous applications when data is transmitted over a wired and/or wireless noisy communication channel before being received and evaluated by a decoder.” *Id.*, 5:46-51; *see id.*, 5:17-19 (“[T]he channel model can be representative of a wireless channel (e.g., a radio frequency (RF) link).”). Holt also teaches an “actual communication channel” can also be used. *Id.*,

17:65-18:2 (“[D]ecoder model 212 can be further trained...to learn how to decode data coming out of the communication channel 302 (e.g., the actual communication channel 302.”).

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



Ex-1006, Figure 2 (annotations added).

170. Figure 13 below further demonstrates its decoder model (implemented in a receiver) receiving data over a “communication channel.” See Ex-1006, 21:57-59 (“[E]xample method 900 can be implemented by a computing device including a receiver (e.g., computing device 180 including a receiver 189 as depicted in FIG.

2).”). “For example, the *communication data obtained at 904* could correspond to a *transmitted encoded version of transmit data*, such as the *data transmitted over a communication channel at 810 in FIG. 12.*” *Id.*, 22:12-14 (emphasis added). In my opinion, this would have confirmed to a POSITA that Holt teaches data being transmitted over a communication channel and received by a receiver. Figure 13 demonstrating Holt method 900 is illustrated below:

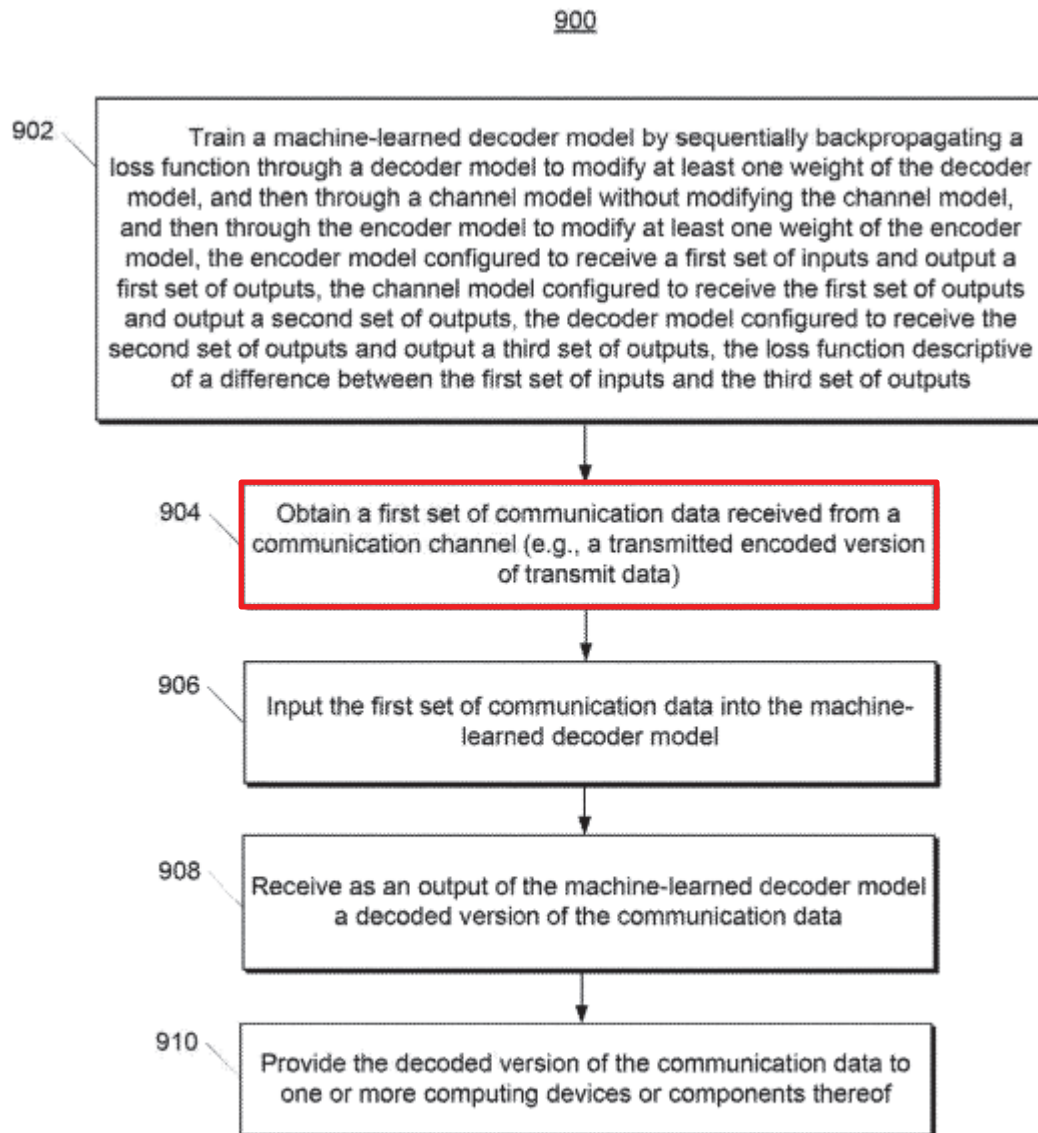


Figure 13

Ex-1006, Figure 13 (annotation added).

171. In my opinion, a POSITA would have understood that Holt's transmitter/encoder model transmitting communication data 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*. And given that Jüschke and Holt both implement ML techniques for improving telecommunications, a POSITA would have been sufficiently motivated to implement Holt's techniques into Jüschke because doing so would have been a simple substitution of known elements in the same field and provided known improvements. *See* Sections VIII.B-C, *supra*. Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches limitation [1D].

6. Jüschke in view of Holt teaches [1E]

172. In my opinion, Jüschke in view of Holt teaches “*updating the model parameters based on a distance metric computed between the transmit radio signal and a received signal that is obtained from the one or more radio receivers, the received signal corresponding to the transmit output radio signal that was transmitted to the one or more radio receivers*” [1E]. For ease of explanation, I break [1E] into two separate limitations and organize them under sections [1E.1]-[1E.2], *infra*.

a) Jüschke in view of Holt teaches [1E.1]

173. In my opinion, Jüschke in view of Holt teaches “*updating the model parameters based on a distance metric computed between the transmit radio signal and a received signal that is obtained from the one or more radio receivers*” [1E.1].

174. First, for *updating the model parameters*, 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] (“The neural network may learn from [the] feedback signal and adapt itself to achieve required results.” “[S]elf learning/optimizing neural networks” have “the parameters...considered and optimized.”). Jüschke’s parameters are used “to fit the feed-forward pre-distortion neural network to a specific downstream hardware component.” *Id.*, [0017]. In my opinion, a POSITA would have understood that Jüschke training and optimizing its parameters teaches *updating the model parameters*.

175. Second, for *updating the model parameters based on a distance metric*, Jüschke presents “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]. A POSITA would have understood that Jüschke's model training uses input and output signals, and therefore a difference or *distance metric* between them. Therefore, in my opinion, a POSITA would have understood that Jüschke training and optimizing its model parameters (*updating the model parameters*) using inputs and outputs (*based on a distance metric*) teaches *updating the model parameters based on a difference or distance metric* between input and output signals.

176. In addition to Jüschke, Holt's uses loss functions to train and update model parameters. *See* Ex-1006, Abstract (referencing "a loss function that describes a difference between input data...and output data"). 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." *Id.*, 18:42-48. As I previously explained, a loss function describes the difference between an ML model's predicted output values and the true output values, and a distance metric is used to quantify this difference. *See* Section V, *supra*. It is my opinion that this would have been consistent with a POSITA's understanding of loss functions and distance metrics. In my opinion, a POSITA would have understood that Holt modifies (*updates*) model parameters using loss function, and 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. Therefore, a POSITA would have therefore understood that Holt training and updating its ML model parameters using loss functions teaches *updating the model parameters based on a distance metric*.

177. Third, for the *distance metric* being *computed between the transmit radio signal and a received signal that is obtained from the one or more radio receivers*, Jüschke teaches transmitting an output radio signal to a radio receiver and implementing a feedback loop of the transmitted signal for training its 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*). The post distortion neural network 43 is trained until its response is “essentially [] the inverse of the transfer or distortion function of the analog transmitter device 14.” *Id.*, [0048]. In my opinion, a POSITA would have understood that Jüschke training post

distortion neural network 43 until the response is an inverse of transmitter device 14 to account for the difference or distance between the undistorted input values and sampled output values used for training. *Id.*, [0048]. Therefore, in my opinion, a POSITA would have understood that Jüschke teaches a *distance metric* that is *computed* using the difference (or distance) between undistorted input samples (*transmit radio signal*) and sample output values (*received signal*) for training its post-distortion network and pre-distortion neural network.

178. In addition, Holt teaches loss functions using 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) connected over a communication channel. Ex-1006, Abstract (“The channel coding model 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.”). As I previously explained, a loss function describes the difference between an ML model’s predicted output values and the true output values, and a distance metric is used to quantify this difference. *See* Section V, *supra*. And as I previously stated, it is my opinion that this would have been consistent with a POSITA’s understanding of loss functions and distance metrics. Hence, in my

opinion, a 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. Therefore, a POSITA would have understood that Holt's loss function, based on difference metrics between encoder input signals and output signals (corresponding to a signal generated by the receiver decoder model) teaches a *distance metric computed between the transmit radio signal and a received signal*. See also '469 patent, 7:34-38 ("The transformation functions can be...one of many other potential forms of nonlinear activation function that works optimally to ***minimize a distance metric or loss function***." (emphasis added)).

179. Figure 6 further below illustrates encoder model 202 with inputs 316 and outputs 304 sent into an actual communication channel 302, which feeds into decoder model 212 of one or more receivers' output signals 312 based on the transmission received over an "Actual [Communication] Channel." Figure 6 also shows the receiver output signal 312 being used to generate the third loss function 314 and/or the further loss function 318. Fourth loss function 318 (*distance metric*) is determined by finding the difference between the input 316 (*transmit radio signal*) at the transmitter and the output 312 (*received radio signal*) obtained from the decoder model 212 of one or more receivers.

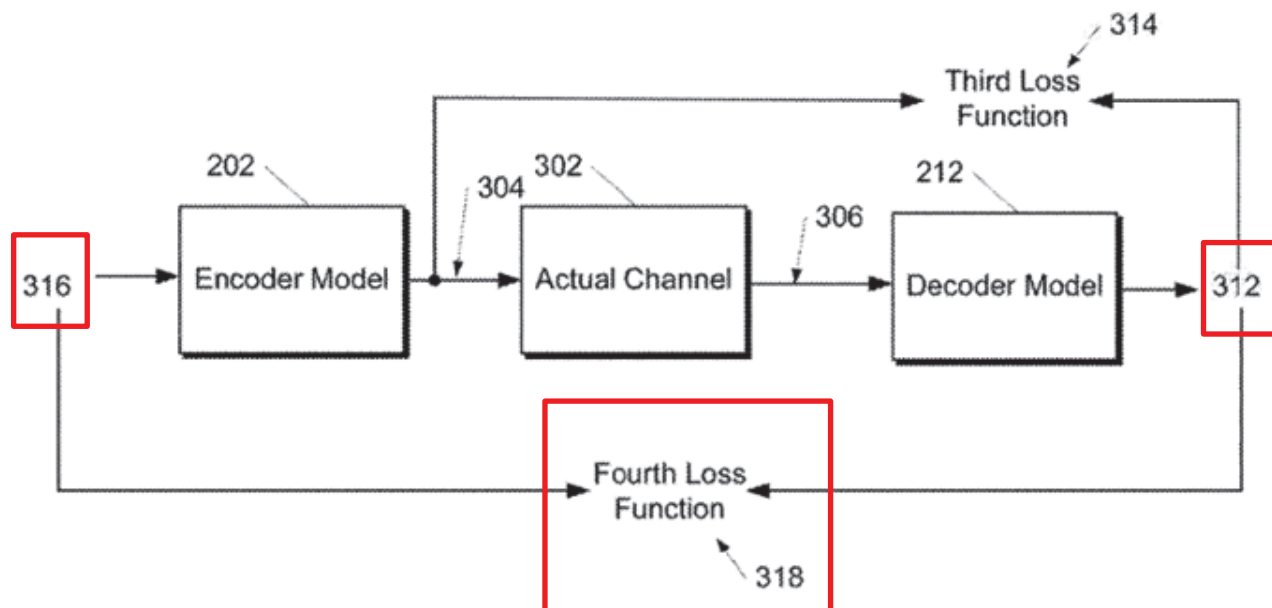


Figure 6

Ex-1006, Figure 6 (annotations added); *see id.*, 18:40-42 (“The fourth loss function 318 can describe a difference between the third set of inputs 316 and the sixth set of outputs 312”); *see also id.*, Figure 3 (first loss function 216 from input 204 and output 214).

180. In my opinion, a 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 VIII.C, *supra*. For example, a POSITA would have understood that Jüschke’s training using transmitter feedback would have enjoyed a known improvement by using Holt’s loss functions that compare the difference between the input signal and *a received signal that is obtained from the one or more radio receivers* that was actually received at a radio

receiver. In my opinion, a POSITA would have understood that implementing Holt's loss function into Jüschke's output feedback sampling would have been an exchange of Jüschke's form of output data sampling used for training neural networks, for the more accurate version of Holt that samples those data actually received and recovered at a receiver. Indeed, a POSITA would have understood that Holt's loss functions, which account for the actual effects of transmission and reception by sampling output data actually received and recovered at a receiver, would have provided a known improvement to Jüschke's by implementing Holt's loss functions and data sampling into Jüschke's output data sampling techniques for neural network training. Therefore, a POSITA would have understood updating parameters using Holt's loss functions in lieu of or in addition to Jüschke's transmission input/output feedback loop would have been a combination known elements, methods, and techniques to achieve predictable results (with a reasonable expectation of success) because doing so would have been exchanging one form of output data sampling (using more accurate sampling data that was actually received and recovered at a receiver) for another form of output data sampling using a transmission output feedback.

181. As shown, a POSITA would have understood that Jüschke training and optimizing of its model parameters using inputs and outputs, in view of Holt's loss

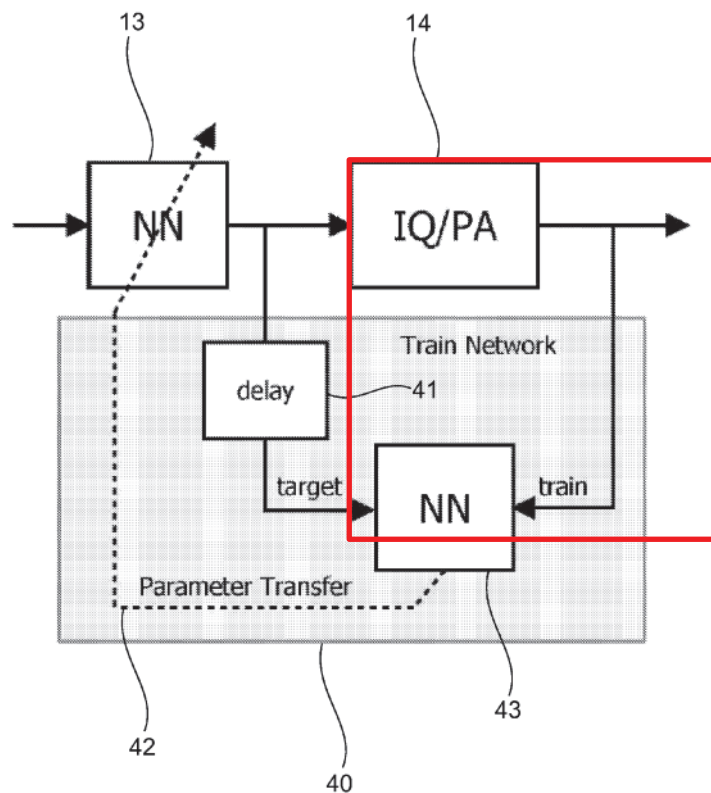
function and backpropagation techniques for updating ML models, teaches *updating the model parameters based on a distance metric*. A POSITA would have understood that Jüschke's use of the difference between undistorted input samples and sample receiver output values for training its pre-distortion neural network, in view of Holt's loss function based on the difference between input data input into the encoder model and output data from the receiver decoder model teaches, teaches *a distance metric being computed between the transmit radio signal and a received signal that is obtained from the one or more radio receivers*. A POSITA also would have been sufficiently motivated to implement Holt into Jüschke at least because doing so would have provided known improvements and would have had a reasonable expectation of success. Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches limitation [1E.1].

b) Jüschke in view of Holt teaches [1E.2]

182. In my opinion, Jüschke in view of Holt teaches “*the received signal corresponding to the transmit output radio signal that was transmitted to the one or more radio receivers*” [1E.2].

183. As I previously explained for limitation [1D], Jüschke teaches transmitting a *transmit output radio signal* to a *radio receiver*. Jüschke further teaches a feedback loop of the signal to be transmitted with the pre-distortion neural

network. Ex-1005, [0072] (“[T]he neural network may be part of the transmitter loop including transmit and feedback path.”). Figure 4 shown below illustrates 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 transmitted to and received by a *radio receiver*:

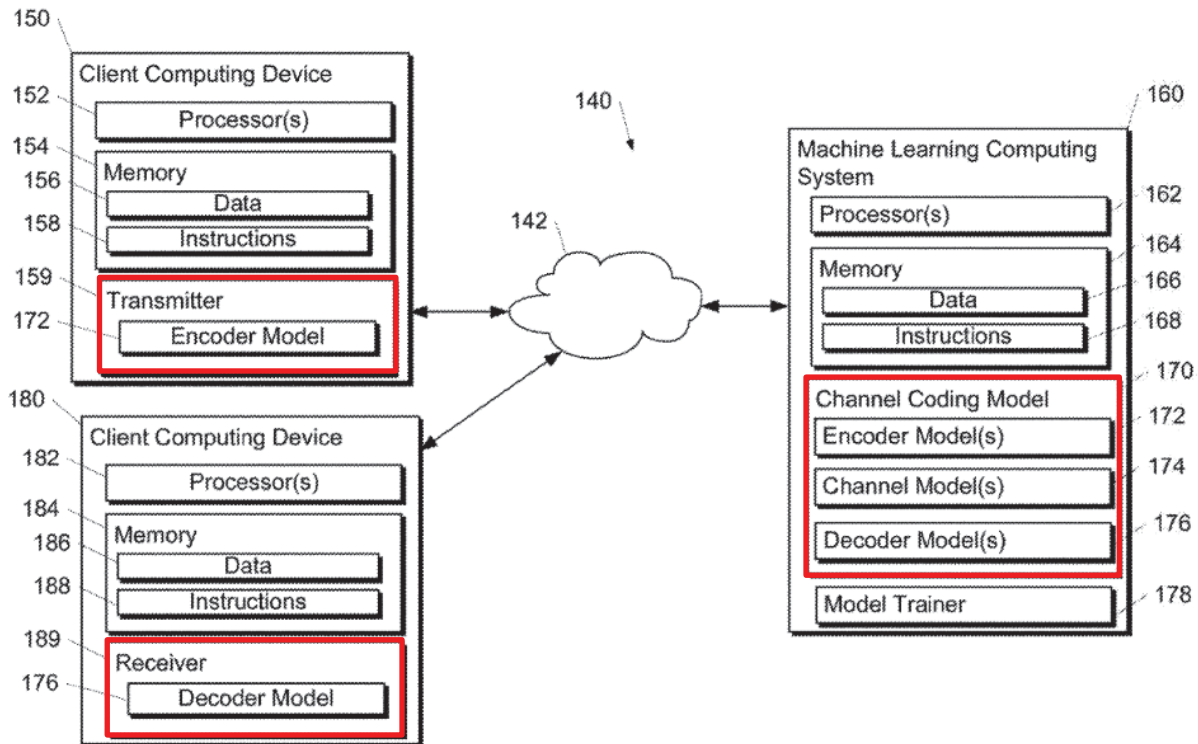


Ex-1005, Figure 4 *See also id.*, Figures 1, 6.

184. In my opinion, a POSITA would have understood that a *transmit output radio signal* being transmitted from transmitter device 14 would be the same signal being received by a *radio receiver* or transceiver. *See* limitation [1D], *supra*.

Therefore, it is also my opinion that a POSITA would have further understood that the feedback from a *transmit output radio signal* that is being transmitted to a *radio receiver* would *correspond to the transmit output radio signal that was transmitted to the one or more radio receivers*.

185. Any gap in the teachings of Jüscke for [1E.2], however, would have been taught in view of Holt in my opinion. As I previously explained for limitation [1D], *see* Section VIII.D.5, *supra*, Holt teaches communication data (*transmit output radio signal*) being transmitted by a transmitter and *received* by a *receiver*. This data is further processed for machine learning and training by Holt's ML models. For example, as shown in Figure 2 below, 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 shown with ML models 172, 174, 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. See Figure 2 below:



Ex-1006, Figure 2 (annotations added). *See id.*, 4:34-39 (“[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” (emphasis added)), 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)).

186. Therefore, in my opinion, a POSITA would have understood that Holt teaching transmitted and received communication data being received by a decoder model implemented in a receiver device teaches a *received signal corresponding to the transmit output radio signal that was transmitted to the one or more radio receivers*.

187. It is also my opinion that a POSITA would have been motivated to combine Holt's techniques for machine learning and training using transmitters (encoders) and receivers (decoders) with Jüschke at least because doing so would have been a simple substitution of known elements in the same field (ML in telecommunications). Additionally, a POSITA would have understood that using transmitted signals actually received and recovered at a receiver, as taught in Holt, would have provided a known improvement to Jüschke by utilizing more accurate training data that accounts for the actual effects of transmission and reception by a receiver into Jüschke's training techniques that sample output signals in feedback. Adding Holt's received and recovered transmission data into Jüschke's feedback loop that already implements transmitter output data would have implemented more accurate training data into Jüschke's neural network training techniques and therefore would have had a reasonable expectation of success implementing and improving Jüschke's form of output data sampling for Holt's output data sampling

that samples more accurate output data that was actually received and recovered at a receiver. *See* Section VIII.C, *supra*.

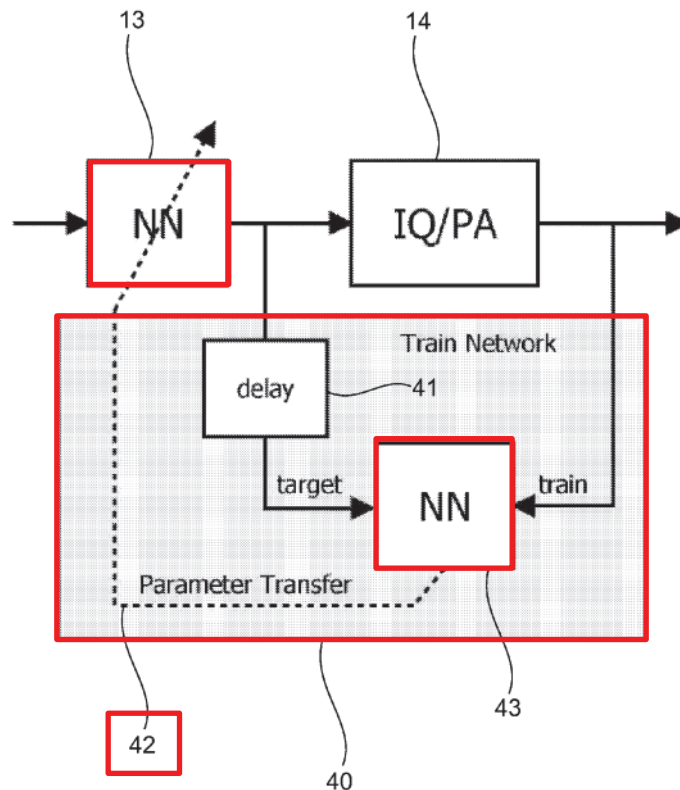
188. As shown, a POSITA would have understood that Jüschke's analog transmitter device 14 and the radio signals it transmits and feedback loop, in view of Holt's communication data sent from an encoder model (in a transmitter) and received by a decoder model (in a receiver), teach *the received signal corresponding to the transmit output radio signal that was transmitted to the one or more radio receivers*. A POSITA also would have been sufficiently motivated to implement Holt into Jüschke at least because doing so would have been a simple substitution of known elements in the same field, a known improvement, and had a reasonable expectation of success. Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches limitation [1E.2].

E. Jüschke in view of Holt teaches claim 2

189. In my opinion, Jüschke teaches “[t]he method of claim 1, wherein the model parameters are updated with values for a particular deployment scenario for a radio communication system upon initial deployment of the nonlinear pre-distortion machine learning model into the radio communication system” [2]. I incorporate by reference my previous analysis for claim 1. *See* Sections VIII.D.1-6, *supra*.

190. First, for *model parameters* being *updated with values*, as I previously explained for limitation [1B.2], a POSITA would have understood that 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. 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.*

191. 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:



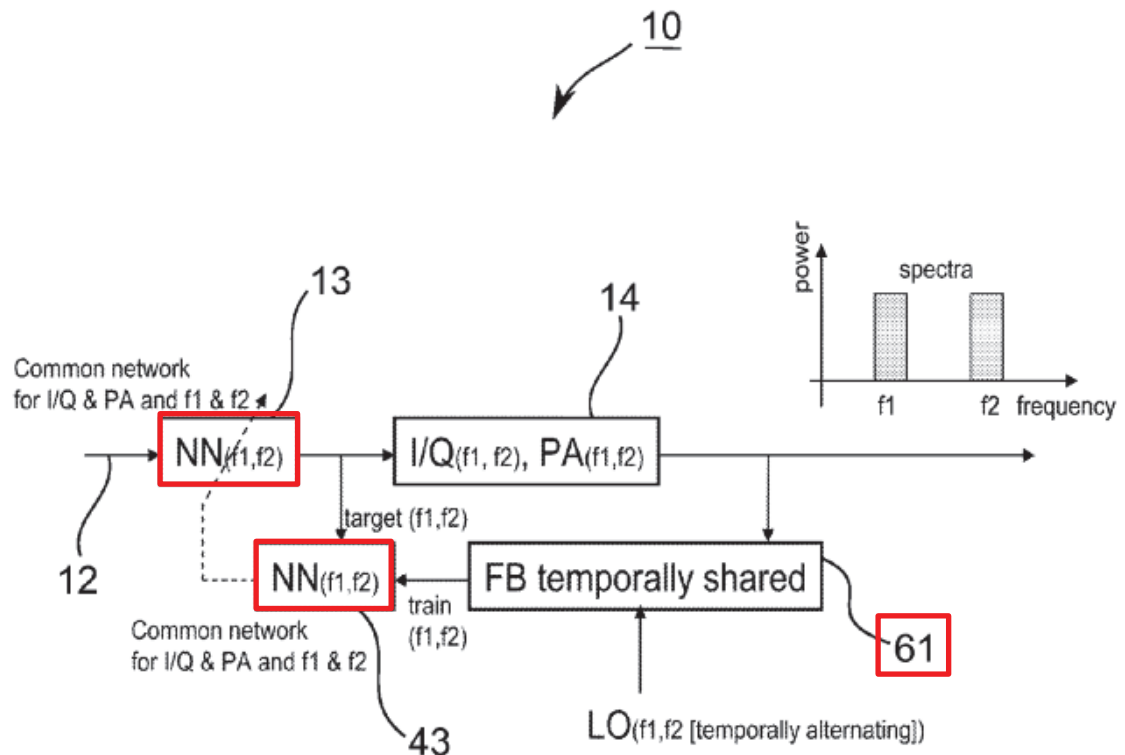
Ex-1005, Figure 4 (annotations added).

192. As the non-ideal transmitter device 14 functions change over time, the Jüschke post-distortion neural network 43 “frequently” transfers the trained weight and/or bias parameters to the pre-distortion neural network 13 to track or otherwise correspond the parameters to the “time varying transfer function of the non-ideal transmitter devices 14.” Ex-1005, [0058]. In my opinion, a POSITA would have understood that Jüschke **tracking** a time-varying transfer function of the non-ideal transmitter device 14 **updates** the neural network parameters to address ongoing I/Q-imbalance and PA non-linearity of the non-ideal transmitter device 14 as the device output characteristics change over time. Therefore, in my opinion, a POSITA would

have understood that frequently transferring weight and/or bias parameters (*model parameters*) in order to track (*update*) a time-varying transfer function for the non-ideal transmitter device teaches *model parameters* being *updated*.

193. Second, for *model parameters* having values for a particular deployment scenario for a radio communication system, Jüschke's weight and bias parameters are implemented in deployments for cellular communications such as GSM (2G), WCDMA (UMTS or 3G), and LTE (4G). In my opinion, a POSITA would have understood that Jüschke's weight and bias parameters are trained and adjusted (*updated with values*) to best fit the unique needs of a specific environment or transmission configuration (*particular deployment scenario*). For instance, "apparatus 10 may also be applied to fragmented spectrum applications, e.g., when using a single transceiver device for transmitting in spectrally separated frequency bands with center frequencies f_1 , f_2 of possibly different mobile communication systems, like e.g. GSM, WCDMA and/or LTE." Ex-1005, [0059]. "By use of a self learning/optimizing neural network according to embodiments, the 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) may commonly be considered and optimized by embodiments of the present invention in a more advanced manner compared to use dedicated, non-self learning algorithms." *Id.*

194. For implementing the aforementioned fragmented spectrum applications, Jüschke describes “transmitting [] spectrally separated frequency bands” with a transmitter device 14 in a “parallel operation of fragmented frequency bands f_1 , f_2 .” Ex-1005, [0059]. Jüschke obtains feedback from the transmitter device 14 in “a time multiplexed manner, e.g. by temporally changing the LO-frequency of an IQ up-conversion mixer from f_1 to f_2 in the feedback path 61.” Ex-1005, [0060]. As shown in Figure 6 below, the time-multiplexed feedback may be provided as a training information to the post distortion neural network 43:



Ex-1005, Figure 6 (annotations added).

195. Jüschke's post distortion neural network processing unit 43 receives, "temporally alternately," the "spectra of the two frequency bands," which are shown in Figure 6 as "train (f_1 , f_2)." Ex-1005, [0061]. The post distortion neural network processing unit 43 may also receive "target (f_1 , f_2)" in support of the post distortion neural network processing unit 43 generating a time multiplexed parameter transfer (shown in dashed line in Figure 6 above) to the pre-distortion neural network processor 13, for example, by "alternately consider[ing] two or more frequency bands." *Id.* The Jüschke "pre-distortion neural network processor 13 may be adapted to pre-distort the input digital baseband transmit signal to obtain a plurality of pre-distorted digital baseband signals, each for a different one of a plurality of analog transmitter devices corresponding to a different one of the plurality fragmented frequency bands f_1 and f_2 ." *Id.*, [0068].

196. Therefore, in my opinion, a POSITA would have understood that Jüschke's weight and/or bias parameters (*model parameters*) have values corresponding to the respective fragmented frequency bands f_1 and f_2 because Jüschke describes pre-distorting the input signal in a manner adapted for each such band. *See* Ex-1005, [0059], [0068]. Jüschke teaches that the pre-distortion processor 13 can operate over a plurality fragmented frequency bands f_1 and f_2 , such that "the pre-distortion neural network processor 13 may be adapted to pre-distort the input

digital baseband transmit signal to obtain a plurality of pre-distorted digital baseband signals, each for a different one of a plurality of analog transmitter devices corresponding to a different one of the plurality fragmented frequency bands f_1 and f_2 .” *Id.*, [0068]. Because the multiplexed frequency bands f_1 and f_2 occur in a multiplexed, temporally alternating manner, a POSITA would have understood each frequency band, or each combination of the fragmented frequency bands f_1 and f_2 , as a deployment scenario encountered by apparatus 10 that Jüschke’s weight and/or bias parameters are trained and fitted for. Hence, POSITA would have understood that Jüschke having weight and/or bias parameter values adapted for respective fragmented frequency bands teaches *model parameters having values for a particular deployment scenario for a radio communication system*.

197. Third, for the model parameters being updated with values for a particular deployment scenario for a radio communication system *upon initial deployment of the nonlinear pre-distortion machine learning model into the radio communication system*, Jüschke discloses “at least one switch for switching between the frequency paths [e.g., frequencies f_1 , f_2]” to facilitate transitioning from a first, initial frequency (f_1) to a second, subsequent frequency (f_2). Ex-1005, [0061]. As I explained above, Jüschke’s weight and bias parameters (*model parameters*) have values that are updated corresponding to the respective fragmented frequency bands

f_1 and f_2 . Ex-1005, [0059], [0068]. In my opinion, a POSITA would have understood that on an initial deployment or use of Jüschke's fragmented spectrum applications, Jüschke switches from a first frequency to a second frequency and the neural network parameters are updated for the same. Given a POSITA would have understood that Jüschke's weight and bias parameters are *updated* for a particular frequency on said switching, POSITA would have also understood that switching to the second frequency teaches updating the parameters *upon initial deployment of the nonlinear pre-distortion machine learning model into the radio communication system*. A POSITA would have further understood that switching to a new second frequency different from the initial second frequency teaches updating the parameters *upon initial deployment of the nonlinear pre-distortion machine learning model into the radio communication system*. See *id.*, [0027] ("That is to say, a transmitter may comprise digital and non-ideal analog circuitry for supporting two or more frequency bands in parallel, e.g., using two or more mixers (one per frequency band) and a power amplifier (one per frequency band or a common one).").]

198. Therefore, in my opinion, a POSITA would have understood that Jüschke training its weight and bias parameters for I/Q imbalance and PA non-linearity to track (or *update* for) time-varying transfer functions of non-ideal

transmitter devices 14 teaches *model parameters are updated with values*. Further, a POSITA would have understood that that Jüschke's configuration deployments, including fragmented spectrum applications for which the parameters are adjusted for, teach *for a particular deployment scenario for a radio communication system*. And a POSITA would have understood that Jüschke's weight and bias parameters being updated for a pre-distortion neural network 13 when switching between from a first frequency to a second frequency for Jüschke's fragmented spectrum applications teaches *upon initial deployment of the nonlinear pre-distortion machine learning model into the radio communication system*.

199. Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches claim 2.

F. Jüschke in view of Holt teaches claim 3

200. In my opinion, Jüschke in view of Holt teaches “[t]he method of claim 2, wherein the model parameters are updated depending on at least one environment condition within the radio communication system” [3]. I incorporate by reference my previous analysis for claim 2. See Sections VIII.E, *supra*.

201. As I previously explained for claim 2, Jüschke trains the weight and bias parameters of pre-distortion neural network 13 for I/Q imbalance and PA non-linearity. In addition, the weight and bias parameters have values for specific

configuration deployments such as a first frequency and a secondary frequency in fragmented spectrum applications. Jüschke discloses “fragmented spectrum applications, e.g., when using a single transceiver device for transmitting in spectrally separated frequency bands with center frequencies f_1 , f_2 of possibly different mobile communication systems, like e.g. GSM, WCDMA and/or LTE.” Ex-1005, [0059]. “I/Q-imbalance and PA non-linearity in parallel for different frequency bands” via a “self learning/optimizing neural network” and the “network parameters, i.e. weights w_{kj} , bias values b_k ” are “considered and optimized.” *Id.* Further, Jüschke discloses transmitting various signal types (e.g., 2G or GSM, 3G or WCDMA, and/or 4G or LTE), which may be associated with one or more frequency bands (f_1 , f_2) accounted for by Jüschke’s pre-distortion neural network. *Id.*, [0003].

202. In my opinion, a POSITA would have understood that Jüschke’s pre-distortion neural network, by supporting fragmented spectrum applications for different frequencies f_1 and f_2 , would have understood that the same pre-distortion neural network accounts for the differences in the parameter values for predistorting and transmitting the different frequencies in fragmented spectrum applications such as for a first frequency f_1 and for a second frequency f_2 . Ex-1005, [0059]. Because Jüschke’s neural network and parameters account for both frequencies f_1 and f_2 , including for GSM (2G), WCDMA (3G), and/or LTE (4G) signal types for different

mobile communication systems, it is my opinion that a POSITA would have understood that Jüschke's pre-distortion neural network accounts for GSM (2G), WCDMA (3G), and/or LTE (4G) signal types (*environment conditions*) associated with a respective frequency. Hence, it is my opinion that a POSITA would have understood that Jüschke teaches accounting for *at least one environment condition within the radio communication system* when updating its parameters.

203. Therefore, it is my opinion that a POSITA would have understood that Jüschke's pre-distortion neural network and its parameters accounting for fragmented frequencies and associated with respective signal types teach *model parameters are updated depending on at least one environment condition within the radio communication system*. Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches claim 3.

G. Jüschke in view of Holt teaches claim 4

204. In my opinion, Jüschke in view of Holt teaches “[t]he method of claim 3, wherein the environment condition is one of: temperature, humidity, antenna impedance, power levels, or candidate transmit signal type” [4]. I incorporate by reference my previous analysis for claim 3. *See* Sections VIII.F, *supra*.

205. As I previously explained for claim 3, Jüschke's accounting for the signal types of different mobile communication systems such as GSM (2G),

WCDMA (3G), and/or LTE (4G) for respective fragmented frequency bands f_1 , f_2 teaches *model parameters are updated depending on at least one environment condition within the radio communication system*. In my opinion, a POSITA would have further understood that by accounting for GSM (2G), WCDMA (3G), and/or LTE (4G) signal types, that Jüschke's pre-distortion neural network and its parameters account for a *candidate transmit signal type* for at least the same reasons as I explained for Jüschke teaching claim 3.

206. Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches claim 4.

H. Jüschke in view of Holt teaches claim 5

207. In my opinion, Jüschke in view of Holt teaches “[t]he method of claim 2, wherein the particular deployment scenario comprises **one of**: [5A] sending a radio transmission from a cellular base station that includes the nonlinear pre-distortion machine learning model inside a tower mounted amplifier in an inline mode; [**or**]...[5C] sending a radio transmission from a cellular base station that includes the nonlinear pre-distortion machine learning model inside a remote radio unit in an inline mode” (emphasis added). For ease of explanation, I break claim 5 into limitations [5A] and [5C] and analyze them below, *infra*. I incorporate by reference my previous analysis for claim 2. See Sections VIII.E, *supra*.

208. I have been informed by counsel that claim 5 is written in a Markush format (*wherein the particular deployment scenario comprises **one of**:*), and I have been instructed to only analyze limitations [5A] and [5C]. I have further been informed by counsel that for a Markush claim like claim 5, only one of the limitations after the language “*one of*.” need be shown taught or anticipated.

1. Jüschke in view of Holt teaches [5A]/[5C]

209. In my opinion, Jüschke in view of Holt teaches the specific *deployment scenarios* of limitations [5A] and [5C] which read on radio transmissions being sent from a base station with a NPDMLM being inside either a tower mounted amplifier [5A] or a remote radio unit [5C] in an inline mode [5A]/[5C]. As I previously explained for claim 2, a 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.

210. For *sending a radio transmission from a cellular base station that includes the nonlinear pre-distortion machine learning model* [5A]/[5C], Jüschke teaches that “it is desirable to have a more flexible radio transmitter that can support

as many standards and signals as possible,” applicable to “macro-cell base stations”⁷ and “small-cell base stations,”⁸ Ex-1005, [0006], and “systems for mobile communications.” *Id.*, [0059]. As I previously explained, a POSITA would have understood that Jüschke’s transmitter can be “flexibl[y]” implemented in a wide range of hardware configurations and scenarios and structures for cellular communication systems including *base stations*, related transceiver infrastructure, and remote radio units (RRUs) or remote radio heads (RRHs). *See* Section V, *supra*. Furthermore, a POSITA would have understood that Jüschke’s flexible transmitter, including apparatus 10 and pre-distortion neural network processor 13, can be implemented within a *cellular base station* and that input transmit signal 12 can be obtained at and ultimately transmitted from such a base station.

⁷ In my opinion, a POSITA would have understood that a macro-cell base station is a cellular base station that sends and receives radio signals through high-power outdoor towers and antennas.

⁸ In my opinion, a POSITA would have understood that a small-cell base station is a low-power cellular base station (possibly indoor), covering a limited area such as malls, hotels, or transportation hubs.

211. For *inside a tower mounted amplifier* [5A], in my opinion, a POSITA further would have understood that a “macro-cell base station” is a cellular base station that sends and receives radio signals to and from mobile stations, respectively, over a larger coverage area. Ex-1005, [0006]. A POSITA would have understood that such a cellular base station includes all necessary equipment such as power amplifiers, including tower mounted amplifiers or mast head amplifiers as I previously explained in Section V, *supra*. Therefore, it is my opinion that a 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* [5A].

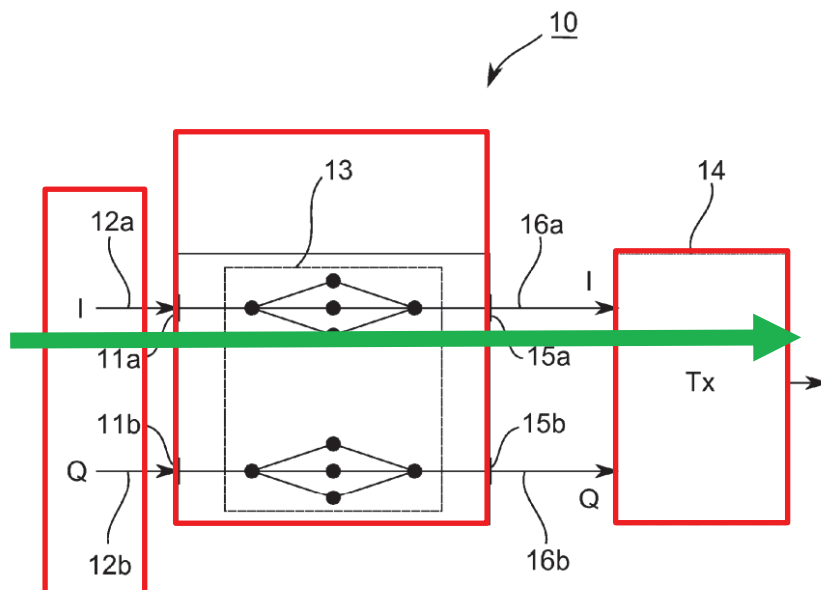
212. For *inside a remote radio unit* [5C], it is also my opinion that a POSITA would have would have understood, at the time of the critical date of the '469 patent, that state-of-the-art “macro-cell base stations,” Ex-1005, [0006], often consisted of a baseband unit (BBU) and at least one remote radio unit (RRU) or remote radio head (RRH). As I previously explained in Section V, *supra*, a remote radio unit (RRU) refers to a radio frequency (RF) transceiver component, typically collocated with an antenna of an extended base station, that directly connects to an operator’s BBU or radio control panel and processing unit via a high-speed link to provide

effective and broad coverage service to wireless devices (e.g., cellular phones) in a wireless network (e.g., cellular network). As an example, the work of Wang, Hossain, and Bhargava briefly describes the history of certain base station infrastructure that a POSITA would have been familiar with, including a “separated baseband unit (BBU)-remote radio unit (RRU) structure of BS design to provide more flexible coverage and capacity solutions with improved processing resource utilization.” Ex-1013, 42; *see also* Ex-1014 (Lanzani discussing remote radio heads (a/k/a remote radio units) as part of a cellular base station infrastructure.). A POSITA would have known that remote radio units are connected to, and work together with, baseband units (BBUs) that are in charge of processing the baseband signals for transmission and reception. Therefore, it is my opinion that a POSITA also 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 in *remote radio units* [5C].

213. It is also my opinion that Jüschke teaches a *deployment scenario* for an *inline mode* [5A]/[5C]. As I previously explained for claim 2, *supra*, a POSITA would have understood that Jüschke’s neural network parameters are trained and fitted to best fit particular deployment scenarios.

214. Jüschke discloses “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 (typically through a backpropagation training process). *Id.* In my opinion, a POSITA would have understood that the deployment configuration of apparatus 10 was applicable to Jüschke’s “flexible radio transmitter” and base station and remote radio unit (or RRH) hardware configuration *deployment scenarios*. Ex-1005, [0006]. In my opinion, a 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.

215. Figure 1, shown below, illustrates apparatus 10 for a transmitter hardware configuration (*deployment scenario*) with a pre-distortion neural network 13 processor (*nonlinear pre-distortion machine learning model*) inline between an input transmit signal 12 and a power amplifier in a transmitter device 14 (*inline mode*). A green arrow has been added to highlight the components being in tandem:



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

216. As I previously explained, the *deployment scenarios* of the ’469 patent are basic hardware configurations. *See* Section V, *supra*. And as I previously explained for claim 2, a POSITA would have understood that the deployment configuration of apparatus 10 illustrated above was applicable to specific *deployment scenarios* in Jüschke’s fragmented spectrum applications. *See* Ex-1005, [0059]; *see also id.*, [0006] (“flexible radio transmitter”). In my opinion, a 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.

217. For comparison, the '469 patent discloses nearly identical inline-mode configurations as compared to Figure 1 of Jüschke. As shown below, “example configuration[s] for in-line mode,” '469 patent, 10:39, are illustrated in Figures 6A (tower-mounted amplifier [5A]) and 7A (remote radio unit [5C]) with a nonlinear pre-distortion system 103 placed in line between an input 601/701 and an amplifier 603/703. A green arrow has been added against to highlight the '469 patent components being in tandem (i.e., in-line):

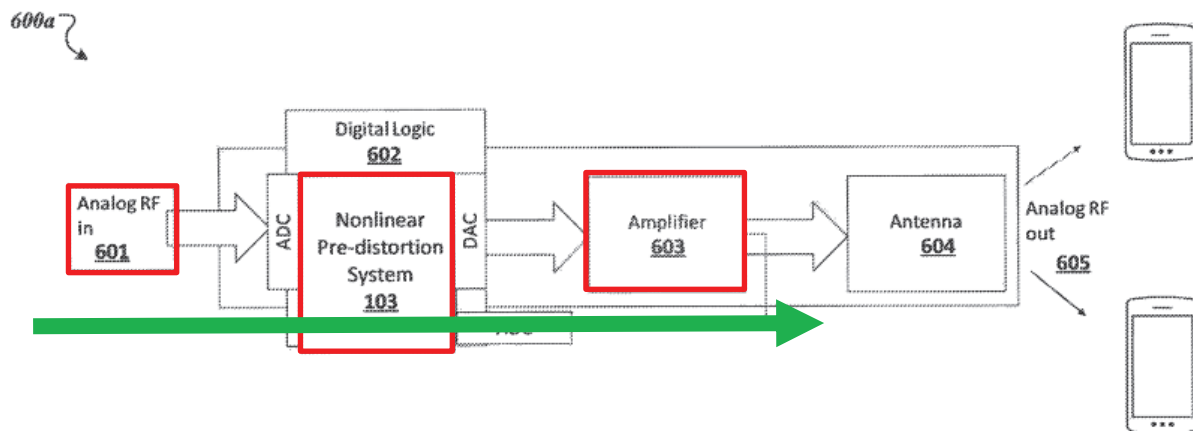


FIG. 6A

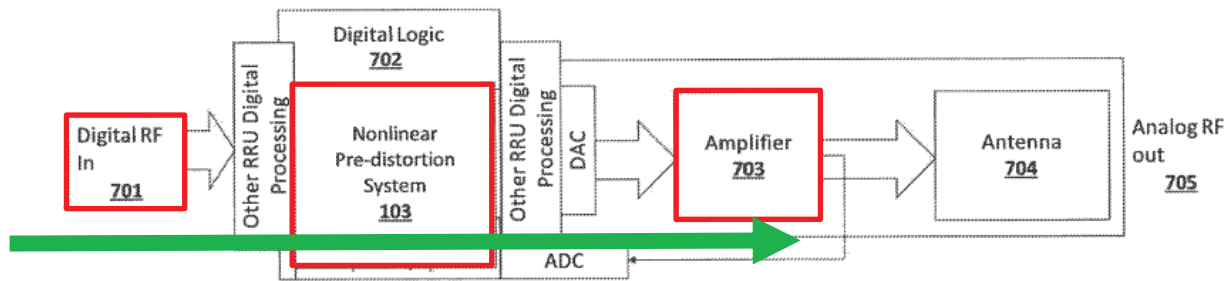


FIG. 7A

'469 patent, Figures 6A, 7A (annotations added).

218. As shown, a POSITA would have understood that Jüschke teaches *inline mode deployment scenarios*, both *inside a tower-mounted amplifier* and *inside a remote radio unit* based on Jüschke's flexible radio transmitter, because Jüschke teaches 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*). Therefore, in my opinion, a POSITA would have understood that Jüschke teaches limitations [5A] and [5C].

I. Jüschke in view of Holt teaches claim 6

219. In my opinion, Jüschke in view of Holt teaches "[t]he method of claim 1, wherein the model parameters are updated through a management process to

minimize the distance metric” [6]. I incorporate by reference my previous analysis for claim 1. *See* Sections VIII.D.1-6, *supra*.

220. As I previously explained for limitation [1E.1], both Jüschke and Holt teach a *distance metric* from training neural networks using input and output signals (Jüschke) and training and updating ML model parameters using loss functions (Holt). Jüschke implements a training system in which training processor 40 trains, in Jüschke’s *management process* to update model parameters, the trained parameters are transferred from neural network 43 to neural network 13. *See* Ex-1005, [0048]-[0049].

221. In addition, Jüschke teaches techniques for mitigating transmit signal impairments due to imperfections in wireless communications. *See* limitation [1P], *supra*; Ex-1005, [0001] (“mitigating impairments of a transmit signal due to transmitter imperfections”). Jüschke further teaches using weight and bias parameters for addressing I/Q-imbalance and PA non-idealities or non-linearity. *See* limitation [1B.2], *supra*. Furthermore, Jüschke’s weight and bias parameters are trained and updated to fit specific hardware using in wireless communications. *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.”).

222. It is therefore my opinion that a POSITA that Jüschke in view of Holt teaches *the model parameters are updated through a management process to minimize the distance metric*. A POSITA would have understood that Jüschke teaches that its weight and bias parameters (*model parameters*) are constantly updated via a training processor for fitting to specific hardware and used for addressing I/Q-imbalance and PA non-linearity in wireless communications (*updated through a management process to minimize the distance metric*). A POSITA would have understood that both Jüschke and Holt teach updating/training their model parameters through a *management process to minimize the distance metric*.

223. Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches claim 6.

J. Jüschke in view of Holt teaches claim 7

224. In my opinion, Jüschke in view of Holt teaches “[t]he method of claim 1, wherein correcting distortion of radio signals includes correcting interference of radio signals” [7]. I incorporate by reference my previous analysis for claim 1. See Sections VIII.D.1-6, *supra*.

225. As I previously explained for limitation [1P], Jüschke teaches “concept[s] for mitigating impairments of a transmit signal due to transmitter

imperfections” and “concept[s] for mitigating impairments due to transmit noise.” Ex-1005, [0001], Abstract. Jüschke discloses that “[a]nalog components in the transmit chain, which will also be referred to as transmitter devices in the sequel, typically cause various impairments because of their imperfect behavior. Most prominent non-idealities are carrier-frequency and sampling-rate offset, phase-noise, IQ-imbalance (I = In-phase signal component, Q = Quadrature signal component), and Power Amplifier (PA) nonlinearities. Especially the *mixer after the digital to analog converter*, which is responsible for mixing the baseband signal with the carrier signal, may cause so called IQ imbalance imperfections.” Ex-1005, [0004] (emphasis added). As I previously explained for limitation [1B.2], Jüschke’s weight and bias parameters address and correct IQ imbalance (and PA non-linearity).

226. Jüschke identifies such mixers after the digital to analog converter as a source of said IQ imbalance that it corrects. *See* Ex-1005, [0027] (Jüschke’s “parameters []addressing I/Q-imbalance and PA non-linearity.”). Indeed, “the input layer” of pre-distortion neural network 13 “may also comprise a plurality (e.g. three) delay elements connected in series to from a delay line, for delaying input samples of the in-phase and/or the quadrature component of the digital baseband transmit signal, respectively. This is particularly advantageous for mitigating IQ imbalance

impairments due to *analog transmitter devices*, as e.g. *IQ up-conversion mixers*, with memory effects.” Ex-1005, [0015] (emphasis added).

227. In my opinion, a POSITA would have understood that Jüschke’s pre-distortion neural network 13 addressing and correcting IQ imbalance teaches *correcting interference of radio signals*. As I previously explained, IQ imbalance is a form circuitry error in direct conversion radio receivers in which the in-phase (I) and quadrature (Q) signal components have mismatched phase and/or amplitude. *See* Section VI.A, *supra*. In my opinion, a POSITA would have understood that the IQ imbalance caused by IQ mixers after the digital to analog converter (that Jüschke corrects for) to be a form of signal interference. This understanding would have been reinforced by the ’469 patent, which states: “sources of distortion or *interference* include reflectors; bad connectors; *imperfections in the digital to analog conversion process*; unintended coupling, impedance mismatch or emissions for other electronics; unintended tones, intermodulation products or sidelobes; responses from other analog electronic components such as filters, power supplies, matching networks, junctions, attenuators, and *mixers*.” ’469 patent 5:43-50 (emphasis added). Hence, a POSITA would have understood the IQ mixers in Jüschke to be a source of signal interference that Jüschke’s pre-distortion neural network 13 corrects.

228. In addition, the '469 patent states that “[c]orrecting distortion of radio signals may include correcting interference of radio signals.” *Id.*, 2:64-65. This would have further reinforced to a POSITA that Jüschke’s pre-distortion neural network 13 *corrects* for distortion of radio signal distortion ***and interference of radio signals***. Therefore, it is my opinion that a POSITA would have understood that Jüschke’s distortion neural network 13, by correcting for sources of IQ-imbalance including imperfect analog transmitter devices such as, e.g., IQ up-conversion mixers, digital-to-analog converters (source of *interference*), teaches that *correcting distortion of radio signals includes correcting interference of radio signals*. Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches claim 7.

K. Jüschke in view of Holt teaches claim 8

229. As shown in the chart below, claim 8 includes limitations that are substantially the same as claim 1 and claim 6, except as discussed below. Therefore, I incorporate and apply the analysis of those prior claim limitations to claim 8. Highlighting has been added to show the shared limitations more clearly. In my opinion, a POSITA would have understood that that the teaching of Jüschke in view of Holt apply to claim 8 as they do to claims 1 and 6.

Claim 8	Claim Limitations Analysis Incorporated by Reference
[8P] A method for training a nonlinear pre-distortion model to correct distortion of radio signals, the method comprising:	[1P] A method for correcting distortion of radio signals, the method comprising:
[8A] obtaining a candidate transmit radio signal corresponding to output of a transmitting radio signal processing system;	[1A] obtaining a transmit radio signal corresponding to an output of a transmitting radio signal processing system;
[8B.1] processing the candidate transmit radio signal using a nonlinear pre-distortion machine learning model that executes using a set of model parameters to generate a pre-distorted candidate transmit radio signal;	[1B.1] generating a pre-distorted radio signal by processing the transmit radio signal using a nonlinear pre-distortion machine learning model, [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;
[8B.2] wherein the nonlinear pre-distortion machine learning model includes at least one nonlinear function to correct one or more of radio signal distortion or interference;	[1B.2] wherein the nonlinear pre-distortion machine learning model includes model parameters and at least one nonlinear function to correct radio signal distortion;
[8C] obtaining a received radio signal corresponding to input of a radio receiver by processing the pre-distorted candidate transmit radio signal using one or more sources of distortion or interference,	[1C] obtaining a transmit output radio signal by processing the pre-distorted radio signal through the transmitting radio signal processing system; [1D] transmitting the transmit output radio signal to one or more radio receivers; and [1E.1] updating the model parameters based on a distance metric computed between the transmit radio signal and a

Claim 8	Claim Limitations Analysis Incorporated by Reference
	received signal that is obtained from the one or more radio receivers,
[8D] computing a distance metric between the candidate transmit radio signal and the received radio signal,	[1E.1] updating the model parameters based on a distance metric computed between the transmit radio signal and a received signal that is obtained from the one or more radio receivers,
[8E.1] determining updated values for the set of model parameters of the nonlinear pre-distortion machine learning model that minimize the distance metric computed between the candidate transmit radio signal and the received radio signal;	[1E.1] updating the model parameters based on a distance metric computed between the transmit radio signal and a received signal that is obtained from the one or more radio receivers, [6] The method of claim 1, wherein the model parameters are updated through a management process to minimize the distance metric.
[8E.2] to correct for one or more sources of alteration of the candidate transmit radio signal in the transmitting radio signal processing system; and	[1P] A method for correcting distortion of radio signals, the method comprising: [1C] obtaining a transmit output radio signal by processing the pre-distorted radio signal through the transmitting radio signal processing system;
[8F] updating the values of the set of model parameters of the nonlinear pre-distortion machine learning model in order to train the nonlinear pre-distortion machine learning model.	[1E.1] updating the model parameters based on a distance metric computed between the transmit radio signal and a received signal that is obtained from the one or more radio receivers,

1. Jüschke in view of Holt teaches claim [8P]

230. In my opinion, Jüschke in view of Holt teaches “[a] *method for training a nonlinear pre-distortion model to correct distortion of radio signals*” [8P]. I incorporate by reference my previous analysis for claim 1.

231. As I previously explained for limitation [1P], Jüschke teaches *methods for correcting distortion of radio signals*, including “concept[s] for mitigating impairments of a transmit signal due to transmitter imperfections” and “transmit noise.” *Id.*, [0001], Abstract. In addition, as I explained for limitation [1E.1], Jüschke and Holt teach training techniques for neural networks. Therefore, in my opinion, Jüschke in view of Holt teaches [8P].

2. Jüschke in view of Holt teaches claim [8A]

232. In my opinion, Jüschke in view of Holt teaches “*obtaining a candidate transmit radio signal corresponding to output of a transmitting radio signal processing system*” [8A]. I incorporate by reference my previous analysis for claim 1. See Sections VIII.D.1-6, *supra*.

233. Claim limitation [8A] is nearly identical to [1A] and is taught by Jüschke in view of Holt for the reasons discussed for limitation [1A], *supra*. In my opinion, a POSITA would have understood that Jüschke’s input digital baseband transmit signal 12 teaches a *candidate transmit radio signal* for the same reasons it

teaches a *transmit radio signal*. Therefore, in my opinion, Jüschke in view of Holt teaches [8A].

3. Jüschke in view of Holt teaches claim [8B.1]

234. In my opinion, Jüschke in view of Holt teaches “*processing the candidate transmit radio signal using a nonlinear pre-distortion machine learning model that executes using a set of model parameters to generate a pre-distorted candidate transmit radio signal*” [8B.1]. I incorporate by reference my previous analysis for claim 1. See Sections VIII.D.1-6, *supra*.

235. Claim limitation [8B.1] shares many of the same limitations as [1B.1] (*generating a pre-distorted radio signal by processing the transmit radio signal using a nonlinear pre-distortion machine learning model*) and [1B.2] (*model parameters*) and is taught by Jüschke in view of Holt for the reasons discussed for limitations [1B.1] and [1B.2], *supra*. In my opinion, a POSITA would have understood that Jüschke’s pre-distorted baseband signal 16 teaches *generate a pre-distorted candidate transmit radio signal* for the same reasons it teaches *generating a pre-distorted radio signal*.

236. As explained for limitation [1B.2], Jüschke’s pre-distortion neural network model includes “weight and/or bias parameters (w_{kj} , b_k) of neurons.” Ex-1005, [0017]; see Ex-1005, [0013], [0014] (“A neuron is an information-processing

unit that is fundamental to the operation of the neural network. This the basic elements of the neuronal model are a set of synapses or collecting links, each of which is characterized by a weight was drawings of its own.”), [0042] (“The pre-distortion neural network processor 13 may hence comprise an input layer 21..., and an output layer 23 for each of the I and Q components of the pre-distorted digital baseband signal 16.”). In my opinion, a POSITA would have understood that these parameters are the weighted values of the neurons within the layers of pre-distortion neural network 13 used for generating pre-distorted baseband signal 16 (*executes using a set of model parameters to generate a pre-distorted candidate transmit radio signal*).

237. Therefore, in my opinion, Jüschke in view of Holt teaches [8B.1].

4. Jüschke in view of Holt teaches claim [8B2]

238. In my opinion, Jüschke in view of Holt teaches “*wherein the nonlinear pre-distortion machine learning model includes at least one nonlinear function to correct one or more of radio signal distortion or interference*” [8B.2]. I incorporate by reference my previous analysis for claim 1. *See* Sections VIII.D.1-6, *supra*.

239. Claim limitation [8B.2] is nearly identical to [1B.2] and is rendered obvious by Jüschke in view of Holt for the reasons discussed for limitation [1B.2], *supra*. In my opinion, a POSITA would have understood that Jüschke’s pre-

distortion neural network includes a nonlinear function for compensating nonlinear power amplifiers (*correcting radio signal distortion*). It is also my opinion that a POSITA would have understood that Jüschke's techniques for compensating nonlinear power amplifiers would also prevent interference caused by nonlinear signal distortion, such as signal splattering and interference caused by intermodulation (*nonlinear function to correct...interference*). See also '469 patent, 2:64-65 ("Correcting distortion of radio signals may include correcting interference of radio signals."); claim 7, *supra*.

240. Therefore, in my opinion, Jüschke in view of Holt teaches [8B.2].

5. Jüschke in view of Holt teaches claim [8C]

241. In my opinion, Jüschke in view of Holt teaches "*obtaining a received radio signal corresponding to input of a radio receiver by processing the pre-distorted candidate transmit radio signal using one or more sources of distortion or interference*" [8C]. Claim limitation [8C] shares many of the same limitations as [1C] (*obtaining a transmit output radio signal by processing the pre-distorted radio signal through the transmitting radio signal processing system*), [1D] (*to one or more radio receivers*), and [1E.1] (*received signal that is obtained from the one or more radio receivers*) and is taught by Jüschke in view of Holt for the reasons

discussed for limitations [1C], [1D], and [1E.1], *supra*. I incorporate by reference my previous analysis for claim 1. *See* Sections VIII.D.1-6, *supra*.

242. As I previously explained for limitations [1C] and [1D], Jüschke's analog transmitter device 14 obtains pre-distorted digital baseband signal 16 and converts and transmits the signal as an analog radio signal that is received by a receiver or transceiver. As explained for limitations [1E.1], Holt's transmitted communication data received by a decoder model implemented in a receiver device teaches a *received radio signal corresponding to input of a radio receiver*. *See also* [1E.2] (*received signal corresponding to the transmit output radio signal*). In my opinion, a POSITA also would have understood that the radio signal sent from Jüschke's transmitter device 14 would be the same signal received by a *radio receiver*.

243. Therefore, in my opinion, Jüschke in view of Holt teaches [8C].

6. Jüschke in view of Holt teaches claim [8D]

244. In my opinion, Jüschke in view of Holt teaches “*computing a distance metric between the candidate transmit radio signal and the received radio signal*” [8D]. Claim limitation [8D] shares many of the same limitations as [1E.1] (*updating the model parameters based on a distance metric computed between the transmit radio signal and a received signal*) and is taught by Jüschke in view of Holt for the

reasons discussed for limitation [1E.1], *supra*. I incorporate by reference my previous analysis for claim 1. *See* Sections VIII.D.1-6, *supra*.

245. In my opinion, a POSITA would have understood that from model parameters being *updated* by a distance metric that said distance metric would be calculated or *computed* as taught, e.g., by Holt’s loss function. Therefore, in my opinion, Jüschke in view of Holt teaches [8D].

7. Jüschke in view of Holt teaches claim [8E.1]

246. In my opinion, Jüschke in view of Holt teaches “*determining updated values for the set of model parameters of the nonlinear pre-distortion machine learning model that minimize the distance metric computed between the candidate transmit radio signal and the received radio signal*” [8E.1]. Claim limitation [8E.1] shares many of the same limitations as [1E.1] (*updating the model parameters based on a distance metric computed between the transmit radio signal and a received signal that is obtained from the one or more radio receivers*) and claim 6 (*minimize the distance metric*) and is taught by Jüschke in view of Holt for the reasons discussed for limitation [1E.1] and claim 6, *supra*. I incorporate by reference my previous analysis for claim 1 and claim 6. *See* Sections VIII.D.1-6, VIII.I, *supra*.

247. For example, Jüschke discloses “fit[ting] the feed-forward pre-distortion neural network to a specific downstream hardware component” using “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] (“[T]raining may take several epochs” until the model response is “the inverse of the transfer or distortion function of the analog transmitter device 14.”), [0019] (“As the training of the post distortion neural network processor converges, its response or transfer function will approximate the inverse transfer function of the non-ideal transmitter device's transfer function. In other words, the output layer of the post distortion neural network processor may be coupled to the in-phase and the quadrature component of the (undistorted) digital baseband transmit signal, while its input layer may be coupled to the output of the non-ideal analog transmitter device.”). Furthermore, Holt discloses that fourth loss function 318, which describes a difference between encoder model inputs and decoder model outputs, “backpropagate[s] through” and “modif[ies] 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, 18:43-48.

248. Therefore, in my opinion, Jüschke in view of Holt teaches [8E.1].

8. Jüschke in view of Holt teaches claim [8E.2]

249. In my opinion, Jüschke in view of Holt teaches “*to correct for one or more sources of alteration of the candidate transmit radio signal in the transmitting radio signal processing system*” [8E.2]. Claim limitation [8E.2] shares many of the same limitations as [1P] (*correcting distortion of radio signals*) and [1C] (*radio signal through the transmitting radio signal processing system*) and is taught by Jüschke in view of Holt for the reasons discussed for limitations [1P] and [1C], *supra*. I incorporate by reference my previous analysis for claim 1. *See* Sections VIII.D.1-6, *supra*.

250. For example, Jüschke discloses that the “pre-distortion neural network processor 13 may be adapted to compensate for a transmitter transfer function of a non-ideal analog transmitter device 14 which is causing an IQ imbalance,” Ex-1005, [0040], and “additionally or alternatively compensate a nonlinear transmitter transfer function of at least one analog amplifier device in the transmitter device 14,” *id.*, [0050]. *See id.*, [0009]. In my opinion, a POSITA would have understood that Jüschke’s neural network compensating for transmitter device imperfections (e.g., IQ imbalance and amplifier nonlinear distortion) to be *correcting for sources of alteration* from said sources of transmitter imperfections (e.g., circuitry errors in

direct conversion radio receivers causing IQ imbalance, nonlinear power amplifiers causing PA nonlinearities).

251. Therefore, in my opinion, Jüschke in view of Holt teaches [8E.2].

9. Jüschke in view of Holt teaches claim [8F]

252. In my opinion, Jüschke in view of Holt teaches “*updating the values of the set of model parameters of the nonlinear pre-distortion machine learning model in order to train the nonlinear pre-distortion machine learning model*” [8F]. Claim limitation [8F] shares many of the same limitations as [1E.1] (*updating the model parameters*) and is taught by Jüschke in view of Holt for the reasons discussed for limitation [1E.1], *supra*. I incorporate by reference my previous analysis for claim 1. See Sections VIII.D.1-6, *supra*.

253. As I previously explained for limitation [1E.1], Jüschke training its pre-distortion neural using undistorted inputs and sampled outputs and Holt training and updating ML model parameters using loss functions teaches *updating the model parameters based on a difference or distance metric*. In my opinion, a POSITA would have understood that Jüschke and Holt updating their model parameters likewise *trains their nonlinear pre-distortion machine learning models*. Therefore, in my opinion, Jüschke in view of Holt teaches [8F].

L. Jüschke in view of Holt teaches claim 10

254. In my opinion, Jüschke in view of Holt teaches “[t]he method of claim 8, wherein the model parameters are optimized through a model optimization process to minimize radio signal distortion or interference” [10]. I incorporate by reference my previous analysis for claim 8. *See* Sections VIII.K.1-9, *supra*.

255. As I previously explained for limitation [1E.1], Jüschke discloses “training weight and/or bias parameters (w_{kj} , b_k)” to fit the pre-distortion neural network to downstream hardware such as an amplifier. Ex-1005, [0017]. The parameters are trained “such that a transfer function of the post distortion neural network 43 approximates an inverse of the transmitter transfer function, including a *power amplifier's AM-AM and/or AM-PM distortion functions*, as well as an IQ up-conversion mixer's transfer function.” *Id.*, [0058]; *see also id.*, [0009] (The “neural network may [be] applied to the mitigation of transmitter impairments of a transmit signal.”).

256. To implement the foregoing, Jüschke discloses that the “training may take several epochs and may be terminated by MSE (Minimum Squared Error).” Ex-1005, [0048]. In my opinion, a POSITA would have understood that the Jüschke MSE termination criterion is an *optimization process* in support of convergence of the response of Jüschke’s neural networks for correcting radio signal distortion

caused by PA nonlinearity and IQ imbalance. *See* Ex Ex-1005, [0048] (“As the response of the post distortion neural network 43 converges, its response will essentially be the inverse of the transfer or distortion function of the analog transmitter device 14, which may comprise one or several IQ up-converter mixers, for example.”); [0046] (“As this distorts the pre-distorted IQ samples inversely, the output of the analog mixer device has approximately IQ balanced components.”), [0050] (“[T]he pre-distortion neural network processor 13 may be adapted to additionally or alternatively compensate a nonlinear transmitter transfer function of at least one analog amplifier device in the transmitter device 14.”). And as I previously explained for claim 7, a POSITA would have understood that Jüschke’s pre-distortion neural network 13 addressing and correcting IQ imbalance from mixers (a source of interference) to address and correct (or *minimize*) interference.

257. Therefore, in my opinion, Jüschke in view of Holt teaches claim 10.

M. Jüschke in view of Holt teaches claim 12

258. In my opinion, Jüschke in view of Holt teaches “[t]he method of claim 8, wherein the distance metric is selected from one of: a mean squared error, absolute error, or computed between the candidate transmit radio signal and the received radio signal to form a quantitative measurement of an error value” [12]. I

incorporate by reference my previous analysis for claim 8. *See* Sections VIII.K.1-9, *supra*.

259. As I previously explained for limitations [1E.1] and [1E.2], 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*. In addition, Jüschke uses the Levenberg-Marquardt algorithm, an algorithm for nonlinear least squares optimization, for training. *Id.*, [0049]. In my opinion, a POSITA would have understood that least squares optimization implements a *mean squared error*.

260. In addition to Jüschke, as I previously explained for limitations [1E.1] and [1E.2], Holt determines loss functions to calculate the distance between transmitted inputs and received outputs. *See* Ex-1006, 18:40-42 (“The fourth loss function 318 can describe a difference between the third set of inputs 316 and the sixth set of outputs 312.”). As I previously explained, a POSITA would have understood that mean squared error is a common example of a loss function. *See* Section V, *supra*; *see* Ex-1012, 55-57 (describing cost functions, equivalent to loss functions, including mean-square error, absolute error, and uniform cost function); Ex-1017, 2 (Discussing cost and loss functions. “Many neural network learning

algorithms explicitly minimize a cost function. The backpropagation technique, for example, uses a gradient descent algorithm for minimizing the ‘Mean Squared Error’ criterion.”). Therefore, in my opinion, a POSITA would have understood that Holt’s loss function in view of Jüschke teaches a *mean squared error* and an *absolute error*. Moreover, a POSITA would have understood that a fourth loss function 318, by describing a 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*.

261. Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches claim 12. A POSITA also would have been motivated to use Holt’s loss function with Jüschke for the reasons explained for limitations [1E.1] and [1E.2].

N. Jüschke in view of Holt teaches claim 13

262. In my opinion, Jüschke in view of Holt teaches “[t]he method of claim 8, wherein initial values for the model parameters are computed using a distance metric in a pre-deployment environment for a specific device or type of device” [13]. I incorporate by reference my previous analysis for claim 8. See Sections VIII.K.1-9, *supra*.

263. As I previously explained for limitation [1E.1], Jüschke trains its pre-distortion neural network model using a difference or *distance metric* between undistorted inputs and sampled outputs. Jüschke discloses training the network parameters using a training processor “to fit the feed-forward pre-distortion neural network to a specific downstream hardware component.” Ex-1005, [0017]. Only the pre-distortion neural network processor 13 is needed “[f]or initial offline training and online use of a fixed trained pre-distortion network.” Ex-1005, [0049]. In addition, the inverse transfer function of a transmitter device is approximated by post distortion neural network 43 and the parameters trained by network 43 may “be transferred to the pre-distortion neural network processor 13...once (e.g. ***at the beginning of operation***) or frequently.” Ex-1005, [0058] (emphasis added). In my opinion, a POSITA would have understood that Jüschke’s offline training is a *pre-deployment environment* given that training is off operating hardware, and hence, it computes the initial values *for the model parameters*. It is also my opinion that a POSITA also would have understood training approximates the inverse of a transmitter device, which accounts for the difference between input and output values.

264. Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches claim 13.

O. Jüschke in view of Holt teaches claim 14

265. In my opinion, Jüschke in view of Holt teaches “[t]he method of claim 8, wherein the candidate transmit radio signal meets a desired use case for a particular deployment scenario” [14]. I incorporate by reference my previous analysis for claim 8. *See* Sections VIII.K.1-9, *supra*.

266. As I previously explained for claim 2, Jüschke trains the weight and bias parameters for respective fragmented spectrum applications (e.g., f_1 and f_2) that “support[s] two frequency bands in parallel, using two mixers per frequency band, each and a power amplifier (dedicated or common).” *Id.*, [0072]; *see id.* [0028] (“Hence, the apparatus may be adapted for parallel operation of spectrally fragmented frequency bands, wherein the pre-distortion neural network processor is adapted to pre-distort the input digital baseband transmit signal to obtain a plurality of pre-distorted digital baseband signals, wherein each of the pre-distorted digital baseband signals corresponds to a different one of the fragmented frequency bands. In some embodiments each of the pre-distorted digital baseband signals may correspond to a different one of a plurality of analog transmitter devices, each of the analog transmitter devices corresponding to a different one of the plurality fragmented frequency bands.”). In my opinion, a POSITA would have understood that Jüschke’s digital baseband transmit signal 12 (*candidate transmit radio signal*)

is a mobile communications signal (e.g., 2G or GSM, 3G or WCDMA, and/or 4G or LTE) that meets the needs of downstream hardware equipment associated with said signal. *See* Ex-1005, [0059]. It is further my opinion that a POSITA also would have understood that Jüschke's flexible transmitter for fragmented spectrum applications processes signals for distinct deployments (e.g., different frequencies) associated with a given signal type, and therefore the input digital baseband transmit signal meets at least one of two distinct deployments as assessed by the neural network (*candidate transmit radio signal meets a desired use case for a particular deployment scenario*).

267. Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches claim 14.

P. Jüschke in view of Holt teaches claim 15

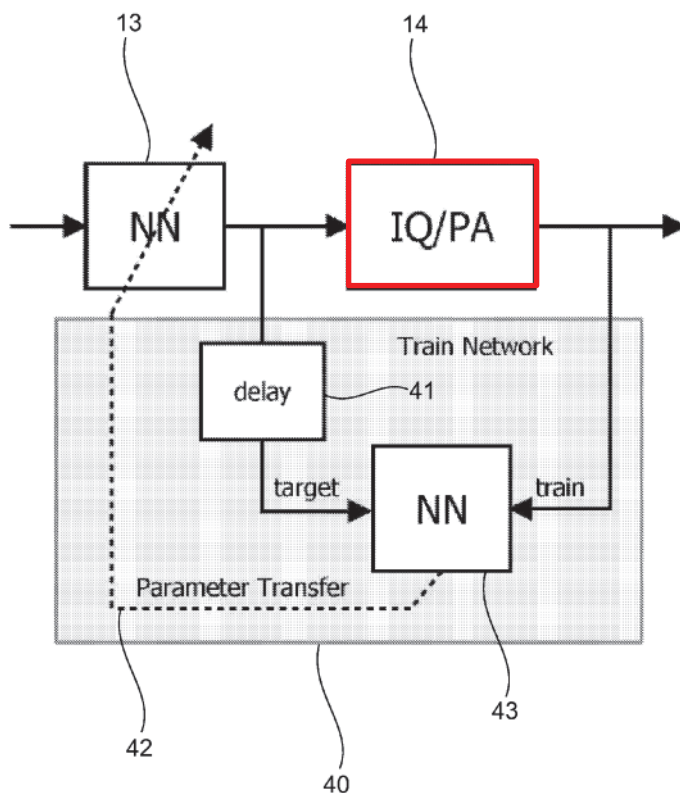
268. As shown in the chart below, claim 15 includes limitations [15A] and [15C] that are substantially the same as limitations [5A] and [5C] of claim 5. Claim 15, like claim 5, is also written in a Markush form, and I have been instructed by counsel to only analyze limitations [15A] and [15C]. Therefore, I incorporate and apply the analysis of those prior claim limitations to claim 15. For at least these reasons, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches claim 15.

Claim 15	Claim Limitations Analysis Incorporated by Reference
[15A] The method of claim 14, wherein the particular deployment scenario comprises <i>one of</i> : sending a radio transmission from a cellular base station that includes the nonlinear pre-distortion machine learning model inside a tower mounted amplifier in an inline mode;	[5A] The method of claim 2, wherein the particular deployment scenario comprises <i>one of</i> : sending a radio transmission from a cellular base station that includes the nonlinear pre-distortion machine learning model inside a tower mounted amplifier in an inline mode;
[15C] sending a radio transmission from a cellular base station that includes the nonlinear pre-distortion machine learning model inside a remote radio unit in an inline mode;	[5C] sending a radio transmission from a cellular base station that includes the nonlinear pre-distortion machine learning model inside a remote radio unit in an inline mode;

Q. Jüschke in view of Holt teaches claim 17

269. In my opinion, Jüschke in view of Holt teaches “[t]he method of claim 8, wherein the received radio signal is received after a wireless transmission from an antenna, or after a relay system receives, amplifies and retransmits the pre-distorted candidate transmit radio signal” [17]. I incorporate by reference my previous analysis for claim 8. *See* Sections VIII.K.1-9, *supra*.

270. As I previously explained for limitations [1C] and [1D], Jüschke’s analog transmitter device 14 obtains pre-distorted digital baseband signal 16 and converts and transmits the signal as an analog radio signal that is received by a receiver or transceiver. As shown in Figure 4 below, the transmitter device 14 includes a power amplifier (PA) module:



Ex-1005, Figure 4 (annotations added).

271. Jüschke recognizes that “higher-order modulated signals of current mobile communications standards like LTE or Wireless Local Area Networks (WLANs) pose high demands on radio systems regarding signal requirements...yields less efficiency in power amplifier modules.” Ex-1005, [0003]. In my opinion, a POSITA would have understood that the power amplifier (PA) module of the transmitter device 14 supports wireless radio system transmission, such as via LTE or WLAN communication standards. *See* Ex-1005, [0003]-[0004]. As I previously explained for limitation [1E.2], a POSITA would have understood that the radio signal sent from Jüschke’s transmitter device 14 would be the same

signal received by a *radio receiver*. Therefore, it is my opinion that a POSITA would have understood that Jüschke, on its own or in combination with Holt, sends a (pre-distorted) radio signal from a transmitter (*wireless transmission from an antenna*) that is received by a receiver or transceiver (*received radio signal is received*).

272. Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches claim 17.

R. Jüschke in view of Holt teaches claim 18

273. In my opinion, Jüschke in view of Holt teaches “[t]he method of claim 8, 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” [18]. I incorporate by reference my previous analysis for claim 8. *See* Sections VIII.K.1-9, *supra*.

274. As I previously 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...”). In my opinion, a POSITA would have understood that artificial neural networks include deep neural networks, convolutional neural

networks, and recurrent neural networks. Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches claim 18.

S. Jüschke in view of Holt teaches claim 19

275. In my opinion, Jüschke in view of Holt teaches “[t]he method of claim 8, wherein determining updated values for the model parameters includes using Stochastic Gradient Descent style optimizers that iteratively minimize the distance metric based on an exact or approximate gradient of the distance metric with regard to the model parameters” [19]. I incorporate by reference my previous analysis for claim 8. See Sections VIII.K.1-9, *supra*.

276. As I previously explained for limitation [1E.2], Jüschke determines the difference between input samples and output values and Holt determines loss functions to calculate the distance between transmitted encoder inputs and received decoder outputs. Jüschke also teaches the use of “[g]radient-based training algorithms, utilizing back-propagation” and the “Levenberg-Marquardt algorithm” which can also apply back-propagation. Ex-1005, [0018]. In my opinion, a POSITA would have understood that the Stochastic Gradient Descent is one of the most common gradient descent types used in backpropagation for optimizing and learning the model parameters, providing a basic approach for updating neural networks, as well as the preferred gradient type for neural network backpropagation. As shown

in the work of Ernst, Stochastic Gradient Descent provides a “more economical method for minimizing Q ... the error function associated with a network.” Ex-1016, 1. And as shown in the work of Bottou, “[t]he *stochastic gradient descent algorithm* [] has been shown to be faster, more reliable, and less prone to reach bad local minima than standard gradient descent” for minimizing the cost function (or loss function) for neural networks. Ex-1017, 2. A POSITA also would have understood that the Levenberg-Marquardt algorithm, as disclosed in Jüschke, uses both the Gradient and second order differentiations (Hessian). *See* Ex-1005, [0018] (“Hessian-based algorithms allow the neural network to learn more subtle features of a complicated mapping.”). A POSITA would have known that the Levenberg-Marquardt algorithm can also be utilized together with back-propagation. In addition, because both Jüschke and Holt teach backpropagation to train and update the parameters of inner layers in neural network models based on loss functions, a POSITA would have understood that the combination of Jüschke and Holt would implement Stochastic Gradient Descent without issue. *See* limitation [4C], *supra*. Therefore, in my opinion, a POSITA would have understood that Stochastic Gradient Descent, being the preferred gradient descent type used in backpropagation for training neural networks, to have been taught by Jüschke’s teaching the use of

“[g]radient-based training algorithms, like back-propagation” for neural networks in view of Holt. Ex-1005, [0018].

277. Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches claim 19. A POSITA would have been motivated to use Holt’s loss function with Jüschke for the reasons explained for limitation [1E.2].

T. Jüschke in view of Holt teaches claim 20

278. As shown in the chart below, claim 20 is substantially the same as claim 7. Therefore, I incorporate and apply the analysis of claim 7 to claim 20. For at least these reasons, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches claim 20.

Claim 20	Claim Limitations Analysis Incorporated by Reference
[20] The method of claim 8, wherein correcting distortion of radio signals includes correcting interference of radio signals.	[7] The method of claim 1, wherein correcting distortion of radio signals includes correcting interference of radio signals.

U. Jüschke in view of Holt teaches claims 21-22

279. As shown in the chart below, claims 21-22 are substantially the same as claims 1-2, except as discussed below. Therefore, I incorporate and apply the analysis of those prior claims to claims 21-22. For at least these reasons, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches claims 21-22. In my opinion, a POSITA would have understood that that Jüschke’s

teachings apply to system claims in addition to method claims. *See* Ex-1005, [0006] (“[A] high flexibility within the transceiver system has the advantage, that different parts of the system can be exchanged and adapted easily.”), claim 14 (“A method (70) for mitigating transmitter impairments of a transmit signal...”).

Claims 21-22	Claim Limitations Analysis Incorporated by Reference
[21P] A system comprising:	[1P] A method for correcting distortion of radio signals, the method comprising:
[21A] See below	
[21B] See below	
[21C] obtain a transmit radio signal corresponding to an output of a transmitting radio signal processing system;	[1A] obtaining a transmit radio signal corresponding to an output of a transmitting radio signal processing system;
[21D.1] generate a pre-distorted radio signal by processing the transmit radio signal using a nonlinear pre-distortion machine learning model,	[1B.1] generating a pre-distorted radio signal by processing the transmit radio signal using a nonlinear pre-distortion machine learning model,
[21D.2] wherein the nonlinear pre-distortion machine learning model includes model parameters and at least one nonlinear function to correct radio signal distortion;	[1B.2] wherein the nonlinear pre-distortion machine learning model includes model parameters and at least one nonlinear function to correct radio signal distortion;
[21E] obtain a transmit output radio signal by processing the pre-distorted radio signal through the transmitting radio signal processing system;	[1C] obtaining a transmit output radio signal by processing the pre-distorted radio signal through the transmitting radio signal processing system;
[21F] transmit the transmit output radio signal to one or more radio receivers; and	[1D] transmitting the transmit output radio signal to one or more radio receivers; and
[21G.1] update the model parameters based on a distance metric computed between the transmit radio signal and a	[1E.1] updating the model parameters based on a distance metric computed between the transmit radio signal and a

Claims 21-22	Claim Limitations Analysis Incorporated by Reference
received signal that is obtained from the one or more radio receivers,	received signal that is obtained from the one or more radio receivers,
[21G.2] the received signal corresponding to the transmit output radio signal that was transmitted to the one or more radio receivers.	[1E.2] the received signal corresponding to the transmit output radio signal that was transmitted to the one or more radio receivers.
[22] The system of claim 21, wherein the model parameters are updated with values for a particular deployment scenario for a radio communication system upon initial deployment of the nonlinear pre-distortion machine learning model into the radio communication system.	[2] The method of claim 1, wherein the model parameters are updated with values for a particular deployment scenario for a radio communication system upon initial deployment of the nonlinear pre-distortion machine learning model into the radio communication system.

V. Jüschke in view of Holt teaches claim 21

280. In my opinion, Jüschke in view of Holt teaches claim 21. As shown in the table above, *supra*, my previous analysis for [1P] applies to [21P] and my previous analysis for [1A]-[1E.2] applies [21C]-[21G.2]. For ease of explanation, I break the remaining limitations of claim 21 into separate limitations and organize them under sections [21A]-[21B], *infra*.

1. Jüschke in view of Holt teaches [21A]

281. In my opinion, Jüschke teaches “*one or more processors*” [21A].

282. As I previously explained for limitation [1P], Jüschke’s teachings implement processors. *See also* Ex-1005, Abstract (“pre-distortion neural network

processor (13) for pre-distorting the input digital baseband transmit signal (12; 12a; 12b)").

283. Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches limitation [21A].

2. Jüschke in view of Holt teaches [21B]

284. In my opinion, Jüschke teaches “*one or more storage devices storing instructions that are operable, when executed on the one or more processors, to cause the one or more processors to*” [21B].

285. As I previously explained for limitation [21A], Jüschke teaches processors implemented in its devices, including pre-distortion neural network processor 13. Jüschke further states that its teachings “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 that “[f]unctions of various elements shown in the figures, including any functional blocks 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].

286. Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches limitation [21B].

W. Jüschke in view of Holt teaches claims 23-24

287. As shown in the chart below, claims 23-24 are substantially the same as claims 1-2 and 21. Therefore, I incorporate and apply the analysis of those prior claims to claims 23-23. For at least these reasons, in my opinion, a POSITA would have understood that Jüschke in view of Holt teaches claims 23-24. In my opinion, a POSITA would have understood that that Jüschke’s teachings apply to apparatus claims in addition to method claims. *See* Ex-1005, [Abstract] (“Provided is apparatus (10) for mitigating transmitter impairments of a transmit signal.”), claim 1 (“An apparatus (10) for mitigating transmitter impairments of a transmit signal...”), claim 14 (“A method (70) for mitigating transmitter impairments of a transmit signal...”). I further note that the *or interference* limitation in [23B.2] is shared in [8B.2] and rendered obvious by Jüschke in view of Holt for the reasons discussed for limitation [8B.2]. *See* Section VIII.K.4, *supra*.

Claims 23-24	Claim Limitations Analysis Incorporated by Reference
[23P] One or more non-transitory computer-readable storage media comprising instructions stored thereon that are executable by a processing device and upon such execution cause the processing device to:	[1P] A method for correcting distortion of radio signals, the method comprising: [21A] one or more processors; and

Declaration of Zhi Ding, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 10,581,469

Claims 23-24	Claim Limitations Analysis Incorporated by Reference
	[21B] one or more storage devices storing instructions that are operable, when executed on the one or more processors, to cause the one or more processors to:
[23A] obtain a transmit radio signal corresponding to an output of a transmitting radio signal processing system;	[1A] obtaining a transmit radio signal corresponding to an output of a transmitting radio signal processing system;
[23B.1] generate a pre-distorted radio signal by processing the transmit radio signal using a nonlinear pre-distortion machine learning model,	[1B.1] generating a pre-distorted radio signal by processing the transmit radio signal using a nonlinear pre-distortion machine learning model,
[23B.2] wherein the nonlinear pre-distortion machine learning model includes model parameters and at least one nonlinear function to correct at least one of radio signal distortion or interference;	[1B.2] wherein the nonlinear pre-distortion machine learning model includes model parameters and at least one nonlinear function to correct radio signal distortion; [8B.2] wherein the nonlinear pre-distortion machine learning model includes at least one nonlinear function to correct one or more of radio signal distortion or interference;
[23C] obtain a transmit output radio signal by processing the pre-distorted radio signal through the transmitting radio signal processing system;	[1C] obtaining a transmit output radio signal by processing the pre-distorted radio signal through the transmitting radio signal processing system;
[23D] transmit the transmit output radio signal to one or more radio receivers; and	[1D] transmitting the transmit output radio signal to one or more radio receivers; and
[23E.1] update the model parameters based on a distance metric computed between the transmit radio signal and a received signal that is obtained from the one or more radio receivers,	[1E.1] updating the model parameters based on a distance metric computed between the transmit radio signal and a received signal that is obtained from the one or more radio receivers,

Declaration of Zhi Ding, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 10,581,469

Claims 23-24	Claim Limitations Analysis Incorporated by Reference
[23E.2] the received signal corresponding to the transmit output radio signal that was transmitted to the one or more radio receivers.	[1E.2] the received signal corresponding to the transmit output radio signal that was transmitted to the one or more radio receivers.
[24] The one or more non-transitory computer-readable storage media of claim 23, wherein the model parameters are updated with values for a particular deployment scenario for a radio communication system upon initial deployment of the nonlinear pre-distortion machine learning model into the radio communication system.	[2] The method of claim 1, wherein the model parameters are updated with values for a particular deployment scenario for a radio communication system upon initial deployment of the nonlinear pre-distortion machine learning model into the radio communication system.

IX. GROUND 2: CLAIMS 2, 22, AND 24 WERE OBVIOUS OVER JÜSCHKE IN VIEW OF HOLT AND DZIERWA.

288. As shown above for Ground 1, *supra*, in my opinion, Jüschke in view of Holt teaches claims 1-8, 10, 12-15, and 17-24. *See* Sections VIII.C-W, *supra*. It is also my opinion that, to the extent that Jüschke, on its own or in combination with Holt, does not teach claim 2 and similar claims 22 and 24, they and their respective dependent claims (3-5) would have been taught over Jüschke in view of Holt and Dzierwa, *infra*.

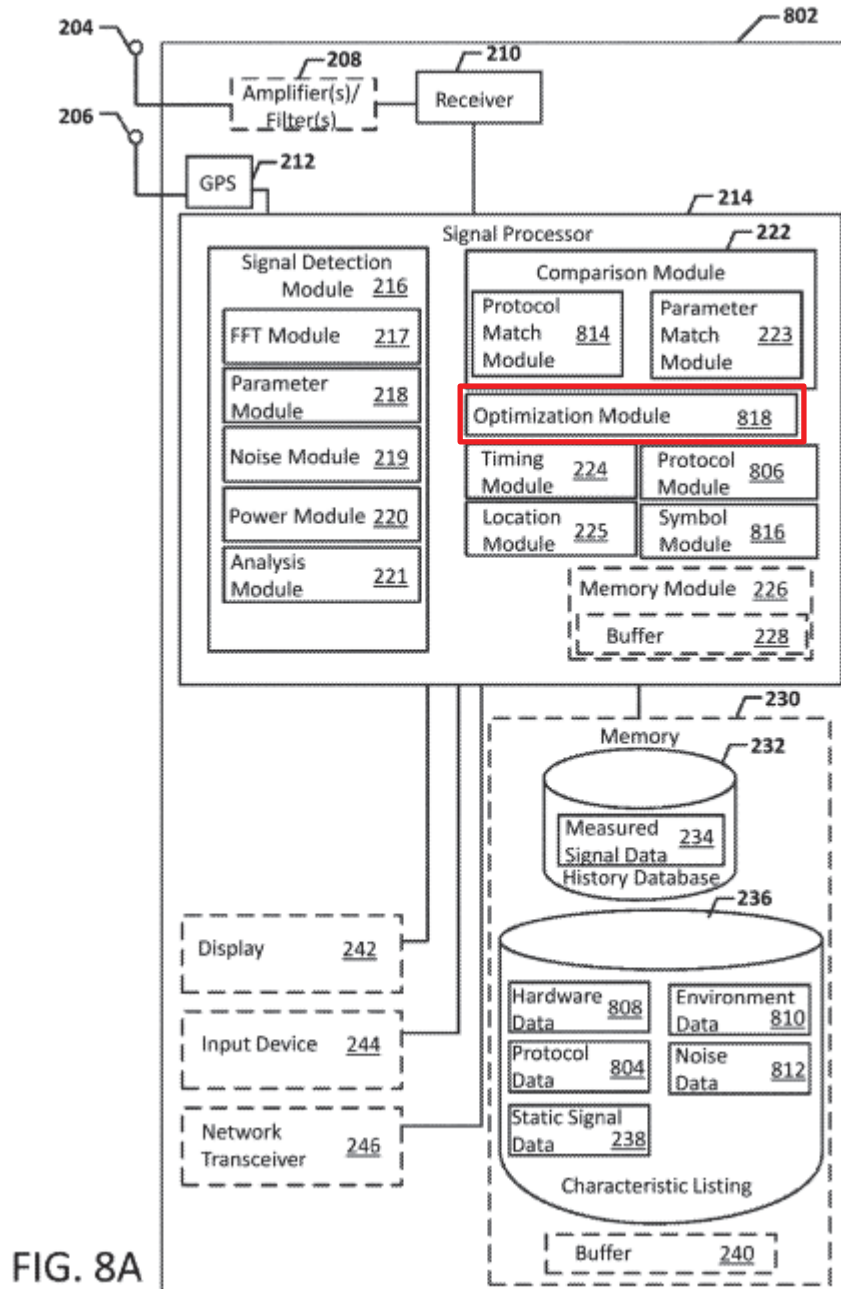
A. The Combined Neural Network Predistortion System: Modifying Jüschke’s Neural Network with Dzierwa’s Environment Data and Optimization.

289. The combined neural network predistortion system comprises implementing Dzierwa’s environment data, used for signal processing enhancements, into Jüschke’s data sampling for its pre-distortion neural network for mitigating transmit noise and signal impairments. In my opinion, this would have been a simple substitution well within the skill and knowledge of a POSITA.

290. As previously discussed, Dzierwa implements machine learning for “observ[ing] and learn[ing] an RF environment over time and identify[ing] temporal features,” and generating “[a] knowledge map...based on learning data.” Ex-1007, 47:38-41. Dzierwa further teaches an “optimization module 818” for “gather[ing] information from the characteristic listing, such as noise figure parameters, antenna

hardware parameters, and environmental parameters correlated with an identified signal of interest to calculate a degradation value for the 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 is illustrated below demonstrating optimization module 818:



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

291. In addition, Dzierwa teaches “a “processor 214 may output the tracking information” used to make signal transmission predictions in a wireless

environment. Ex-1007, 33:20. The “location of emitters, time and duration of transmission at a location may be stored in the history database such that historical information may be used to perform and predict movement of signal transmission,” and “the environmental factors may be considered to further reduce the measured error and generate a more accurate measurement of the location.” *Id.*, 33:23-29. Furthermore, Dzierwa “provide[s] for leveraging the use of 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 each of the at least one apparatus or device unit.” *Id.* 37:45-49.

B. Motivation for Modifying Jüschke’s Neural Network

292. In my opinion, a POSITA would have been motivated to modify Jüschke’s neural network using Dzierwa’s environment data and optimization module.

293. Jüschke and Dzierwa already use ML in wireless environments to improve signal processing performance (e.g., reduce distortion, noise, etc.). *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 figure parameters.”) *and id.*, 33:15-36 (“[T]he processor may filter the data to reduce interference.”) *and id.*, 27:33-34

(“communication protocol (e.g., LTE, CDMA, etc.) associated with a signal of interest”). Therefore, in my opinion, because Jüschke and Dzierwa are in identical fields and solve similar problems, a POSITA also would have been familiar with both and would have recognized that their elements could be interchanged and combined for optimizing the predistortion functionality of Jüschke’s neural networks.

294. In my opinion, a POSITA also would have recognized that combining Dzierwa’s RF environment data into Jüschke’s neural network and data sampling would have been a combination of familiar elements according to known methods and a simple substitution of one known element for another. A POSITA would have understood that Dzierwa’s “knowledge map...based on the *learning data*” and “extract[ion] [of] temporal features of [an] *RF environment* from the knowledge map,” Ex-1007, 9:19-21 (emphasis added), could be implemented in Jüschke’s input/output data sampling techniques and neural network. Therefore, a POSITA would have understood that implementing Dzierwa’s knowledge map and RF environment data would have predictably modified Jüschke’s sampling techniques for its neural networks.

295. Moreover, in my opinion, a 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). A POSITA would have understood that Dzierwa’s optimization module gathers “noise figure parameters, antenna hardware parameters, and environmental parameters correlated with an identified signal of interest” for a user “to optimize a signal of interest.” Ex-1007, 27:53-60. A 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” and neural network. Ex-1005, [0006]. Therefore, in my opinion, a POSITA would have been sufficiently motivated to implement Dzierwa’s optimization module and RF environment data in Jüschke’s transmitter and neural networks because doing so would have predictably improved transmitter performance using RF environment data for similar ML model devices.

C. Jüschke in view of Holt and Dzierwa teaches claim 2

296. As I previously stated, in my opinion, Jüschke in view of Holt teaches claim 2 (Ground 1). Should Jüschke be found not to teach claim 2 in view of Holt, however, it is also my opinion that claim 2 would have been taught in view of Dzierwa. I incorporate by reference my previous analysis for claim 2 under Ground 1. *See* Sections VIII.D-E, *supra*.

1. Jüschke in view of Holt and Dzierwa teaches claim 2

297. 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. In my opinion, a POSITA would have understood that Dzierwa’s knowledge map, being based on machine learning data, would include values used for deep learning and hence would be used to *update model parameters*. It is further my opinion that a POSITA would have understood that Dzierwa’s 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 nonlinear deep neural network to optimize predistortion *for a particular deployment scenario*.

298. A POSITA would have been sufficiently motivated to combine Jüschke and Dzierwa because they are both in the fields of wireless communications and machine learning and signal processing. Hence, a POSITA would have been familiar with both and combining them would have been a simple substitution of known elements. A 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 IX.B, *supra*. A POSITA also would have

understood that Dzierwa could be implemented in a combination of Jüschke and Holt for the same reasons that Dzierwa could be combined with Jüschke on its own. This would have been reinforced by Jüschke, Holt, and Dzierwa being in similar fields (wireless communications and machine learning and signal processing). Therefore, in my opinion, a POSITA would have understood that Jüschke in view of Holt and Dzierwa teaches claim 2.

D. Jüschke in view of Holt and Dzierwa teaches claim 3

299. As I have previously stated, in my opinion, Jüschke in view of Holt teaches claim 3. Therefore, it is also my opinion that Jüschke in view of Holt and Dzierwa teaches claim 3 for the same reasons. *See* Section VIII.F, *supra*.

E. Jüschke in view of Holt and Dzierwa teaches claim 4

300. As I previously stated, in my opinion, Jüschke in view of Holt teaches claim 4. Therefore, it is also my opinion that Jüschke in view of Holt and Dzierwa teaches claim 4 for the same reasons. *See* Section VIII.G, *supra*.

F. Jüschke in view of Holt and Dzierwa teaches claim 5

301. As I previously stated, in my opinion, Jüschke in view of Holt teaches limitations [5A] and [5C]. Therefore, it is also my opinion that Jüschke in view of Holt and Dzierwa teaches [5A] and [5C] for the same reasons. *See* Section VIII.H.1, *supra*.

G. Jüschke in view of Holt and Dzierwa teaches claims 22 and 24

302. As shown in the charts at Sections VIII.U and VIII.W, *supra*, claims 22 and 24 are substantially the same as claim 2. Therefore, I incorporate and apply the analysis of those prior claim limitations to those claims under Ground 2. I refer to charts at Section VIII.U (*comparing* claims 21-22 *to* claims 1-2) and Section VIII.W (*comparing* claims 23-24 *to* claims 1-2 and 21).

X. NO SECONDARY CONSIDERATIONS OF NON-OBVIOUSNESS

303. I understand from counsel that there is no underlying district court litigation, and that the Patent Owner has not otherwise yet identified any evidence with respect to secondary considerations of non-obviousness. I have independently performed an internet search on the Patent Owner and the '469 patent and found no publicly-available information indicating that the any of the claims of the '469 patent are non-obvious. I am unaware of any factual nexus supporting any of the claims of the '469 patent being novel.

304. To the extent the Patent Owner cites any evidence of sales or any praise or any industry recognition of products that the Patent Owner asserts to implement the claimed invention, I am not aware of any information demonstrating that any increased sales, commercial success, praise, recognition, or any other secondary factor (that the Patent Owner may assert) was a result of the particular features recited in the '469 patent's claims.

305. I refer to and incorporate my opinions stated throughout this Declaration, including my analysis showing that the '469 patent is directed to techniques known in the prior art and does not provide any inventive technology. *See* Sections VIII-IX, *supra*. As I have previously stated, a POSITA would have found combining prior art references Jüschke, Holt, and Dzierwa to have involved a

simple substitution of known elements, techniques, and methods in the same field and would have predictably provided known improvements with a reasonable expectation of success. *See* Sections VIII.C, IX.B, *supra*.

306. To the extent the Patent Owner at a later date cites or provides any other evidence regarding secondary considerations, including evidence of commercial success of the claimed invention, long felt but unsolved need, failure of others, copying of the claimed invention, unexpected results of the claimed invention, praise of the claimed invention, and/or surprised or skepticism of at the claimed intention and any expert opinions, I reserve the right to supplement my analysis and opinions to comment on it.

XI. CONCLUSION

307. For at least the preceding reasons, in my opinion, claims 1-8, 10, 12-15, and 17-24 of the '469 patent would have been obvious to a person of ordinary skill in the art based on the prior art references that I have cited and analyzed above.

308. In signing this Declaration, I recognize that the Declaration will be filed as evidence in a contested case before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I also recognize that I may be subject to cross-examination in this proceeding. If required, I will appear for cross-examination at the appropriate time. I reserve the right to offer opinions relevant to the invalidity

Declaration of Zhi Ding, Ph.D. in Support of
Petition for *Inter Partes* Review of
U.S. Patent No. 10,581,469

of the '469 patent claims at issue and/or offer testimony in support of this Declaration.

I hereby declare that all statements made herein of my own knowledge are true and that all statements are made on information and belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. §1001.

Dated: October 14, 2024

Respectfully submitted,



Digitally signed by Zhi
Ding
Date: 2024.10.14 15:09:02
-07'00'

Zhi Ding, Ph.D.