

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

SAMSUNG ELECTRONICS CO., LTD.
Petitioners

v.

NETWORK SYSTEM TECHNOLOGIES, LLC
Patent Owner

Patent No. 8,086,800

Inter Partes Review No. IPR2024-00339

DECLARATION OF DR. ERIK CHMELAR

TABLE OF CONTENTS

I. INTRODUCTION.....	5
II. EXPERT QUALIFICATIONS AND BACKGROUND	5
III. UNDERSTANDING OF LEGAL PRINCIPLES.....	10
A. Validity Generally	10
B. Prior Art.....	11
C. Anticipation	12
D. Obviousness.....	13
IV. PERSON OF ORDINARY SKILL IN THE ART.....	17
V. OVERVIEW OF THE TECHNICAL BACKGROUND AND THE ‘800 PATENT	17
A. Technical Background.....	17
B. Challenges of On-Chip Networks	26
C. Overview of the ’800 Patent.....	35
D. Claim Constructions	41
VI. ANALYSIS OF THE ’800 PATENT	43
A. Ground 1: Radulescu in View of Bahadiroglu Does Not Render ’800 Patent Claim 12 Obvious.....	43
1. Radulescu in view of Bahadiroglu fails to disclose “wherein the first interconnect module includes a first memory for buffering first data from the first processing module, the first interconnect module further including a first determination unit, wherein the first determination unit is configured to determine a first optimal amount of data to be buffered by the first memory, and the second interconnect module includes a second memory for buffering second data from the second processing module, the second interconnect module further including a second determination unit, wherein the second determination unit is configured to determine a second optimal amount of data to be buffered by the second memory”	43
2. Radulescu in view of Bahadiroglu fails to disclose “the first interconnect module transferring the first data to the second processing module when the first data buffered in the first memory reaches the first optimal amount, and the	

second interconnect module transferring the second data to the first processing module when the second data buffered in the second memory reaches the second optimal amount”.....	53
3. Radulescu in view of Bahadiroglu fails to disclose “wherein at least one of the first determination unit and the second determination unit is further configured to determine an optimal moment for sending the data in said first wrapper or said second wrapper according to communication properties of the communication between the master and the slave”	58
4. Analysis of Motivation to Combine.....	64
B. Ground 2: Radulescu in View of Petersen Does Not Render ’800 Patent Claim 12 Obvious.....	67
1. Radulescu in view of Petersen fails to disclose “wherein the first interconnect module includes a first memory for buffering first data from the first processing module, the first interconnect module further including a first determination unit, wherein the first determination unit is configured to determine a first optimal amount of data to be buffered by the first memory, and the second interconnect module includes a second memory for buffering second data from the second processing module, the second interconnect module further including a second determination unit, wherein the second determination unit is configured to determine a second optimal amount of data to be buffered by the second memory”	67
2. Radulescu in view of Petersen fails to disclose “the first interconnect module transferring the first data to the second processing module when the first data buffered in the first memory reaches the first optimal amount, and the second interconnect module transferring the second data to the first processing module when the second data buffered in the second memory reaches the second optimal amount”.....	75
3. Radulescu in view of Petersen fails to disclose “wherein at least one of the first determination unit and the second determination unit is further configured to determine an optimal moment for sending the data in said first wrapper or said second wrapper according to communication properties of the communication between the master and the slave”	81
4. Analysis of Motivation to Combine.....	85
C. Ground 3: Radulescu in View of Bahadiroglu and/or Radulescu-WO Does Not Render ’800 Patent Claim 12 Obvious.....	88

1. Radulescu-WO fails to disclose “wherein the first interconnect module includes a first memory for buffering first data from the first processing module, the first interconnect module further including a first determination unit, wherein the first determination unit is configured to determine a first optimal amount of data to be buffered by the first memory, and the second interconnect module includes a second memory for buffering second data from the second processing module, the second interconnect module further including a second determination unit, wherein the second determination unit is configured to determine a second optimal amount of data to be buffered by the second memory”	89
2. Radulescu-WO fails to disclose “the first interconnect module transferring the first data to the second processing module when the first data buffered in the first memory reaches the first optimal amount, and the second interconnect module transferring the second data to the first processing module when the second data buffered in the second memory reaches the second optimal amount”	89
3. Radulescu-WO in view of Petersen fails to disclose “wherein at least one of the first determination unit and the second determination unit is further configured to determine an optimal moment for sending the data in said first wrapper or said second wrapper according to communication properties of the communication between the master and the slave”	89
D. Ground 4: Radulescu in View of Peterson and/or Radulescu-WO Does Not Render ’800 Patent Claim 12 Obvious	90
VII. CONCLUSION	90

I. INTRODUCTION

1. I have been retained on behalf of Plaintiff Network System Technologies, LLC (“NST”), as an expert consultant.

2. As part of this work, I have been requested by counsel for NST to provide this declaration concerning the technical subject matter relevant to the *inter partes* review (“IPR”) petition of U.S. Patent No. 8,086,800 (the “’800 Patent”) filed by Samsung Electronics Co., Ltd. (“Samsung”) (“Petitioner”).

3. I have provided an exemplary, but not exhaustive, summary of materials that I considered or relied upon in preparing this report, besides my personal knowledge and experience, in Appendix 2. The paragraphs below may identify additional materials considered or relied upon not listed in Appendix 2. I reserve the right to rely on additional documents as I become aware of them and to respond to documents relied on by Defendant’s expert.

4. I am being compensated for my time in the amount of \$600 per hour. I have no financial interest in NST or the Asserted Patents, and my compensation is not contingent upon the substance of my declaration, any statements or opinions made, or the outcome of this matter.

II. EXPERT QUALIFICATIONS AND BACKGROUND

5. In 2000, I received a B.S. degree in Electrical Engineering from Michigan Technological University with the highest distinction (summa cum

laude). In 2001, I received an M.S. degree in Electrical Engineering from Stanford University, with a focus on computer networks (EE284 Introduction to Computer Networks, EE384A Internet Protocols, EE384S Network Architectures, and EE384S Multimedia Networks). In 2004, I received a Ph.D. degree in Electrical Engineering from Stanford University, with a primary focus on very large-scale integrated (VLSI) systems that culminated in a dissertation entitled “The Test and Diagnosis of Field-programmable Gate Arrays, advised by Dr. Edward J. McCluskey, and with conclusion of a secondary focus on computer networks (EE384X Packet Switch Architectures). While studying at Stanford University from 2000–2004, I taught undergraduate and graduate level courses as a Teaching Assistant (EE284 Introduction to Computer Networks, EE384A Internet Protocols and Standards, Digital Design Laboratory EE121, and E40 Introduction to Electronics), and I was employed as a Research Assistant in the Center for Reliable Computing. In 2011, I received a second B.S. degree in Chemical Physics from Saginaw Valley State University in 2011 with the highest distinction (summa cum laude).

6. During my undergraduate and graduate studies, I was additionally employed at several product-technology companies. From 1999–2002, I was employed as a Manufacturing Engineer at Cisco Systems, Inc, where I oversaw functional testing of layer-2 and layer-3 network switches and determined root-cause

failures thereof. During the summers of 2002 and 2003, I was employed as a Product Test Engineer intern at Xilinx, Inc., where I researched and developed software tools to detect and diagnose physical hardware defects in FPGAs.

7. After earning my Ph.D., I was employed at LSI Corp. (now Broadcom, Inc.) from 2004–2012 as a Staff Software Engineer in the Advanced Technology Development department, where I conducted R&D for very large-scale integrated (VLSI) systems and serializer-deserializer (SerDes) circuits for high-speed data communications, for which I was granted approximately 20 U.S. patents. During this period of employment at LSI Corp., I also taught at several Universities. From 2006–2008, I was a consulting Assistant Professor in the Electrical Engineering and Computer Science department at Stanford University, where I advised PhD students in the Center for Reliable Computing on the test and reliability of VLSI systems, and created and taught a PhD-level course on clock and data recovery (CDR) for digital communications (EE391 Clock and Data Recovery Architectures). From 2009–2012, I was an Adjunct Professor in the Electrical and Computer Engineering department at Saginaw Valley State University, where I taught several undergraduate courses in data communications and computer hardware (ECE455 Data Communications, ECE368 Computer Engineering, ECE368L Computer Engineering Lab, ECE216 Digital Circuits, and ECE216L Digital Circuits Lab).

8. From 2012–2015, I was employed at Stryker Instruments as a Senior Principal Engineer in the Advanced Development department, where I conducted and supervised R&D of computationally intelligent medical instruments and surgical technologies, and I delivered working prototypes for bio-photonic devices and boundary-constraint surgical robotics.

9. From 2015–2018, I was employed at Interdigital, Inc. as a Senior Manager of Innovation Partners, where I conducted and supervised R&D of 5th generation wireless (5G) network-enabled technologies including the Internet of Things, connected/autonomous vehicles, and augmented/virtual reality infrastructure, and I managed scientists and engineers to create over 100 new-invention disclosures for patent protection.

10. From 2019–present, I have provided consulting services to a major patent-licensing corporation as a Technical Expert in the areas of semiconductor circuits, microchips, and data communications. My consulting services have included determining whether and how specific patent claims within the aforementioned areas read on various products or product features. I analyze patents, technical specifications, product white papers, and product marketing materials to produce detailed claim charts used by attorneys to formulate infringement opinions and licensing decisions. I have also performed technology market analysis and

developed detailed technology landscapes and roadmaps in the aforementioned areas.

11. I have written technical publications dating back to 2003 on topics in the areas of semiconductor circuits, microchips, and data communications. I have written technical articles published in peer reviewed professional journals, trade journals, and proceedings of peer-reviewed conferences. I have presented at several technical symposia and technical conferences on topics in the aforementioned areas.

12. I am or have been a member or senior member of several professional organizations, including the Institute of Electrical & Electronic Engineers (IEEE), the Association for Computing Machinery (ACM), the American Society of Mechanical Engineers (ASME), the International Society of Optics & Photonics (SPIE), and the American Chemical Society (ACS).

13. I served as a Program Evaluator for the Accreditation Board of Engineering and Technology (ABET) from 2013–2018. As a Program Evaluator, I was responsible for determining ABET accreditation for university engineering programs categorized as Electrical Engineering programs and Computer Engineering programs.

14. I served as an Industry Mentor for National Science Foundation (NSF) I-Corps program in 2013, and served as an Instructor for the same in 2015. In these

roles, I provided industry and technical guidance to engineers, inventors, and entrepreneurs launching technology startups.

15. I hold approximately 30 U.S. Patents, many of which concern technologies in the areas of semiconductor circuits, microchips, and data communications.

16. Additional details on my background and experience can be found in my *curriculum vitae*, attached hereto as Appendix 1.

III. UNDERSTANDING OF LEGAL PRINCIPLES

A. Validity Generally

17. I have a general understanding of validity based on my experience with patents, my career experience, and my discussions with counsel. I have also reviewed the legal principles used to determine invalidity, such as those set forth in the AIPLA Model Patent Jury Instructions. *See, e.g.,* <https://www.aipla.org/home/news-publications/model-patent-jury-instructions>).

18. I understand that the claims of a patent are presumed valid. I further understand that Defendant has the burden of proving invalidity by clear and convincing evidence and that Defendant cannot meet this burden without at least providing with specificity the basis of Defendant's invalidity contention.

B. Prior Art

19. I have a general understanding of prior art and priority dates based on my experience with patents, my career experience, and my discussions with counsel.

20. I understand that inventors may be entitled to a priority date earlier than an actual date of filing of a patent application that provides written description support for a particular claim to the extent that they can show a complete possession of such a particular claimed invention at such an earlier priority date, and reasonable diligence to reduce such a particular claimed invention to practice between such an earlier priority date and such an actual date of filing. I understand that if Plaintiff contends that particular claims are entitled to an earlier priority date than the actual date of filing, then Plaintiff has the burden to prove this contention with specificity.

21. I understand that an invention by another must be made before the priority date of a particular patent claim in order to qualify as “prior art” under 35 U.S.C. § 102 or 35 U.S.C. § 103. I also understand that a printed publication or a product usage must be publicly available before the priority date of a particular patent claim in order to qualify as “prior art” under 35 U.S.C. § 102(a), that a printed publication or a product usage or offer for sale must be publicly available in the United States more than one year prior to the actual date of filing of a patent application that provides written description support for a particular claim in order to qualify as “prior art” under 35 U.S.C. § 102(b), and that the invention by another

must be described in an application for patent filed in the United States before the priority date of a particular patent claim in order to qualify as “prior art” under 35 U.S.C. § 102(e). I understand that Defendant has the burden of proving that any particular reference or product usage or offer for sale is prior art.

C. Anticipation

22. I have a general understanding of anticipation based on my experience with patents, my career experience, and my discussions with counsel.

23. I understand that the anticipation analysis is a two-step process. The first step is to determine the meaning and scope of the asserted claims. Each claim must be viewed as a whole, and it is improper to ignore any element of the claim. For a claim to be anticipated under U.S. patent law: (1) each and every claim must be identically disclosed, either explicitly or inherently, in a single prior art reference; (2) the claim elements disclosed in the single prior art reference must be arranged in the same way as in the claims; and (3) the identical invention must be disclosed in the single prior art reference, in as complete detail as forth in the claim. Where even one element is not disclosed in a reference, the anticipation contention fails. Moreover, to serve as an anticipatory reference, the reference itself must be enabled. In other words, the reference must provide enough information so that a person of ordinary skill in the art (“POSITA”) can practice the subject matter of the reference without undue experimentation.

24. I further understand that where a prior art reference fails to explicitly disclose a claim element, the prior art reference inherently discloses the claim element only if the prior art reference must necessarily include the undisclosed claim element. Inherency may not be established by probabilities or possibilities. The fact that an element may result from a given set of circumstances is not sufficient to prove inherency. I have applied these principles in forming my opinions in this matter.

25. I further understand that a dependent claim incorporates each and every limitation of the claim from which it depends. Thus, my understanding is that if a prior art reference fails to anticipate an independent claim, then that prior art reference also necessarily fails to anticipate all dependent claims that depend from that independent claim.

D. Obviousness

26. I have a general understanding of obviousness based on my experience with patents, my career experience, and my discussions with counsel.

27. I understand that a patent claim is invalid under 35 U.S.C. § 103 as being obvious only if the differences between the claimed invention and the prior art such that the subject matter as a whole would have been obvious at the time of invention to a POSITA. An obviousness analysis requires consideration of four factors: (1) scope and content of the prior art relied upon to challenge patentability; (2) differences between the prior art and the claimed invention; (3) level of ordinary skill

in the art at the time of the invention; and (4) the objective evidence of non-obviousness, such as commercial success, unexpected results, the failure of others to achieve the results of the invention, a long-felt need which the invention fills, copying of the invention by competitors, praise for the invention, skepticism for the invention, or independent development.

28. I understand that a prior art reference is proper to use in an obviousness determination if the prior art reference is analogous art to the claimed invention. I understand that a prior art reference is analogous art if at least one of the following two circumstances is met. First, a prior art reference is analogous art if it is from the same field of endeavor as the claimed invention, even if the prior art reference addresses a different problem and/or arrives at a different solution. Second, a prior art reference is analogous art if the prior art reference is reasonably pertinent to problem faced by the inventor, even if it is not in the same field of endeavor as the claimed invention.

29. I understand that it must be shown that a POSITA at the time of the invention would have had a reasonable expectation that a modification or combination of one or more prior art references would have succeeded. Furthermore, I understand that a claim may be obvious in view of a single prior art reference, without the need to combine references, if the elements of the claim that are not found in the reference can be supplied by the knowledge or common sense of one of

ordinary skill in the relevant art. However, I understand that it is inappropriate to resolve obviousness issues by a retrospective analysis or hindsight reconstruction of the prior art and that the use of such hindsight reconstruction is improper in analyzing the obviousness of a patent claim.

30. I further understand that the law recognizes several specific guidelines that inform the obviousness analysis. First, I understand that a hindsight reconstruction approach is not permitted. In other words, it is improper to use post-invention information to help perform the selection and combination of prior art references, and it is also improper to use the listing of elements in a claim as a blueprint to identify selected portions of different prior art references in an attempt to show that the claim is obvious.

31. Second, I understand that any prior art that specifically teaches away from the claim subject matter points to non-obviousness. In other words, prior art that would lead a POSITA to a specifically different solution than the claimed invention points to non-obviousness. Conversely, I understand that any prior art that contains any teaching, suggestion, or motivation to modify or combine such prior art references points to obviousness of such a modification or combination.

32. Third, while many combinations of the prior art may be “obvious to try,” I understand that any obvious-to-try analysis will not render a patent invalid unless it is shown that the possible combinations are: (1) sufficiently small in number so as to

be reasonable to include that the combination would have been selected; and (2) such that the combination would have been believed to be one that would produce predictable and well understood results.

33. Fourth, I understand that if a claimed invention that arises from the modification or combination of one or more prior art references uses known methods or techniques that yield predictable results, then that factor also points to obviousness.

34. Fifth, I understand that if a claimed invention that arises from the modification or combination of one or more prior art references is the result of known work in one field prompting variations of it for use in the same field or a different one based on design incentives or other market forces that yield predictable variations, then that factor also points to obviousness.

35. Sixth, I understand that if a claimed invention that arises from the modification or combination of one or more prior art references is the result of routine optimization, then that factor also points to obviousness.

36. Seventh, I understand that if a claimed invention that arises from the modification or combination of one or more prior art references is the result of a substitution of one known prior art element for another known prior art element to yield predictable results, then that factor also points to obviousness.

37. I further understand that a dependent claim incorporates each and every limitation of the claim from which it depends. Thus, my understanding is that if a

prior art reference fails or a combination of prior art references fail to render obvious an independent claim, then that prior art reference fails or that combination of prior art references fail to render obvious all dependent claims that depend from the independent claim.

IV. PERSON OF ORDINARY SKILL IN THE ART

38. I understand that, in assessing the level of ordinary skill, one may consider several factors, including: (1) the education level of the inventor; (2) the education level of active workers in the field; (3) the types of problems encountered in the art; (4) prior art solutions to these problems; (5) the rate of innovation in the field; and (6) the sophistication of the technology.

39. I understand that the Court adopted the following identification of a person of ordinary skill in the art (“POSITA”): “a POSITA would have at least a Bachelor’s degree in electrical engineering, computer engineering, or a related field, and as at least two years of work experience in the field of SoC and NoC technologies.” EX2022, 18. I met and exceeded this level of skill as of 2002. I have applied the Court’s definition of a POSITA throughout this Report.

V. OVERVIEW OF THE TECHNICAL BACKGROUND AND THE ‘800 PATENT

A. Technical Background

40. A network-on-chip (“NoC”) is a data communication network on a microchip. In some ways, NoCs are like traditional computer networks, such as local

area networks (“LANs”) and/or wide area networks (“WANs”). For example, both traditional computer networks and NoCs interconnect devices or components: traditional computer networks commonly interconnect discrete devices, such as computers, printers, servers, and so on, whereas NoCs commonly interconnect integrated components, such as processors, cores, caches, memories, input/output (I/O) interfaces, peripherals, and so on. However, implementation of a NoC on a SoC presents many unique challenges, as discussed in more detail below. *Infra*, § VI.B. The components interconnected by a NoC are typically integrated on a very large-scale integrated (“VLSI”) circuit, such as a system-on-chip (“SoC”).

41. An SoC is an integrated circuit in which most or all components of a system are integrated, including memories, processors, high-speed I/O interfaces, and dedicated application-specific logic. *See, e.g.*, EX2003, 1; *see also* EX2004, 4-5. By the early 2000s, certain limitations of statically wired interconnections between components of an SoC were known, and as a result, NoCs were already emerging. *See, e.g.*, EX2004, 7-8, *see also* EX2005. Rather than considering on-chip interconnections as point-to-point wires, they can be abstracted as communication channels over which to transmit data within a “quality-of-service” setting. *See id.*

42. Before describing details of a NoC, it is important to describe how components can be interconnected in an SoC that does not implement a NoC. Interconnections, sometimes called interconnects, traces, wires, and signal lines, are

electrical connections between a first component and a second component. Interconnects that carry similar signals are grouped together into buses, for example interconnects that carry address signals are grouped into an address bus and interconnects that carry data signals are grouped into a data bus. *See* EX2006, 4. A bus can be considered a direct interconnection between components (e.g., the path between a first component and a second component is always the same), where the interconnection may be dedicated or shared. A dedicated bus comprises point-to-point interconnection between components, typically between two components. A shared bus comprises interconnections that can be shared between many components, and requires some type of bus arbitration to resolve conflicts when multiple components attempt to access the shared bus simultaneously.

43. There are several types of bus arbitration methods for shared buses, including centralized and distributed (sometimes called decentralized). In centralized arbitration, a single device, known as a bus controller or bus arbiter, is responsible for managing access to the bus. In distributed arbitration, components compete for access to the bus by sending a request signal to the bus arbiter and waiting for a grant signal. In some distributed arbitration methods, e.g., decentralized, each component has its own priority level, and the component with the highest priority is given access to the bus. A component capable of originating a bus transaction (e.g., requesting control of the bus) is a master component, and a component that cannot originate a

bus transaction (e.g., only listens on the bus for communications intended for it) is a slave component. In large and/or complex SoCs, such shared buses interconnecting master devices and slave devices via bus arbitration may not be sufficiently scalable nor provide enough bandwidth. *See* EX2007, 7, 235.

44. In particular, with increasing quantities of components integrated in SoCs, increasing transistor densities, and increasing clock frequencies, it becomes very difficult to operate all the components synchronously within a same clock domain. This creates a need for a modular approach, where the processing system of the SoC comprises a plurality of relatively independent complex modules. In conventional processing systems, the system modules usually communicate with each other via a shared bus as described above. But as the number of modules increases, shared-bus communication is no longer practical at least because: (1) the large quantity of modules creates too high of an electrical load on the shared bus, and (2) the shared bus forms a communication bottleneck because only one component can control the bus at a time. Consequently, some SoCs implement shared buses that utilize crossbar switches, also called crossbars, cross-point switch, matrix switch, or switch matrix, to achieve higher throughput, with STBus, IBM CoreConnect, and ARM AMBA being several examples of bus standards that may utilize crossbars. *See* EX2007, 320. A crossbar has a plurality of input lines and output lines that form a crossed pattern of intersecting lines between which a connection may be established

by closing a switch located at a given intersection. An n -input, m -output crossbar, i.e., an $n \times m$ crossbar, is the equivalent to a set of m multiplexers (one per output port), each having n inputs (one per input port). *See* EX2007, 332–333. However, a shared bus that utilizes crossbars still has disadvantages. For example, the physical implementation of complex SoCs require large and complex crossbars that are difficult to place-and-route during physical design on the SoC in the areas between the components that implement functional logic of the SoC. *See* EX2008, 3. An on-chip communication network, e.g., a NoC, is an effective way to overcome some of the aforementioned disadvantages.

45. A NoC interconnect paradigm radically changes the on-chip communication compared to direct interconnects, such as buses or switches. This is because of the multi-hop nature of a network, where components are not directly connected, but separated by one or more network nodes. This is in contrast with buses, where components are directly connected as described above. Several implications of this change include arbitration methods (which can no longer be centralized), and communication properties (e.g., ordering, flow control, and so on). In an SoC that implements a NoC, certain design-specific wiring structures of the SoC, typically the longer “global” wires that interconnect components, are replaced with a general-purpose on-chip communication network. EX2005, 2. Instead of connecting these top-level modules by routing dedicated wires, they are

connected to a network that routes packets between them. *See id.* In doing so, these network-connected modules, or components, can be thought of as “network clients” that communicate with each other through communication channels of a network that provide one or more levels of service, e.g., as may be achieved by various communication properties. *See id.*

46. The 2001 paper by Dally and Towles provides a basic example of a NoC implementation. *See* EX2005, 2–3. It describes dividing a chip into 16 “tiles,” where each tile comprises client logic, such as processors, DSPs, peripheral controllers, memory subsystems, and so on. The client logic within a tile (i.e., nearby logic) may communicate via direct interconnections, but client logic in different tiles (i.e., far-apart logic) must communicate only over the network. In other words, there are no top-level (global) bus-based interconnections, only interconnections provided by the network. Figure 1 of the Dally and Towles paper is reproduced below. *See* EX2005, 2.

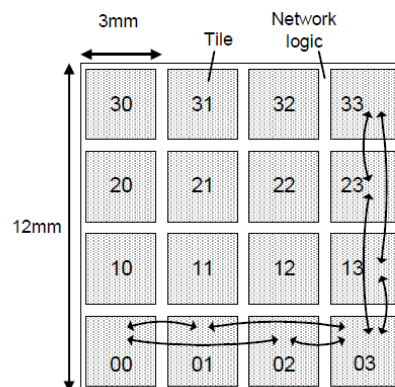


Figure 1 Partitioning the die into module tiles and network logic

47. Each tile interfaces with the network over an input port that is used to insert packets into the network and an output port over which the network delivers packets to the tile (where input and output here refer to packets going in and out of the network, not the tile). Each packet includes a virtual channel field that specifies which virtual channel it may be routed on, to provide different levels of service, and route field to specify a route through the network, i.e., the hops taken from source tile to destination tile. Such a port interface can present a simple and reliable level of service to the network clients. *See* EX2005, 3.

48. Logic local to the network clients can layer higher-level services on top of this interface. For example, this local logic could present a memory read/write service, a flow-controlled data stream, or a logical wire to the client. Local logic can also provide a translation from a destination node to a route. *See* EX2005, 3.

49. The router needed at each tile consists of five input controllers (one for each direction and one for input from the tile) and five output controllers. The router for each tile is distributed across the four edges of a tile with the input and output controller for a given direction (e.g., west) located on that edge of the tile. Figure 2 of the Dally and Towles paper, reproduced below, shows the west input controller and its connection to four output controllers (one for each of the other directions and one for the tile). *See* EX2005, 3.

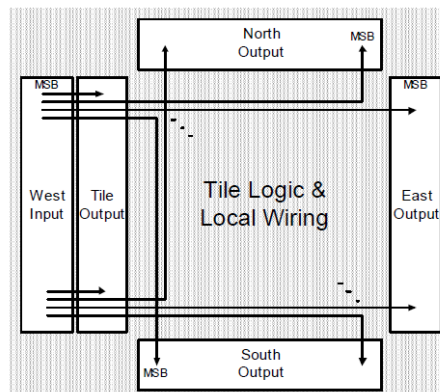


Figure 2 West input connections to output controllers

50. Details of an input controller and output controller are shown in Figure 3 of the Dally and Towles paper, reproduced below. Each input controller has an input buffer and input state logic for each virtual channel. When a head flit (flow control digit, e.g., part of a packet) arrives at each tile, the input controller strips the next entry off the route field and uses these two bits to select one of four output ports. The flit then arbitrates with the other virtual channels on that input port and, if it wins the arbitration, is forwarded to the output port. This arbitration and forwarding takes place in parallel with allocating a virtual channel and checking available buffer space to reduce latency. The output port provides a single stage of buffering for each input port connection. The flits in these buffers arbitrate for the link to the input controller on the next tile. Credits for buffer allocation are piggybacked on flits travelling in the reverse direction. *See* EX2005, 4.

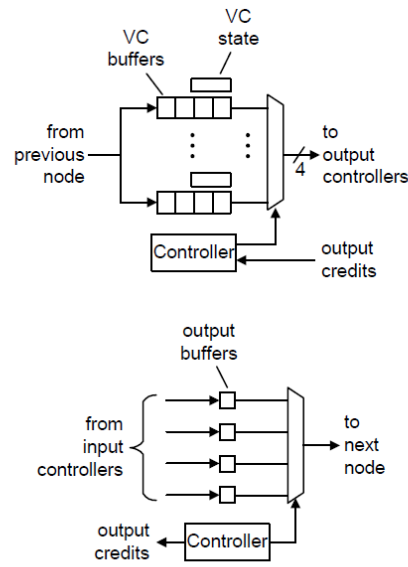


Figure 3 Input (top) and output controllers

51. As indicated earlier, the communication between modules (e.g., components, network clients, etc.) are performed over connections. Some examples of connection properties, which are sometimes called communication properties, are provided below.

The connection properties may include ordering (data transport in order), flow control (a remote buffer is reserved for a connection, and a data producer will be allowed to send data only when it is guaranteed that space is available for the produced data), throughput (a lower bound on throughput is guaranteed), latency (upper bound for latency is guaranteed), the lossiness (dropping of data), transmission termination, transaction completion, data correctness, or data delivery.

EX1001, 5:27–35.

52. With such connection properties, on-chip networks are typically able to provide varying levels of communication service, which may be referred to as a quality of service (QoS) levels. “The QoS is a very important feature in the

interconnect infrastructures because it provides a regulation mechanism allowing specification of guarantees on some of the parameters related to the traffic.” EX2007, 325. Traffic here refers to a flow of data between a first module and a second module, for example, a series of data packets transmitted from the first module to the second module. With QoS, different traffic classes can be supported, such as:

- Real time and low latency (RTLL)—Traffic flows that require the lowest possible latency.
- Guaranteed throughput (GT)—Traffic flows that must maintain their throughput over a relatively long time interval.
- Guaranteed bandwidth (GBW)—Traffic flows that require a guaranteed amount of bandwidth over a relatively long time interval.
- Best effort (BE)—Traffic flows that do not require guaranteed latency or throughput but have an expectation of fairness.

EX2007, 326.

B. Challenges of On-Chip Networks

53. The issues and problems of implementing a multi-hop on-chip network interconnect are complex. Prior to the invention of a network-on-chip interconnect, functional blocks of an integrated circuit were typically

interconnected using buses or crossbars. However, both buses and crossbars have respective drawbacks, as discussed in more detail below.

54. Buses, such as ARM's Advanced Microcontroller Bus Architecture (AMBA), were an early solution for interconnecting on-chip functional blocks, e.g., initiators and targets. However, "the integration of multiple cores within chips" caused "too many initiators trying to access different targets simultaneously" which "created bottlenecks on the bus." EX2009, 3. Buses also had "latency issues." *Id.* As a result, "[t]he industry wanted to enable concurrent access and more overall system data throughput." *Id.*

55. Following the use of buses, crossbars were introduced as a next interconnect solution. While crossbars cure some of the deficiencies caused by buses, "crossbars require complex control logic to keep track of outstanding transactions" and require "wide data paths for the initiators and targets that need the most throughput as well as signaling for address and control information." EX2009, 3. "To support continued system scaling, crossbars were cascaded with bridges between fabrics" which "carried a significant cost in silicon area" and "limited clock frequencies and added significant latency to transactions." *Id.* at 4. Moreover, "during the 2000s, the initiators and targets became so numerous and widely distributed in the physical floorplan of a chip that the crossbar became a physical wiring nightmare." *Id.*

56. Due to the drawbacks of buses and crossbars, “Network-on-Chip (NoC) technology emerged as a solution to the wiring problem.” EX2009, 5. NoC technology packetizes address and data information on the same wires, thereby allowing transactions to be sent over a smaller number of wires compared to prior interconnect solutions “while maintaining high quality of service for each transmission.” *Id.* NoC technology also contributed to decreased die density by “distributing the interconnect logic throughout the chip rather than having bridges as chokepoints.” *Id.*

57. The ‘800 patent describes the distinction between prior interconnects (e.g., buses and crossbars) and a multi-hop interconnect.

NoCs help resolve the electrical problems in new deep-submicron technologies, as they structure and manage global wires. At the same time they share wires, lowering their number and increasing their utilization. ... NoCs also decouple computation from communication, which is essential in managing the design of billion-transistor chips. NoCs achieve this decoupling because they are traditionally designed using protocol stacks, which provide well-defined interfaces separating communication service usage from service implementation.

EX1001 at 1:36–47. Moreover, using multi-hop interconnects “radically changes the communication when compared to direct interconnects, such as buses or switches.” *Id.* at 2:3–10 (“This is because of the multi-hop nature of a network, where communication modules are not directly connected, but separated by one or

more network nodes. This is in contrast with the prevalent existing interconnects (i.e., buses) where modules are directly connected.”).

58. The ’800 Patent describes that “on-chip communication protocols (e.g., Device Transaction Level DTL and AXI-Protocol) operate on a split and pipelined basis with transactions consisting of a request and a response, and the bus is released for use by others after a request issued by a master is accepted by a corresponding slave.” EX1001 at 2:11–16. However, these “protocols have been designed mainly for buses with small latency” and “have been designed based on the assumption that read operations are always urgent and should therefore be completed as soon as possible without unnecessary buffering.” *Id.* at 3:18–22. “[A]s systems grow larger ... the latency grows as well,” thus, “the communication granularity become coarser and the latency requirements become less strict.” *Id.* at 3:23–26. The ’800 patent further teaches that the aforementioned protocols “force some of the currently buffered data to be transferred although the optimal burst length has not been reached,” which causes “deadlock.” *Id.* at 3:27–30. The ’800 Patent improves how the bus-based protocols were implemented and discloses a method for buffering data that “improve[s] interconnect utilization.”

59. The authors of the ’800 Patent identified specific examples of some differences between a multi-hop interconnect and a bus include: (1) media arbitration; (2) data routing; (3) data formatting; (4) latency; and (5) flow control

for input buffers. Regarding media arbitration, “[i]n a bus, master modules request access to the interconnect, and the arbiter grants the access for the whole interconnect at once,” whereas in a NoC, “the arbitration is distributed, because it is performed in every router, and is based only on local information.” EX2017 at 5:25–42.

60. Regarding data routing, “[f]or a bus, the command, address, and data are broadcasted on the interconnect. They arrive at every destination, of which one activates based on the broadcasted address, and executes the requested command,” whereas in a NoC, “it is not feasible to broadcast information to all destinations, because it must be copied to all routers and network interfaces. This floods the network with data. The address is better decoded at the source to find a route to the destination module.” EX2017 at 5:43–54.

61. Regarding data formatting, “[i]n modern bus interfaces the data format is defined by separate wire groups for the transaction type, address, write data, read data, and return acknowledgements/errors (e.g., VCI, OCP, AMBA, or CoreConnect),” whereas in a NoC, “there is typically no distinction between different kinds of data. Data is treated uniformly, and passed from one router to another.” EX2017 at 6:8–25.

62. Regarding latency:

[L]atency is caused by two factors: (a) the access time to the bus, which is the time until the bus is granted, and (b) the latency introduced by the interconnect to transfer the data. For a bus, where the arbitration is centralized the access time is proportional to the number of masters connected to the bus. The transfer latency itself typically is constant and relatively fast, because the modules are linked directly. However, the speed of transfer is limited by the bus speed, which is relatively low. In a NoC, arbitration is performed at each router for the following link. The access time per router is small. Both end-to-end access time and transport time increase proportionally to the number of hops between master and slave. However, network links are unidirectional and point to point, and hence can run at higher frequencies than buses, thus lowering the latency. From a latency prospective, using a bus or a network is a trade off between the number of modules connected to the interconnect (which affects access time), the speed of the interconnect, and the network topology.

EX2017 at 5:55–6:7.

63. Multi-hop interconnects also differ from buses with respect to flow control for input buffers.

For buses, the source and destination are directly linked, and, destination can therefore signal directly to a source that it cannot accept data. This information can even be available to the arbiter, such that the bus is not granted to a transaction trying to write to a full buffer. In a NoC, however, the destination of a transaction cannot signal directly to a source that its input buffer is full. Consequently, transactions to a destination can be started, possibly from multiple sources, after the destinations input buffer has filled up. If an input buffer is full, additional incoming transitions are not accepted, and stored in the network. However, this approach can easily lead [sic] to network congestion, as the data could be eventually stored all the way to the sources, blocking the links in between.

EX2017 at 6:48–62.

64. The ‘800 patent and the ‘449 patent also emphasize that “NoC’s premises are different from off-chip networks, and, therefore, most of the network design choices must be reevaluated” because “[o]n-chip networks have different properties (e.g., tighter link synchronization) and constraints (e.g., higher memory cost) leading to different design choices, which ultimately affect the network services.” EX1001 at 1:66–2:5; EX2017 at 1:66–2:5. Another distinction between an on-chip network compared to off-chip networks is that “[a]n off-chip network interface usually contains a dedicated processor to implement the protocol stack up to network layer or even higher, to relieve the host processor from the communication processing;” however, “[i]ncluding a dedicated processor in a network interface is not feasible on chip, as the size of the network interface will become comparable to or larger than the IP to be connected to the network.” EX2017 at 2:16–25.

65. Further, data sent via a NoC may be subject to out-of-order arrival, diverse routing paths, and retransmission after dropping.

In a network, data sent from a source to a destination may arrive out of order due to reordering in networks nodes, following different routes, or retransmission after dropping. For off-chip networks out-of-order data deliver is typical. However, for NoCs where no data is dropped, data can be forced to follow the same path between a source and a destination (deterministic routing) with no reordering. This in-order data transportation requires less buffer-space, and reordering modules are no longer necessary.

EX2017 at 3:8–17.

66. On-chip networks also differ from off-chip networks with respect to time-related guarantees.

Off-chip networks typically use packet switching and offer best-effort services. Contention can occur at each network node, making latency guarantees very hard to offer. Throughput guarantees can still be offered using schemes such as rate-based switching or deadline-based packet switching, but with high buffering costs. An alternative to provide such time-related guarantees is to use time-division multiple access (TDMA) circuits, where every circuit is dedicated to a network connection. Circuits provide guarantees at a relatively low memory and computation cost. Network resource utilization is increased when the network architecture allows any left-over guaranteed band width to be used by best-effort communication.

EX2017 at 3:34–46.

67. On-chip networks can also provide more reliable communication compared to off-chip networks.

A consequence of the tight on-chip resource constraints is that the network components (i.e., routers and network interfaces) must be fairly simple to minimize computation and memory requirements. Luckily, on-chip wires provide a reliable communication medium, which can help to avoid the considerable overhead incurred by off-chip networks for providing reliable communication. Data integrity can be provided at low cost at the data link layer. However, data loss also depends on the network architecture, as in most computer networks data is simply dropped if congestion occurs in the network.

EX2017 at 2:43–53.

68. On-chip networks also differ from off-chip networks because of their constraints and synchronization.

Typically, resource constraints are tighter on chip than off chip. Storage (i.e., memory) and computation resources are relatively more expensive, whereas the number of point-to-point links is larger on chip than off chip. Storage is expensive, because general-purpose on-chip memory, such as RAMS, occupy a large area. Having the memory distributed in the network components in relatively small sizes is even worse, as the overhead area in the memory then becomes dominant.

EX2017 at 2:6–15.

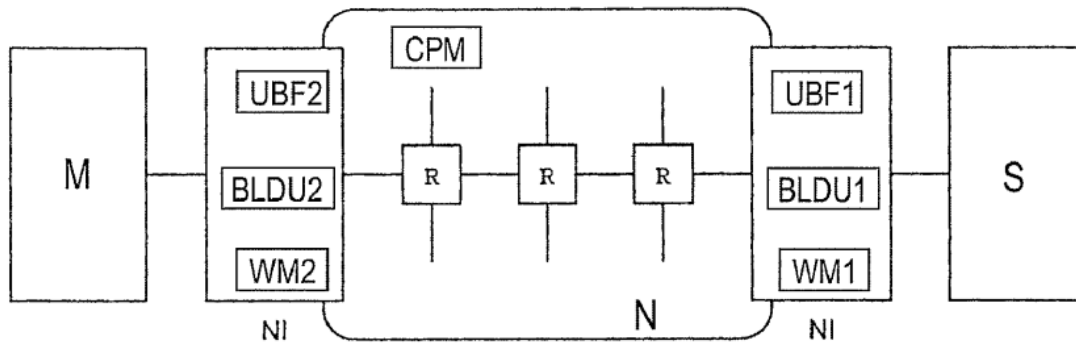
69. Given the technical challenges, I do not agree with Dr. Mowry's simplistic assertion that the prior art technologies (e.g., bus-based interconnects and traditional computer networks) could be easily combined without undue experimentation by a person of skill in the art. A POSITA would have been dissuaded from attempting to modify or combine different prior art to arrive at the respective claimed inventions. My observations and opinions in this respect apply to all alleged modifications and combinations proposed by Dr. Mowry.

70. I note that all of the secondary references suffer from the issues stated immediately above. Dr. Mowry does not sufficiently show how the prior art could be modified or combined to arrive at the claimed inventions in view of these disclosures. In my opinion, arriving at the invention would require extensive re-design, which is neither suggested nor alluded to, and would raise the difficulties and challenges discussed herein.

C. Overview of the '800 Patent

71. The '800 patent generally relates to integrated circuits having a plurality of circuitry modules that are communicatively interconnected via NoCs. The '800 patent discloses an integrated circuit including an on-chip network where interconnect utilization is improved by buffers that prevent data from entering the network until an optimal amount of data is available for transmission on the network. Specifically, the '800 Patent discloses “an integrated circuit having a plurality of processing modules and an interconnect means for coupling said plurality of processing, a method for buffering and a data processing system.” EX1001, 1:7–10. The '800 Patent also discloses “FIG. 2 shows a schematic representation of a network on chip” where “a first wrapper means WM1 is arranged in the network interface NI associated to the slave S and a second wrapper means WM2 is arranged in the network interface NI associated to the master M.” *Id.* at 6:7–15. “[T]he wrapper means WM1, WM2 serve to buffer data to be sent over the network N until the conditions for an optimal transfer are satisfied.” *Id.* 6:15–17. “In addition, a first and second determination means BLDU1, BLDU2 are arranged in the network interfaces NI associated to the slave S and master M, respectively.” *Id.* 6:47–49. “The first and second determination units BLDU1, BLDU2 serve to determine the optimal burst length for transferring

data over the network N according to the current communication or connection properties.” *Id.* 6:49-53. FIG. 2 is shown below:



72. The '800 Patent provides that it is “an objection of the invention to provide an integrated circuit, a method for buffering as well as a data processing system with an improved interconnect utilization.” EX1001, 3:35–37. To achieve this end, “at least one of said first and second wrapper means comprise a determination unit BLDU for determining the optimal first or second amount of data to be buffered in said first or second wrapper means before said data is transferred ...,” and thus, “the packet the size of the data transferred over the interconnect can be adapted according to the properties of the actual communication and thereby the utilization of the interconnect can be improved.” *Id.* at 4:18–28.

73. “The invention is based on the idea to buffer data until the buffered data is sufficiently large to be transferred optimally over an interconnect means in a packet. ... data is only transferred when sufficient data for an optimal packet size

has been buffered even when data can be sent earlier.” *Id.* 4:49–57. The ’800 patent discloses NoC buffers to increase network efficiency and resolve synchronization issues between fast and slow communication channels. The ’800 Patent concerns aspects of data storage and buffering. The ’800 Patent discloses a NoC network interface that has buffers to store data until an optimal amount of data is available for transmission on the network. EX1001, 4:18–28, 4:50–57. Rather than inefficiently transmitting data in small packages as it is generated by the circuitry modules, the buffers aggregate data until an optimal amount of data is ready to be transmitted over the network, increasing overall network utilization and efficiency. *Id.*, 3:1–3, 7:44–46.

74. The ’800 patent discloses systems for Quality of Service (“QoS”), called “connection properties” in the ’800 patent, which are mechanisms to prioritize and manage communication between different circuitry modules. QoS allows for the allocation of network resources according to the specific requirements of different applications or tasks, ensuring, for example, that critical communications get higher priority while lower-priority communications do not unnecessarily consume resources. Examples of QoS include minimum throughput or bandwidth, maximum latency, and priority.

75. The ’800 Patent teaches an invention whose object is improving—indeed, optimizing—interconnect utilization in a network, where the improvement

is achieved by buffering data until conditions for an optimal transfer are satisfied, as detailed in the following passages of the '800 Patent.

It is therefore an object of the invention to provide an integrated circuit, a method for buffering as well as a data processing system with **an improved interconnect utilization.**

EX1001, 3:35–37.

According to a preferred aspect of the invention at least one of said first and second wrapper means comprise a determination unit BLDU for determining the optimal first or second amount of data to be buffered in said first or second wrapper means before said data is transferred according to the communication properties of said communication between said first and second processing module. Accordingly, the packet size of the data transferred over the interconnect can be adapted according to the properties of the actual communication and thereby the utilization of the interconnect can be improved.

EX1001, 4:18–28.

The invention is based on the idea to buffer data until the buffered data is sufficiently large to be transferred optimally over an interconnect means in a packet. The larger a packet is, the smaller is the amount of packet headers and therefore the overhead is reduced and the interconnect is utilized more efficiently. The data is only transferred when sufficient data for an optimal packet size has been buffered even when data can be sent earlier. The data is only transferred from the buffer when the conditions for an optimal transfer are satisfied.

EX1001, 4:49–57.

As in the first embodiment, the wrapper means WM1, WM2 serve to buffer data to be sent over the network N until the conditions for an optimal transfer are satisfied.

EX1001, 6:15–17.

The first and second wrapper means WM1, WM2 are adapted to immediately transfer the buffered data when the first or second unbuffer-flag is set in the first or second unbuffer-flag unit UBF1, UBF2, respectively. By providing this possibility to flush the buffered data a deadlock caused by data being buffered can be prevented. In such case, all data being buffered are transferred as fast as possible without waiting for a packet size to suit the optimal burst length for the network N.

EX1001, 6:22–30.

The flushing of the buffered data may also be achieved by a first or second unbuffer signal received by the first or second wrapper means WM1, WM2. Therefore, if the first or second wrapper means WM1, WM2 receives an unbuffer signal, the data currently buffered are transferred as fast as possible, i.e. without waiting for the optimal burst length to be acquired.

EX1001, 6:31–36.

The data is buffered until the buffered data reach an optimal packet size such that the buffered data can be transferred over the network. If the amount of buffered data has not yet reached the optimal packet size, none of the data is transferred over the network N even if transferring a smaller packet size is possible.

EX1001, 7:23–28.

The requests and responses are aggregated in bursts with a length optimal for the particular interconnect. As in the case of a request like a write, the buffering can be prevented with an unbuffered flag to the response part of the transaction. Hence, all responses or requests pending and including the current one are transferred as fast as possible without being buffered to form an optimal burst length for the interconnect.

EX1001, 7:44–51.

Besides the optimal burst length determination, also the transmission moment may be determined in a similar way by the first and second

determination means BLDU1, BLDU2. For a guaranteed throughput GT connection, data should wait for one of the slots reserved for its connection. Using a “flush”-like signal/flag may force the sending of the data in advance. Alternatively, for a best effort BE connection, a round robin arbitration across connections in the NI can be used. A “flush”-like signal may force a temporarily higher priority for a connection.

EX1001, 7:4–14.

76. The '800 Patent, teaches (1) buffering of data improves and/or optimizes interconnect utilization (EX1001, 3:35–37, 4:49–57, 6:15–17, 7:23–28, 7:44–51); (2) the optimal amount of data to be buffered can be based on communication properties of a communication (*id.*, 4:18–28); (3) the optimal amount of data to be buffered can be specific to a particular interconnect (*id.*, 7:44–51) and (4) the transmission moment is based on communication properties (*id.*, 7:4–14). In other words, an optimal amount of data to be buffered may not be fixed or constant; rather, it may be according to communication properties and/or interconnect characteristics, i.e., based on certain conditions of a connection. This is typical for engineering optimization, for example, optimizing power consumption of a SoC, optimizing throughput of a CPU, optimizing manufacturing yield of an integrated circuit, and so on. A solution for each of these optimizations depends on properties and characteristics of the system, and the solution varies as the properties and characteristic vary.

D. Claim Constructions

77. I understand that the following terms of the Asserted Patents have been construed by agreement or are currently in dispute:

Term/Phrase	Agreed Upon Construction
“flow control”	“controlling flow whereby a remote buffer is reserved for a connection, and a data producer will be allowed to send data only when it guaranteed that space is available for the produced data”
“integrated circuit”	“interconnected circuitry on a chip”
“said first wrapper” / “said second wrapper”	“said first memory” / “said second memory”

Term/Phrase	Petitioner’s Proposed Construction	Patent Owner’s Proposed Construction
“optimal amount of data to be buffered”	Indefinite	Plain and ordinary meaning
“optimal moment for sending the data”	Indefinite	Plain and ordinary meaning
“determination unit”	Subject to Section 112, ¶ 6. Function: “determine a first/second amount of data to be buffered by the first/second memory” and for at least one of the first determination unit and the second determination unit “determine an optimal moment for sending data.” Structure: Indefinite.	“circuit that makes determinations based on stored or received information” Patent Owner disagrees that this term is governed by Section 112, ¶ 6. Nevertheless, if the Board holds that this term is so governed, Patent Owner proposes the

		following construction: Function: “determine an optimal amount of data to be buffered” and “determine an optimal moment for sending the data” Structure: “circuit”
--	--	---

EX2033, Appendix 1; *Id.*, Appendix 2; EX2035, 2, n. 2.

78. I understand that in the pending TI litigation, the district court also construed “network” to mean “multi-hop interconnection between modules separated by one or more network nodes.” EX2022, 58.

79. For all other claim terms in the asserted claims of the Asserted Patents that are not specifically agreed upon by the parties or disputed, I have applied the plain and ordinary that a POSITA would understand for such terms and/or constructions at the time of the invention.

VI. ANALYSIS OF THE '800 PATENT

A. Ground 1: Radulescu in View of Bahadiroglu Does Not Render '800 Patent Claim 12 Obvious

1. Radulescu in view of Bahadiroglu fails to disclose “wherein the first interconnect module includes a first memory for buffering first data from the first processing module, the first interconnect module further including a first determination unit, wherein the first determination unit is configured to determine a first optimal amount of data to be buffered by the first memory, and the second interconnect module includes a second memory for buffering second data from the second processing module, the second interconnect module further including a second determination unit, wherein the second determination unit is configured to determine a second optimal amount of data to be buffered by the second memory”

80. Sec. IX.A.1.e of Dr. Mowry’s Declaration alleges that Radulescu discloses the limitation of claim 12 of the '800 patent reciting the first and second determination units determining first and second optimal amounts of data, respectively. I disagree.

81. In relevant part, claim 12 of the '800 Patent recites a first interconnect module for providing a communication connection between a first processing module and a network. The first interconnect module includes a first memory for buffering first data from the first processing module and a first determination unit configured to determine a first optimal amount of data to be buffered by the first memory. Claim 12 recites similar elements and functions with respect to a second interconnect module, a second processing module, a second memory, a second

determination unit, and a second optimal amount of data. To be clear, the determination unit *determines* the optimal amount of data to be buffered. The specification of the '800 Patent, which uses terms like “optimal burst length,” “optimal packet,” “optimal transfer,” and “optimal first or second amount of data” when discussing an optimal amount of data to be buffered, explains that “determination of the optimal burst length can be performed statically or dynamically, during the initial phases of the required processing or during predetermined time intervals.” EX1001, 6:52–56. The specification provides additional disclosure as follows.

Alternatively, the optimal burst length may be determined dynamically (1) every time a connection is set up for transferring data over the network, (2) for every transaction, (3) whenever an IP switches to another connection, (4) for every packet, (5) when the state of the network on chip changes (e.g., reconfiguration, NoC load, buffer fillings, etc.). The determination of the optimal burst length may be performed based on information stored in the determination units BDLU 1, 2 or on information received from the network N or IP blocks.

EX1001, 6:56–65.

82. Paragraph 85 of Dr. Mowry’s Declaration alleges that “the scheduler included in the NI2 (the claimed first interconnect module) and the scheduler included in the NI1 (the claimed second interconnect module) are each a “determination unit” included in each respective interconnect module at least because each scheduler makes a determination about whether to transmit buffered

data **based on a configurable threshold that is stored or received by the scheduler.**” EX1003, ¶ 85 (emphasis added). However, claim 12 of the ’800 Patent does not say that the determination unit (e.g., a scheduler) makes a decision *based on* some predetermined (e.g., stored or received) optimal amount of data to be buffered (e.g., a configurable threshold). Claim 12 specifies that the determination unit “is configured to determine” the optimal amount of data to be buffered, (EX1001, 11:67–12:6).

83. While Radulescu discloses a “configurable threshold,” it does not disclose or suggest any circuit or device capable of configuring the threshold, e.g., capable of determining an optimal amount of data to be buffered. *See, e.g.,* EX1005, 4 (“[W]e implemented a configurable threshold mechanism.”); *see also id.*, 4 (“[A] queue becomes eligible for scheduling when either the amount of sendable data are above a first threshold, or when the amount of credits is above a second threshold”).

84. For at least the above reasons, Radulescu does not disclose or suggest either the first determination unit or the second determination unit as recited in claim 12 of the ’800 Patent.

85. Paragraph 89 of Dr. Mowry’s Declaration alleges that Bahadiroglu’s disclosure of determining an optimal amount of data in a sending or receiving node is a determination of a first or second optimal amount of data, respectively, as

recited in claim 12 of the '800 Patent. Paragraph 90 of Dr. Mowry's Declaration further alleges that "a POSITA would have looked to Bahadiroglu's detailed disclosures of an adaptive packet mechanism as a way to configure Radulescu's threshold mechanism, and would have had a reasonable expectation of success in configuring Radulescu's mechanism as an adaptive packet mechanism." EX1003, ¶ 90. I disagree.

86. First, a POSITA would lack motivation to combine Radulescu and Bahadiroglu because Radulescu is directed to on-chip networks for communication between IP modules where "the size of IP modules attached to the NoC is relatively small," (EX1005, 1), while Bahadiroglu is directed to computer networks for communication between nodes, such as "a computer, network server or router, a network or, for example, a more specialized data display or presentation device such as a telephone, mobile phone or processing device, television or audio system," (EX1006, ¶ 73). Bahadiroglu's "adaptive packet mechanism" is directed to computer networks, not on-chip networks (Radulescu, and the '800 Patent, are directed to on-chip networks). FIG. 2 of Bahadiroglu (reproduced below) is "an exemplary Network 10 in which the present invention may be implemented." EX1006, ¶ 73.

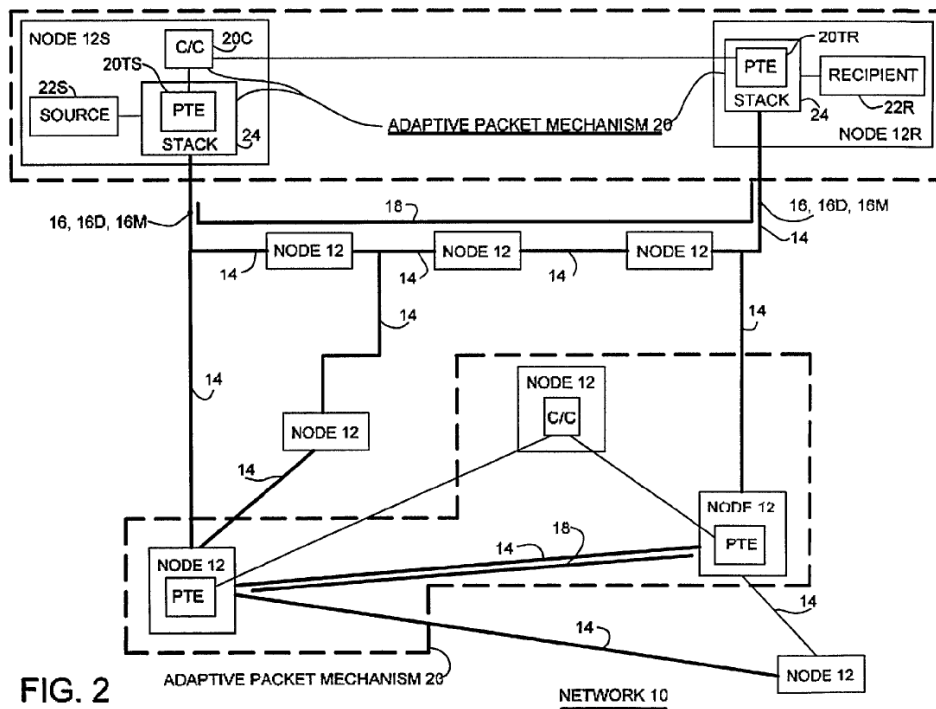


FIG. 2

EX1006, FIG. 2

87. Bahadiroglu explains that “networks specifically oriented to intercomputer communications typically provide individual computers shared access to a common communication backbone having relatively broad bandwidth, such as fiber optic cables or coaxial or twisted pair cable.” EX1006, ¶ 6.

A typical example of such a computer network is a personal computer (“PC”) network as PCs are used for a wide range of applications requiring the PCs to communicate with each other over a computer network. Typical computer networks connecting PCs, as well as other computers, nodes and related devices and systems, include local area networks (“LANs”) interconnecting PCs located in relatively close proximity and wide area networks (“WANs”), which are typically comprised of a network of separate LANS and provide communications over wider areas. The “Web”, for example, is a very extensive WAN interconnecting LANS as well as PCs, larger computers, nodes and related devices.

EX1006, ¶ 6. Referring to FIG. 2 of Bahadiroglu, “[a] Network 10 is comprised of a plurality of Nodes 12 A Node 12, in turn, may be comprised of a computer, network server or router, a network or, for example, a more specialized data display or presentation device such as a telephone, mobile phone or processing device, television or audio system.” EX1006, ¶ 73. For the reasons explained in Sec. V.B above, a POSITA would lack motivation to combine elements developed for off-chip networks into an on-chip network environment.

88. Second, a POSITA would lack motivation to combine Radulescu and Bahadiroglu because Radulescu emphasizes the necessity to keep on-chip network interfaces (where the scheduler is located) small and simple, while Bahadiroglu teaches an extremely complex adaptive packet mechanism, as explained below.

89. Radulescu explains that “[t]he pressure to keep the protocol stack small is higher on-chip than off-chip, because the size of the IP modules attached to the NoC is relatively small.” To that end, Radulescu discusses the need for simple network interfaces. EX1005, 1 (“On-chip network interfaces must provide a low-area overhead, because the size of IP modules attached to the NoC is relatively small.”); *see also id.*, 1 (“Our NI is intended for systems on chip (SoC), hence, it must have low area. . . . we also have to provide a simple and flexible configuration.”). In contrast, Bahadiroglu teaches “[a]n adaptive packet mechanism

and method for optimizing data packet transmission through a network [where] measurements of latency and jitter are used to determine an optimum packet size and an optimum inter-packet interval,” (EX1006, Abstract), which is extremely complex and where the number of possible combinations of values of network variable is extremely large:

[T]he latency and jitter of a network are extremely complex or virtually impossible to predict or determine.”

EX1006, ¶ 68 (emphasis added).

[T]he latency and jitter of a network or connection are often not linear functions with respect to such variables as time, data type, traffic load, packet size or inter-packet interval, and the latency and jitter of a network are extremely complex or virtually impossible to predict.”

EX1006, ¶ 70 (emphasis added).

It will be apparent from the above discussions that Network Conditions 42 directly or indirectly represent a significant number of network variables and that the relationship of these variables to the current optimum packet size and inter-packet interval is complex, often nonlinear and variable with time. In addition, and as a consequence, the number of possible combinations of values of these network variable will be extremely large. In addition, and according to the present invention, it is necessary for the Adaptive Packet Mechanism 20, and in particular the Network Analyzer 48, to be able to determine the optimum packet size and inter-packet interval for current connection conditions in ‘real time’ because of the rapid variability of connection transmissioin [sic] conditions.”

EX1006, ¶ 153. (emphasis added).

[illegible]

EX1006, FIG. 6A

50

Abstract), in Radulescu's simple network interface as a way to configure Radulescu's threshold mechanism as proposed in paragraph 90 of Dr. Mowry's Declaration. EX1003, ¶ 90. Moreover, because of the stark differences between on-chip and off-chip environments including those indicated above, a POSITA would not have a reasonable expectation of success to implement Bahadiroglu's complex adaptive packet mechanism, e.g., the Network Analyzer 48, in Radulescu's relatively simple network interface as a way to configure Radulescu's threshold mechanism as proposed in paragraph 90 of Dr. Mowry's Declaration.

92. Third, a POSITA would lack motivation to combine Radulescu and Bahadiroglu because Radulescu teaches that “[t]o optimize the NoC utilization, it is preferable to send longer packets,” (EX1005, 4), while Bahadiroglu's adaptive packet mechanism determines an optimum packet size that yields the highest probability of successful delivery. EX1006, ¶ 77 (“As described, this optimum packet size and inter-packet interval will, as a consequence, yield the highest probability of packets arriving successfully at a given Receiving Node 12R.”). Bahadiroglu explicitly states that the optimum packet size could be a minimally sized packet that contains only the control or header fields; or the optimum packet size could be a maximally sized packet that contains the maximum allowable payload for a given communication protocol; or the optimum packet could have a size anywhere within that range.

It has been described that an Adaptive Packet Mechanism 20 controls the size of Packets 16 to be an optimum for the current network conditions of a Connection 18, and it will therefore be apparent that an Adaptive Packet Mechanism 20 does so by controlling the size of the data fields of the Packets 16 as the control or header fields are typically determined by the communications protocol employed by the Sending Node 12S and Receiving Node 12R. **The range of Packet 16 sizes will thereby range from a minimum determined primarily by the size of the control or header fields of the Packets 16 to a maximum determined by the communications protocol employed by the Sending Node 12S and Receiving Node 12R 24 executed by Network Interface Mechanism 26 and will include both the control or header fields and the maximum capacity of the data fields.** As such, it will be understood that within the context of the present invention the size of a Packet 16, whether a Data Packet 16C or a Monitor Packet 16M, will be a function of both the control or header fields and the data fields.”

EX1006, ¶ 126 (emphasis added).

93. Accordingly, a POSITA would lack motivation to implement Bahadiroglu’s adaptive packet mechanism, where the optimum packet size is based on probability of successful delivery, to configure Radulescu’s configurable threshold for optimizing NoC utilization based on sending longer packets. Further, a POSITA would not have a reasonable expectation of success at least because Bahadiroglu’s optimum packet size, which could be a minimally sized packet, would not yield the results alleged in Dr. Mowry’s Declaration. Indeed, the implementation of Bahadiroglu’s adaptive packet mechanism to configure Radulescu’s configurable threshold would frustrate the purpose of Radulescu’s optimization of NoC utilization.

94. For at least the above reasons, Radulescu in view of Bahadiroglu does not disclose or suggest either “the first determination unit is configured to determine a first optimal amount of data to be buffered by the first memory” or the “the second determination unit is configured to determine a second optimal amount of data to be buffered by the second memory” as recited in claim 12 of the ’800 Patent.

2. Radulescu in view of Bahadiroglu fails to disclose “the first interconnect module transferring the first data to the second processing module when the first data buffered in the first memory reaches the first optimal amount, and the second interconnect module transferring the second data to the first processing module when the second data buffered in the second memory reaches the second optimal amount”

95. Claim 12 of the ’800 Patent further recites that the first interconnect module transfers the first data buffered in the first memory to a second processing module when the first data buffered in the first memory reaches the first optimal amount. Claim 12 recites similar elements and functions with respect to the second interconnect module, the second memory, and the second optimal amount.

Paragraph 92 of Dr. Mowry’s Declaration alleges that Radulescu discloses this claim limitation, and paragraph 93 of Dr. Mowry’s Declaration alleges that Radulescu in view of Bahadiroglu renders this claim limitation obvious. EX1003, ¶¶ 92-93. I disagree.

96. First, Radulescu describes a system in which channels, or queues, are scheduled for transmission by a scheduler. Radulescu at 3. When the scheduler selects a channel (either when a guaranteed-throughput channel's time slot arrives or via an arbitration scheme for best-effort channels), the selected channel is granted transmission if it has data to send and there is space in the destination queue. Radulescu at 4. To optimize NoC utilization, Radulescu explains that it is preferable to send longer packets. EX1005, 4. A configurable threshold mechanism causes the scheduler to skip a channel if the "sendable data" for the channel is below the threshold, where sendable data refers to how much data can be sent before overflowing the destination queue, and skipping a channel refers to not granting transmission to a guaranteed-throughput channel when its transmission slot arrives or not granting transmission to a best-effort channel when it is selected by the arbitration scheme. EX1005, 3–4.

97. In Radulescu's system, no channels, or queues, are included in the scheduler's scheduling process unless the channel's sendable data exceeds the threshold (or there are enough credits). EX1005, 3. A channel becomes "eligible for scheduling" when its sendable data exceeds the threshold (or there are enough credits). Radulescu at 4. Eligible for scheduling means the channel is selectable by the scheduler for scheduling; eligible for scheduling does not mean that sendable data (or any data in the queue) of the channel is sent. EX1005, 4. Rather, the

scheduler may eventually select the channel for transmission according to the scheduler's scheduling (e.g., guaranteed-throughput time slots or best-effort arbitration). EX1005, 3–4. When the channel is eventually selected, a packet is created that contains the largest possible amount of data (and credits). Radulescu at 4. The size of this packet is the lesser of the channel's sendable data or a predetermined maximum packet length. EX1005, 4. This is materially different from what is recited in claim 12 of the '449 Patent for at least two reasons.

98. First, in claim 12, when the data in the buffer reaches the optimal amount, the data in the buffer will be sent (i.e., transferred to a processing module via the network). On the other hand, Radulescu teaches that when the amount of sendable data in a queue for a channel exceeds the configurable threshold, that channel is merely *eligible for scheduling*. As I explained above, eligible for scheduling means that the channel may eventually be selected by the scheduler according to the selection, or arbitration, scheme.

99. Second, in claim 12, when the data in the buffer reaches the optimal amount, *that data* in the buffer will be sent. On the other hand, Radulescu teaches that when the amount of sendable data in a queue for a channel exceeds the configurable threshold, that channel becomes *eligible for scheduling*, which does not mean data in the queue will be sent or that any specific data in the queue will be sent. A POSITA would recognize that while the eligible channel is waiting to be

selected by the scheduler, the eligible channel's queue will continue to fill up with data it receives from the IP block to be sent across the network. Thus, when the eligible channel is eventually selected by the scheduler, the amount of data in the queue is likely more than the amount that was in the queue when it became eligible for scheduling. To be clear, claim 12 of the '800 Patent specifies that *the specifically buffered optimal amount of data will be sent* after that optimal amount of data is reached; Radulescu specifies that a channel is eligible for scheduling when the configurable threshold is exceeded, but doesn't specify which data may eventually be sent. Indeed, in Radulescu, the size of the packet that is formed for transmission when a channel is eventually selected is the lesser of the channel's sendable data or a predetermined maximum packet length. Such packet likely contains less than the amount in the queue when the channel was finally selected, and therefore, it is unknown what data from the queue is actually sent. Put simply, while Radulescu discloses sending a packet whose data is taken from the eligible channel's queue, Radulescu does not disclose or suggest that the data buffered according to the configurable threshold is the data that is packetized and sent.

100. For at least the above reasons, Radulescu does not disclose or suggest "the first interconnect module transferring the first data to the second processing module when the first data buffered in the first memory reaches the first optimal amount, and the second interconnect module transferring the second data to the

first processing module when the second data buffered in the second memory reaches the second optimal amount” as recited in claim 12 of the ’800 Patent.

101. Next, as explained in Sec. VI.A.1.i above, Bahadiroglu’s adaptive packet mechanism determines an optimum packet size that yields the highest probability of successful delivery. EX1006, ¶ 77 (“As described, this optimum packet size and inter-packet interval will, as a consequence, yield the highest probability of packets arriving successfully at a given Receiving Node 12R.”). Bahadiroglu’s “optimum packet size and an optimum inter-packet interval for transmission of packet data between the sending node and the receiving node . . . are used in the transmission of data packets from the sending node to the receiving node,” (Bahadiroglu at Abstract). However, as I explain in Sec. VI.A.1.i, Bahadiroglu’s optimum packets that are sent are not “longer packets” for optimizing NoC utilization as specified by Radulescu. Thus, a POSITA would lack motivation to combine Radulescu and Bahadiroglu, a POSITA would not have a reasonable expectation of success, and the combination would frustrate the purpose of Radulescu’s optimization of NoC utilization.

102. For at least the above reasons, Radulescu in view of Bahadiroglu does not disclose or suggest “the first interconnect module transferring the first data to the second processing module when the first data buffered in the first memory reaches the first optimal amount, and the second interconnect module transferring

the second data to the first processing module when the second data buffered in the second memory reaches the second optimal amount” as recited in claim 12 of the ’800 Patent.

3. Radulescu in view of Bahadiroglu fails to disclose “wherein at least one of the first determination unit and the second determination unit is further configured to determine an optimal moment for sending the data in said first wrapper or said second wrapper according to communication properties of the communication between the master and the slave”

103. Claim 12 of the ’800 Patent further recites that “at least one of the first determination unit and the second determination unit is further configured to determine an optimal moment for sending the data in said first wrapper or said second wrapper according to communication properties of the communication between the master and the slave.” Paragraphs 94–96 of Dr. Mowry’s Declaration allege that Radulescu discloses this claim limitation, and paragraphs 97–98 of Dr. Mowry’s Declaration allege that Radulescu in view of Bahadiroglu renders this claim limitation obvious. EX1003, ¶¶ 94–98. I disagree.

104. As I explained in Sec. VI.A.1.i, Radulescu’s schedulers do not disclose the first and second determination units because no scheduler is “configured to determine an optimal amount of data to be buffered.” For at least this reason, Radulescu does not disclose the claim limitation discussed in this section.

105. As I detailed in Sec. VI.A.1.ii, Radulescu's scheduler schedules, or arbitrates between, channels, or queues, for transmission. Radulescu at 3. When the scheduler selects a channel (either when a guaranteed-throughput channel's time slot arrives or via an arbitration scheme for best-effort channels), the selected channel is granted transmission if it has data to send and there is space in the destination queue. EX1005, 4. Arbitration schemes include round-robin, weighted round-robin, and based on queue filling. EX1005, 4.

106. In Radulescu's system, no channels, or queues, are included in the scheduler's scheduling process unless the channel's sendable data exceeds the threshold (or there are enough credits). EX1005, 3. A channel becomes "eligible for scheduling" when its sendable data exceeds the threshold (or there are enough credits). EX1005, 4. Eligible for scheduling means the channel is selectable by the scheduler for scheduling; eligible for scheduling does not mean that sendable data (or any data in the queue) of the channel is sent. EX1005, 4. Rather, the scheduler may eventually select the channel for transmission according to the scheduler's scheduling (e.g., guaranteed-throughput time slots or best-effort arbitration). EX1005, 3–4. When the channel is eventually selected, a packet is created that contains the largest possible amount of data (and credits). EX1005, 4. The size of this packet is the lesser of the channel's sendable data or a predetermined maximum packet length. EX1005, 4.

107. First, in claim 12, when the data in the buffer reaches the optimal amount, the data in the buffer will be sent (i.e., transferred to a processing module via the network) at the optimal moment. On the other hand, Radulescu teaches that when the amount of sendable data in a queue for a channel exceeds the configurable threshold, that channel is merely *eligible for scheduling*, and whenever that channel is eventually selected, a packet is sent.

108. Second, in claim 12, when the data in the buffer reaches the optimal amount, *that data* in the buffer will be sent at the optimal moment. On the other hand, Radulescu teaches that when the amount of sendable data in a queue for a channel exceeds the configurable threshold, that channel becomes *eligible for scheduling*, which does not mean data in the queue will be sent or that any specific data in the queue will be sent. A POSITA would recognize that while the eligible channel is waiting to be selected by the scheduler, the eligible channel's queue will continue to fill up with data it receives from the IP block to be sent across the network. Thus, when the eligible channel is eventually selected by the scheduler, the amount of data in the queue is likely more than the amount that was in the queue when it became eligible for scheduling. To be clear, claim 12 of the '800 Patent specifies that *the specifically buffered optimal amount of data will be sent at the optimal moment after that optimal amount of data is reached*. EX1001, 12:18–19 (claiming “an optimal moment for sending **the data** in said first wrapper or said

second wrapper” (emphasis added))¹. In contrast, Radulescu specifies that a channel is eligible for scheduling when the configurable threshold is exceeded, but does not specify which data may eventually be sent. Indeed, in Radulescu, the size of the packet that is formed for transmission when a channel is eventually selected is the lesser of the channel’s sendable data or a predetermined maximum packet length. Such packet likely contains less than the amount in the queue when the channel was finally selected, and therefore, it is unknown what data from the queue is actually sent. Put simply, while Radulescu discloses sending a packet whose data is taken from the eligible channel’s queue, Radulescu does not disclose or suggest that the data buffered according to the configurable threshold is the data that is packetized and sent. Thus, even if the moment at which transmission is granted for the eligible channel (e.g., when the channel is selected) is an “optimal moment for sending the data” as alleged in Dr. Mowry’s Declaration, Radulescu does not disclose or suggest “an optimal moment for sending the [specifically buffered optimal amount of] data in said first wrapper or said second wrapper.”

109. For at least the above reasons, Radulescu does not disclose or suggest “wherein at least one of the first determination unit and the second determination

¹ As indicated in Sec. V.D, Patent Owner has proposed a construction for “said first wrapper” / “said second wrapper” to mean “said first memory” / “said second memory.”

unit is further configured to determine an optimal moment for sending the data in said first wrapper or said second wrapper” as recited in claim 12 of the ’800 Patent.

110. As I explained in Sec. VI.A.1.i, the combination of Bahadiroglu’s adaptive packet mechanism with Radulescu’s scheduler does not disclose or suggest the first and second determination units because, inter alia, using Bahadiroglu’s adaptive packet mechanism, where the optimum packet size is based on probability of successful delivery, to configure Radulescu’s configurable threshold for optimizing NoC utilization based on sending longer packets, would frustrate the purpose of Radulescu’s optimization of NoC utilization. For at least this reason, Radulescu in view of Bahadiroglu does not disclose or suggest the claim limitation discussed in this section.

111. Bahadiroglu’s adaptive packet mechanism determines an optimum packet size and an optimum inter-packet interval that yields the highest probability of successful delivery. EX1006, ¶ 77 (“As described, this optimum packet size and inter-packet interval will, as a consequence, yield the highest probability of packets arriving successfully at a given Receiving Node 12R.”). Paragraph 97 of Dr. Mowry’s Declaration alleges that “[a] POSITA would have understood that by determining the optimum interval between transmission of data, Bahadiroglu’s mechanism determines the optimal moment to transmit each data packet.” EX1003, ¶97. I disagree.

112. First, a POSITA understand a difference between an optimal moment and an optimal interval. First, a moment is “an indefinitely short period of time; instant.” *See* <https://www.dictionary.com/browse/moment>. In contrast, an interval is “an intervening period of time.” <https://www.dictionary.com/browse/interval>. A POSITA would recognize that there are many moments within an interval. Accordingly, there would be many optimal moments within an optimal interval. However, that would be contrary to the concept of an optimal moment, because any moment could be an optimal moment, and therefore, no one moment is any better than any other moment (hence, no moment is an optimal moment).

113. Bahadiroglu explains that the “inter-packet interval” is “the interval between packets that are transmitted consecutively through the connection.” EX1006, ¶ 68; *see also* EX1006, ¶ 69 (describing “inter-packet delay of 10 msec” and “inter-packet delay of 20 ms”). However, Bahadiroglu does not disclose or suggest *when* the inter-packet interval starts, i.e., when the first packet of a quantity of packets is sent. Therefore, even if Bahadiroglu’s inter-packet interval were somehow incorporated into Radulescu’s scheduler, a POSITA would immediately recognize that such interval would be of no use in informing the scheduler as to when data should be sent. Indeed, the once Radulescu’s scheduler finally grants transmission to a channel whose sendable data exceeded the configurable threshold, a *single* packet is transmitted. EX1005, 4 (“[O]nce a queue is selected, a

packet containing the largest possible amount of credits and data will be produced.”). An inter-packet interval requires, by definition, at least two packets.

114. For at least the above reasons, Radulescu in view of Bahadiroglu does not disclose or suggest “wherein at least one of the first determination unit and the second determination unit is further configured to determine an optimal moment for sending the data in said first wrapper or said second wrapper” as recited in claim 12 of the ’800 Patent.

4. Analysis of Motivation to Combine

115. A POSITA would not have been motivated to combine Radulescu with Bahadiroglu to arrive at the claimed invention and could not have done so.

116. As explained in Sec. V.B above, a POSITA would have been dissuaded from attempting to modify the prior art to arrive at the claimed invention. Modifying prior art systems that do not provide the characteristics described above, for example, quality of service guarantees, described in Sec. V.B would require extensive redesign to arrive at the invention. Nothing in Radulescu or Bahadiroglu suggests or alludes to how such a redesign could be implemented. I also incorporate the reasons a POSITA would not have modified or combined the references in the previous sections (Sec. VI.A.1–3) and incorporate those into this section.

117. Dr. Mowry opines that it would have been obvious to implement Radulescu 's NoC using Bahadiroglu's adaptive control and threshold mechanism implementation. EX1003, ¶¶ 57–63. I note that Dr. Mowry provides no details of why a POSITA would have any motivation to combine these references. Dr. Mowry's discussion treats the references at a generic level without looking at the specific disclosures. Radulescu and Bahadiroglu are fundamentally different.

118. For example, as discussed above, Radulescu is directed to a NoC, while Bahadiroglu is directed to a computer network. The '800 Patent details the differences between a NoC compared to a bus-based interconnect and a multi-hop interconnect compared to an off-chip computer network. For example, the '800 Patent teaches: "NoC's premises are different from off-chip networks, and, therefore, most of the network design choices must be evaluated" because "[o]n-chip networks have different properties (e.g., tighter link synchronization) and constraints (e.g., higher memory cost) leading to different design choices, which ultimately affect the network services." EX1001, 1:66–2:5. An on-chip multi-hop network differs from an off-chip computer network with respect to: (1) data ordering; (2) time-related guarantees; (3) communication reliability; and (4) constraints and synchronization. *Supra* § V.B. In my opinion, a POSITA would not look to an off-chip computer network (such as the network disclosed in Bahadiroglu) to arrive at the claimed inventions.

119. Dr. Mowry opines that “a POSITA would have reasonably expected that using Bahadiroglu’s known technique for optimizing data communication in a network through its adaptive packet mechanism would improve aRadulescu’s NoC in the same way.” EX1002, ¶ 61. However, Dr. Mowry ignores that implementation of end-to-end performance guarantees in a multi-hop interconnect in an on-chip environment is drastically different compared to an off-chip computer network. *See* ’449 Patent at 3:34–46 (“latency guarantees [are] very hard to offer” in off-chip networks); *id.* at 5:65–6:7 (In a multi-hop interconnect, as opposed to a bus, “[b]oth end-to-end access time and transport time increase proportionally to the number of hops between master and slave. However, networks links are unidirectional and point to point, and hence can run at higher frequencies than buses, thus lowering the latency. From a latency perspective, using a bus or a network is a trade off between the number of modules connected to the interconnect (which affects access time), the speed of the interconnect, and the network topology.”). Therefore, in my opinion, the type of end-to-end performance guarantees, and the implementation of such guarantees, differs drastically in a multi-hop interconnect in an on-chip environment compared to a computer network.

120. As explained above, Dr. also Mowry fails to provide any explanation for how the combination of Radulescu and Bahadiroglu would be implemented.

Moreover, the '800 Patent teaches that the considerations for implementing end-to-end performance guarantees in a multi-hop interconnect compared to a computer network are drastically different.

121. Due to the differences between buses and computer networks compared to multi-hop interconnects, a person of skill in the art would have been dissuaded from attempting to modify a bus or a computer network to arrive at the claimed invention.

B. Ground 2: Radulescu in View of Petersen Does Not Render '800 Patent Claim 12 Obvious

1. Radulescu in view of Petersen fails to disclose “wherein the first interconnect module includes a first memory for buffering first data from the first processing module, the first interconnect module further including a first determination unit, wherein the first determination unit is configured to determine a first optimal amount of data to be buffered by the first memory, and the second interconnect module includes a second memory for buffering second data from the second processing module, the second interconnect module further including a second determination unit, wherein the second determination unit is configured to determine a second optimal amount of data to be buffered by the second memory”

122. Sec. IX.B.1.e of Dr. Mowry’s Declaration alleges that Radulescu discloses the limitation of claim 12 of the '800 patent reciting the first and second determination units determining first and second optimal amounts of data, respectively. I disagree.

123. In relevant part, claim 12 of the '800 Patent recites a first interconnect module for providing a communication connection between a first processing module and a network. The first interconnect module includes a first memory for buffering first data from the first processing module and a first determination unit configured to determine a first optimal amount of data to be buffered by the first memory. Claim 12 recites similar elements and functions with respect to a second interconnect module, a second processing module, a second memory, a second determination unit, and a second optimal amount of data. To be clear, the determination unit *determines* the optimal amount of data to be buffered. The specification of the '800 Patent, which uses terms like “optimal burst length,” “optimal packet,” “optimal transfer,” and “optimal first or second amount of data” when discussing an optimal amount of data to be buffered, explains that “determination of the optimal burst length can be performed statically or dynamically, during the initial phases of the required processing or during predetermined time intervals.” EX1001, 6:52–56. The specification provides additional disclosure as follows.

Alternatively, the optimal burst length may be determined dynamically (1) every time a connection is set up for transferring data over the network, (2) for every transaction, (3) whenever an IP switches to another connection, (4) for every packet, (5) when the state of the network on chip changes (e.g., reconfiguration, NoC load, buffer fillings, etc.). The determination of the optimal burst length may be performed based on information stored in the determination

units BDLU 1, 2 or on information received from the network N or IP blocks.

EX1001, 6:56–65.

124. Dr. Mowry’s Declaration refers to Sec. IX.A.1.e (Ground 1) for support. Paragraph 85 of Dr. Mowry’s Declaration alleges that “the scheduler included in the NI2 (the claimed first interconnect module) and the scheduler included in the NI1 (the claimed second interconnect module) are each a “determination unit” included in each respective interconnect module at least because each scheduler makes a determination about whether to transmit buffered data **based on a configurable threshold that is stored or received by the scheduler.**” EX1003, ¶ 85 (emphasis added). However, claim 12 of the ’800 Patent does not say that the determination unit (e.g., a scheduler) makes a decision *based on* some predetermined (e.g., stored or received) optimal amount of data to be buffered (e.g., a configurable threshold). Claim 12 specifies that the determination unit “is configured to determine” the optimal amount of data to be buffered, (EX1001, 11:67–12:6).

125. While Radulescu discloses a “configurable threshold,” it does not disclose or suggest any circuit or device capable of configuring the threshold, e.g., capable of determining an optimal amount of data to be buffered. *See, e.g.,* EX1005, 4 (“[W]e implemented a configurable threshold mechanism.”); *see also*

EX1005, 4 (“[A] queue becomes eligible for scheduling when either the amount of sendable data are above a first threshold, or when the amount of credits is above a second threshold”).

126. For at least the above reasons, Radulescu does not disclose or suggest either the first determination unit or the second determination unit as recited in claim 12 of the '800 Patent.

127. Paragraph 160 of Dr. Mowry's Declaration alleges that Radulescu in view of Petersen discloses a determination unit configured to determine an optimal amount of data to be buffered in a network interface memory as recited in claim 12 of the '800 Patent. EX1003, ¶ 160. Paragraph 162 of Dr. Mowry's Declaration further alleges that “[a] POSITA would have understood that Petersen's optimal “threshold value” represents an optimal amount of data to be buffered, because the threshold determines the amount of data that must be buffered in the network interface memory before the data can be transmitted across the network”). *Id.*, ¶162. I disagree.

128. First, a POSITA would lack motivation to combine Radulescu and Petersen because Radulescu is directed to on-chip networks for communication between IP modules where “the size of IP modules attached to the NoC is relatively small,” (EX1005, 1), while Petersen is directed to computer networks and “the interfaces between communications networks and data processing

systems,” (EX1007, 1:17–18), where “[t]he host computer or sending system 30 communicates with a host interface logic 31,” (EX1007, 4:26–27). Peterson explains that representative prior art systems include a local-area network (LAN) network interface controller microchip and a LAN coprocessor microchip. EX1007, 2:3–6. (“Representative prior art systems include the National Semiconductor DP83932B, a systems-oriented network 5 interface controller (SONIC) and the Intel 82586 local area network coprocessor.”). Petersen’s “threshold logic” is directed to computer networks, not on-chip networks (Radulescu, and the ’800 Patent, are directed to on-chip networks). FIG. 2 of Petersen (reproduced below) is “a schematic functional block diagram of the early transmit system according to the present invention.” EX1007, 3:1–3.

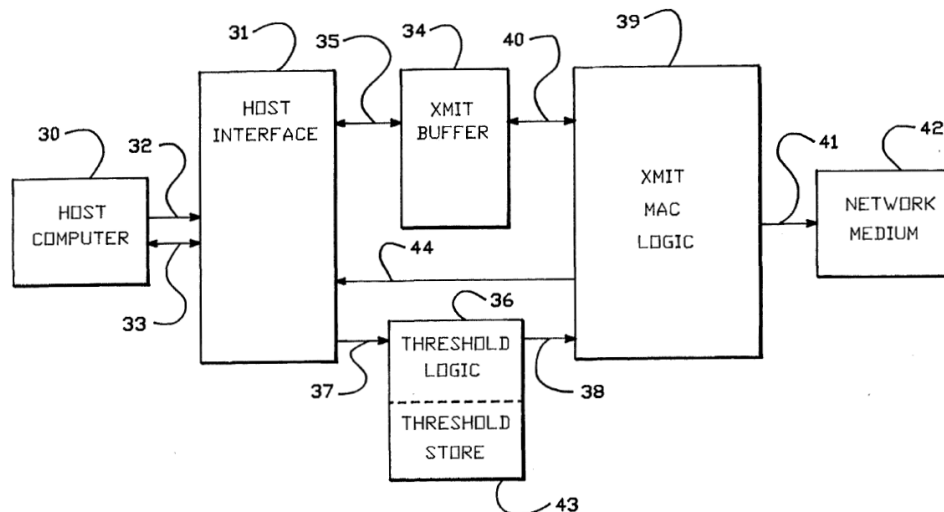


FIG. -2

129. Petersen explains that “[a] network interface adapter 6 is coupled to the host bus 2. The adapter 6 is coupled to a transceiver 7 which is, in turn, connected to a network medium 8, such as coaxial cable, a twisted pair conductor, a fiber optic, satellite, wireless, or other communication medium.” EX1007, 4:2–7. For the reasons explained in Sec. V.B above, a POSITA would lack motivation to combine elements developed for off-chip networks into an on-chip network environment.

130. Second, a POSITA would lack motivation to combine Radulescu and Petersen because Radulescu teaches that “[t]o optimize the NoC utilization, it is preferable to send longer packets,” (EX1005, 4), while Petersen’s transmit logic uses a threshold value that seeks to balance an input rate of a transmit buffer, for data of a frame provided by a host computer, to an output rate of the data of the frame to a communication network. EX1007, 4:56–5:13 (“The transmit logic 39 then begins retrieving data from the buffer 34 to support transmission of the frame on the medium 42. This operation begins before the entire frame has been transferred from the host computer 30 into the buffer 34, if the transmit logic 39 is available to transmit the frame subject of the ongoing download from the host computer 30, the frame being downloaded into the buffer 34 is larger than the threshold set by the threshold store 43, and the host computer 30 indicates that immediate transmission of the data is desired.”). Petersen explains that “the

threshold value may be set by the host system to optimize performance using the alterable threshold store and the posted status information.” EX1007, 2:31–34; *see also id.*, 29:39–40 (“The value for this [threshold] register may be programmed by the host to optimize performance.”). Petersen further explains that “[i]f set too low, system latencies or bandwidth limitations may cause the adapter to underrun the network during transmission, causing a partial frame with a guaranteed bad CRC to be transmitted. If the value is set too high, then unnecessary delays will be incurred before the start of transmission.” EX1007, 29:40–44. A POSITA would recognize that, while Petersen’s transmit-buffer rate balancing technique is not divorced from the properties of the communication network, it is not commensurate to optimizing network utilization as taught by Radulescu. Thus, a POSITA would lack motivation to combine elements developed for transmit-buffer rate balancing into a system for optimizing network utilization, and a POSITA would not have a reasonable expectation of success.

131. Recall that claim 12 of the ’800 Patent recites, in first and second counterparts, an interconnect module that includes a memory for buffering data from a processing module, wherein the interconnect module further includes a determination unit, wherein the determination unit is configured to determine an optimal amount of data to be buffered by the memory. To be clear, the determination unit, which *determines* the optimal amount of data, is included in the

interconnect module, not in the processing module. Further, Radulescu explains the importance to “separate the IP requirements and their implementation, and make application quality of service independent of the IP and NoC implementations.”

EX1005, 1; *see also id.*, 2 (“Decoupling computation from communication is a key ingredient in managing the complexity of designing chips with billions of transistors, because it **allows the IP modules and the interconnect to be designed independently.**” (emphasis added)). Further still, Radulescu explains that “[t]he pressure to keep the protocol stack small is higher on-chip than off-chip, because the size of the IP modules attached to the NoC is relatively small.” To that end, Radulescu discusses the need for simple network interfaces. EX1005, 1 (“On-chip network interfaces must provide a low-area overhead, because the size of IP modules attached to the NoC is relatively small.”); *see also id.*, 1 (“Our NI is intended for systems on chip (SoC), hence, it must have low area. . . . we also have to provide a simple and flexible configuration.”).

132. However, Petersen teaches a system where the host computer, e.g., a processing module, performs the function of determining the value of the threshold for used by the transmit logic (external to the host computer) in the transmit-buffer rate balancing. EX1007, 29:39–40 (“[T]he threshold value may be set by the host system to optimize performance using the alterable threshold store and the posted status information.”); *id.*, 29:39–40 (“The value for this [threshold] register may be

programmed by the host to optimize performance.”); *id.*, 4:46–47 (“The threshold store 43, in a preferred system, is dynamically programmable by the host computer 30.”); *id.*, 4:56–60 (“The transmit logic 39 also supplies status information across line 44 to the host interface logic 31, for posting to the host system. The status information includes indications of underrun conditions and may be used by the host to optimize the value in the threshold store 43.”).

133. Accordingly, a POSITA would lack motivation to incorporate a technique for determining a threshold value for transmit-buffer rate matching that is implemented by a host computer into Radulescu’s small and simple network interfaces, and further, a POSITA would not have a reasonable expectation of success in doing so.

2. Radulescu in view of Petersen fails to disclose “the first interconnect module transferring the first data to the second processing module when the first data buffered in the first memory reaches the first optimal amount, and the second interconnect module transferring the second data to the first processing module when the second data buffered in the second memory reaches the second optimal amount”

134. Claim 12 of the ’800 Patent further recites that the first interconnect module transfers the first data buffered in the first memory to a second processing module when the first data buffered in the first memory reaches the first optimal amount. Claim 12 recites similar elements and functions with respect to the second interconnect module, the second memory, and the second optimal amount.

Paragraph 164 of Dr. Mowry's Declaration alleges that Radulescu in view of Petersen renders this limitation obvious. EX1003, ¶164. I disagree.

135. First, as I explained in Sec. VI.A.1.ii above, Radulescu does not disclose or suggest an interconnect module transferring a specifically buffered optimal amount of data as recited in claim 12 of the '800 Patent.

136. Radulescu describes a system in which channels, or queues, are scheduled for transmission by a scheduler. EX1005, 3. When the scheduler selects a channel (either when a guaranteed-throughput channel's time slot arrives or via an arbitration scheme for best-effort channels), the selected channel is granted transmission if it has data to send and there is space in the destination queue. EX1005, 4. To optimize NoC utilization, Radulescu explains that it is preferable to send longer packets. EX1005, 4. A configurable threshold mechanism causes the scheduler to skip a channel if the "sendable data" for the channel is below the threshold, where sendable data refers to how much data can be sent before overflowing the destination queue, and skipping a channel refers to not granting transmission to a guaranteed-throughput channel when its transmission slot arrives or not granting transmission to a best-effort channel when it is selected by the arbitration scheme. EX1005, 3–4.

137. In Radulescu's system, no channels, or queues, are included in the scheduler's scheduling process unless the channel's sendable data exceeds the

threshold (or there are enough credits). EX1005, 3. A channel becomes “eligible for scheduling” when its sendable data exceeds the threshold (or there are enough credits). *Id.*, 4. Eligible for scheduling means the channel is selectable by the scheduler for scheduling; eligible for scheduling does not mean that sendable data (or any data in the queue) of the channel is sent. *Id.*, 4. Rather, the scheduler may eventually select the channel for transmission according to the scheduler’s scheduling (e.g., guaranteed-throughput time slots or best-effort arbitration). *Id.*, 3–4. When the channel is eventually selected, a packet is created that contains the largest possible amount of data (and credits). *Id.*, 4. The size of this packet is the lesser of the channel’s sendable data or a predetermined maximum packet length. *Id.*, 4. This is materially different from what is recited in claim 12 of the ’449 Patent for at least two reasons.

138. First, in claim 12, when the data in the buffer reaches the optimal amount, the data in the buffer will be sent (i.e., transferred to a processing module via the network). On the other hand, Radulescu teaches that when the amount of sendable data in a queue for a channel exceeds the configurable threshold, that channel is merely *eligible for scheduling*. As I explained above, eligible for scheduling means that the channel may eventually be selected by the scheduler according to the selection, or arbitration, scheme.

139. Second, in claim 12, when the data in the buffer reaches the optimal amount, *that data* in the buffer will be sent. On the other hand, Radulescu teaches that when the amount of sendable data in a queue for a channel exceeds the configurable threshold, that channel becomes *eligible for scheduling*, which does not mean data in the queue will be sent or that any specific data in the queue will be sent. A POSITA would recognize that while the eligible channel is waiting to be selected by the scheduler, the eligible channel's queue will continue to fill up with data it receives from the IP block to be sent across the network. Thus, when the eligible channel is eventually selected by the scheduler, the amount of data in the queue is likely more than the amount that was in the queue when it became eligible for scheduling. To be clear, claim 12 of the '800 Patent specifies that *the specifically buffered optimal amount of data will be sent* after that optimal amount of data is reached; Radulescu specifies that a channel is eligible for scheduling when the configurable threshold is exceeded, but doesn't specify which data may eventually be sent. Indeed, in Radulescu, the size of the packet that is formed for transmission when a channel is eventually selected is the lesser of the channel's sendable data or a predetermined maximum packet length. Such packet likely contains less than the amount in the queue when the channel was finally selected, and therefore, it is unknown what data from the queue is actually sent. Put simply, while Radulescu discloses sending a packet whose data is taken from the eligible

channel's queue, Radulescu does not disclose or suggest specifically what data is actually packetized and sent.

140. For at least the above reasons, Radulescu does not disclose or suggest “the first interconnect module transferring the first data to the second processing module when the first data buffered in the first memory reaches the first optimal amount, and the second interconnect module transferring the second data to the first processing module when the second data buffered in the second memory reaches the second optimal amount” as recited in claim 12 of the '800 Patent.

141. Next, as explained in Sec. VI.B.1.i above, Petersen's transmit logic uses a threshold value that seeks to balance an input rate of a transmit buffer, for data of a frame provided by a host computer, to an output rate of the data of the frame to a communication network. EX1007, 4:56–5:13; *Id.*, 2:31–34; *Id.*, 29:39–40; 29:40–44. Petersen's rate-balancing technique relies on the frames being of a predetermined length. *Id.*, 14:43–46 (“The FRAME LENGTH field is computed and updated by the adapter. The frame length is calculated by summing all of the XMIT DATA LEN values and the XMIT IMMED LEN value.”) The XMIT DATA LEN and XMIT IMMED LEN values are provided by the host computer. *Id.*, 16:1–2 (“The XMIT DATA LEN field, one of two entries per host data buffer descriptor.”); *Id.*, 14:27–29 (“once the last XMIT DATA LEN value has been written to the adapter, the frame length can be calculated.”); *Id.*, 14:19–20 (“XMIT

IMMED LEN values have been written to the adapter.”); *Id.*, 15:35–36 (“The XMIT IMMED LEN field, defined by a write from the host.”). Accordingly, a POSITA would lack motivation to incorporate Petersen’s threshold logic, which relies on frames, e.g., packets, being of a predetermined length dictated by the host computer, into Radulescu’s configurable threshold mechanism, which relies on increasing the length of packets to optimize NoC utilization. Further, a POSITA would not have a reasonable expectation of success in doing so. For example, a POSITA would not understand how to modify Radulescu’s channel scheduling scheme, which selects queues for transmission, to accommodate Petersen’s partially-filled transmit-buffers.

142. For at least the above reasons, Radulescu in view of Petersen does not disclose or suggest “the first interconnect module transferring the first data to the second processing module when the first data buffered in the first memory reaches the first optimal amount, and the second interconnect module transferring the second data to the first processing module when the second data buffered in the second memory reaches the second optimal amount” as recited in claim 12 of the ’800 Patent.

3. Radulescu in view of Petersen fails to disclose “wherein at least one of the first determination unit and the second determination unit is further configured to determine an optimal moment for sending the data in said first wrapper or said second wrapper according to communication properties of the communication between the master and the slave”

143. Claim 12 of the ’800 Patent further recites that “at least one of the first determination unit and the second determination unit is further configured to determine an optimal moment for sending the data in said first wrapper or said second wrapper according to communication properties of the communication between the master and the slave.” Dr. Mowry’s Declaration refers to Sec. IX.A.1.g (Ground 1) for support. Paragraphs 94–96 of Dr. Mowry’s Declaration allege that Radulescu discloses this claim limitation and paragraphs 97–98 of Dr. Mowry’s Declaration allege that Radulescu in view of Bahadiroglu renders this claim limitation obvious. EX1003, ¶¶94–98. I disagree.

144. As I explained in Sec. VI.B.1.i, Radulescu’s schedulers do not disclose the first and second determination units because no scheduler is “configured to determine an optimal amount of data to be buffered.” For at least this reason, Radulescu does not disclose the claim limitation discussed in this section.

145. As I detailed in Sec. VI.B.1.ii, Radulescu’s scheduler schedules, or arbitrates between, channels, or queues, for transmission. EX1005, 3. When the

scheduler selects a channel (either when a guaranteed-throughput channel's time slot arrives or via an arbitration scheme for best-effort channels), the selected channel is granted transmission if it has data to send and there is space in the destination queue. *Id.*, 4. Arbitration schemes include round-robin, weighted round-robin, and based on queue filling. *Id.*, 4.

146. In Radulescu's system, no channels, or queues, are included in the scheduler's scheduling process unless the channel's sendable data exceeds the threshold (or there are enough credits). *Id.*, 3. A channel becomes "eligible for scheduling" when its sendable data exceeds the threshold (or there are enough credits). *Id.*, 4. Eligible for scheduling means the channel is selectable by the scheduler for scheduling; eligible for scheduling does not mean that sendable data (or any data in the queue) of the channel is sent. *Id.*, 4. Rather, the scheduler may eventually select the channel for transmission according to the scheduler's scheduling (e.g., guaranteed-throughput time slots or best-effort arbitration). *Id.*, 3–4. When the channel is eventually selected, a packet is created that contains the largest possible amount of data (and credits). *Id.*, 4. The size of this packet is the lesser of the channel's sendable data or a predetermined maximum packet length. *Id.*, 4.

147. First, in claim 12, when the data in the buffer reaches the optimal amount, the data in the buffer will be sent (i.e., transferred to a processing module

via the network) at the optimal moment. On the other hand, Radulescu teaches that when the amount of sendable data in a queue for a channel exceeds the configurable threshold, that channel is merely *eligible for scheduling*, and whenever that channel is eventually selected, a packet is sent.

148. Second, in claim 12, when the data in the buffer reaches the optimal amount, *that data* in the buffer will be sent at the optimal moment. On the other hand, Radulescu teaches that when the amount of sendable data in a queue for a channel exceeds the configurable threshold, that channel becomes *eligible for scheduling*, which does not mean data in the queue will be sent or that any specific data in the queue will be sent. A POSITA would recognize that while the eligible channel is waiting to be selected by the scheduler, the eligible channel's queue will continue to fill up with data it receives from the IP block to be sent across the network. Thus, when the eligible channel is eventually selected by the scheduler, the amount of data in the queue is likely more than the amount that was in the queue when it became eligible for scheduling. To be clear, claim 12 of the '800 Patent specifies that *the specifically buffered optimal amount of data will be sent at the optimal moment after that optimal amount of data is reached*. EX1001, 12:18–19 (claiming “an optimal moment for sending **the data** in said first wrapper or said

second wrapper” (emphasis added))². In contrast, Radulescu specifies that a channel is eligible for scheduling when the configurable threshold is exceeded, but does not specify which data may eventually be sent. Indeed, in Radulescu, the size of the packet that is formed for transmission when a channel is eventually selected is the lesser of the channel’s sendable data or a predetermined maximum packet length. Such packet likely contains less than the amount in the queue when the channel was finally selected, and therefore, it is unknown what data from the queue is actually sent. Put simply, while Radulescu discloses sending a packet whose data is taken from the eligible channel’s queue, Radulescu does not disclose or suggest that the data buffered according to the configurable threshold is the data that is packetized and sent. Thus, even if the moment at which transmission is granted for the eligible channel (e.g., when the channel is selected) is an “optimal moment for sending the data” as alleged in Dr. Mowry’s Declaration, Radulescu does not disclose or suggest “an optimal moment for sending the [specifically buffered optimal amount of] data in said first wrapper or said second wrapper.”

149. For at least the above reasons, Radulescu does not disclose or suggest “wherein at least one of the first determination unit and the second determination

² As indicated in Sec. V.D, Patent Owner has proposed a construction for “said first wrapper” / “said second wrapper” to mean “said first memory” / “said second memory.”

unit is further configured to determine an optimal moment for sending the data in said first wrapper or said second wrapper” as recited in claim 12 of the ’800 Patent.

4. Analysis of Motivation to Combine

150. A POSITA would not have been motivated to combine Radulescu with Petersen to arrive at the claimed invention and could not have done so.

151. I note that, while Dr. Mowry opines generally that there is a motivation to combine Radulescu with Petersen, he never opines on the motivation to take a specific teaching or feature from Petersen and combine it with Radulescu. In my opinion, Dr. Mowry does not provide sufficient basis for a motivation to combine any specific teaching or feature from Petersen with Radulescu.

152. As explained in Sec. V.B above, a POSITA would have been dissuaded from attempting to modify the prior art to arrive at the claimed invention. Modifying prior art systems that do not provide the characteristics described above, for example, quality of service guarantees, described in Sec. V.B would require extensive redesign to arrive at the invention. Nothing in Radulescu or Petersen suggests or alludes to how such a redesign could be implemented. I also address additional reasons why a POSITA would not have combined the references in the previous sections (Secs. VI.B.1–3) and incorporate them here.

153. Dr. Mowry opines that it would have been obvious to implement Radulescu’s NoC using Petersen’s threshold mechanism. EX1003, ¶¶ 147–153. I

note that Dr. Mowry provides no details of why a POSITA would have any motivation to combine these references. Dr. Mowry's discussion treats the references at a generic level without looking at the specific disclosures. Radulescu and Petersen are fundamentally different.

154. For example, as discussed above, Radulescu is directed to a NoC, while Petersen is directed to a SoC. The '800 Patent details the differences between a NoC compared to SoC. "However, NoC's premises are different from off-chip networks, and, therefore, most of the network design choices must be reevaluated. On-chip networks have different properties (e.g., tighter link synchronization) and constraints (e. g., higher memory cost) leading to different design choices, which ultimately affect the network services. Introducing networks as on-chip interconnects radically changes the communication when compared to direct interconnects, such as buses or switches. This is because of the multi-hop nature of a network, where communication modules are not directly connected, but separated by one or more network nodes. This is in contrast with the prevalent existing interconnects (i.e., buses) where modules are directly connected." EX1001, 1:65–2:10. An on-chip multi-hop network differs from a bus-based interconnect with respect to: (1) media arbitration; (2) data routing; (3) data formatting; (4) latency; and (5) flow control for input buffers. *Supra* § V.B. Therefore, in my opinion, a POSITA would not implement the teachings of a direct

interconnect (such as Radulescu ’s communication subsystem) to arrive at a multi-hop interconnect.

155. Dr. Mowry opines that a POSITA would combine a network on chip with an off-chip interconnection. However, Dr. Mowry ignores that implementation of end-to-end performance guarantees in a multi-hop interconnect in an on-chip environment is drastically different compared to a bus-based interconnect. *See* EX1001, 3:18–26 (“The above-mentioned on-chip protocols have been designed mainly for buses with a small latency. In addition, these protocols have been designed based on the assumption that read operations are always urgent and should therefore be completed as soon as possible Without unnecessary buffering. However, as systems grow larger and multi-hop interconnects like networks or buses with bridges, the latency grows as well. In these cases the communication granularity become coarser and the latency requirements become less strict.”). Therefore, in my opinion, the type of end-to-end performance guarantees, and the implementation of such guarantees, differs drastically in a NoC compared to a SoC.

156. Dr. Mowry also opines that a “POSITA would have reasonably expected that use of Petersen’s known technique for optimizing data communication in a network through its dynamically programmable threshold would improve Radulescu’s NoC in the same way.” EX1003, ¶ 152. As explained

above, Dr. Mowry fails to provide any explanation for how the combination of Radulescu and Petersen would be implemented. Moreover, the '800 Patent teaches that the considerations for implementing end-to-end performance guarantees in a NoC to a SoC are drastically different.

157. Due to the differences between SoCs and NoCs, a person of skill in the art would have been dissuaded from attempting to modify Radulescu with Petersen to arrive at the claimed invention.

C. Ground 3: Radulescu in View of Bahadiroglu and/or Radulescu-WO Does Not Render '800 Patent Claim 12 Obvious

158. As discussed in Sec. VI.A.1 above, Radulescu and Bahadiroglu fail to disclose several limitations of claim 12. For at least these reasons, Radulescu in view of Bahadiroglu does not render claim 12 obvious.

159. The disclosure of Radulescu-WO is similar to that of Radulescu. Indeed, with reference to claim limitations 12[d], 12[e] and 12[f] (as numbered in Dr. Mowry's Declaration) Dr. Mowry's Declaration merely refers to earlier sections thereof without provided any citations to Radulescu-WO or any additional discussion. Accordingly, for the reasons I explained above with reference to those sections, Radulescu in view of Radulescu-WO does not render claim 12 obvious.

1. Radulescu-WO fails to disclose “wherein the first interconnect module includes a first memory for buffering first data from the first processing module, the first interconnect module further including a first determination unit, wherein the first determination unit is configured to determine a first optimal amount of data to be buffered by the first memory, and the second interconnect module includes a second memory for buffering second data from the second processing module, the second interconnect module further including a second determination unit, wherein the second determination unit is configured to determine a second optimal amount of data to be buffered by the second memory”

160. For the reasons explained in Sec. VI.A.1.i with respect to Radulescu, Radulescu-WO fails to disclose this claim limitation.

2. Radulescu-WO fails to disclose “the first interconnect module transferring the first data to the second processing module when the first data buffered in the first memory reaches the first optimal amount, and the second interconnect module transferring the second data to the first processing module when the second data buffered in the second memory reaches the second optimal amount”

161. For the reasons explained in Sec. VI.A.1.ii with respect to Radulescu, Radulescu-WO fails to disclose this claim limitation.

3. Radulescu-WO in view of Petersen fails to disclose “wherein at least one of the first determination unit and the second determination unit is further configured to determine an optimal moment for sending the data in said first wrapper or said second wrapper according to communication properties of the communication between the master and the slave”

162. For the reasons explained in Sec. VI.A.1.iii with respect to Radulescu, Radulescu-WO fails to disclose this claim limitation.

D. Ground 4: Radulescu in View of Peterson and/or Radulescu-WO Does Not Render '800 Patent Claim 12 Obvious

163. As discussed above, Radulescu, Petersen and/or Radulescu-WO fail to disclose all of the limitations of claim 12. A POSITA would not combine Radulescu, Petersen and/or Radulescu-WO for at least the same reasons discussed above.

VII. CONCLUSION

164. In addition to the materials identified in Appendix 2, I may also review other documents in this matter. I may also review and comment upon the testimony of others that has been or will be given in this matter.

165. I reserve the right to add or to modify this expert report in the event that inaccuracies or omissions are discovered or if new or additional information is provided to me. I also reserve the right to make any such other changes as may be deemed necessary or appropriate.

166. I have not yet developed any demonstratives or illustrations to illustrate my opinions and the bases for those opinions; however, I reserve the right to present such demonstratives or illustrations at trial.

167. I reserve the right to supplement and/or amend my analyses and opinions expressed herein in response to any new, additional, or modified positions taken by Defendants, its experts or witnesses in reports, declarations, affidavits, or

testimonies, whether in response to my report herein or in their coincident reports (or amendments thereto), to the extent such positions are relevant to the subject matter I have addressed in my report herein, or to the extent such positions cover topics relevant to this report for which Defendants have the burden of proof.

I declare under penalty of perjury under the laws of the United States that the foregoing is a true and correct reflection of my opinions. This declaration is executed this 27th day of March, 2024.

A handwritten signature in black ink, appearing to read "Erik Chmelar", written in a cursive style.

Dr. Erik Chmelar

APPENDIX 1

ERIK CHMELAR

(650) 533-7952 | echmelar@yahoo.com

Subject matter expert in semiconductor hardware and software with over two decades of experience developing, implementing, and protecting advanced technologies including digital and mixed-signal systems, high-speed serial communications, 5G/wireless telecommunications, and smart medical devices.

CREDENTIALS

EDUCATION —

- **JD, Intellectual Property**, Univ. Michigan Law School, Ann Arbor, MI 2015-2017
- **MBA, Management**, Western Michigan Univ., Kalamazoo, MI 2012-2013
- **BS, Chemical Physics**, Saginaw Valley State Univ., University Center, MI 2008-2011
- **PhD, Electrical Engineering**, Stanford Univ., Stanford, CA 2002-2004
Dissertation: “Test and Diagnosis of Field-programmable Gate Arrays.”
- **MS, Electrical Engineering**, Stanford Univ., Stanford, CA 2000-2001
- **BS, Electrical Engineering**, Michigan Technological Univ., Houghton, MI 1997-2000

LICENSES AND CERTIFICATIONS —

- **Agile Certified Practitioner** (PMI-ACP, PMI) 2015-pres.
- **Certified Project Management Professional** (PMP, PMI) 2014-pres.
- **Licensed Professional Engineer** (PE, State of Michigan LARA) 2010-pres.
- **Licensed Patent Agent/Attorney** (State Bar of Michigan, USPTO) 2007-pres.

EXPERIENCE

Technical Expert, Semiconductor Hardware/Software, Midland, MI 08/2007-pres.

- Serve as a technical expert witness for litigation in court and inter partes review (IPRs)
- Mine and analyze patent portfolios and develop assertion strategies for non-practicing entities (NPEs)
- Recent technology areas include memory, programmable hardware, microcontrollers, and AI devices

Senior Counsel, Young Basile Hanlon & MacFarlane, PC, Troy, MI 06/2023-pres.

- Draft and prosecute patent applications in a variety of semiconductor, software, and related arts
- Draft legal opinions for non-infringement, invalidity, and freedom-to-operate (FTO)
- Interview inventors, harvest inventions, and manage patent portfolios

Senior Associate, Dilworth IP, New Haven, CT 10/2020-06/2023

- Drafted and prosecuted patent applications in a variety of semiconductor, software, and related arts
- Drafted legal opinions for non-infringement, invalidity, freedom-to-operate (FTO), and export controls
- Interview inventors, harvest inventions, manage patent portfolios, and create technology landscapes

Senior Associate, Murphy, Bilak & Homiller, PLLC., Cary, NC 01/2020-06/2020

- Prosecuted patent applications for industry-leading wireless and semiconductor corporations
- Drafted approx. 100 office-action responses for complex 5G technologies
- Routinely referenced ETSI 3GPP specifications for prosecution of standard essential patents (SEPs)

Senior Manager, Innovation Partners, InterDigital, Inc., Wilmington, DE 09/2015-10/2018

- Created patent portfolios for 5G, IoT, connected/autonomous vehicles, and augmented/virtual reality
- Managed scientists and engineers to create over 100 new invention disclosures / patent applications
- Forged research partnerships with commercial and academic organizations to develop patent portfolios

- Co-founder, Managing Member**, Dera Industries LLC., Midland, MI 05/2016-07/2017
- Developed and productized the award-winning One-Tie™ re-usable tie strap consumer product
 - Achieved \$1 million in annual sales prior to licensing the intellectual property
- Senior Principal Engineer, Adv. Dev.**, Stryker Corp., Kalamazoo, MI 02/2012-03/2015
- Supervised R&D of disruptive medical instruments and surgical technologies
 - Formalized new business opportunities, developed budgets and timelines, and staffed project teams
 - Collaborated with several top-tier universities to bring necessary expertise into the organization
 - Delivered working prototypes for biophotonic devices and boundary-constraint surgical robotics
- Staff SW Engineer, Adv. Tech. Dev.**, LSI Corp. (Broadcom), Milpitas, CA 07/2004-01/2012
- Conducted R&D for very large scale integrated (VLSI) and serializer-deserializer (SerDes) circuits
 - Granted approx. 20 patents for physical (PHY) interfaces, SerDes, and VLSI test architectures
 - Made cutting-edge contributions in SerDes and test that were implemented in customer designs
- Adjunct Professor, ECE Dept.**, SVSU, University Center, MI 01/2009-01/2012
- Taught courses in digital circuits, computer architecture and organization, and data communications
- Consulting Assistant Professor, EE/CS Dept.**, Stanford University, Stanford, CA 09/2006-08/2008
- Advised PhD students on the test and reliability of VLSI circuits
 - Created and taught a PhD-level course on clock and data recovery (CDR) for digital communications
- Product Test Engineer**, Xilinx, Inc., Los Gatos, CA (PhD. Intern, two summers) 06/2002-09/2003
- Engineered tools to detect and diagnose defects in FPGAs, which reduced diagnosis time by 75%
- Manufacturing Engineer**, Cisco Systems, Inc., San Jose, CA 11/1999-04/2002
- Oversaw functional testing of layer-2/3 network switches and determined root-cause failures

SERVICE AND MENTORING

- Admissions Representative**, Univ. Michigan Law School, Ann Arbor, MI 2018
- Advised prospective law students at 30 law fairs and forums
- Instructor**, National Science Foundation, Western Michigan Univ. I-Corps, Kalamazoo, MI 2015
- Taught Lean Startup and Business Canvas fundamentals to a cohort of 30 entrepreneurs
- Engineering Program Evaluator**, Accred. Board of Eng. & Tech. (ABET), Baltimore, MD 2013-2014
- Evaluated electrical engineering and computer engineering university educational programs
- Industry Mentor**, National Science Foundation, Univ. Michigan I-Corps, Ann Arbor, MI 2013
- Mentored teams of medical-device entrepreneurs participating in the I-Corps program
- Industry Mentor**, Western Michigan Univ., Starting Gate Accelerator, Kalamazoo, MI 2013
- Mentored entrepreneurs participating in the university startup-accelerator program

SELECT PROFICIENCIES

- | | |
|--|--|
| • Hardware – Verilog, VHDL, ASICs, FPGAs | • Software – C/C++, Perl, Java, MatLab |
| • Engineering – COMSOL, ANSYS | • CAD – SolidWorks, AutoCAD, Blender |

Select US Patents

US11446109B2, "Metal detection system for use with medical waste container," issued 2022-09-23.

US11389556B2, "Sterilization container capable of providing an indication regarding whether or not surgical instruments sterilized in the container were properly sterilized," issued 2022-07-19.

US11357877B2, "Sterilization enclosure for surgical instruments," issued 2022-06-14.

US§ 11225996B1, "Dual hook clasp," issued 2022-01-18.

US11002308B2, "Hook device with opposing jaws," issued 2021-05-11.

US11000341B2, "Metal detection system for use with medical waste container," issued 2021-05-11.

US10994906B2, "Easy-to-remove cable tie," issued 2021-05-04.

US10731698B2, "Hook device with rotatable opposing jaws," issued 2020-08-04.

US10604317B2, "Reusable tie strap with multiple apertures," issued 2020-03-31.

US10258706B2, "Sterilization container capable of providing an indication regarding whether or not surgical instruments sterilized in the container were properly sterilized," issued 2019-04-16.

US10010636B2, "Sterilization container capable of providing an indication regarding whether or not surgical instruments sterilized in the container were properly sterilized," issued 2018-07-03.

US9292644B2, "Row based analog standard cell layout design and methodology," issued 2016-03-22.

US9087157B2, "Low-loss transmission line TDM communication link and system," issued 2015-07-21.

US9014313B2, "Error signature analysis for data and clock recovery in a communication system," issued 2015-04-21.

US8982941B2, "Predictive selection in a fully unrolled decision feedback equalizer," issued 2015-03-17.

US8929497B2, "Dynamic deskew for bang-bang timing recovery in a communication system," issued 2015-01-06.

US8923382B2, "Tap adaptation with a fully unrolled decision feedback equalizer," issued 2014-12-30.

US8615062B2, "Adaptation using error signature analysis in a communication system," issued 2013-12-24.

US8527912B2, "Digitally obtaining contours of fabricated polygons," issued 2013-09-03.

US8432250B2, "Process variation based microchip identification," issued 2013-04-30.

US8121186B2, "Systems and methods for speculative signal equalization," issued 2012-02-21.

US7956790B2, "Systems and methods for synchronous, retimed analog to digital conversion," issued 2011-07-05.

US7973692B2, "Systems and methods for synchronous, retimed analog to digital conversion," issued 2011-06-07.

US7779320B2, "Low power scan shifting with random-like test patterns," issued 2010-08-17.

US7696915B2, "Analog-to-digital converter having reduced number of activated comparators," issued 2010-04-13.

US7656340B2, "Systems and methods for pipelined analog to digital conversion," issued 2010-02-02.

US7656339B2, "Systems and methods for analog to digital conversion," issued 2010-02-02.

US7328386B2, "Methods for using checksums in X-tolerant test response compaction in scan-based testing of integrated circuits," issued 2008-02-05.

US7293312B2, "Multipurpose skateboard tool," issued 2007-11-13.

US7210083B2, "System and method for implementing postponed quasi-masking test output compression in integrated circuit," issued 2007-04-24.

US7206983B2, "Segmented addressable scan architecture and method for implementing scan-based testing of integrated circuits," issued 2007-04-17.

US7093842B2, "Skateboard truck assembly," issued 2006-08-22.

US6979142B1, “Retractable tip mechanical pencil assembly,” issued 2005-12-27.

US6920621B1, “Methods of testing for shorts in programmable logic devices using relative quiescent current measurements,” issued 2005-07-19.

Select Publications

- E. Chmelar, "Error signature analysis (ESA) receiver architecture for data communication," DesignCon, 2012.
- E. Chmelar and C. Ito, "Mostly digital SerDes (MDS): a comprehensive low power receiver architecture," DesignCon, 2012.
- I. Park, D. Lee, E. Chmelar, and E. McCluskey, "Inconsistent fails due to limited tester timing accuracy," Proc. 26th VLSI Test Symp., 2008.
- A. Al-Yamani, N. Devta-Prasanna, E. Chmelar, M. Grinchuk, and A. Gunda, "Scan test cost and power reduction through systematic scan reconfiguration," IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems, vol. 26, pp. 907–918, 2007.
- A. Al-Yamani, E. Chmelar, and M. Grinchuk, "Segmented addressable scan architecture," Proc. 23rd VLSI Test Symp., 2005.
- E. Chmelar, "Minimizing the number of test configurations for FPGAs," Int. Conf. Computer Aided Design, 2004.
- E. Chmelar, "The test and diagnosis of FPGAs," Ph.D. thesis, Stanford University, 2004.
- E. Chmelar and S. Toutounchi, "FPGA bridging fault detection and location via differential IDDQ," Proc. 22nd VLSI Test Symp., 2004.
- E. Chmelar, "Subframe multiplexing: FPGA manufacturing test time reduction," CRC TR-04-01, 2004.
- E. Chmelar, "Subframe multiplexing for FPGA manufacturing test," Proc. Int. Symp. FPGAs, 2004.
- E. Chmelar, "FPGA interconnect delay fault testing," Proc. Int. Test Conf., pp. 1239–1247, 2003.

Select Professional Organizations

State Bar of Michigan
Member (2018-pres.)

Project Management Institute
Certified Agile Practitioner (2014-pres.)
Certified Project Management Professional (2015-pres.)

Association for Computing Machinery
Member (2018-2019)

Accreditation Board of Engineering and Technology
Program Evaluator (2013-2018)

Institute of Electrical & Electronic Engineers
Senior Member (2004-2018)

American Chemical Society
Member (2013-2016)

National Science Foundation I-Corps
Instructor (2015)
Industry Mentor (2013)

American Society of Mechanical Engineers
Member (2014-2015)

International Society of Optics & Photonics
Member (2015)

APPENDIX 2--LISTING OF MATERIALS CONSIDERED

Exhibit No.	Description
2003	Neil Weste & David Harris, CMOS VLSI DESIGN: A CIRCUIT AND SYSTEMS PERSPECTIVE (4th ed. 2010)
2004	Jan Rabaey, Anantha Chandrakasan & Borivoje Nikolic, DIGITAL INTEGRATED CIRCUITS: A DESIGN PERSPECTIVE (2nd ed. 2003)
2005	William J. Dally & Brian Towles, <i>Route Packets, Not Wires: On-Chip Interconnection Networks</i> , PROC. 38TH DESIGN AUTOMATION CONF. (2001)
2006	John F. Wakerly, DIGITAL DESIGN: PRINCIPLES AND PRACTICES (4th ed. 2005)
2007	Fayez Gebali, Haytham Elmiligi & Mohamed Watheq El-Kharashi, NETWORKS-ON-CHIPS: THEORY AND PRACTICE (2009)
2008	Kurt Shuler, <i>Busses, Crossbars and NoCs: The 3 Eras of SoC Interconnect History</i> , ARTERISIP (Oct. 22, 2010), https://www.arteris.com/blog/busses-crossbars-and-nocs-the-3-eras-of-soc-interconnect-history/
2009	Kurt Shuler, <i>The SoC Interconnect Fabric: A Brief History</i> , ARTERISIP (Aug. 6, 2013), https://www.arteris.com/blog/the-soc-interconnect-fabric-a-brief-history/
2017	U.S. Patent No. 7,373,449
2022	Claim Construction Order, <i>Network System Technologies, LLC v. Texas Instruments Incorporated</i> , Civil Action No. 2:22-cv-00482-RWS (E.D. Tex.), Dkt. 150 (Jan. 3, 2024)
2033	P.R. 4-3 Claim Construction and Prehearing Statement, <i>Network System Technologies, LLC v. Samsung Electronics Co., Ltd.</i> , Civil Action No. 2:22-cv-00482-RWS (E.D. Tex.), Dkt. 91 (Jan. 9, 2024)
2040	Defendants' Joint Opening Claim Construction Brief, <i>Network System Technologies, LLC v. Qualcomm Incorporated et al.</i> , Civil Action No. 1:22-cv-1331-DAE (W.D. Tex.), Dkt. 77 (Oct. 12,

Exhibit No.	Description
	2023)
	2023-12-18 - Paper 2 - Petition for Inter Partes Review of US 8,086,800
1001	U.S. Patent No. 8,086,800
1002	Prosecution History of U.S. Patent No. 8,086,800
1003	Declaration of Todd Mowry
1005	<i>Andrei Rădulescu et al., An Efficient On-Chip Network Interface Offering Guaranteed Services, Shared-Memory Abstraction, and Flexible Network Configuration</i> , 2 PROC. DESIGN, AUTOMATION AND TEST IN EUR. CONF. & EXHIBITION 878 (2004) (“Rădulescu”)
1006	U.S. Patent Application Publication No. 2002/0186660 to Murat I. Bahadiroglu <i>et al.</i> (“Bahadiroglu”)
1007	U.S. Patent No. 5,434,872 to Brian Petersen <i>et al.</i> (“Petersen”)
1008	International Patent Publication No. WO 2004/034176 A2 to Rădulescu <i>et al.</i> (“Rădulescu-WO”)