

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC,
Petitioner,

v.

VALTRUS INNOVATIONS LIMITED,
Patent Owner.

IPR2022-01418
Patent 7,523,454 B2

Before NORMAN H. BEAMER, KEVIN C. TROCK, and
SHARON FENICK, *Administrative Patent Judges*.

TROCK, *Administrative Patent Judge*.

JUDGMENT
Final Written Decision
Determining All Challenged Claims Unpatentable
35 U.S.C. § 318(a)

I. INTRODUCTION

We have authority to hear this *inter partes* review under 35 U.S.C. § 6. This Final Written Decision issues pursuant to 35 U.S.C. § 318(a) and 37 C.F.R. § 42.73. For the reasons discussed herein, we determine that Petitioner, Google LLC, (“Petitioner” or “Google”) has shown by a preponderance of the evidence that claims 1–21 of U.S. Patent No. 7,523,454 B2 (Ex. 1001, “the ’454 patent”) are unpatentable. *See* 35 U.S.C. § 316(e) (2018); 37 C.F.R. § 42.1(d) (2019).

II. BACKGROUND

A. *Procedural History*

Google filed a Petition requesting *inter partes* review of claims 1–21 of the ’454 patent (the “challenged claims”). Paper 2 (“Pet.”). Valtrus Innovations Limited (“Patent Owner” or “Valtrus”) filed a Preliminary Response opposing institution. Paper 6 (“Prelim. Resp.”). With the Board’s permission, Petitioner filed a Preliminary Reply (Paper 7, “Prelim. Reply”) and Patent Owner filed a Preliminary Sur-reply (Paper 8, “Prelim. Sur-reply”). Based on the record at that time, we issued a Decision granting *inter partes* review (Paper 9, “Dec.”).

After institution, Patent Owner filed a corrected Response (Paper 19, “PO Resp.”), Petitioner filed a Reply (Paper 28, “Pet. Reply”), and Patent Owner filed a Sur-reply (Paper 35, “PO Sur-reply”).

On January 29, 2024, an oral hearing was held. A transcript of the hearing is made part of the record. Paper 40.

B. *Related Proceedings*

As required by 37 C.F.R. § 42.8(b)(2), Petitioner and Patent Owner identify the judicial or administrative matters that would affect or be affected

by a decision in this proceeding. Petitioner and Patent Owner state that the '454 patent is the subject of (1) *Valtrus Innovations, Ltd. v. Google LLC*, No. 3:22-cv-00066-N (N.D. Tex.) (the "Texas district court litigation") and (2) *Valtrus Innovations Ltd. v. Google LLC*, No. 4:22-cv-00020 (N.D. Tex.) (transferred to the Dallas Division of N.D. Texas and captioned 3:22-cv-00066-N). Pet. x; Paper 4, 1.

C. Real Parties-in-Interest

As required by 37 C.F.R. § 42.8(b)(1), each party identifies the real party-in-interest. Petitioner identifies itself as a real party-in-interest. Pet. x; *see id.* at n.1 (noting that Petitioner is a subsidiary of XXVI Holdings Inc., which is itself a subsidiary of Alphabet Inc., but asserting that neither of these are real parties-in-interest). Patent Owner states that itself, Key Patent Innovations Limited, and Valinvest LLC "are financially interested in the outcome of the litigation between Valtrus and Google." Paper 4, 1.

D. The '454 Patent (Ex. 1001)

The '454 patent is titled "Apparatus and Method for Routing a Transaction to a Partitioned Server." Ex. 1001, code (54). The '454 patent describes "an apparatus and method for routing a transaction to a partition on a server based at least in part on the configuration thereof." *Id.* at 1:7–9.

According to the '454 patent, individual servers may be used to process transactions or "multiple servers may be pooled, and the incoming transactions directed to an available server in the server pool using a load balancer. As such, the traffic is balanced across all of the servers in the pool, and the transactions are processed more quickly." Ex. 1001, 1:22–26.

The '454 patent recognizes that using multiple servers may be inefficient and one solution to this problem "is a partitioned server."

Ex. 1001, 1:27–34. The ’454 patent states that a “partitioned server is a single server or aggregation of server resources subdivided to perform as multiple servers.” *Id.* at 1:34–36. “[A] partitioned server continues to service multiple locations, multiple departments, and/or multiple transactions.” *Id.* at 1:36–38. With partitioned servers, “fewer individual servers are needed,” “operations are centralized,” multiple individual server characteristics are maintained, and “utilization of the server resources is enhanced.” *Id.* at 1:38–42.

The ’454 patent states that because “[t]he partitions of partitioned servers . . . may not be equally weighted,” some partitions may be better able to process transactions than other partitions, based on the configuration of the partition. Ex. 1001, 1:43–44. Some partitions, for example, may be allocated more server resources (such as CPUs or number of processors) and the allocation may be dynamic. *Id.* at 1:44–56.

To address these issues, the ’454 patent describes a load balancer which recognizes each partition of a partitioned server and its associated configuration, and routes transactions to partitions based at least in part on those configurations of the partitions. Ex. 1001, Abstract. “The configuration may be determined in response to an event, on a continuous basis, etc. In addition, the load balancer may also use other factors, in combination with the configuration, to route the transaction to the partition, such as a characteristic of the transaction.” *Id.* The patent also states configuration comprises the configuration parameters of the respective partitions, including, for example, the resource allocation, the resource utilization, the average response time, or any other suitable information regarding the partitions. *Id.* at 7:34–39.

The partitions and the associated configuration may be stored in a configuration profile provided to the load balancer. Ex. 1001, Fig. 3.

Figure 3 is reproduced below:

FIG. 3

PARTITION PROFILE			
PARTITION ID	NETWORK ADDRESS	CORRESPONDING CONFIGURATION	ASSIGNED RANK
A	1.1.1.1	RESOURCE ALLOCATION = 20% RESOURCE UTILIZATION = 50% AVG RESPONSE TIME = 0.5 ms	3
B	1.1.1.2	RESOURCE ALLOCATION = 50% RESOURCE UTILIZATION = 35% AVG RESPONSE TIME = 0.2 ms	1
C	1.1.1.3	RESOURCE ALLOCATION = 30% RESOURCE UTILIZATION = 40% AVG RESPONSE TIME = 0.3 ms	2
.	.	.	.
.	.	.	.
PARTITION n	1.1.1.n	CONFIGURATION n	RANK n

Figure 3 illustrates a partition profile for maintaining the configuration of the partitions. *Id.* at 3:13–14. The partitions can be identified by “a name, serial number, etc.” *Id.* at 7:46–48. Based at least in part on the associated configurations, the partitions may be assigned a rank. *Id.* at 7:53–56. The ranks can be indicated by “a numerical assignment (e.g., ‘1’, ‘2’, ‘3’), an empirical assignment (e.g., ‘Fast’, ‘Slow’), a specific parameter (e.g., ‘0.5 ms’), etc.” *Id.*

E. Illustrative Claim

Petitioner challenges claims 1–21 of the '454 patent. Independent claim 1 is reproduced below.¹

1. [1PRE] A method for routing a transaction to a partitioned server, comprising:

[1A] identifying a plurality of partitions on said partitioned server;

[1B] determining a configuration of each of said plurality of partitions;

[1C](1) providing said configurations to a load balancer,

[1C](2) wherein said transaction is routed by said load balancer to one of said plurality of partitions based at least in part on said configurations;

[1D] assigning a rank to each of said plurality of partitions based at least in part on said configurations, wherein said transaction is routed based on said ranking; and

[1E] storing said configurations in a partition profile, wherein said configurations are provided to said load balancer via said partition profile.

Ex. 1001, 8:62–9:10.

F. Evidence and Asserted Grounds

Petitioner relies upon the following evidence:

(1) US 6,006,264, issued December 21, 1999 (“Colby” Ex. 1005);

(2) US 6,587,938 B1, issued July 1, 2003 (“Eilert” Ex. 1006); and

(3) US 6,092,178, issued July 18, 2000 (“Jindal” Ex. 1007).

¹ For ease of reference, we use Petitioner’s labelling scheme shown in brackets.

Petitioner relies on the declarations of Vijay K. Madiseti. (Exs. 1003; 1046). Patent Owner relies on the declaration of Hans van Rietschote (Ex. 2007).

Petitioner challenges the patentability of the '454 patent claims based on the following grounds (Pet. 2):

Ground	Claim(s) Challenged	35 U.S.C. § ²	Reference(s)/Basis
1	1–6, 8, 10, 13, 16–21	103(a)	Colby
2	1–6, 8–10, 13–21	103(a)	Colby, Eilert
3	1–21	103(a)	Colby, Eilert, Jindal

III. ANALYSIS

A. *Legal Standards*

Petitioner bears the burden of persuasion to prove unpatentability of the claims challenged in the Petition by a preponderance of the evidence. 35 U.S.C. § 316(e). A claim is unpatentable as obvious under 35 U.S.C. § 103(a) (2010) “if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains.” The question of obviousness is resolved on the basis of underlying factual determinations, including: (1) the scope and content of the prior art; (2) any differences between the claimed subject matter and the prior art; (3) the level of skill in the art; and (4) so-called “secondary considerations,” including commercial

² The relevant sections of the Leahy-Smith America Invents Act (“AIA”), Pub. L. No. 112–29, 125 Stat. 284 (Sept. 16, 2011), took effect on March 16, 2013. Because the '454 patent claims priority to an application filed before this date, our citations to 35 U.S.C. § 103(a) in this Decision are to the pre-AIA version.

success, long-felt but unsolved needs, failure of others, and unexpected results. *Graham v. John Deere Co.*, 383 U.S. 1, 17–18 (1966). Neither party has presented any evidence on the fourth *Graham* factor.

B. Level of Ordinary Skill in the Art

The level of ordinary skill in the art is “a prism or lens” through which we view the prior art and the claimed invention. *Okajima v. Bourdeau*, 261 F.3d 1350, 1355 (Fed. Cir. 2001). The person of ordinary skill in the art is a hypothetical person presumed to have known the relevant art at the time of the invention. *In re GPAC Inc.*, 57 F.3d 1573, 1579 (Fed. Cir. 1995). In determining the level of ordinary skill in the art, we may consider certain factors, including the “type of problems encountered in the art; prior art solutions to those problems; rapidity with which innovations are made; sophistication of the technology; and educational level of active workers in the field.” *Id.*

Petitioner states a person of ordinary skill in the art would have had at least a bachelor’s degree in computer science, computer engineering, or a related field, with 3 years of experience in the area of distributed computer systems, including performance and/or resource optimizations. A higher level of education may substitute for less experience.

Pet. 6 (citing Ex. 1003 ¶¶ 38–42).

Patent Owner disagrees, and asserts that a person of ordinary skill in the art

would be someone with a general science or engineering degrees, including those with a bachelor’s degree in physics, applied mathematics, computer science, computer engineering, or electrical engineering, but with three or more years of experience in computer networking, computer software, data management, virtualization and related fields.

PO Resp. 18 (citing Ex. 2007 ¶¶ 18–20). Patent Owner states that “[t]he analysis, however, would not change under either definition.” PO Resp. 18 (citing Ex. 2007 ¶ 21).

For the purposes of this Decision, we continue to use the same assessment offered by Petitioner that we adopted in our Institution Decision, as it is consistent with the ’454 patent and the asserted prior art. *See Okajima v. Bourdeau*, 261 F.3d 1350, 1355 (Fed. Cir. 2001) (the prior art itself may reflect an appropriate level of skill).

C. Claim Construction

In an *inter partes* review, the claims are construed using the same claim construction standard that would be used to construe the claim in a civil action under 35 U.S.C. § 282(b). *See* 37 C.F.R. § 42.100(b) (2021). This claim construction standard includes construing the claim in accordance with the ordinary and customary meaning of such claims as would have been understood by one of ordinary skill in the art. *Id.*; *see Phillips v. AWH Corp.*, 415 F.3d 1303, 1312–13 (Fed. Cir. 2005). In construing claims in accordance with their ordinary and customary meaning, we consider intrinsic evidence such as the specification and the prosecution history of the patent. *Phillips*, 415 F.3d at 1315–17. Extrinsic evidence, including expert and inventor testimony, dictionaries, and treatises, may also be used but is less significant than the intrinsic record. *Id.* at 1315. Usually, the specification is dispositive, and it is the single best guide to the meaning of a disputed term. *Id.* Any special definitions for claim terms must be set forth in the specification with reasonable clarity, deliberateness, and precision. *See In re Paulsen*, 30 F.3d 1475, 1480 (Fed. Cir. 1994).

In its Petition, Petitioner states that it construes the claim terms “in accordance with the ordinary and customary meaning as understood by a [person of ordinary skill in the art] and the patent’s prosecution history,” except for the means plus function terms in claim 16, which it construes under 35 U.S.C. § 112 ¶ 6. Pet. 6–7.

In its Response, Patent Owner does not appear to take a general position on claim construction, but instead presents specific construction arguments concerning the preamble (PO Resp. 18–19), the term “configuration” (*id.* at 19–29) and independent claim 16 (*id.* at 64). Patent Owner also raises an argument with respect to 37 CFR § 42.104(b). *Id.* at 68–69.

We address Patent Owner’s arguments concerning the preamble, the term “configuration,” and 37 CFR § 42.104(b) immediately below. We address Patent Owner’s arguments concerning claim 16 in Section III.D.4.

1. Preamble

The preamble of independent claims 1, 6, 16, and 17 all recite “routing a transaction to a partitioned server.” Ex. 1001, 8:62 (claim 1), 9:20–21 (claim 6), 10:15–16 (claim 16), 10:28–29 (claim 17).

With respect to claim 1, Petitioner asserts that “[t]o the extent [the preamble] is limiting,” Colby’s “‘method for servicing a content request’ constitutes a ***method for routing a transaction to a partitioned server.***” Pet. 16–17 (citing Ex. 1005, 4:20–22, Fig. 3; Ex. 1003 ¶¶ 132, 141) (emphasis by Petitioner). Petitioner also relies on this position with respect to the preambles of independent claims 6, 16, and 17. *See* Pet. 37, 41–42.

Patent Owner argues that the parties agreed in the Texas district court litigation that the preamble is limiting. PO Resp. 18 (citing Ex. 2002, 2).

Patent Owner also argues that “the terms ‘transaction’ and ‘said partitioned server’ recited in the body of the claims rely on ‘a transaction’ and ‘a partitioned server’ in the preamble for antecedent basis” and that consequently, “the Board should find the preambles limiting.” PO Resp. 19 (citing *Shoes by Firebug v. Stride Rite Children’s Group*, 962 F.3d 1362, 1368 (Fed. Cir. 2020)).

In its Reply, Petitioner does not dispute that the parties agreed in their Joint Claim Construction Statement submitted in the Texas district court litigation that the preambles of claims 1 and 17 are “limiting in both claims.” *See generally* Pet. Reply; *see also* Ex. 2002, 2. Petitioner also does not dispute Patent Owner’s contention that the preamble is limiting. *See generally* Pet. Reply.

Based on this record, there does not appear to be a cognizable dispute between the parties as to whether the preamble is limiting. Additionally, we need not make a determination on that issue here because only those terms that are in controversy need be construed, and only to the extent necessary to resolve the controversy; as discussed below (*see, e.g.*, Section III.D.2.a), Petitioner makes a sufficient showing that the preamble is taught in the asserted grounds. *See Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.*, 868 F.3d 1013, 1017 (Fed. Cir. 2017); *Vivid Techs., Inc. v. Am. Sci. & Eng’g, Inc.*, 200 F.3d 795, 803 (Fed. Cir. 1999).

2. “configuration”

Independent claims 1, 6, 16, and 17 all recite “determining a configuration of each of said plurality of partitions.” Ex. 1001, 8:66–67 (claim 1), 9:27–28 (claim 6), 10:19–20 (claim 16), 10:32–33 (claim 17).

The parties disagree about the scope of the term “configuration” as that term is used in the ’454 patents claims.

Petitioner asserts that “[t]he ’454 patent refers to a partition’s ‘configuration’ as including ‘any suitable information’ about the partition, including ‘resource allocation,’ ‘resource utilization,’ and ‘average response time.’” Pet. 19 (citing Ex. 1001, 2:22, 7:34–39, 7:50–53; Ex. 1015 ¶ 133; Ex. 1003 ¶¶ 147, 151). Petitioner also asserts that “[e]ach piece of information regarding a server’s load, health, and content is independently *a configuration* as claimed in the ’454 patent.” Pet. 19 (citing Ex. 1003 ¶ 151) (emphasis by Petitioner).

Patent Owner contends that “a partition’s configuration refers to a specific set of parameters, settings, and/or resource allocations that define how the partition is configured or portioned out of an overall server,” and “would encompass hardware and logical resources, operating systems, type of devices the partition is emulating, connectivity options, startup and shutdown parameters associated with the partition.” PO Resp. 19 (citing Ex. 2007 ¶ 86). Patent Owner relies on the testimony of Mr. van Rietschote (Ex. 2007 ¶¶ 86–100) and also points to several dictionary definitions of the term “configure” to support its position. PO Resp. 22–23 (citing Exs. 2034–2039, 2041, 2042, 2044, 2046); *see also* PO Sur-reply 2–4 (“Configuration” is the noun form of “configure,” and is the result of the act of configuring).

Patent Owner argues that “[t]he specification draws a distinction between configuration . . . and behavior of the configuration.” PO Resp. 24–25. Patent Owner contends that “[l]oad, health and content . . . cannot be considered configuration . . . because they are not parameters that a system administrator can provision.” PO Resp. 28, *see also* PO Sur-reply 3–4.

Patent Owner also argues, among other things, that “the claims expressly distinguishes a characteristic of a transaction on the one hand and a configuration of the partition on the other hand.” *Id.* at 28 (citing Ex. 1001, claim 17).

In construing a claim term in accordance with its ordinary and customary meaning, we consider intrinsic evidence such as the specification and the prosecution history of the patent to be the best guide to provide an understanding of the disputed term. *Phillips*, 415 F.3d at 1315–17. While extrinsic evidence, such as expert testimony and dictionaries, may also be used, they are often less significant than the intrinsic record. *Id.* at 1315. Usually, the specification is dispositive, and is the single best guide to the meaning and scope of a disputed term. *Id.*

We begin our analysis by considering the passages of the ’454 patent cited by Petitioner to support its view that a partition’s “configuration” includes “‘any suitable information’ about the partition, including ‘resource allocation,’ ‘resource utilization,’ and ‘average response time.’” Pet. 19 Petitioner relies on the following passages.

The corresponding configuration 330 comprises the configuration parameters of the respective partition 160–162. For example, the corresponding configuration 330 may include the resource allocation . . . , the resource utilization . . . , the average response time . . . , or any other suitable information regarding the partitions 160–162.

Ex. 1001, 7:34–39.

[A]ny suitable information may be included as part of the corresponding configuration 300, including specifics of the various resources and the corresponding allocation thereof.

Id. at 7:50–53.

[T]he partition manager 150 may reside on a single partition 160, or have modules on each partition 160 for monitoring *the configuration thereof (e.g., the resource utilization, **load**, etc.)*.

Id. at 6:4–7 (emphasis added).

These passages indicate that the '454 patent uses the term “configuration” and “suitable information” broadly. These passages thus support Petitioner’s view that

a partition’s “configuration” includes “any suitable information” about the partition, including “resource allocation,” “resource utilization,” and “average response time”

and

information regarding a server’s load.

See Pet. 19.

As to information regarding a server’s “health and content,” Petitioner does not identify any portion of the '454 patent which addresses explicitly a server’s “health and content.” Petitioner, however, relying on Dr. Madisetti’s testimony, argues that a server’s “health and content” is “suitable information” about a partition’s configuration. Pet. 19 (citing Ex. 1003 ¶ 151). Dr. Madisetti testifies that a person of ordinary skill in the art would have understood that “the '454 patent refers to a ‘configuration’ of a partition as including ‘any suitable information’ about the partition, including ‘resource allocation,’ ‘resource utilization,’ and ‘average response time.’” Ex. 1003 ¶ 147 (citing Ex. 1001, 2:22, 7:34–39, 7:50–53). Dr. Madisetti also states that his opinion is confirmed by Patent Owner’s complaint in the Texas district court litigation which asserts that “[a] partition’s ‘health’ is a configuration of the partition.” Ex. 1003 ¶ 147

(citing Ex. 1015 ¶ 133).³ Dr. Madisetti further states that “information regarding a server’s load, health, and content, is independently a configuration, as that term is used in the ’454 patent.” Ex. 1003 ¶ 151 (citing Ex. 1001, 7:34–39, 7:50–53).

Patent Owner disagrees with Petitioner’s position, and instead contends that “a partition’s configuration refers to a specific set of parameters, settings, and/or resource allocations that define how the partition is configured or portioned out of an overall server.” PO Resp. 19.

Patent Owner asserts that “[t]he specification draws a distinction between configuration . . . and the characteristics of the transaction . . . and behavior of the configuration,” citing to Mr. van Rietschote’s testimony. PO Resp. 24–25 (citing Ex. 2007 ¶¶ 88, 97–98). Patent Owner also argues that “the claims expressly distinguishes a characteristic of a transaction on the one hand and a configuration of the partition on the other hand.” *Id.* at 28 (citing Ex. 1001, claim 17).

Patent Owner’s argument with respect to claim distinction is largely unavailing because claim 17, cited by Patent Owner to support this argument, is an independent claim and thus provides little insight into the proper scope of the other independent claims of the ’454 patent.

Mr. van Rietschote’s testimony, however, is more probative on this point because it refers to a portion of the ’454 specification that explains

³ Exhibit 1015, cited by Petitioner, is a complaint (“the ’020 Complaint”) filed by Patent Owner in a related proceeding alleging infringement of the ’454 patent against Petitioner asserting that “[a] partition’s ‘health’ is a configuration of the partition.” *See* Sec. II.B; Ex. 1015 ¶ 133, *see also id.* ¶¶ 118–142.

the load balancer routes the transaction to one of the partitions ***based at least in part on the configuration*** thereof. Optionally, ***one or more characteristics of the transaction*** may also be determined and used for routing the transaction to the appropriate portion.

For example, the characteristic(s) may include a type of transaction (e.g., video streams, backup requests, etc.), a requested level of service (e.g., high priority, low priority), a combination thereof, etc.

Ex. 2007 ¶ 97 (citing Ex. 1001, 1:66–2:4; 5:26–30) (emphasis by Patent Owner).

Patent Owner also argues that “[c]onfiguration’ does not, however, include ‘load,’ ‘health’ or ‘content’ . . . because those are not metrics used to create/configure a partition,” and “are not parameters that a system administrator can provision.” PO Resp. 20, 28 (citing Ex. 2007 ¶¶ 86, 88). Patent Owner’s arguments on this point, however, rely on less significant extrinsic evidence, as Mr. van Rietschote’s testimony does not cite to any specific passages from the ’454 specification to support his opinion. *See* Ex. 2007 ¶¶ 86, 88. Also, the collection of Patent Owner’s external dictionary definitions cited in support of its position constitute less significant extrinsic evidence. *See* PO Resp. 21–23 (citing Exs. 2034–2039, 2041, 2042, 2044, 2046).

With respect to the issue of whether a partition’s configuration may include suitable information about a partition’s “load,” Patent Owner concedes that “[l]oad is discussed,” in the ’454 patent specification, but argues that it is “always in combination with features that can be set when forming or altering a partition.” PO Resp. 23 (citing Ex. 2007 ¶¶ 89, 96).

In its Reply, Petitioner provides rebuttal challenging each of Patent Owner’s arguments in its Response. *See* Pet. Reply 2–5. Patent Owner also

addresses some of the issues in its Sur-reply, reiterating some of its same arguments. *See, e.g.* Sur-reply 3 (“health, ‘resource utilization’ and ‘load’ are not examples of ‘configuration’ because, as Mr. van Rietschote explains, they cannot be provisioned as part of the configuring process.”).

We do not agree with Patent Owner’s proposal that “a partition’s configuration refers to a specific set of parameters, settings, and/or resource allocations that define how the partition is configured or portioned out of an overall server.” PO Resp. 19. Patent Owner’s proposal is not fully supported by the ’454 patent specification and relies in significant part on extrinsic evidence provided by the testimony of Mr. van Rietschote and external dictionary definitions.

Neither do we fully agree with Petitioner’s proposal that a partition’s configuration includes “‘any suitable information’ about the partition, including ‘resource allocation,’ ‘resource utilization,’ and ‘average response time,’” and that “[e]ach piece of information regarding a server’s load, health, and content is independently **a configuration** as claimed in the ’454 patent.” Pet 19 (emphasis by Petitioner). A portion of Petitioner’s proposal, namely Dr. Madisetti testimony regarding a server’s health and content being independent configurations, is not described expressly in the ’454 patent specification.

Based on the complete record, we determine that the appropriate scope of “configuration” is best described by the intrinsic evidence provided by the ’454 patent specification, namely that a partition’s configuration “may include the resource allocation . . . the resource utilization . . . the average response time . . . or any other suitable information regarding the partitions (Ex. 1001, 7:34–39), including specifics of the various resources

and the corresponding allocation thereof” (*id.* at 7:50–53), such as “e.g., the resource utilization, load, etc.” (*id.* at 6:4–7). We apply this understanding of the scope of term “configuration” to the claims of the ’454 patent.

3. 37 CFR § 42.104(b)

Patent Owner contends that

Petitioner has not shown how the claims are invalid under its own proposed construction in the district court. The non-limiting examples include failure to show unpatentability under Petitioner’s proposed constructions for “determining a configuration of each of said plurality of partitions” (which did not list “health,” “content” or “load”), “partition” (no showing that VWHs are not sharing resources with other VWHs), or “assigning a rank to each of said plurality of partitions based at least in part on said configurations”(no showing of how “load,” “health” or “content” relate to “relative weight of resources allocated to each partition, frequency with which the partition changes, or response time of the partition” as proposed by Google). Just as “[c]laims may not be construed one way to obtain their allowance and in a different way against accused infringers,” claims should not [be] construed one way to avoid infringement and a different way to challenge validity.

PO Resp. 68–69. “Consequently,” Patent Owner concludes, “Petitioner has not met the requirements under 37 C.F.R. § 42.104(b) and the Petition fails.” *Id.* at 69.

Petitioner responds by arguing that “Rule 42.104 **does not require** Petitioner to address constructions proposed in litigation or advance uniform constructions in proceedings addressing **different** invalidity challenges (EX1043, 64-69 (35 U.S.C. §§101, 112)), evidence (EX2004, 10-11 (confidential invention disclosure)), and issues (*e.g.*, infringement).” Pet. Reply 1. “Instead,” Petitioner argues, it may “pursue an IPR based on a broader construction (***such as one offered by the patent owner***) while also

taking a narrower position in [litigation].” *Id.* at 1–2 (quoting *Crowdstrike v. Open Text*, IPR2023-00124, Paper 11 at 12–13 (June 15, 2023)).

37 C.F.R. § 104(b) requires a petition, among other things, to “[p]rovide a statement of the precise relief requested for each claim challenged,” including identifying “[h]ow the challenged claim is to be construed” and “[h]ow the construed claim is unpatentable” by “specify[ing] where each element of the claim is found in the prior art patents or printed publications relied upon.” 37 C.F.R. § 104(b)(3)(4) (2023).

Contrary to Patent Owner’s argument, 37 C.F.R. § 104(b) does not require a Petitioner to show “how the claims are invalid under its own proposed construction in the district court.” 37 C.F.R. § 104(b) does not preclude a petitioner from advancing one construction in a district court as an infringement defendant and a different construction as a petitioner before the Board. It is not uncommon for litigation defendants to seek a narrow construction in the district court in an effort to avoid infringement, but file a petition before the Board presenting a broader construction to prove unpatentability. Similarly, patent owners may seek a broader construction in district court to prove infringement but present a narrower construction before the Board to avoid unpatentability. This, however, does not mean that the Board cannot take into consideration a party’s position on construction before a different tribunal.

Here, however, we determine that Petitioner has satisfied the requirements of 37 C.F.R. § 104(b) for the reasons discussed in Sections III.C–F.

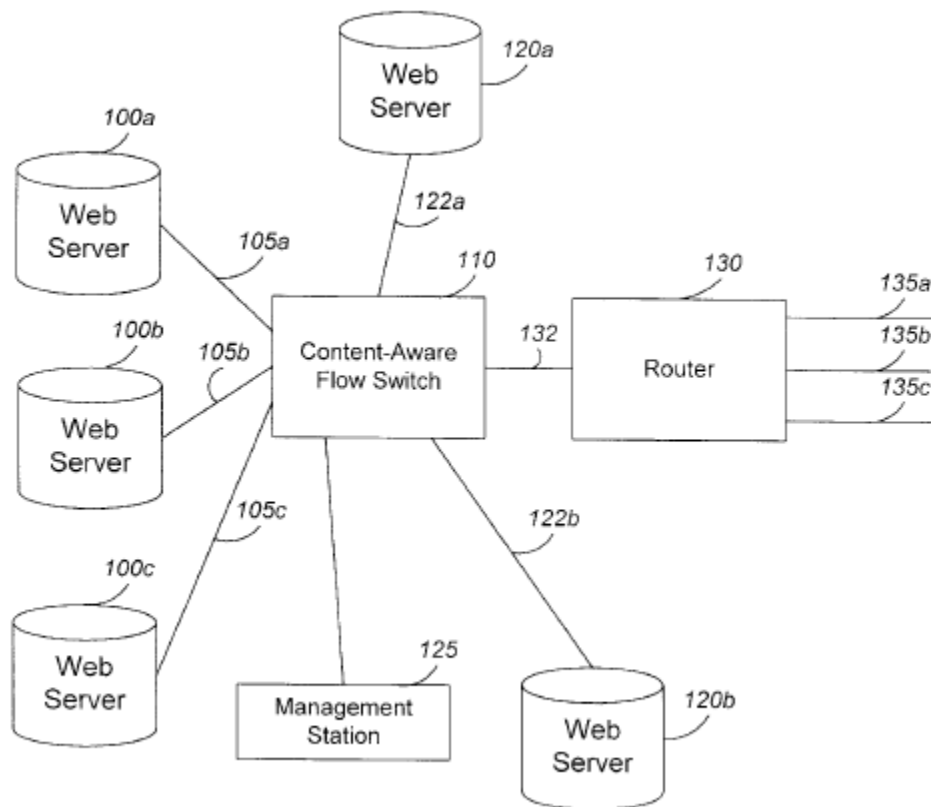
D. Ground 1: Obviousness of Claims 1–6, 8, 10, 13, and 16–21 over Colby

For Ground 1, Petitioner contends that claims 1–6, 8, 10, 13, and 16–21 are unpatentable under 35 U.S.C. § 103(a) as being obvious over Colby (Ex. 1005). Pet. 7–50. We present a brief overview of Colby and then address the parties’ contentions.

1. Colby (Ex. 1005)

Colby is a U.S. patent titled “Method and system for Directing a Flow between a Client and a Server.” Ex. 1005, code (54). Colby describes a content-aware flow switch (sometimes referred to as CAFS) for directing content requests to a best-fit server among a selection of servers. *Id.* at code (57). The best-fit server is chosen based on the type of content requested, the quality of service requirements implied by the content request, the degree of load on available servers, network congestion information, and the proximity of the client to available servers. *Id.* The flow switch also provides the functionality of multiple physical web servers on a single web server in a way that is transparent to the client, through the use of virtual web hosts (VWH) and flow pipes. *Id.* Each virtual web host provides the functionality of a single server, and many virtual web hosts may reside on a single physical server. *Id.* at 2:27–33, 16:20–24.

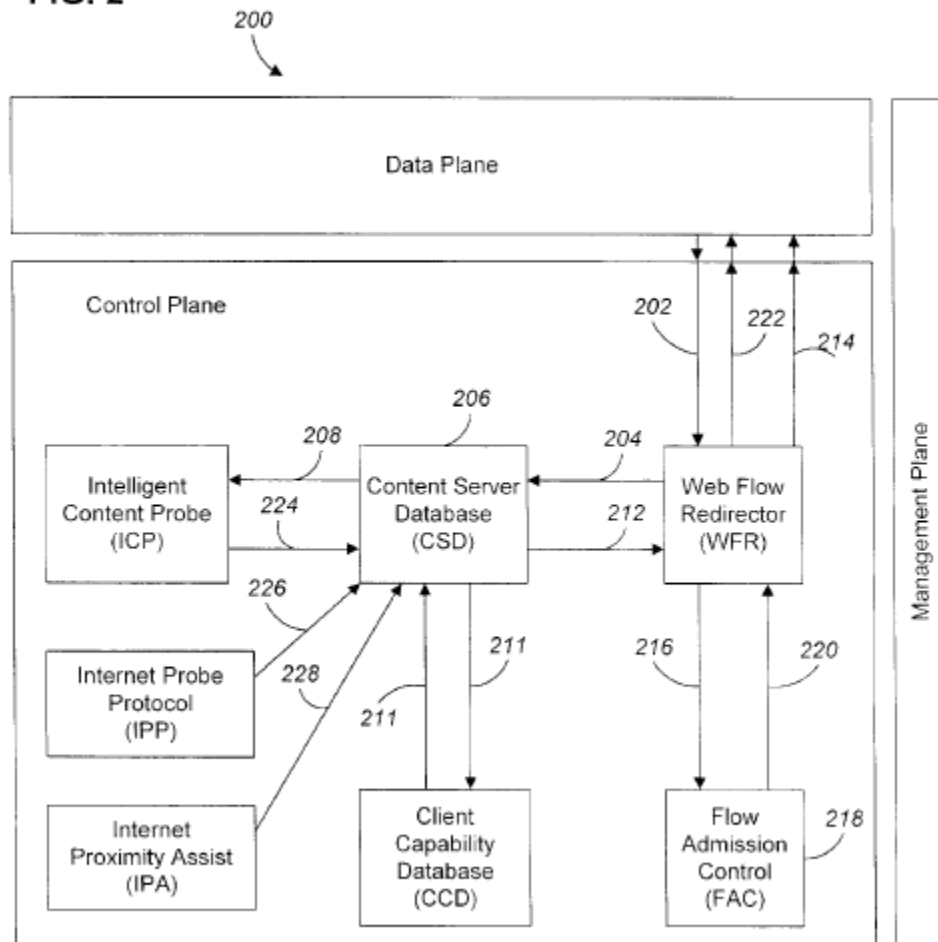
Colby’s Figure 1b is shown below.



Colby's Figure 1b is a block diagram of a segment of a network employing a content-aware flow switch, showing web servers 100a–c and 120a–b connected to content-aware flow switch 110 (Ex. 1005, 5:22–25) that “intercepts a client content request in an IP network, and transparently directs the content request to a best-fit server.” *Id.* Abstract.

Colby's Figure 2 is shown below.

FIG. 2



Colby's Figure 2 is a block diagram illustrating operations performed by and communications among components of a content-aware flow switch during flow setup, including Intelligent Content Probe (ICP), Internet Probe Protocol (IPP), Internet Proximity Assistant (IPA), Content Server Database (CSD) 206, Client Capability Database (CCB), Web Flow Redirector (WFR), and Flow Admission Control (FAC) 218. Ex. 1005, 4:27–29.

2. *Claim 1*

Petitioner provides a limitation-by-limitation comparison of Colby to independent claim 1. Pet. 15–29.

a) *[IPRE] A method for routing a transaction to a partitioned server,*

Petitioner asserts that Colby's "method for servicing a content request using a [content aware flow switch]" is the recited "method for routing a transaction." Pet. 16 (citing Ex. 1005, 4:20–22, Fig. 3; Ex. 1003 ¶ 132). Petitioner explains that Colby's "content" is "loosely defined as any information that a client application is interested in receiving," including "an HTML page." Pet. 16 (citing Ex. 1005, 1:59–65). Petitioner further explains that a person of ordinary skill in the art would have understood that the "content request" referenced in Colby's Figure 3 and related discussion is an example of the recited "transaction." Pet. 16 (citing Ex. 1003 ¶¶ 133–136; Ex. 1005, 16:5–9; Ex. 1001, 3:61–4:4 ("transaction" may be "a uniform resource locator (URL) requesting a hypertext markup language (HTML) page"))).

Petitioner also asserts that "Colby routes transactions *to a partitioned server*," as claimed. Pet. 16–17 (citing Ex. 1003 ¶¶ 138–141) (emphasis by Petitioner). Petitioner explains that according to the '454 patent, "a partitioned server is a single server or aggregation of server resources subdivided to perform as multiple servers," and this is "the same as Colby's physical web server with multiple VWHs [virtual web hosts]," which "are implemented on and 'share a single physical web server,' with each VWH 'provid[ing] the functionality of a single physical web server.'" Pet. 17 (citing Ex. 1001, 1:34–37; Ex. 1003 ¶¶ 139; Ex. 1005, 2:29–44).

Patent Owner argues that "[t]here is no evidence, however, that Colby expressly or inherently discloses that its VWHs would be run on . . . a partitioned server as Petitioner asserts." PO Resp. 30. Patent Owner also argues that "Colby never suggests that the three VWHs [Colby Figures 21a–

b] should be run as three partitions on a partitioned server (comprised of aggregated resources of web servers 100a-c).” PO Resp. 30 (citing Ex. 1005, 16:19–17:36). Patent Owner points to Mr. van Rietschote’s testimony that “[r]unning a single instance of an application for providing virtual web hosting for different web sites would implicate only a single partition or an un-partitioned server because a single instance of an application can’t be run on more than one partition.” PO Resp. 33 (citing Ex. 2007 ¶ 118).

Patent Owner further argues that Colby’s teaching of a VWH as a logical entity “applies equally to shared virtual web hosting service that [runs] all virtual hosts in a single partition or unpartitioned server.” PO Resp. 34 (citing Ex. 2007 ¶ 133; Ex. 1011, 2:20–25); *see also* PO Resp 35–36 (citing Ex. 1005, 16:20–24, 2:36–39). A prior art reference such as Colby, however, may be relied upon for all that it would have reasonably suggested to one having ordinary skill in the art. Merely because Colby’s teaching might “appl[y] equally” to a shared virtual web hosting service running on an “unpartitioned server,” as Patent Owner ostensibly argues, does not mean that Colby’s descriptions of a VWH as a “logical entity” that “may span multiple physical web servers” or “may share a single physical web server,” and that “[e]ach of the physical web servers . . . may embody one or more virtual web hosts” are somehow vitiated as prior art teachings. *See* Ex. 1005, 2:26–30, 16:20–24. Similarly, Patent Owner’s arguments that VWHs would commonly be run on a single instance of an Apache server application are equally unavailing in the face of these descriptions. *See* PO Resp. 31–34.

The ’454 patent explains that a “partitioned server is a single server or aggregation of server resources subdivided to perform as multiple servers”

and describes “three partitions appear[ing] on the network (e.g., via IP address) as three distinct servers.” Ex. 1001, 1:34–36, 4:24–32. The prosecution history of the ’454 patent application also shows that applicants described a “partitioned server” as “a single server that ‘perform[s] as multiple servers.’” See Ex. 1002, 110–111 (“Applicants define what they mean by a ‘partitioned server’ on page 2, lines 3–5, of their specification, which states, ‘[a] partitioned server is a single server or aggregation of server resources subdivided to perform as multiple servers.’”); see also Ex. 1001, 1:34–36.

We credit Dr. Madisetti’s testimony that Colby teaches how its virtual web hosts share one physical server. Ex. 1003 ¶¶ 119–129. We also agree with Dr. Madisetti and credit his testimony that Colby’s VWHs on one physical server are themselves “a plurality of partitions” on a “partitioned server” as the intrinsic record uses those terms. See Ex. 1003 ¶¶ 138–143; EX1001, 1:34–37, 4:24–32; Ex. 1002, 110–111; Ex. 1046 ¶¶ 45–63.

Based on the complete record, we determine that Petitioner has established that Colby’s description of servicing a content request using a single physical web server implemented with multiple virtual web hosts teaches a “*method for routing a transaction to a partitioned server*” as recited in the preamble.

b) *[1A] identifying a plurality of partitions on said partitioned server*

Petitioner contends that Colby teaches this limitation

[b]ecause each of Colby’s VWHs “provides the functionality of a single physical web server” and “has one or more public virtual IP address[es] that clients use to access content on the [VWH],” and because multiple VWHs may “share a single physical web server” . . . , a [person of ordinary skill in the art] would have

understood each VWH is an example of what the '454 patent calls a *partition* on a *partitioned server*, as claimed.

Pet. 17–18 (citing Ex. 1005, 2:23–44; Ex. 1003 ¶¶ 142–143).

According to Petitioner, Colby's CAFS (Content Aware Flow Switch) has a "Content Server Database (CSD)" including "server records *identifying* servers" with information such as each server's IP address. Pet. 18 (citing Ex. 1005, 6:35–63; Ex. 1003 ¶ 144). Petitioner explains that these "server records" contain data collected from "servers 100a-c" by the CAFS's "Intelligent Content Probe (ICP)" component. Pet. 18 (citing Ex. 1005, 6:37–38, 8:19–26; Ex. 1003 ¶ 144). The identifying information (*e.g.*, IP address), Petitioner explains is "stored in Colby's server records mirrors the identifying information stored in the '454 patent's exemplary 'partition profile,' which likewise may include a 'network address,' *e.g.*, 'an IP address.'" Pet. 18 (citing Ex. 1001, 7:45–50, Fig. 3; Ex. 1003 ¶ 145).

Petitioner asserts that "Colby's method . . . includes the step of using Colby's 'server records' that 'identify' 'servers 100a-c'—with identifying information such as each server's IP address (Ex. 1005, 6:35–63) to create a list of servers (*i.e.*, Colby's VWHs) that have the content requested in a client request." Pet. 18 (citing Ex. 1005, 8:39–42, 9:1–4, 9:57–61, 10:1–13, Fig. 4; Ex. 1003 ¶ 154). Petitioner argues that "[t]his process of identifying a subset of VWHs in Colby constitutes *identifying a plurality of partitions on said partitioned server*, as claimed." Pet. 18 (citing Ex. 1003 ¶ 146).

According to Petitioner, Colby's virtual web host (VWH) implementation is a physical web server with partitions for two reasons. First, Petitioner contends that it is expressly taught by Colby. Pet. 12. Colby states that a "virtual web host" ("VWH")—referred to as a "logical entity"—may "provide[] the functionality of a single physical web server,"

and that “**several** virtual web hosts . . . may share a single physical web server.” *Id.* (citing Ex. 1005, 1:8, 2:22–36; *see also id.*, 16:20–24 (“Each of the physical web servers 100a-c may embody one or more [VWHs].”); Ex. 1003 ¶ 119). Each “virtual web host has one or more public virtual IP address that clients use to access content,” and is “uniquely identified by its public IP address.” Pet. 12 (citing Ex. 1005, 2:36–39; Ex. 1003 ¶ 120).

Petitioner states that, given these disclosures, a person of ordinary skill in the art would have understood that (1) “Colby teaches a switch that is **not** restricted to operating solely with physical servers, and discloses a CAFS used with ‘web servers’ (*e.g.*, in Fig. 1b) implemented as multiple VWHs on a ‘single physical web server’” and (2) “the ‘single physical web server’ acts as a ‘partitioned server’ for the VWHs.” Pet. 12 (citing Ex. 1001, 1:34–36 (describing “partitioned server” as “a single server or aggregation of server resources subdivided to perform as multiple servers”); (Ex. 1005, 2:29–44; Ex. 1003 ¶¶ 121–122).

Second, Petitioner argues that, alternatively, a person of ordinary skill in the art would have been motivated to implement Colby’s web servers as multiple VWHs on a single physical web server. Pet. 13 (citing Ex. 1003 ¶ 123). Petitioner asserts the following reasons:

First, Colby expressly teaches using its CAFS with physical web servers that “embody one or more virtual web hosts” and discloses that VWHs may each “provide[] the functionality of a single physical web server.” Ex. 1005, 16:21–24, 2:27–33; Ex. 1003 ¶ 124.

Second, such an implementation would have conferred the many well-known benefits (acknowledged in the ’454 patent)—*e.g.*, flexibility, efficiency, ease of management—of having multiple servers implemented on a single physical server. Ex.

1003 ¶ 125; Kauffman (Ex. 1010), 2:7–18; Shahabuddin (Ex. 1011), 2:31–40; ’454 patent, 1:34–42.

Third, such an implementation would have been nothing more than a simple substitution of one known element (multiple physical web servers) for another (multiple VWHs on a single physical web server each operating as a web server), yielding the predictable result of providing the functionality of multiple web servers in a content delivery network environment like the one described in Colby. Ex. 1005, 2:26–37; Ex. 1003, ¶ 126.

Pet. 13.

In its Response, Patent Owner argues that “Petitioner has not presented any evidence that more than one VWH would have the requested content” and “there is no evidence that two or more web sites hosted on a single web server would have replicated content.” PO Resp. 29–30. We disagree.

Dr. Madisetti testifies, and we agree, that Colby’s servers/VWHs 100a–100c have the requested content. *See* Ex. 1046 ¶¶ 30–32 (citing Ex. 1003 ¶ 171 (“Colby starts with a **list** of possible server candidates, which have the requested **content**”) (citing, *e.g.*, Ex. 1005, 11:60–67, 12:45–47, 13:3–10, 13:19–32); *see also* Ex. 1003 ¶ 181 (“The CSD then ‘identif[ies] the nature of the requested content.’ (Ex. 1005, 9:2–4) . . . The CSD then ‘queries its database for records of candidate servers **containing the requested content**’ (*id.*, 9:58–61).”) (emphasis original), 189 (“Suppose that the list of ‘candidate servers containing the requested content’ . . . contains **all three servers** 100a–c”). Dr. Madisetti’s testimony amply supports the position taken in the Petition. *See, e.g.*, Pet. 23 (“Colby starts with a list of possible server candidates which have the requested content”) (citing, *e.g.*, Ex. 1005, 11:60–67, 12:45–46, 13:3–33); *see also* Pet. 25 (citing Ex. 1005, 9:58–61), 27 (citing Ex. 1005, Fig. 8).

Patent Owner also asserts some of the same arguments it made with respect to the “partitioned server” recited in the preamble. *See* Sec.III.D.2.a, above. For example, Patent Owner argues that Petitioner has not shown that “there would be more than one partition to identify” because “[r]unning multiple virtual web hosts does not imply multiple partitions on a partitioned server” (PO Resp. 29–30; *see also* PO Sur-reply 10), and Colby’s disclosure of Virtual Web Hosts and IP-based virtual hosting applies equally as well to shared virtual hosting services that run all virtual hosts on a single partition or on an unpartitioned server (*id.* at 34–36). We find these arguments unavailing for the same reasons we discussed with respect to the preamble.

Patent Owner also argues that name-based virtual hosts cannot be identified by their IP addresses (PO Resp. 36–37) and that Colby’s server records do not store the IP addresses of its VWHs (PO Resp. 37–40). For example, Patent Owner argues that “Colby does not disclose identifying VWHs based on their VIPs because it does not disclose associating a VIP with a VWH in the server records.” PO Resp. 39. We find these arguments unavailing as well.

As we noted in the Decision, Colby’s Fig. 21a is “only one way of implementing multiple VWHs in a server environment.” Dec. 26–27. Dr. Madisetti testifies, and we agree, that another way to implement multiple VWHs in a server environment is to have the VWHs share one physical server. *See* Ex. 1003 ¶¶ 119–129; Ex. 1046 ¶ 48. Moreover, Dr. Madisetti testifies, and we agree, that “because Colby’s VWHs provide server functionality ‘transparent[ly],’ and because Colby’s virtual IP address (‘VIP’) is a ‘public’ address ‘uniquely identif[y]ing’ a VWH and ‘use[d]’ by

‘clients . . . to access content,’ the ‘server records’ in Colby’s VWH implementation associate a VIP with a VWH just as they associate an IP address with a server.” Ex. 1046 ¶ 66 (citing Pet. 12–14, 18, 34; Ex. 1005, Abstr., 2:22–39; Ex. 1003 ¶¶ 119–128, 146, 213).

Patent Owner further argues that “[i]t would not have been obvious to implement VWHs on separate partitions in Colby,” for a number of reasons. Patent Owner arguments include, “there is no evidence that VWHs were implemented on separate partitions” and “small VWHs were commonly run on a single instance of the http application” (PO Resp. 41 (citing Ex. 2007 ¶ 139); “Petitioner has not shown that a POSITA would have set up the VWHs using IP-based virtual hosting” (PO Resp 42); “Petitioner has also not shown that a POSITA would have run IP-based virtual hosting with multiple copies of the web server, one for each web site” (*id.* at 42–43); “Petitioner has not shown that its proposed configuration would have offered ‘flexibility, efficiency, [or] ease of management’ over the then-prevailing way of virtual hosting” (*id.* at 43–47); “Multiple partitions result in dynamic data movement that is difficult to track, making it difficult to implement with Colby’s content-aware flow switch” (*id.* at 47–48); and “it would have been more economical to purchase separate servers from Sun Microsystems than creating partitioned servers.” (*id.* at 48–50 (quoting Ex. 2007 ¶ 157)).

We have considered all of Patent Owner’s arguments, as well as the testimony of Mr. van Rietschote that Patent Owner relies on to support its arguments. *See, e.g.*, Ex. 2007 ¶¶ 137–164. In weighing the testimony of Mr. Rietschote and Dr. Madisetti on these issues, we find that Dr. Madisetti’s testimony has more probative value. We address the arguments and the evidence below.

First, Dr. Madisetti testifies, and we agree, that Colby's Virtual Web Hosts (VWHs) are the recited partitions which are identified by their IP addresses. Ex. 1046 ¶ 69 (citing Ex. 1005, 2:22–44, 6:35–62; Ex. 1003 ¶¶ 142–144). Mr. van Rietschote testifies that name-based virtual hosting was favored over IP-based hosting (Ex. 2007 ¶ 120), but regardless of whether name-based virtual hosting was more favorable, Colby still expressly teaches IP-based virtual hosting. *See* Ex. 1005, 2:36–38 (“A virtual web host has one or more public virtual IP address that clients use to access content on the virtual web host. A web host is uniquely identified by its public IP address.”); *see also* Ex. 1003 ¶¶ 142–144).

Second, Patent Owner's arguments and evidence that other implementations *could* have achieved similar or other results are beside the point, are not responsive to Petitioner's evidence, and are ultimately unavailing. That is because an obviousness determination “does not require that a particular combination must be the preferred, or the most desirable, combination described in the prior art in order to provide motivation for the current invention.” *In re Fulton*, 391 F.3d 1195, 1200 (Fed. Cir. 2004). For example, Patent Owner's arguments concerning Shahabuddin (Ex. 1011; PO Resp. 45; *see also* PO Sur-reply 14) are not helpful because Shahabuddin discloses the benefits of multiple servers implemented on a single physical server, which supports Colby's VWH Implementation. *See* Ex. 1003 ¶¶ 125, 271–272. Similarly, Patent Owner's arguments regarding running separate Apache instances without partitions is not directly responsive to, and does not negate, Colby's teaching of VWH implementation. *See* PO Resp. 46; PO Sur-reply 9–12; Ex. 2007 ¶¶ 140, 150; *see also* Ex. 1003 ¶¶ 123–124, 269–272.

We also disagree with Patent Owner's argument regarding Colby's "flow pipes." *See* PO Resp. 46 n.8. We instead credit Dr. Madisetti's testimony on this issue. Dr. Madisetti testifies, and we agree, that

Colby's flow pipes allocate **CAFS** bandwidth to VWHs using "round robin" scheduling—not server bandwidth. POR, 46 n.8 (citing EX1005, 3:7-9, 16:48-61; EX2007, ¶151. These disclosures in Colby of course do not *preclude or negate* the utility of using Colby's load-balancing techniques *for servers implemented as VWHs*—because a POSA, for example, understood that Colby's load-balancing disclosures would have still been valuable in scenarios where *VWHs* became overloaded or unresponsive.

Ex. 1046 ¶ 74.

Finally, Patent Owner's arguments that implementation of Colby's Virtual Web Hosts would be difficult to develop and deploy do not countermand Petitioner's obviousness showing. *See* PO Resp. 47–48 (citing Exs. 2028, 2029). Whether it was more economical or not to use separate or partitioned servers does not adequately address whether a person of ordinary skill in the art would have had a sufficient reason to implement Colby's Virtual Web Hosts, because "[u]nder the correct [obviousness] analysis, any need or problem known in the field of endeavor at the time of invention and addressed by the patent can provide a reason for combining the elements in the manner claimed." *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 420 (2007).

Here, the '454 patent acknowledges that partitions were already developed, without mentioning or describing any particular difficulty in deploying them. *See* Ex. 1001, 1:34–42, 4:22–24. Dr. Madisetti testifies, and we agree, that a person of ordinary skill in the art would have "understood that IBM's servers *were* used for web hosting, and even if they

were more expensive to ***purchase*** than Sun’s” servers, Mr. van Rietschote’s testimony notwithstanding, *See* Ex. 1046 ¶ 80; Ex. 2007 ¶¶ 157, 163. We credit Dr. Madisetti’s testimony, which provides a reasoned analysis and explanation of why a person of ordinary skill in the art would have implemented Colby’s VWHs in the manner described in the Petition. Ex. 1003 ¶¶ 123–127.

Based on the complete record, we determine that Petitioner has established that Colby’s description of server records identifying a subset of VWHs (Virtual Web Hosts) implemented on a single physical web server with client requested content teaches “*identifying a plurality of partitions on said partitioned server*” as recited in limitation [1A].

c) *[1B] determining a configuration of each of said plurality of partitions*

Petitioner asserts that Colby teaches this limitation by disclosing that its “ICP [Intelligent Content Probe] ‘probes the servers’ to determine ‘information regarding server status and content’ for each server” and then “stores this information — ‘status’ data like ‘load metrics’ and server ‘health,’ as well as ‘content’ data — about each server.” Pet. 19 (citing Ex. 1005, 6:58–63, 7:4–15, 8:19–25; Ex. 1003 ¶¶ 148–149). Relying on the testimony of Dr. Madisetti, Petitioner asserts that “[e]ach piece of information regarding a server’s load, health, and content is independently ***a configuration*** as claimed in the ’454 patent.” *Id.* (citing Ex. 1003 ¶ 151; Ex. 1001, 7:34–39, 7:50–53). According to Petitioner, a person of ordinary skill in the art would have understood “each piece of information (load, health, content) is ***determined*** by Colby’s [content-aware flow switch] when the ICP ‘probes’ the servers and ‘updates the CSD’ with information for use in

Colby’s best-fit server selection process.” *Id.* (citing Ex. 1005, 6:58–63, 7:4–15, 8:19–25; Ex. 1003 ¶ 151).

Petitioner further contends that because servers 100a–100c are the “plurality of partitions” in Colby’s VWH Implementation, the ICP’s collection of server “status and content” information (citing Ex. 1005, 8:19–21) used during Colby’s method “for servicing a content request using a [content-aware flow switch]” (citing *id.* at 4:20–22), constitutes “determining a configuration of each of said plurality of partitions” as recited in limitation [1B]. Pet. 20 (citing Ex. 1003 ¶ 152).

Patent Owner argues that “[t]he configurations that Petitioner maps to are those associated with a web server, and not with a partition as required by the claims.” PO Resp. 51. Patent Owner asserts that “Colby describes its method in reference to ‘servers,’ with no indication that Colby uses the term ‘server’ to encompass a partition of the physical server.” *Id.* (citing Pet. 19; Ex. 1005, 6:58–63, 7:4–15, 8:19–25; Ex. 1003 ¶¶ 148–151). Patent Owner asserts that “Petitioner’s argument rests on the [incorrect] assumption that [Colby’s] VWHs are each run as separate partitions in a partitioned server.” PO Resp. 52 (citing Pet. 12–14).

Patent Owner also argues that in Colby, “VWHs and servers are distinct and not co-extensive,” and “Colby does not treat the web servers 100a-c the same way as VWHs 100d-f.” PO Resp. 53–54 (citing Pet. 20; Ex. 2007 ¶ 170). PO Resp. 52. Patent Owner argues that “Petitioner has not pointed to anything in Colby that shows Colby would load balance with respect to VWHs in addition to or in lieu of load balancing with respect to physical servers.” PO Resp. 53 (citing Pet. 19–20).

Patent Owner further argues that “Petitioner has not offered competent evidence to show why each of ‘load, health, content’ is a claimed configuration,” in that “‘load,’ ‘health’ or ‘content’ is not something that can be setup (i.e. configured) by a system administrator, and so cannot satisfy the definition of ‘configurations.’” PO Resp. 55–56 (citing Ex. 2007 ¶¶ 86–89, 172).

We disagree with Patent Owner. We credit Dr. Madisetti’s testimony that “Colby’s teachings about best-fit ‘server’ selection apply regardless of whether ‘servers’ are implemented as separate physical servers [or] VWHs.” Ex. 1046 ¶ 109 (citing Pet. 12–15; Ex. 1003 ¶¶ 119–129, 296–276). Dr. Madisetti testifies, and we agree, that when Colby’s “servers” 100a–100c are implemented as VWHs, they meet the claimed “partitions,” because “‘configurations’ of servers 100a–100c are ‘configurations’ of partitions.” Ex. 1046 ¶ 109 (citing Pet. 17–18, 54; Ex. 1003 ¶¶ 142–143, 279).

We also disagree with Patent Owner’s argument that Colby’s teaching of a “flow pipe” somehow excludes or negates Colby’s teaching of load balancing. *See* PO Resp. 53–54; Ex. 2007 ¶¶ 167–170. As Dr. Madisetti testifies, and we agree, Colby’s load-balancing teaching and flow pipe teaching are not mutually exclusive arrangements. *See* Ex. 1046 ¶ 74. Moreover, Dr. Madisetti also persuasively testifies that Colby’s VWH teachings are not restricted to Figures 21a–b as Patent Owner argues in its Response. *See* PO Resp. 54; Ex. 2007 ¶ 170; Ex. 1046 ¶ 69.

Based on the complete record, we determine that Petitioner has established that Colby’s description of an ICP (Intelligent Content Probe) collection of VWH (Virtual Web Host) server status and content information

using a content-aware flow switch teaches “*determining a configuration of each of said plurality of partitions*” as recited in limitation [1B].

d) [1C](1) *providing said configurations to a load balancer*

Petitioner contends that Colby teaches this limitation. Pet. 20–21. According to Petitioner, Colby’s “content-aware flow **switch** [CAFS]” includes a Web Flow Redirector (WFR), Content Server Database (CSD), and Flow Admission Control (FAC), “which together ‘are responsible for selecting a server to service a content request based on a variety of criteria,’ including ‘load’ on the servers.” *Id.* at 20 (citing Ex. 1005, 6:24–28, 6:35–44, 6:65–67, 8:33–54; 2:48–58, claims 1–2; Ex. 1003 ¶¶ 155–157). Petitioner explains that the process uses “load metrics indicating [server] load,” and “load balance algorithm[s] to use when accessing [particular] content.” *Id.* (citing Ex. 1005, 6:42–63; Ex. 1003 ¶ 158). Thus, Petitioner contends that Colby’s WFR, CSD, and FAC collectively comprise the recited “load balancer.” *Id.* at 20–21 (citing Ex. 1003 ¶ 159).

Petitioner further explains that “Colby’s ICP collects information comprising *said configurations*—*i.e., a configuration of each of said plurality of partitions* recited in [1B], such as load, health, or content” and the “‘ICP’s job is to ‘update[] the CSD’ with these configurations — that is, to ‘populate the CSD with server and content information,’ *e.g.*, server ‘status’ information like ‘load’ or ‘health,’ and ‘content’ information.” Pet. 21 (citing Ex. 1005, 6:58–63, 7:4–12, 8:19–26; Ex. 1003 ¶¶ 160–161). Because the CSD is part of the load balancer, Petitioner contends that this process constitutes providing the configurations to a load balancer. *Id.* (citing Ex. 1003 ¶ 161).

Patent Owner does not specifically dispute Petitioner's evidence and arguments with respect to limitation [1C](1), except to assert generally that Colby does not teach limitations that recite a "configuration" term. *See, e.g.,* PO Resp. 59; Ex. 2007 ¶¶ 180–181.

Based on the complete record, we determine that Petitioner has established that Colby teaches "*providing said configurations to a load balancer*" as recited in limitation [1C](1).

- e) [1C](2) wherein said transaction is routed by said load balancer to one of said plurality of partitions based at least in part on said configurations

Petitioner contends that Colby teaches this limitation. Pet. 21–24. According to Petitioner, Colby's content-aware flow switch "services" content requests received from clients and "forwards" them to a best-fit server (VWH) for each request where each "content request" is the recited "transaction." *Id.* at 21 (citing Ex. 1005, Abstract, 2:48–53, 4:21–22, Fig. 3; Ex. 1003 ¶ 162). Petitioner explains that forwarding the transaction constitutes "routing" the transaction, as the term "routing" is used in the '454 patent, and that Colby's transactions (i.e., "content requests") are routed by Colby's load balancer (i.e., Colby's WFR, CSD, and FAC). *Id.* (citing Ex. 1003 ¶ 163). Thus, Petitioner contends that "said transaction is routed by said load balancer" to Colby's VWH (i.e., a claimed partition) on Colby's multiple-VWH web server (i.e., a claimed partitioned server). *Id.* (citing Ex. 1003 ¶ 169).

According to Petitioner, a person of ordinary skill in the art would have understood that each of Colby's "content requests" (i.e., transactions) is routed based at least in part on the configurations, as claimed. Pet. 22–23 (citing Ex. 1003 ¶ 170). Petitioner explains that (1) the "content"

information, “health” information, and “load metrics” collected for each partition are each configurations (i.e., a configuration of each of said plurality of partitions recited in limitation 1B) and (2) each content, health, load from each partition is used in various steps of Colby’s method “for servicing a content request using a [CAFS].” Pet. 23 (citing Ex. 1005, 4:20–22).

Patent Owner does not specifically dispute Petitioner’s evidence and arguments with respect to limitation [1C](2)⁴, except to assert generally that Colby does not teach limitations that recite a “configuration” term. *See, e.g.*, PO Resp. 59; Ex. 2007 ¶¶ 180–181.

Based on the complete record, we determine that Petitioner has established that Colby teaches “*wherein said transaction is routed by said load balancer to one of said plurality of partitions based at least in part on said configurations*” as recited in limitation [1C](2).

f) [1D] assigning a rank to each of said plurality of partitions based at least in part on said configurations, wherein said transaction is routed based on said ranking

Petitioner asserts that a person of ordinary skill in the art would have understood Colby’s method for servicing a content request involves “assigning a rank to each of said plurality of partitions based at least in part on said configurations” for at least two independent reasons. Pet. 24 (citing Ex. 1003 ¶ 178).

⁴ Patent Owner asserts in the Title to Section VI.C of its Response that “The combinations do not disclose or suggest . . . routing a transaction to one of the plurality of partitions based at least in part on the configurations,” but Section VI.C does not contain any arguments directed specifically to the “routing” limitation. *See* PO Resp. 58–59.

First, Petitioner argues that Colby's method includes selecting a "best-fit server" based at least in part on the "server load metrics" of each of the "candidate servers" and the skilled artisan would have understood that "selecting a 'best' server from among multiple candidates based on load metrics is an 'empirical assignment' (Ex. 1001, 7:52-56), *i.e.*, **rank**, based on evaluating such metrics from all candidates (to determine the 'best')." Pet. 24–25 (citing Colby, Abstract, claims 1–2; Ex. 1003 ¶ 179; Ex. 1019, 507). Thus, Petitioner asserts that because Colby's various "load metrics" are each a "configuration" of a partition, Colby's rankings are "based at least in part on said configurations," as claimed. *Id.* at 24 (citing Ex. 1003 ¶ 179).

Second, Petitioner argues that "the exemplary request servicing method Colby describes with respect to Figure 3 includes ranking servers using server content, health, and load data in the CSD that are each said configurations." Pet. 25 (emphasis omitted). According to Petitioner, "[i]n Colby's method, the WFR 'querie[s]' the CSD 'for a **list** of available servers to serve the content request.'" *Id.* (citing Ex. 1005, 8:33–40, Fig. 3; Ex. 1003 ¶¶ 180–181). Petitioner explains that "[t]he CSD then 'identif[ies] the nature of the requested content' (Ex. 1005, 9:2–4) and 'queries its database for records of candidate servers **containing the requested content**' (*id.*, 9:58–61). Pet. 25. "If it finds such servers," Petitioner explains, "Colby's CSD '**sorts** the list of candidate servers.'" Pet. 25 (citing Ex. 1005, 10:1–13; Ex. 1003 ¶ 181). According to Petitioner, a person of ordinary skill in the art

would have understood Colby's query/sorting method constitutes **assigning a rank** based on server content, health, and load; for example, when handling content requests that are not required to be served by a specific server, Colby's CSD creates a 'list of candidate servers,' *i.e.*, servers with the requested

content, and ‘filters’ and ‘orders’ it using the process shown in Colby Figure 6.

Pet. 25–26 (citing Ex. 1005, 10:51–11:3; Ex. 1003 ¶ 182).

Petitioner explains that “[i]n this process, the CSD ‘filters the candidate servers’ by removing servers that are, *e.g.*, ‘unresponsive’ or ‘unreachable.’” Pet. 26 (citing Ex. 1005, 11:4–6, 11:60–65, 12:6–13, Figs. 6, 8–9; Ex. 1003 ¶¶ 183–184). According to Petitioner, “[t]he CSD then ‘orders (sorts)’ the ‘*remaining*’ servers.” Pet. 26 (citing Ex. 1005, 11:6–8). “In particular,” Petitioner explains that

the remaining servers are “evaluated” and “ordered” by applying “ordering rules.” Pet. 26–27 (citing Ex. 1005, 11:60–12:5, Fig. 8; Ex. 1003 ¶¶ 185–186). The “evaluation” process “flags” certain conditions for each server, such as whether the “port bandwidth” is low. Ex. 1005, 12:13–50. The “ordering” process considers these flags along with other information, like whether the requested content is “high priority” or “frequently requested,” and assigns to each a server a status of “OKAY” or “not optimal”; “not optimal” servers are “placed at the bottom of the candidate server list.” *Id.*, 11:26–59, 12:51–13:33; Ex. 1003 ¶¶ 187–188.

Pet. 26–27.

According to Petitioner, a person of ordinary skill in the art would thus have understood that a server’s eventual place on the “candidate servers” list depends at least in part on content information (presence of requested content), health information (*e.g.*, whether the server is “unresponsive” or “unreachable”) and load information (*e.g.*, whether port bandwidth is low) of *all* candidate servers, *i.e.*, on *said configurations* (whose antecedent basis is *a configuration of each of the plurality of partitions* from Limitation [1B]). Ex. 1003 ¶ 189; Ex. 1020, 1:27–37; Ex. 1021, 2:48–50; Ex. 1017 [0003]; Ex. 1018 1:13–20. This is because, as a POSA would have understood, whenever an unhealthy server is filtered from the list of candidates with the requested content, or a “not optimal” server is moved to the

bottom, the other servers move closer to the head of the list. Ex. 1003 ¶ 189; Ex. 1005, Fig. 8. Thus, Colby's server list ordering process ***assigns a rank to each of said plurality of partitions*** (Colby's VWHs, *supra* § VI.B.1.b) ***based at least in part on said configurations*** because each partition's eventual place in the server list is dependent on each of content, health, and load for that partition ***and*** the other partitions. Ex. 1003 ¶ 190. Because the above-discussed rank assignment affects which partition the content request (***said transaction***) is forwarded to (***routed to***), *see supra* §§ VI.A.2, VI.B.1.e, ***said transaction is routed based on said ranking*** as required. Ex. 1003 ¶ 191.

Pet. 27–28.

Patent Owner argues that Petitioner has not shown that “Colby discloses assigning a rank to each of the plurality of partitions, let alone that the ranking is based at least in part on the configurations of the partitions.” PO Resp. 58. Patent Owner asserts that “Petitioner has not shown that when a content request comes in, more than one VWH would have the requested content,” because “[t]here would then be no need to perform the additional steps of ranking possible candidates and choosing the best-fit servers because there would be only one candidate server.” *Id.* at 58–59. Patent Owner also argues with respect to Colby's Figure 21a that “Colby does not teach how to rank 100e versus 100f as its evaluation is with respect to the underlying physical servers 100b and 100c,” and “the Petition [does not explain] how Colby's method should be extended to VWHs serviced by web servers.” *Id.* at 59 (citing Pet. 21–24; Ex. 2007 ¶ 179).

We disagree with Patent Owner. Patent Owner's arguments rely on the premise that Colby's VWHs cannot access the same content and can only be deployed as shown in Colby's Figures 21a and 21b. As we previously discussed, Colby's Figures 21a and 21b represent only one way of implementing multiple Virtual Web Hosts. *See* Sec. III.D.2.b, above. Dr.

Madisetti testifies, and we agree, that Colby teaches another implementation, where VWHs share one physical server. *See* Ex. 1003 ¶¶ 119–129 (citing Ex. 1005, 2:22–44; 16:20–24, 16:29–34); *see also* Pet. 12–15.

We credit Dr. Madisetti’s testimony that a person of ordinary skill in the art “would have understood Colby to disclose a CAFS used in a system with ‘web servers’ (such as those shown in Figure 1b) implemented as multiple VWHs on a ‘single physical web server.’” Ex. 1003 ¶ 122 (citing Ex. 1005, 2:29–44). Patent Owner also acknowledges that the “evidence shows that multiple virtual web hosts were run on a single server.” *See* PO Response, 35 (citing Ex. 2007 ¶ 135; Ex. 1011, 2:20–25) (“The shared services model is a popular method . . . especially for virtual web hosting . . . In this approach, multiple customers and services are hosted on a single server”).

Dr. Madisetti provides credible, supported testimony that Colby’s VWH implementation identifies a plurality of partitions where multiple VWHs in the implementation would have the requested content, and the ranking that Petitioner and Dr. Madisetti describe of choosing a best-fit server from among multiple servers having the requested content satisfies the recited limitation. *See* Ex. 1045 ¶¶ 27–44; Ex. 1003 ¶¶ 176–191.

Based on the complete record, we determine that Petitioner has established that Colby teaches “*assigning a rank to each of said plurality of partitions based at least in part on said configurations, wherein said transaction is routed based on said ranking*” as recited in limitation [1D].

- g) *[1E] storing said configurations in a partition profile, wherein said configurations are provided to said load balancer via said partition profile*

Petitioner contends that Colby teaches this limitation. Pet. 28–29. Petitioner explains that the recited “configurations” are the server “load,” “health,” and “content” information collected by the ICP, and that Colby expressly states “load” information for each server is stored in the CSDs “server record” for each server. *Id.* at 28 (citing Ex. 1005, 6:58–63).

Petitioner also explains that a person of ordinary skill in the art would have understood that “health” information is also stored in “server records,” or would have found it obvious to implement Colby this way, “because:

- (1) Colby’s ‘server record’ stores information about individual servers; and
- (2) Colby says the ICP ‘probes’ servers to ‘monitor’ their ‘health,’ and then ‘updates the CSD [Content Server Database]’—which stores the ‘server records’ — with the data it collects.” *Id.* (citing Ex. 1005, 6:56–58, 7:7–12, 8:19–25; Ex. 1003 ¶¶ 192–194).

Petitioner also explains that Colby’s Content Server Database (CSD), which contains server and content records with information regarding specific servers (i.e., VWHs), is the recited “partition profile,” and Colby’s request servicing process includes storing the configurations in that profile. Pet. 29 (citing Ex. 1005, 6:38; Ex. 1003 ¶¶ 195–196). Petitioner contends that, “[l]ike the ‘partition profile’ in the ’454 patent, which is ‘a computer readable file, such as a database’ ‘for maintaining the configuration of the partitions’” (Ex. 1001, 7:16–20), Colby’s CSD “maintains several databases” (Colby, 6:41–42) that contain configuration information. Thus, because Colby’s CSD is part of the “load balancer” in limitation [1C](1), Petitioner states that the “configurations are provided to said load balancer

via said partition profile,” as claimed. *Id.* (citing Ex. 1003 ¶¶ 197–199; Ex. 1022, 1315 (“via” means “by way of” or “by means of”).

Patent Owner does not specifically dispute Petitioner’s evidence and arguments with respect to limitation [1E]. *See generally* PO Resp.

Based on the complete record, we determine that Petitioner has established that Colby teaches “*storing said configurations in a partition profile, wherein said configurations are provided to said load balancer via said partition profile*” as recited in limitation [1E].

h) Conclusion as to Claim 1

Based on the complete record before us, and for the reasons discussed above, we are persuaded that Petitioner has demonstrated by a preponderance of the evidence that Colby teaches the subject matter of independent claim 1.

3. Independent Claims 6 and 17

Independent claims 6 and 17 recite limitations similar to the limitations recited in independent claim 1 and Petitioner relies on its analysis addressing independent claim 1 to address independent claims 6 and 17. *See* Ex. 1001, 9:30–38 (claim 6), 10:28–40 (claim 17); Pet. 37–38 (comparing limitations in claims 1 and 6), 42–44 (comparing limitations in claims 1 and 17). Patent Owner’s arguments addressing independent claims 6 and 17 rely on the same arguments addressed above in the context of independent claim 1. *See* PO Resp. 50.

Accordingly, based on the complete record before us, and for the same reasons as those explained above with respect to independent claim 1, we are persuaded that Petitioner has demonstrated by a preponderance of the

evidence that Colby teaches the subject matter of independent claims 6 and 17.

4. *Independent Claim 16*

37 CFR § 42.104(b)(3) (2023) states in pertinent part that

Where the claim to be construed contains a means-plus-function or step-plus-function limitation as permitted under 35 U.S.C. 112(f), the construction of the claim must identify the specific portions of the specification that describe the structure, material, or acts corresponding to each claimed function.

Patent Owner argues that Petitioner failed to identify the specific portions of the specification that describe the structure, material or acts corresponding to each claimed function of claim 16. PO Resp. 64 (citing 37 C.F.R. § 42.104(b)(3)). Patent Owner argues, for example, that “there is no comparison of the ‘configurations’ that are determined in the ‘454 patent to the ‘configurations’ relied on in the petition. Petitioner does not explain where ‘content,’ ‘health’ or ‘load’ appear in the ‘454 patent.” *Id.*

Independent claim 16 reads as follows:

[16PRE] An apparatus having a processor for routing a transaction to a partitioned server, comprising:

[16A] means for identifying a plurality of partitions on said partitioned server;

[16B] means for determining a configuration of each of said plurality of partitions in response to an event;

[16C] means for routing said transaction to one of said plurality of partitions based at least in part on said configuration thereof; and

[16D] means for assigning a rank to each of said plurality of partitions based at least in part on said configuration thereof, wherein said transaction is routed based on said rank.

Ex. 1001, 10:15–27 (bracketed labelling by Petitioner; *see* Pet. 77).

Petitioner states, in part, that “[l]imitations [16A]–[16D] recite ‘means for’ performing the same functions performed by ‘program code’ recited in limitations [6C]–[6F], respectively, and ‘the load balancer that routes transactions as recited in [6E] performs the required function of *routing said transaction* in [16C].” Pet. 41 (citing Ex. 1003 Decl. ¶¶ 232–233). We understand that Petitioner contends that the structure, material, and acts in limitations 6C–6F and limitations 6E correspond to the functions in limitations 16A–16D and limitation 16C, respectively.

Petitioner provides a discussion of claim 6’s limitations, complete with a chart mapping claim 6’s limitations to the corresponding limitations in claims 1 and 2, as well as the corresponding sections of the Petition where these limitations are discussed. See Pet. 37–38. Moreover, Petitioner’s discussion of claim 6 provides citations to Colby (Ex. 1005 19:44–59) and to Dr. Madisetti’s supporting testimony (Ex. 1003 ¶¶ 220–223).

Dr. Madisetti provides a separate analysis of independent claim 16. See Ex. 1003 ¶¶ 231–234 (cited and discussed at Pet. 41–42). Dr. Madisetti testifies that

Limitations [16A]–[16D] recite “means for” performing functions that are the same functions performed by “program code” recited in limitations [6C]–[6F], respectively. Those functions are all performed by Colby’s CAFS “program product,” as I explained in § V.A.2.f.ii above. The only variation between the functions recited in Limitations [16A]–[16D] and [6C]–[6F] is that [16C] recites “routing said transaction to...” whereas [6E] recites “providing said configuration of each of said plurality of partitions to a load balancer, wherein said transaction is routed by said load balancer to...” A POSA would have understood that a load balancer that routes transactions as recited in [6E] performs the required function of “routing said transaction” in [16C].

Ex. 1003 ¶ 233.

Dr. Madisetti also testifies that

I understand that if a claim element is expressed as a “means” for performing a specified function, it covers the corresponding structure described in the specification and equivalents of the described structure. I have also reviewed the prosecution history of the ’454 patent. During prosecution, the applicants represented that the recited “means” in the means-plus-function limitations in as-filed claim 19 (which issued as claim 16) was simply “program code,” without specifying a particular algorithm that the code had to perform. *See* ’454 FH, 126-127 (referring to “means (e.g., ‘program code’; p. 5, lines 11-13; p. 6, lines 16-21)” when describing the limitations of claim 19), 183-184 (same), 261 (same), 301 (showing that claim 19 was renumbered to 16). Therefore, I have assumed for the purposes of my analysis of claim 16 that “program code” is the corresponding structure for all “means” limitations. Under this construction, the corresponding structure for Limitations [16A]-[16D] is met by Colby’s CAFS “program product,” which I discussed in § V.A.2.f.i above, which a POSA understood was implemented by machine-readable *program code* that was executable by a processor. Colby’s CAFS thus meets [16A]-[16D].

Ex. 1003 ¶ 234.

Given this record, we disagree with Patent Owner’s contention that Petitioner failed to comply with 37 CFR § 42.104(b)(3). In particular, we point out the Petitioner’s discussion and Dr. Madisetti’s cited testimony identifying the ’454 patent’s “program code” as the corresponding structure for all the means limitations recited in claim 16, explaining that

[d]uring prosecution, the applicants represented that the recited “means” in the means-plus-function limitations in as-filed claim 19 (which issued as claim 16) was simply “program code,” without specifying a particular algorithm that the code had to perform. *See* ’454 FH, 126-127 (referring to “means (e.g., ‘program code’; p. 5, lines 11-13; p. 6, lines 16-21)” when

describing the limitations of claim 19), 183-184 (same), 261 (same), 301 (showing that claim 19 was renumbered to 16).

Ex. 1003 ¶ 234; Pet. 41–42.

Based on the complete record before us, and for the same reasons explained above with respect to independent claim 1, we are persuaded that Petitioner has demonstrated by a preponderance of the evidence that Colby teaches the subject matter of independent claim 16.

5. *Dependent Claims 2–5, 8, 10, 13, and 18–21*

Claims 2–5, 8, 10, 13, and 18–21 depend directly from independent claims 1, 6 and 17. *See* Ex. 1001, 9:11–19, 9:39–10:14, 10:41–51.

Petitioner provides argument and supporting evidence showing how Colby teaches or suggests each of the limitations recited by these claims. *See* Pet. 29–36, 39–40, 45–50; *see also* Ex. 1003 ¶¶ 200–219, 225, 230, 244–248, 258–260. Patent Owner challenges Petitioner’s contentions with respect to dependent claims 10, 19, and 20. *See* PO Resp. 65–67. We address those arguments specifically made by Patent Owner with respect to these dependent claims.

a) *Dependent Claim 10*

Dependent claim 10 recites: *[a]n apparatus as in claim 6, wherein said program code for identifying said plurality of partitions identifies said plurality of partitions as resource partitions on said server.*

Patent Owner argues that “[t]he plain language of ‘identify . . . as’ requires the program code to identify the partitions ‘as’ possessing the recited characteristics.” PO Resp. 66 (citing Ex. 2007 ¶ 189).

With respect to claim 10, Petitioner explains that

The ’454 patent describes an exemplary “server 140 comprising resource partitions 160–162” as “a single server having

separately allocated resources. For example, a server 140 with resource partitions 160-162 may have a single processor that is allocated 20% to a first partition, 30% to a second partition, and 50% to a third partition.” ’454 patent, 4:24–29. The specification does not describe “identifying” a partition “as” a resource partition or give examples of partitions identified as such; it merely discloses that partitions can *be* resource partitions, *e.g.*, by sharing a processor. ’454 patent, 4:24–29. Given these disclosures, a POSA would have understood claim 10 encompasses merely identifying partitions *that are* resource partitions. Ex. 1003 ¶ 226.

Petitioner also explains that

Colby discloses “*shar[ing]* server resources, such as *CPU [central processing unit] cycles[.]*” Ex. 1005, 2:24–36. A POSA would have had reason to implement Colby’s VWHs on a single-CPU server (and would have reasonably expected success doing so), because multiple servers sharing a single CPU was a known, conventional way to implement servers. Ex. 1003 ¶¶ 227–228 (citing Keshav (Ex. 1024), Abstract, 1:63–2:12, 2:37–39; Arndt (Ex. 1025), 1:21–29, 4:52–62). In such an implementation, Colby’s VWHs are what the ’454 patent refers to as *resource partitions*, since they share processor (CPU) cycles, and the single- CPU server is a *partitioned server* as recited in claim 6, because it performs as multiple servers (*see supra* § VI.B.1.a). Ex. 1003 ¶ 229.

Pet. 39–40. Petitioner’s arguments are supported by the testimony of Dr. Madisetti. Ex. 1003 ¶¶ 226–229.

We agree with Petitioner. The ’454 specification does not describe “identifying” a partition “as” a resource partition or give examples of partitions identified as such; it merely shows that partitions can be resource partitions, *e.g.*, by sharing a processor. *See, e.g.*, Ex. 1001, 4:24–29.

Colby’s VWHs are what the ’454 patent refers to as resource partitions, because they share processor (CPU) cycles, and the single-CPU server is a

partitioned server because it performs as multiple servers. *See* Ex. 1003 ¶ 229.

Accordingly, based on the complete record before us, we are persuaded that Petitioner has demonstrated by a preponderance of the evidence that Colby teaches the subject matter of dependent claim 10.

b) Dependent Claim 19

Dependent claim 19 recites: *[a] method as in claim 17, wherein said at least one characteristic is determined by reading a tag associated with said transaction.*

Patent Owner argues that

the Petition focuses on “type” of requested content, but the FAC cited for claim 19 does not consider content type, but whether “sufficient port and switch resources [are] available on the content-aware flow switch to support the flow.” Ex. 1005, 7:65–8:1. Thus, the resources FAC considers for a given flow request is not that of the partition, but that of the FACS. *See* Ex. 2007, ¶ 187; Ex. 1005, 14:13–15 (“Flow admission control is performed for each content request in order to verify that adequate resources exist to service the content request, and to offer the content request the level of service indicated by its QoS requirements. If sufficient resources are not available, the content request may be redirected to another site capable of servicing the request or simply be rejected.”). This is not surprising because as Petitioner acknowledges, flow admission control is done after the server has been ordered (i.e., has been evaluated that it had sufficient port bandwidth and cache capacities). *Id.*; Pet. 47. As to rejecting a flow request because the server is not local, that is not “content type,” but server location. Petitioner has not shown that claim 19 is obvious. Ex. 2007 ¶ 187.

PO Resp. 65.

With respect to claim 19, Petitioner explains that the ’454 patent does not provide examples of a “tag,” but instead incorporates-by-reference co-

owned U.S. Patent Application Ser. No. 09/751,009, which published as Daoud (Ex. 1026), which it describes as disclosing “modifying” a “load balancer” for “reading a tag.” Pet. 46 (citing Ex. 1001, 5:15–20). According to Petitioner, “Daoud in turn discloses an exemplary ‘tag’ that is a ‘single or multi-bit packet associated with’ a ‘data packet.’” Pet. 26 (citing Ex. 1026 ¶ 26). “Thus,” Petitioner asserts, a “**tag** has its plain and ordinary meaning encompassing a ‘name (label, mnemonic) assigned to a data structure, such as a field’ or ‘file.’” Pet. 46–47 (citing Ex. 1027, 416 (defining “tag”); Ex. 1003 ¶¶ 249–251).

According to Petitioner, a person of ordinary skill in the art “would have understood that in addition to routing requests to partitions based on the requested content . . . Colby routes requests to partitions based on another *characteristic of said transaction*, namely the ‘type’ of the requested content.” Pet. 47 (citing Ex. 1005, Abstr. (“best-fit server is chosen based on the *type* of content requested”), claim 1 (“best-fit server” selected based on, *inter alia*, “content *type* information”); Ex. 1003 ¶ 252). “For example,” Petitioner explains, “after the servers are ordered . . . Colby’s CAFS performs ‘flow admission control’ (FAC) for the content request to ensure that a content request is ‘offer[ed]’ the ‘level of service indicated by its QoS requirements.’” Pet. 47 (citing Ex. 1005, 14:5–18). Petitioner explains that “[t]he FAC unit in the CAFS determines whether the first server on the list is a local server and can ‘satisfy the content request,’ and if not, rejects that server.” Pet. 47 (citing Ex. 1005, 14:20–30; *see also id.*, 2:53–56; Ex. 1003 ¶¶ 253–254). According to Petitioner, “Colby’s CAFS determines the content request’s ‘QoS’ requirements based in part on

the content type.” Pet. 47 (citing Ex. 1005, 9:9–12 (for an “HTTP request,” “*content-type*” is used to “deduce a QoS class”); Ex. 1003 ¶ 255).

Petitioner asserts that a person of ordinary skill in the art “would have understood that when the *at least one characteristic of said transaction* recited in claim 17 (from which claim 19 depends) is content type, Colby’s request servicing method meets claim 17, because the process for *routing said transaction* . . . includes consideration of the content type, and is thus *based on* the content type.” Pet. 47–48 (citing Ex. 1003 ¶ 256). Petitioner asserts that “Colby discloses that content type is *determined* by ‘extract[ing],’ *i.e.*, *reading*, the ‘HTTP content-type’ ‘element’ from the ‘HTTP header’ of the content request.” Pet. 48 (citing Ex. 1005, 9:4–6). According to Petitioner, a person of ordinary skill in the art “would have understood that this ‘element’ is a *tag* as the term is used in the ’454 patent, since it is a piece of information indicating something about the request (its type), and that it is *associated with* the content request—*i.e.*, *said transaction*.” Pet. 48 (citing Ex. 1003 ¶ 257; Ex. 1027, 416; Ex. 1026 ¶¶ 25–26 (exemplary “tag” is “packet associated with the data packet” in the “transaction” whose value “identifies a requested level of service for the transaction”).

We agree with Petitioner that a person of ordinary skill in the art would understand that a “tag” as that term is used in the ’454 patent would be a name, label, or mnemonic assigned to a data structure, such as a field or a file. We also agree with Petitioner that Colby’s CAFS performs flow admission control (FAC) and determines the content request’s requirements based in part on the content type, which is extracted from the content request by reading the HTTP content-type element from the header of the request.

We further agree with Petitioner that a person of ordinary skill in the art would have understood that an HTTP content-type element in a request header would be a “tag” as the term is used in the ’454 patent.

Accordingly, based on the complete record before us, we are persuaded that Petitioner has demonstrated by a preponderance of the evidence that Colby teaches the subject matter of dependent claim 19.

c) Dependent Claim 21

Dependent Claim 21 recites: *[a] method as in claim 17, wherein said at least one characteristic is determined to be a requested level of service for said transaction.*

Patent Owner argues that “there is no showing that the alleged characteristics of ‘requested level of service of said transaction’ was compared to a configuration of the server/partition.” PO Resp. 65–66 (citing Pet. 49–50).

With respect to claim 21, Petitioner asserts that “[t]he ’454 patent does not describe how to **determine** that a **characteristic** is a **level of service** for said transaction, but says exemplary ‘characteristic(s)’ of a transaction ‘may include’ a ‘requested level of service (e.g., low priority, high priority).” Pet. 49 (citing Ex. 1001, 5:8–29; Ex. 1003 ¶ 261). Petitioner asserts that “Colby likewise discloses determining whether ‘the requested content is **high priority** content’ and using that as a factor in ‘order[ing]’ the ‘candidate server list.’” Pet. 49 (citing Ex. 1005, 11:19–29, 12:51–13:11; Ex. 1003 ¶ 262).

According to Petitioner, a person of ordinary skill in the art “would have understood that when the **at least one characteristic of said transaction** recited in claim 17 (from which claim 21 depends) is the

‘priority’ of the requested content, Colby’s request servicing method meets claim 17, because the process for **routing said transaction . . .** includes considering priority.” Pet. 49–50 (citing Ex. 1005, 11:19–39 (“HI_PRIORITY” request flag set for “high priority content”), 12:31–13:11 and Fig. 11 (value of “HI_PRIORITY” affects, inter alia, whether “port bandwidth” is considered in “order[ing]” servers); Ex. 1003 ¶ 263).

We agree with Petitioner that Colby teaches determining whether the requested content is high priority content and uses that as a factor in ordering the candidate server list. We also agree that a person of ordinary skill in the art would have understood that when at least one characteristic of the transaction is a priority of the requested content, Colby’s request servicing method meets the claim because the process for routing the transaction includes considering priority. *See* Ex. 1005, 11:19–39, 12:31–13:11; Ex. 1003 ¶¶ 262–263.

Accordingly, based on the complete record before us, we are persuaded that Petitioner has demonstrated by a preponderance of the evidence that Colby teaches the subject matter of dependent claim 21.

d) Uncontested Claims 2–5, 8, 13, 18, 20

Petitioner presents evidence and arguments demonstrating that Colby teaches the recited limitations of dependent claims 2–5, 8, 13, 18, 20. *See* Pet. 29–36, 39–40, 45–56, 48–49. Dr. Madisetti provides testimonial evidence supporting Petitioner’s position. *See* Ex. 1003 ¶¶ 200–219, 225, 230, 244–248, 258–260.

Patent Owner does not present any arguments or evidence directly challenging Petitioner’s evidence and arguments for these particular claims. *See* PO Resp. 29–68; PO Sur-reply 4–16; Ex. 2007 ¶¶ 117–181, 187–190.

We have considered Petitioner's arguments and evidence directed to these dependent claims, as well as the supporting testimony of Dr. Madisetti. Based upon consideration of the entire record, we are persuaded that Petitioner has demonstrated by a preponderance of the evidence that Colby teaches the subject matter recited by dependent claims 2–5, 8, 13, 18, and 20.

E. Ground 2: Asserted Obviousness over Colby and Eilert

Petitioner contends that claims 1–6, 8–10, and 13–21 are unpatentable under 35 U.S.C. § 103(a) as being obvious over Colby and Eilert. Pet. 50–58. We have determined that these claims are unpatentable either over Colby (Ground 1), or the combined teachings of Colby, Eilert, and Jindal (Ground 3). See Sections III.D and III.F. Accordingly, we need not address Petitioner's alternative Ground 2 based on the combined teachings of Colby and Eilert. See *Boston Sci. Scimed, Inc. v. Cook Grp. Inc.*, 809 F. App'x 984, 990 (Fed. Cir. Apr. 30, 2020) (non-precedential) (recognizing that “the Board need not address issues that are not necessary to the resolution of the proceeding” and that the Board has “discretion to decline to decide additional instituted grounds once the petitioner has prevailed on all its challenged claims”).

F. Ground 3: Asserted Obviousness over Colby, Eilert, and Jindal

For Ground 3, Petitioner contends that claims 1–21 are unpatentable as obvious over Colby, Eilert, and Jindal. Pet. 58–70. For this ground, we consider the evidence and arguments directed to the remaining dependent claims 7, 9, 11, 12, 14, and 15. We begin with an overview of Eilert and Jindal.

1. *Eilert (Ex. 1006)*

Eilert is directed to managing workload in a partitioned system, such as a logically partitioned system used, for example, for providing a server function. Ex. 1006, 1:41–44. Eilert explains that logical partitioning allows the establishment of a plurality of system images within a single physical machine or central processor complex (CPC). Each system image is capable of operating as if it was a separate computer system. That is, each logical partition can be independently reset, initially loaded with an operating system that may be different for each logical partition, and operate with different software programs using different input/output (I/O) devices. *Id.* at 1:48–56.

Eilert explains that one important aspect of a logically partitioned system is the management of workload running within the partitions of the system. In certain systems, for example, workload managers are used to manage the workload within and among the partitions. The workload managers attempt to balance the workload of the partitions by moving work to the physical resources of the system. In order to move the work, however, it is important to ensure that the data needed by the relocated work is at the moved location. *Id.* at 2:8–20.

Eilert describes a “workload manager” (WLM) that dynamically moves resources like CPU resources between logical partitions. *Id.* at 2:41–55. Eilert explains that a WLM can distribute CPU resources across logical partitions by dynamically adjusting the CPU processor weights associated with logical partitions. A WLM understands, for instance, when an important workload is late because the weight of the partition it is running within is too low. A WLM can adjust this workload by, for instance, raising

the weight of this partition and lowering the weight of another partition, thereby providing additional CPU capacity to an important workload. This allows CPU resources to be dynamically moved to needed partitions, as workload requirements change. *Id.* at 2:44–55. Eilert refers to a logical partition by the acronym “LPAR.” *Id.* at 8:20–35.

2. *Jindal (Ex. 1007)*

Jindal is in the field of computer systems and networking and is directed to a system and method for responding to a network resource request. Ex. 1007, 1:12–15. Jindal explains that multiple servers may be configured to offer an application or replicated service (e.g., a service offered simultaneously on each of multiple servers), in which case the client may be directed to any of multiple servers in order to satisfy the client’s request. The multiple servers may be situated in various locations and/or serve different clients. Jindal describes how to distribute clients’ requests among servers and/or among instances of the application or service in a process known as “load balancing.” *Id.* at 1:27–37.

Jindal describes the use of a computer-executable trigger that executes in response to a client request for an application load-balanced among multiple instances of an application operating on multiple servers. *Id.* at 2:66–3:4. Information concerning the status or performance of the servers may include, for example, the server’s response time for a client request, its operational status (e.g., whether it is up or down), the number of clients connected to the instance, the number of client requests pending with the instance, its throughput (e.g., number of client requests handled in a period of time), etc. *Id.* at 3:13–19. Information concerning the status or performance of the servers themselves (e.g., load, capacity, distance from a

central server) may also be collected and analyzed as part of the process of choosing a preferred server. *Id.* at 3:26–36.

Jindal explains that this information is collected by a “framework” including various “executable objects.” *Id.* at 3:59–67 (“In one embodiment a load-balancing framework is employed to collect information from application instances. An illustrative framework consists of multiple executable objects, modules or other collection of executable instructions.”). A first type of object, called a status object, is configured to collect one or more pieces of information from an instance of the application (and/or its server) that is being load-balanced. *Id.*

Each instance of an application is associated with its own status object. *Id.* at 4:1–6. Each instance of the load-balanced application is also associated with an individual monitor object or IMO (e.g., another module or series of executable instructions). Each IMO invokes and stores information from one or more status object(s) collecting information concerning an instance of the application. *Id.*

A replicated monitor object (RMO) or module is employed to collect information from the IMOs associated with the various instances of the load-balanced application. The RMO stores this information, which is then analyzed to identify a preferred server in accordance with the selected policy/policies. *Id.* at 4:7–12.

3. *Rationale to Combine*

Petitioner asserts that “Colby’s CAFS [Content Aware Flow Switch] ‘direct[s]’ datapackets ‘between a client and a best-fit server’ (Ex. 1005, 20:1–24, 1:66–2:7) selected from multiple servers that each have an ‘IP address’ (*id.* at 6:42–63).” Pet. 50 (citing *id.* at Sec. VI.A.2). According to

Dr. Madisetti, “Eilert describes a single physical computing system providing the functionality of multiple such servers.” Ex. 1003 ¶¶ 269–270. Petitioner asserts that “Eilert expressly describes its WLM [Work Load Manager] embodied in a ‘9672 Parallel Enterprise Server’ (Ex. 1006, 4:2–11) and says an ‘example[]’ of ‘physical resources’ its WLM ‘redistribute[s]’ between partitions are ‘network adapter cards’ (*id.* at 5:65–6:6).” Dr. Madisetti testifies that “[i]t was well-known that such ‘adapter cards’ provided network connectivity to individual partitions.” Ex. 1003 ¶ 270 (citing Ex. 1028, 2:27–28; Ex. 1029, 1:8–11). Dr. Madisetti testifies that “[i]t was also well-known for network-connected servers to use TCP/IP, with each server having its own IP address.” Ex. 1003 ¶ 270 (citing Ex. 1030, 3:6–9; Ex. 1031, 60). “Thus,” Dr. Madisetti testifies, a person of ordinary skill in the art “understood Eilert’s partitions and WLM could be used to implement Colby’s servers each having their own IP address.” Ex. 1003 ¶ 270.

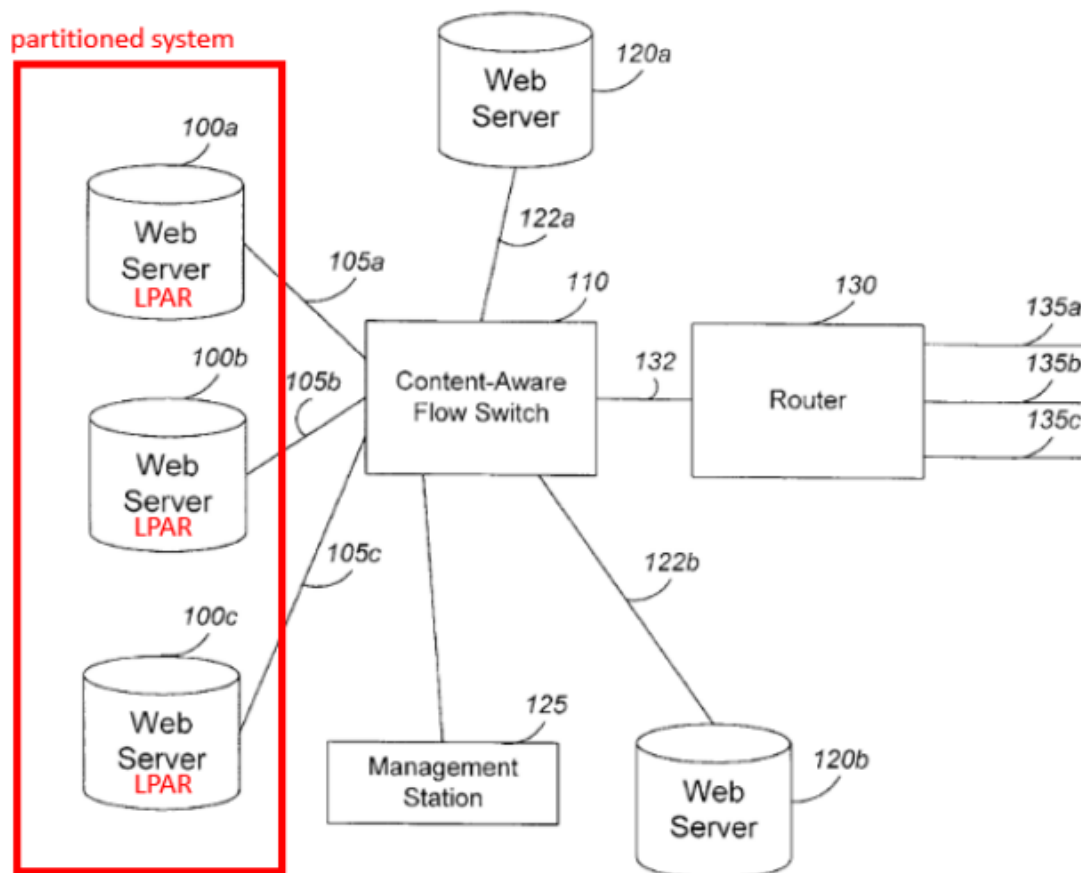
Petitioner asserts that a person of ordinary skill in the art “had multiple reasons to use Colby’s CAFS in a network where Eilert’s partitions and WLM are used to implement Colby’s servers with IP addresses.” Pet. 51. “First,” Petitioner argues, “such an implementation would have provided well-known benefits—*e.g.*, flexibility, efficiency, manageability—of multiple servers implemented on a single physical server (Ex. 1010, 2:7–13; Ex. 1011, 2:31–37; Ex. 1001, 1:34–42) while providing the additional benefit of overcoming ‘shortcomings’ Eilert identified in prior partitioned systems.” Pet. 51–52 (citing Ex. 1006, 2:8–23, 2:41–55; Ex. 1003 ¶¶ 271–272).

“Second,” Petitioner argues, “such an implementation would have been nothing more than a substitution of one known element (multiple physical servers) for another (multiple LPARs with Eilert’s WLM on a single physical system) yielding the predictable result of providing the functionality of multiple web servers for servicing content requests as described in Colby.” Pet. 52 (citing Ex. 1005, 2:22–44, 16:20–23; Ex. 1003 ¶ 273).

“Moreover,” Petitioner argues, “Eilert’s WLM would not have discouraged a Colby/Eilert combination, because Colby’s and Eilert’s teachings are complementary.” Pet. 52. According to Petitioner, “Colby’s load-balancing method considers factors including whether a server has requested content, whether it is responsive, and the server’s load/bandwidth, which remain relevant to Colby’s load-balancing technique even where server resources can be moved between partitions, as Eilert teaches.” *Id.* (citing Ex. 1003 ¶ 274) (internal citations omitted).

Petitioner argues that a person of ordinary skill in the art “would have reasonably expected success using Colby’s CAFS with multiple servers implemented as partitions because partitions provide the same functionality as individual servers, and because Colby’s teachings regarding selecting a best-fit server are compatible with servers implemented as partitions.” Pet. 52–53 (citing Ex. 1005, Abstr., 2:31–34, 16:20–23; Ex. 1003 ¶ 275; Ex. 1009 ¶ 2 (“An example of logical partitions is the partitioning of a multiprocessor computer system into multiple independent servers.”); Ex. 1006, 1:48–56 (“Logical partitioning allows the establishment of a plurality of system images” each “capable of operating as” a “separate computer system.”); Ex. 1030, 1:36–43).

Petitioner provides an annotated version of Colby's Figure 1b, shown below. Pet. 53.



Petitioner explains that “[a]n implementation where Colby’s CAFS directs content requests to servers 100a–100c implemented as LPARs on Eilert’s partitioned system is reflected in annotated Fig. 1b, [above], with a red box representing the partitioned system.” Pet. 53 (citing Ex. 1003 ¶ 276). Petitioner argues that a person of ordinary skill in the art “would have been motivated to implement Colby’s servers as partitions in Eilert’s partitioned server, with each partition having its own IP address.” Pet. 60 (citing Ex. 1003 ¶ 307).

With respect to Jindal, Petitioner argues that

Jindal's invention "is not limited to the use of DNS" and may be used in any "method of satisfying network service requests" (Ex. 1007, 4:55–63), and a POSA understood Jindal's "framework" for collecting and analyzing data using RMOs, IMO, and status objects could be used in devices like Colby's CAFS. Ex. 1003 ¶ 308. Specifically, a POSA understood that Jindal's "object-oriented software programming concepts" (Ex. 1007, 5:4–18) could be implemented in Colby's CAFS "computer program product" (Ex. 1003, 19:43–58; Pet. Sec. VI.B.6.a), and had multiple reasons to add Jindal's framework to Colby's CAFS "program product" in an environment where Colby's CAFS services requests for servers implemented as partitions on Eilert's system.

Pet. 60 (citing Ex. 1003 ¶ 309).

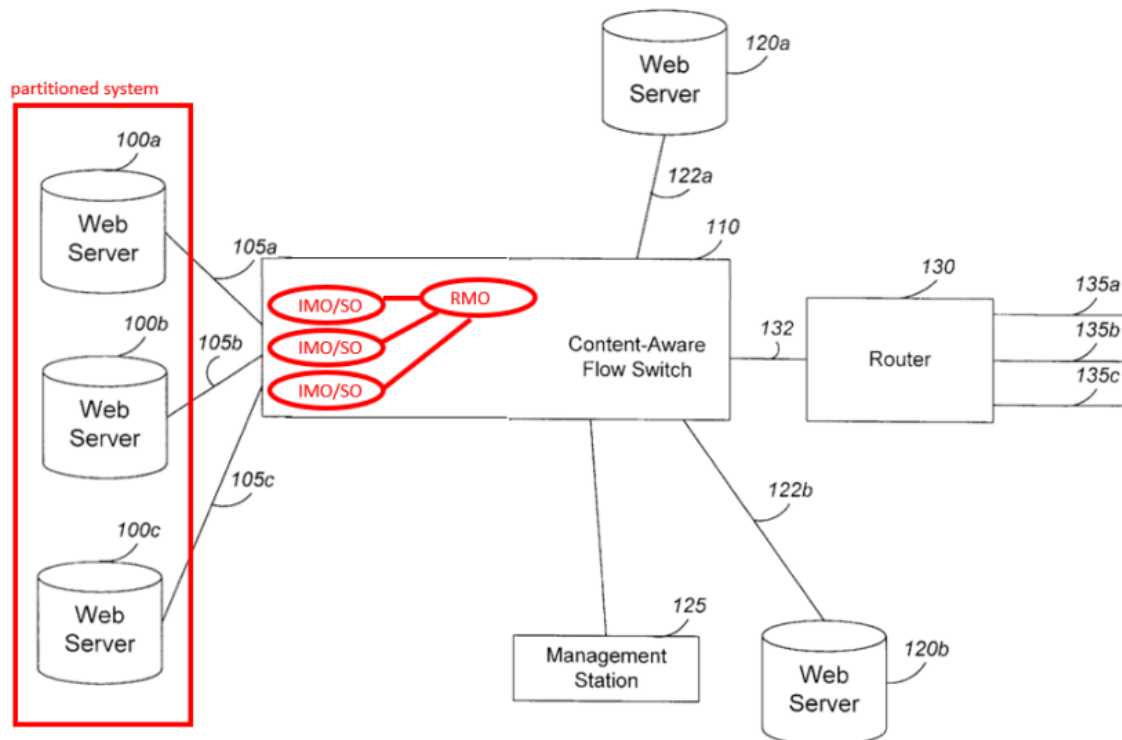
"First," Petitioner argues, "Jindal's teachings would have motivated a POSA to add Jindal's framework objects to Colby's CAFS in an environment where Colby's servers are implemented as Eilert's partitions managed by Eilert's WLM, to achieve a flexible, customizable load-balancing system with 'status objects' that can gather information based on different 'load balancing polic[ies]' 'selected by a system manager or administrator.'" Pet. 60 (citing Ex. 1007, 5:36–38, 7:51–54, 8:30–47; Ex. 1003 ¶ 310). Petitioner argues that "[a]dding Jindal's RMOs, IMO, and status objects to Colby's CAFS would have beneficially enabled Colby's CAFS to collect additional information from servers and implement desirable load-balancing policies based on that information." Pet. 61.

"Second," Petitioner argues, "such an implementation would have merely been the use of a known technique (Jindal's 'framework' implemented with 'object-oriented programming') to improve similar devices (Jindal's DNS server and Colby's CAFS both perform load balancing for 'client requests' (Ex. 1005, 3:10–12; Ex. 1007, Abstr.)) in the

same way (providing increased flexibility in responding to client requests).” Pet. 61 (citing Ex. 1003 ¶ 311; *see also id.* at Sec. VI.B.1.d (describing Colby’s CAFS load-balancing)).

“Third,” Petitioner argues, “such an implementation is merely combining prior art elements (Jindal’s RMO, IMO, and status objects with Colby’s CAFS) according to known methods (‘object oriented-oriented software programming concepts’ (Ex. 1007, 5:14–18)) to yield predictable results (Colby’s CAFS being able to collect additional information from servers).” Pet. 61 (citing Ex. 1003 ¶ 312; Ex. 1032, 4:32–36 (“Object[-]Oriented Programming” is “well known”). According to Petitioner, a person of ordinary skill in the art “would have reasonably expected success adding Jindal’s RMOs, IMOs, and status objects to Colby’s CAFS, because this would have involved the use of basic programming techniques such as object-oriented programing that were well within a POSA’s skill.” Pet. 61–62 (citing Ex. 1003 ¶ 313).

Petitioner provides an annotated version of Colby’s Figure 1b, shown below.



Petitioner provides “[a]n implementation where Colby’s CAFS (implemented with components of Jindal’s framework) directs content requests to servers 100a–100c (implemented as LPARs on Eilert’s partitioned system) is reflected in annotated Fig. 1b, [above], with a red box representing the partitioned system.” According to Petitioner, “the CAFS includes IMOs that invoke status objects to collect information from the servers, and at least one RMO that aggregates this information.” Pet. 62 (citing Ex. 1003 ¶ 314). Petitioner asserts that “this CAFS implementation collects information from servers using Colby’s ICP *and* Jindal’s RMO/IMO/status object framework, and uses *both* types of information to choose a best-fit server.” Pet. 62.

Patent Owner argues that

Eilert acknowledges that as workload of the partitions are rebalanced and work is moved, “it is important to ensure that the data needed by the relocated work is at the moved location.” Ex.

1006, 2:12–16. This reflects the general understanding that in a logical partition, there would be significant amount of content movement. Ex. 2007 ¶ 183. This data movement would not be eliminated by Eilert’s attempt to reallocate CPU resources across partitions. *Id.* This is because when resources, such as memory, were reallocated, the data in those RAM memory associated with a previous partition would be moved as well. *Id.*; *cf.* Ex. 2029 (noting the need to scale RAMs when content is replicated over different network nodes).

PO Resp. 60.

Patent Owner argues that “Colby’s load balancer, however, is a content-aware load balancer. Its algorithm depends on correctly determining where the content resides, and a person of ordinary skill in the art would understand that it would not be compatible with Eilert’s dynamically changing LPAR architectures.” PO Resp. 60–61 (citing Ex. 1005, Figs. 3–4; Ex. 2007 ¶ 184); *see also* PO Sur-reply 18–19. “At a minimum,” Patent Owner argues, “the need to ascertain the dynamic location (partition) of the desired content could mean added latencies in responding to queries, which would discourage a POSITA from pursuing the combination.” PO Resp. 61 (citing Ex. 2007 ¶ 184).

“In fact,” Patent Owner argues, “Sun Microsystems was originally not interested in logical partitions like Eilert’s LPARs and only released logically partitioned servers like those in Eilert in late 2003.” PO Resp. 61 (citing Ex. 2027). “Moreover,” Patent Owner argues, “at the time of the invention, the ability for Eilert to dynamically reallocate resources to deal with changing work load would remove the need for Colby’s CAFS to rank based on load and health, because the server can rebalance the work load to address the shortage in resources needed for a given load and the resulting longer response time.” PO Resp. 61 (citing Ex. 2007 ¶ 185).

Patent Owner also argues that

Petitioner has not even considered how Eilert, with its dynamically moving content (Ex. 2007 ¶ 184), would be compatible with a ‘content-aware’ switch. Petitioner has not shown any examples in the prior art where a content-aware load balancer was used in conjunction with partitioned servers, especially dynamically configurable partitioned servers like those of Eilert. In other words, Petitioner has not shown that combining Eilert’s dynamically configurable partitions with Colby’s content-aware flow switches (as opposed to using partitioned servers in general) would have resulted in the alleged “flexibility,” “efficiency,” “manageability,” or predictability. Pet. 51–52; Ex. 2048 ¶ 4 (“content-aware routing usually incurs significantly more overhead” and the associated benefits “must be significant enough to justify the higher overhead”), ¶ 5 (allegedly as of 3/31/2005, content-aware routing was insufficient to handle dynamic data).”

PO Resp. 61–62.

Patent Owner argues that “Petitioner has also not considered the effects with both the CAFS and the server adjusting for the changing load.”

PO Resp. 62 (citing Pet. 50–53). “For instance,” Patent Owner argues, “whether that could lead to oscillation in the behavior pattern. In the prior art, load balancing is performed by one node of the network (client side, DNS server, dispatcher (load balancer), or server side).” PO Resp. 62 (citing Ex. 2020, 2). Patent Owner argues that “Petitioner has not shown examples where load balancing was performed at multiple nodes of the network (*e.g.*, both the server side and the dispatcher side).” PO Resp. 62.

We disagree with Patent Owner’s argument that Eilert’s dynamic resource reallocation capability “remove[s] the need for Colby’s CAFS to rank based on load and health.” PO Response 61; Ex. 2007 ¶ 185. This argument is not supported, but instead, is contradicted by the evidence of

record. First, Patent Owner does not explain why a server that can rebalance a work load somehow eliminates the need to “rank based on load and health.” Moreover, there is nothing in Colby or Eilert (or any record evidence) that suggests or promotes ignoring a server partition’s content, load, or health (factors considered by Colby’s CAFS in Colby/Eilert) when routing transactions in an implementation where server resources are movable between partitions. On the contrary, Dr. Madisetti testifies that a person of ordinary skill in the art would have understood that such factors should be considered. *See* Ex. 1046 ¶¶ 91–93.

Colby’s load-balancing performed by the CAFS includes the step of filtering non-responsive server partition candidates. *See* Ex. 1003 ¶¶ 99–105; Pet. 9–12 (step 3 in “Best-Fit Server Selection in Colby”). Colby’s load-balancing addresses a factor Eilert’s WLM cannot address by moving resources. *See* Ex. 1003 ¶ 274; Pet. 52. Patent Owner’s own evidence confirms filtering unreachable servers is important. *See* Ex. 2020, 3 (“Knowledge of server state conditions is essential for a highly available Web-server system to exclude servers that are unreachable because of faults or congestion.”). Colby’s CAFS filters such server partitions, regardless of what resources Eilert’s WLM allocates to them. *See, e.g.*, Ex. 1003 ¶¶ 103–104 (explaining that in Colby’s best-fit server selection algorithm, Colby’s CAFS filters out candidates that are unresponsive or unreachable), 279 (explaining that in Colby/Eilert combination, Colby’s CAFS still uses the best-fit server-selection algorithm).

As Dr. Madisetti testifies, and we agree, a person of ordinary skill in the art would have understood that using Colby’s CAFS to load-balance requests going to Eilert’s LPARs would help mitigate any problem of

heavily-loaded servers. Ex. 1003 ¶ 189 (explaining that in Colby’s best-fit server selection algorithm, more heavily-loaded servers are moved to the bottom of the candidate server list), ¶ 279 (explaining that in Colby/Eilert combination, Colby’s CAFS still uses the best-fit server-selection algorithm).

We also disagree with Patent Owner’s arguments that Colby is not compatible with Eilert because in Eilert’s LPARs, “there would be [a] significant amount of content movement,” and Colby’s “algorithm depends on correctly determining where . . . content resides.” PO Resp. 60–61; Ex. 2007 ¶ 155. Rather, Eilert teaches that LPARs access all needed data from shared “main storage 202.” Ex. 1006, 4:52–5:46 (“Logical partitions 108 (LP1-LP4) [in Fig. 1a] and logical partition manager 110 each comprise one or more programs residing in respective portions of central storage associated with the central processors . . . In a further embodiment of a computing environment, two or more central processor complexes are coupled to one another to form a sysplex, as depicted in FIG. 1b . . . In the example shown [in Fig. 1b], CPC 120 includes a plurality of logical partitions 124 (e.g., LP1–LP3), which are managed by a logical partition manager 126).

Moreover, we credit with Dr. Madisetti’s testimony that Colby’s load balancer does not show any incompatibility with Eilert. Ex. 1003 ¶ 98. Mr. van Rietschote’s testimony that ascertaining “the dynamic location (partition) of the desired content could mean added latencies in responding to queries” is not availing. An obviousness determination “does not require that a particular combination must be the preferred, or the most desirable, combination described in the prior art in order to provide motivation for the

current invention.” *In re Fulton*, 391 F.3d 1195, 1200 (Fed. Cir. 2004). Indeed, Dr. Madiseti provides credible testimony explaining the benefits of implementing Colby’s load-balancing CAFS in a network with Eilert’s LPARs. *See* Ex. 1003 ¶¶ 79–118, 264–276.

Patent Owner also argues that combining Eilert’s dynamically configurable partitions with Colby’s [CAFS] would somehow create undesirable “overhead” that is “usually incur[red]” with “content-aware routing.” *See* PO Response, 61–62; *see also* PO Sur-reply 19–20. Patent Owner also argues that “Petitioner has . . . not considered the effects with both the CAFS and the server adjusting for the changing load” and “whether that could lead to oscillation in the behavior pattern.” PO Resp. 62; *see also* PO Sur-reply 20–21. These assertions appear to be unsupported attorney argument, and are entitled to little probative weight. *See Icon Health & Fitness, Inc. v. Strava Inc.*, 849 F.3d 1034, 1046 (Fed. Cir. 2017) (“Attorney argument is not evidence or explanation in support of a conclusion.”); *Invitrogen Corp. v. Clontech Lab., Inc.*, 429 F.3d 1052, 1068 (Fed. Cir. 2005) (“Unsubstantiated attorney argument regarding the meaning of technical evidence is no substitute for competent, substantiated expert testimony.”); *In re Payne*, 606 F.2d 303, 315 (CCPA 1979) (citing *In re Lindner*, 457 F.2d 506, 508 (CCPA 1972)) (“Arguments of counsel unsupported by competent factual evidence of record are entitled to little weight.”).

Based on the complete record, we find that Petitioner has demonstrated by a preponderance of the evidence that one of ordinary skill in the art at the time of the claimed invention would have had sufficient reason to combine the teachings of Colby, Eilert, and Jindal in the manner

described in the Petition and would have had a reasonable expectation of success in that endeavor.

4. *Dependent claims 7, 9, 11, 12, 14, and 15*

Petitioner presents evidence and arguments showing that the combined teachings of Colby, Eilert, and Jindal meet the limitations of dependent claims 7, 9, 11, 12, 14, and 15. *See* Pet. 58–70. Dr. Madisetti provides testimonial evidence supporting Petitioner’s position. *See* Ex. 1003 ¶¶ 200–219, 225, 230, 244–248, 258–260.

Patent Owner presents some arguments and evidence challenging Petitioner’s showing with respect to Ground 3. *See* PO Resp. 66–68; PO Sur-reply 4–16; Ex. 2007 ¶¶ 189, 191–194. We address the arguments and evidence provided by Patent Owner challenging Petitioner’s position with respect to these claims under Ground 3.

a) *Dependent Claim 7*

Dependent claim 7 recites: “[a]n apparatus as in claim 6, wherein said program code for determining said configuration receives said configuration from a partition manager.”

Patent Owner argues that

[c]laim 7 requires that the program code for determining the configuration receive the configuration from a partition manager. But in Eilert, the CPU reallocation is done dynamically by the WLM and not the PR/SM pointed to by Petitioner. And the cited passages in Eilert column 9 do not mention PR/SM as the entity that provides the CPU capacity information. *See* EX1006, 9:21–25, 9:43–46; EX2007, ¶194.

PO Resp. 68.

We disagree with Patent Owner. Petitioner asserts that the combined teachings of Colby, Eilert, and Jindal meet claim 7 because CPU capacity

information is received from Eilert's "PR/SM," which meets the recited "partition manager." Pet. 65–67; Ex. 1003 ¶¶ 324–331. Patent Owner argues that Eilert's PR/SM does not reallocate CPU resources or provide CPU capacity information. See PO Resp. 68; Ex. 2007 ¶ 194. However, claim 7 does not require a "partition manager" to "reallocate" resources as Patent Owner argues. Moreover, whether Eilert expressly mentions the PR/SM providing CPU capacity information is beside the point. Petitioner's argument was that a person of ordinary skill in the art understood that CPU capacity was a "processor resource" managed by Eilert's PR/SM, *i.e.*, partition manager, and that a person of ordinary skill in the art would have implemented Petitioner's combination so that Jindal's status objects retrieve CPU capacity information from Eilert's PR/SM. See Pet. 66–67; Ex. 1003 ¶¶ 324–331.

b) Dependent Claim 9

Dependent claim 9 recites: "[a]n apparatus as in claim 6, wherein said program code for identifying said plurality of partitions identifies said plurality of partitions as logical partitions on said server."

Patent Owner asserts that

Petitioner argues that a POSITA would have used Jindal's objects to obtain such information because "the presence of dedicated processor resources for a partition potentially affects that partitions performance and should be considered when choosing a best-fit partition for a content request." Pet. 67–68.

PO Resp. 67. Patent Owner argues that a person of ordinary skill in the art would not have used Jindal's objects to obtain such information because logical partitions "would not give that information" versus resource partitions. PO Resp. 67 (citing Pet. 67–68). "For example," Patent Owner argues,

3 processors could be divided over ten different partitions. EX2007, ¶191. And Eilert teaches ways of allocating complete CPU's rather than a number of CPU resources to a particular task to avoid the short engine effect (EX1006, 12:50-63). *Id.* Thus, logical or resource partition would not indicate whether there are "dedicated processor" as Petitioner argues. EX2007, ¶191.

PO Resp. 67.

Patent Owner's argument is unavailing because it does not address Petitioner's stated position. Neither the Petition nor Dr. Madisetti stated that a person of ordinary skill in the art would have been motivated to use Jindal's objects to obtain specific divisions of processors, as Patent Owner asserts here. Rather, Petitioner's and Dr. Madisetti's position is that a person of ordinary skill in the art would have used Jindal's objects to determine whether or not processors were shared across partitions. *See* Pet. 67–68; Ex. 1003 ¶¶ 336–338.

5. *Conclusion on Ground 3*

We have considered Petitioner's arguments and evidence directed to dependent claims 7, 9, 11, 12, 14, and 15, as well as the supporting testimony of Dr. Madisetti. Based upon consideration of the entire record, including Patent Owner's arguments and evidence and the testimony of Mr. van Rietschote, we are persuaded that Petitioner has demonstrated by a preponderance of the evidence that the combined teachings of Colby, Eilert, and Jindal teach the subject matter of dependent claims 7, 9, 11, 12, 14, and 15.

IV. CONCLUSION

For the foregoing reasons, we determine that Petitioner has demonstrated by a preponderance of the evidence that claims 1–21 of U.S. Patent 7,523,454 B2 are unpatentable on the bases set forth in the following table.⁵

Claim(s) Challenged	35 U.S.C. §	References	Claims Shown Unpatentable	Claims Not Shown Unpatentable
1–6, 8, 10, 13, 16–21	103	Colby	1–6, 8, 10, 13, 16–21	
1–6, 8–10, 13–21	103	Colby, Eilert ⁶		
1–21	103	Colby, Eilert, Jindal	7, 9, 11, 12, 14, 15	
Overall Outcome			1–21	

⁵ Should Patent Owner wish to pursue amendment of the challenged claims in a reissue or reexamination proceeding subsequent to the issuance of this Final Decision, we draw Patent Owner's attention to the April 2019 *Notice Regarding Options for Amendments by Patent Owner Through Reissue or Reexamination During a Pending AIA Trial Proceeding*. See 84 Fed. Reg. 16,654 (Apr. 22, 2019). If Patent Owner chooses to file a reissue application or a request for reexamination of the challenged patent, we remind Patent Owner of its continuing obligation to notify the Board of any such related matters in updated mandatory notices. See 37 C.F.R. § 42.8(a)(3), (b)(2).

⁶ Because each of these challenged claims is held unpatentable on other grounds, we do not reach Ground 2 in the Petition.

V. ORDER

Upon consideration of the foregoing, it is

ORDERED that Petitioner has demonstrated by a preponderance of the evidence that claims 1–21 of U.S. Patent 7,523,454 B2 are unpatentable; and

FURTHER ORDERED that because this is a Final Written Decision, any party to the proceeding seeking judicial review of this Decision must comply with the notice and service requirements of 37 C.F.R. § 90.2.

FOR PETITIONER:

Anant Saraswat
Adam Wichman
WOLF, GREENFIELD & SACKS, P.C.
asaraswat-ptab@wolfgreenfield.com
awichman-ptab@wolfgreenfield.com

FOR PATENT OWNER:

Hong Zhong
Amy Proctor
Blair Silver
IRELL & MANELLA LLP
hzhong@irell.com
aproctor@irell.com
bsilver@irell.com