Trials@uspto.gov
571-272-7822

Paper 11
Entered: July 5, 2017

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

EMC CORPORATION, LENOVO (UNITED STATES) INC., and NETAPP, INC.,
Petitioner,

v.

INTELLECTUAL VENTURES I LLC,
Patent Owner.

Case IPR2017-00429
Patent 6,775,745 B1

Before GEORGIANNA W. BRADEN, DANIEL J. GALLIGAN, and KRISTI L. R. SAWERT, *Administrative Patent Judges*.

SAWERT, *Administrative Patent Judge*.

DECISION
Institution of *Inter Partes* Review
37 C.F.R. § 42.108

I. INTRODUCTION

EMC Corporation, Lenovo (United States) Inc., and NetApp, Inc. (collectively, "Petitioner") filed a Petition for *inter partes* review of claims 4, 5, and 6 of U.S. Patent No. 6,775,745 B1 (Ex. 1001, "the '745 patent"). Paper 5 ("Pet."). Intellectual Ventures I LLC ("Patent Owner") filed a

Preliminary Response. Paper 9 ("Prelim. Resp."). Institution of an *inter partes* review is authorized by statute when "the information presented in the petition . . . and any response . . . shows that there is a reasonable likelihood that the Petitioners would prevail with respect to at least 1 of the claims challenged in the petition." 35 U.S.C. § 314(a); *see* 37 C.F.R. § 42.108. Upon consideration of the Petition and the Preliminary Response, we institute *inter partes* review as to claims 4 and 5, but not claim 6, of the '745 patent.

A. Related Matters

The parties identify the following district court cases in which the '745 patent has been asserted: *Intellectual Ventures I LLC v. NetApp, Inc.*, No. 1:16-cv-10868-IT (D. Mass); and *Intellectual Ventures I LLC v. Lenovo Grp. Ltd.*, No. 1:16-cv-10860-IT (D. Mass). Pet. 3; Paper 6, 1. The '745 patent is also the subject of an *inter partes* review designated IPR2016-01643.

B. The '745 Patent

The '745 patent is directed to a computer hybrid data caching mechanism "for enhancing system performance." Ex. 1001, 1:6–9. The '745 patent explains as background that computer users "often wait long periods of time for programs to load." *Id.* at 1:12–14. In particular, the '745 patent states that disk input/output (I/O) "can take significant amounts of time while searching for and loading programs and data," especially when a user is attempting to run more than one software program at the same time. *Id.* at 1:16–19. "Much of this wasted time," the '745 patent explains, results from "disk seeks" and "overhead in issuing commands to read/write the many small pieces of data." *Id.* at 1:19–22.

According to the ’745 patent, accessing the data that makes up a file or program via a “disk seek” requires first accessing directory information, which keeps track of file and program locations, and then accessing the file or program itself. *Id.* at 1:23–31. Thus, the ’745 patent explains, “[m]ultiple seeks are typically required to locate a file.” *Id.* at 1:31–33. Also, according to the ’745 patent, “[w]hen more than one program is requesting data, the file system driver can end up reading a small amount of data for a first program, then seek to a different area on the disk to read another small amount for a second program, then seek back to the original area to read the next small amount of data for the first program, and so forth.” *Id.* at 1:41–47.

The ’745 patent explains that to circumvent these problems, known file system drivers in various computer operating systems “do a certain amount of caching.” *Id.* at 2:1–3. According to the ’745 patent, “caching programs strictly use the MRU-LRU (most recently used-least recently used) mechanism as their sole means of deciding what data to keep and what data to discard.” *Id.* at 2:14–17. But, a disadvantage in using the MRU-LRU approach is that “an important file that has not been used recently will be discarded when the cache reaches a target capacity and is forced to free up additional capacity.” *Id.* at 2:18–20. The ’745 patent states that “[s]trictly adhering to the MRU-LRU mechanism fails to ensure that files which have been used often, but may not have been used recently, are maintained in the cache.” *Id.* at 2:20–23.

The ’745 patent seeks to “solve the problems of the prior art” by providing for “an improved caching mechanism to minimize overhead and seeks, thereby improving system performance.” *Id.* at 2:24–27. The caching mechanism “minimizes seeks and reads to a hard drive” by “keep[ing] data in the cache based upon currency of use and the number of times the data is used.” *Id.* at 2:32–35. The ’745 patent states that “system performance is enhanced through the hybrid caching mechanism where a file that has been used often but not recently is maintained in the cache.” *Id.* at 3:47–49.

The caching mechanism disclosed in the ’745 patent generally begins with “reading files in response to a request from an operating system.” *Id.* at 2:41–43. Copies of those read files are then stored in cache, located within the computer’s random access memory. *Id.* at 2:43–45. Next, “frequency factors are assigned to each of the files stored in the cache.” *Id.* at 2:45–46. “Frequency factors,” the ’745 patent explains, “indicate how often each of the corresponding files is accessed by the operating system.” *Id.* at 2:46–48. When the cache needs additional capacity to store files, the “frequency factors are scanned” and “the least frequently and least recently used file is eliminated to liberate capacity of the cache.” *Id.* at 2:49–53. The ’745 patent explains that the “frequency factor” together with “the most recently used-least recently used (MRU-LRU) mechanism,” identify the “least *frequently* used (LFU) file which has not been *recently* used.” *Id.* at 4:39–42 (emphases added).

Figure 2A, shown below, illustrates one embodiment disclosed in the ’745 patent:

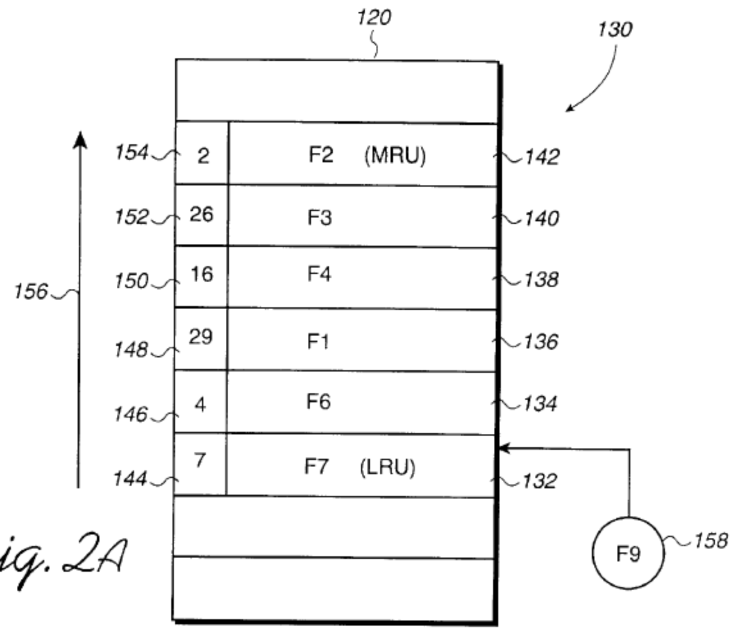

Figure 2A depicts a structure of a cache according to an embodiment of the invention. *Id.* at 4:1–3.

Cache **120** includes files F1 through F7. *Id.* at 5:55–56. The term “file” as used in the ’745 patent “refers to a number of blocks of data, i.e., any group of blocks of data.” *Id.* at 5:60–62. File F7 is the least recently used (LRU) file. *Id.* at 5:56–57. “In other words,” the ’745 patent explains, “file F7 **132** has gone the longest time as compared to the other files in the cache without being used by the operating system.” *Id.* at 5:57–60. File F2 is the most recently used (MRU) file. *Id.* at 5:63. Put differently, “file F2 **142** has gone the shortest time without being used by the operating system.” *Id.* at 5:64–65.

As illustrated in Figure 2A, above, a frequency factor is assigned to each file in the cache. *Id.* at 5:65–66. File F7 has a frequency factor of seven (**144**), file F6 has a frequency factor of four (**146**), file F1 has a frequency factor of twenty-nine (**148**), and so on. *Id.* at 6:4–8. For one embodiment, the frequency factor "is increased with each use of the file"; in another embodiment, the frequency factor "is decreased with the lack of use of the corresponding file." *Id.* at 5:66–6:4.

Next, "once the target capacity of the cache is reached, the frequency factors are scanned, beginning with the frequency factor for the LRU file in the cache and ending with the MRU file in the cache." *Id.* at 6:8–12. Specifically, "the scanning begins with frequency factor **144** and ends with frequency factor **154**" because those frequency factors "correspond to the LRU and the MRU files, respectively." *Id.* at 6:14–17. Arrow **156** shows the "scanning direction from the LRU file to the MRU file." *Id.* at 6:17–18. The scanning of the frequency factors from the LRU file to the MRU file "determines which file . . . is the least *frequently* used (LFU) file which has been least *recently* used." *Id.* at 6:18–21 (emphases added).

"It should be appreciated," the '745 patent states, "that the least frequently used file takes into consideration how *often* a file has been used through the frequency factor." *Id.* at 6:21–24 (emphasis added). Moreover, "how *recently* a file has been used is also taken into consideration" when deciding which file to eliminate from the cache. *Id.* at 6:25–28 (emphasis added). And, in one embodiment, "the frequency factor for the MRU file is not considered when determining the LFU file" because "the MRU file has been so recently placed in the cache and has not had an opportunity to be reused." *Id.* at 6:39–43.

As illustrated in Figure 2A, once cache capacity is reached, the frequency factors are scanned to find the least frequently used (LFU) file that is the least recently used (LRU) file. *Id.* at 6:44–55. File F6 **134** is the

LRU file that has the lowest frequency factor of all the files, i.e., a frequency factor of four (**146**), that is not the most recently used (MRU)[1] file. *Id.* at 6:46–52. Thus, file F6 is the LFU file that has been the LRU file. *Id.* Once identified, file F6 is discarded and a new file, file F9 **158**, may be placed in the cache to replace it. *Id.* at 6:52–55.

The ’745 patent also describes implementing the caching mechanism "in conjunction with reading extending segments of data." *Id.* at 4:32–36. The ’745 patent explains that, "by performing large reads, the data will be present in the cache in order to minimize seeks and reads from the hard drive." *Id.* at 3:49–52. Specifically, "large reads (greater than 64 Kbyte) eliminate the overhead, i.e., seek times, rotational latencies, transfer times, etc., associated with performing two, four or more reads at 32 Kbytes or less as performed in the prior art." *Id.* at 5:38–42. For example, the ’745 patent states that, "when the first block of the file is read[,] an additional 64, 128 or 256 Kbytes of data are read with it." *Id.* at 11:23–26. The ’745 patent further states that "the reads of the extended segments allows for the minimization of seeks and reads from the storage medium, since the files are transferred to cache upon the initial read." *Id.* at 11:45–48.

---

[1] As explained above, file F2 **142** is not considered for elimination in the hybrid caching mechanism of one embodiment disclosed in the ’745 patent. *Id.* at 6:39–43. Even though file F2 has the lowest frequency factor of any file in the cache (**154**), it also is the most recently used. Thus, the ’745 patent states, "the MRU file is not chosen as the LFU file in order to prevent a thrashing situation." *Id.* at 6:49–52.

C. Challenged Claims

Claims 4, 5, and 6 are challenged and read as follows:

> 4. A caching method for enhancing system performance of a computer, comprising:
>
> reading an extended segment of data in response to a request from an operating system;
>
> storing copies of files associated with the extended segment in a cache;
>
> assigning frequency factors to each of the files stored in the cache, the frequency factors indicating how often each of the corresponding files are requested by the operating system;
>
> scanning the frequency factors, the scanning being performed in response to a target capacity of the cache being attained;
>
> identifying a least frequently and least recently used file; and
>
> eliminating the least frequently and least recently used file to liberate capacity of the cache.
>
> 5. The method as recited in claim 4, wherein the files are read from a storage medium containing one of a directory, a file allocation table and a data file.
>
> 6. The method as recited in claim 4, wherein the scanning the frequency factors further includes:
>
> scanning from a frequency factor corresponding to a LRU file to a frequency factor corresponding to a MRU file.

Id. at 12:54–13:11.

D. Asserted Grounds of Unpatentability

Petitioner contends that claims 4, 5, and 6 of the ’745 patent are unpatentable based on the following specific grounds:

References	Challenged Claim(s)	Basis
Burton[2] and Karedla[3]	4–6	35 U.S.C. § 103(a)[4]
Burton and Sweeney[5]	4–6	35 U.S.C. § 103(a)
Lee,[6] Dharap,[7] and Karedla	4 and 5	35 U.S.C. § 103(a)
Lee, Dharap, Karedla, and Robinson[8]	6	35 U.S.C. § 103(a)
Lee, Dharap, and Sweeney	4 and 5	35 U.S.C. § 103(a)
Lee, Dharap, Sweeney, and Robinson	6	35 U.S.C. § 103(a)

Pet. 20–66. In its analysis, Petitioner relies on the declaration testimony of John D. Kubiatowicz, Ph.D. Ex. 1002.

---

[2] David Alan Burton & Erez Webman, U.S. Patent No. 6,738,865 B1 (issued May 18, 2004) ("Burton"). Ex. 1005.

[3] Ramakrishna Karedla et al., *Caching Strategies to Improve Disk System Performance*, COMPUTER, 27(3):38–46 (March 1994) ("Karedla"). Ex. 1006.

[4] The Leahy-Smith America Invents Act ("AIA"), Pub. L. No. 112-29, 125 Stat. 284, 287–88 (2011), revised 35 U.S.C. § 103, effective March 16, 2013. Because the challenged patent was filed before March 16, 2013, we refer to the pre-AIA version of § 103 in this decision.

[5] Adam Sweeney et al., *Scalability in the XFS File System*, 1996 USENIX TECHNICAL CONFERENCE (Jan. 22–26, 1996) ("Sweeney"). Ex. 1012.

[6] Donghee Lee et al., *Implementation and Performance Evaluation of the LRFU Replacement Policy*, IEEE (1997) ("Lee"). Ex. 1003.

[7] Chanda Dharap, U.S. Patent Application Publication No. US2002/0078300 A1 (published June 20, 2002) ("Dharap"). Ex. 1004.

[8] As presented by Petitioner, "Robinson" refers collectively to two references: (1) John T. Robinson & Murthy V. Devarakonda, *Data Cache Management Using Frequency-Based Replacement*, PERFORMANCE EVALUATION REV., 18(1):134–142 (May 1990) ("the Robinson article"), Ex. 1007, and (2) James T. Robinson, U.S. Patent No. 5,043,885 (Aug. 27, 1991) ("Robinson '885"), Ex. 1008.

II. DISCUSSION

As noted above, the ’745 patent is currently the subject of *inter partes* review IPR2016-01643 (“the 2016 IPR”). The Petitioner in the 2016 IPR, Unified Patents Inc., challenged certain claims of the ’745 patent on six grounds of unpatentability. Patent Owner argues that grounds of unpatentability raised in this proceeding present the same (or substantially the same) prior art and the same (or substantially the same) arguments as those presented to the Board in the 2016 IPR, or those considered by the Examiner during prosecution of the application leading to the ’745 patent. Prelim. Resp. 10–30. Thus, Patent Owner argues, we should invoke our discretion under 35 U.S.C. § 325(d) and deny the instant petition in its entirety. Prelim. Resp. 9–30. For the reasons that follow, we agree, in part, with Patent Owner as to Petitioner’s challenge based on Burton and Karedla. But, for the remaining grounds, we decline to exercise our discretion under § 325(d).

A. Alleged Obviousness over Burton in view of Karedla

Petitioner contends that claims 4–6 are unpatentable under 35 U.S.C. § 103(a) as obvious over Burton in view of Karedla. Pet. 20–35. Patent Owner argues that we should deny this challenge under § 325(d). Prelim. Resp. 12–14.

Institution of an *inter partes* review is discretionary. *See* 35 U.S.C. § 314(a) (authorizing institution of an *inter partes* review under particular circumstances, but not requiring institution under any circumstances); 37 C.F.R. § 42.108(a) (“the Board *may* authorize the review to proceed” (emphasis added)); *Harmonic Inc. v. Avid Tech., Inc.*, 815 F.3d 1356, 1367 (Fed. Cir. 2016) (explaining that under § 314(a), “the PTO is permitted, but never compelled, to institute an [*inter partes* review] proceeding"). Even so, we have express discretion under § 325(d) to reject a petition when the same or substantially the same prior art or arguments were presented previously in another proceeding before the Office. Specifically, "[i]n determining whether to institute or order a proceeding under this chapter, chapter 30, or chapter 31, the Director may take into account whether, and reject the petition or request because, the same or substantially the same prior art or arguments previously were presented to the Office." 35 U.S.C. § 325(d).

Petitioner's first challenge presents prior art identical to that presented in the 2016 IPR. *See* Prelim. Resp. at 12–14. Specifically, the petitioner in the 2016 IPR, Unified Patents, alleged as "Ground 2" that the combination of Burton and Karedla rendered obvious claims 1, 4, 6–9, and 11–17. Ex. 2002, 3 (IPR2016-01643, Paper 2). The Board instituted an *inter partes* review on that ground, but only as to claims 1, 4, and 12. IPR2016-01643, Paper 9. Here, Petitioner contends that claims 4–6 are unpatentable as obvious over Burton in view of Karedla. Pet. 20–35. Thus, as Patent Owner correctly notes, both the first challenge of this petition and "Ground 2" of the 2016 IPR rely on the same prior art. Prelim. Resp. 12.

Moreover, as Patent Owner further explains, the instant Petition presents substantially the same arguments as those presented to the Board in the 2016 IPR. *Id.* at 13. Specifically, in the 2016 IPR, petitioner Unified Patents relied on almost identical passages in Burton and Karedla to teach each element of representative claim 4. Ex. 2002, 39; *see also* Prelim. Resp. at 12–13. The only exception, as Patent Owner explains, is that petitioner Unified Patents relied on Burton to disclose the "storing copies of files" limitation of claim 4, whereas Petitioner in this case relies on Karedla. *See*

Prelim. Resp. at 13 (comparing Pet. 27, with Ex. 2002, 40–43). We agree with Patent Owner, however, that Petitioner's arguments represent substantially the same arguments under § 325(d). *Id.* at 13–14.

Because this challenge presents "the same or substantially the same prior art or arguments previously . . . presented to the Office," we exercise our discretion under § 325(d) to deny institution of an *inter partes* review based on this ground as to claims 4 and 6.

As to claim 5, we note that the Board's rules for *inter partes* review proceedings, including those pertaining to institution, are "construed to secure the just, speedy, and inexpensive resolution of every proceeding." 37 C.F.R. § 42.1(b); *see also* 35 U.S.C. § 316(b) (regulations for *inter partes* review proceedings take into account "the efficient administration of the Office" and "the ability of the Office to timely complete [instituted] proceedings"). Although claim 5 was not specifically challenged by the Petitioner in the 2016 IPR on the basis of Burton and Karedla, in view of our instituting *inter partes* review as to claim 5 on another ground, discussed below, we exercise our discretion and do not authorize *inter partes* review of claim 5 based on the combination of Burton and Karedla. *See* 37 C.F.R. § 42.108(a).

B. Alleged Obviousness Over Various Prior Art References Including Sweeney

The Petition presents several obviousness challenges to claims 4–6 of the '745 patent over various combinations of prior art references including Sweeney. Specifically, Petitioner challenges (1) claims 4–6 as unpatentable over Burton in view of Sweeney, Pet. 35–43; (2) claims 4 and 5 as unpatentable over Lee in view of Dharap and Sweeney, *id.* at 62–65; and (3) claim 6 as unpatentable over Lee in view of Dharap, Sweeney, and Robinson, *id.* at 65–66. Patent Owner argues that we should deny these challenges under § 325(d) because "Sweeney is duplicative of Karedla," and because these challenges essentially repeat challenges "already under consideration in IPR2016-01643." Prelim. Resp. 14, 22.

We find it unnecessary to analyze the application of § 325(d) to these challenges because we exercise our discretion to deny their institution under § 314(a). Here, Petitioner proposes the grounds involving Sweeney as alternative unpatentability challenges "[i]n the event that the Patent Owner argues that the term ['extended segment of data'] should be construed . . . [as] something to connote a 'large block of data.'" Pet. 36. Patent Owner, however, represents that it "does not and will not contend that 'an extended segment of data' connotes a 'large block of data' under the broadest reasonable interpretation standard." Prelim. Resp. 8. Thus, because Petitioner's claim construction contingency has not materialized, we do not institute trial on these alternative grounds of unpatentability.

C. Alleged Obviousness over Lee in view of Dharap and further in view of Karedla

Petitioner contends that claims 4 and 5 are unpatentable under 35 U.S.C. § 103(a) as obvious over Lee, Dharap, and Karedla. Pet. 43–61. Relying in part on the testimony of Dr. Kubiatowicz, Petitioner explains how the references allegedly teach or suggest the claim limitations and provides purported reasoning for combining the teachings of the references. *Id.* (citing Ex. 1002 ¶¶ 137–70).

1. 35 U.S.C. § 325(d)

Patent Owner argues that we should deny this challenge under § 325(d) because (1) Lee and Dharap were before the Examiner during prosecution of the application leading to the ’745 patent; (2) “Lee is a duplicative stand-in for Burton”; and (3) this challenge “is essentially identical to the Burton/Karedla combination already under consideration in IPR2016-01643.” Prelim. Resp. 16, 19.

During prosecution, the Examiner rejected certain original claims under 35 U.S.C. § 103(a) for obviousness over Lee in view of Dharap. Ex. 1009, 7. The Examiner relied on the combination of Lee and Dharap to teach all the limitations of the original independent claims except for, *inter alia*, the limitation “reading an extended segment of data.” *Id.* at 7–11. The Examiner noted that certain dependent claims reciting that feature “would be allowable if rewritten in independent form” because the claim limitations, “when taken in combination with the remaining limitations, are not found in and are not obvious in view of, the prior art of record.” *Id.* at 11. In response, the Applicants amended the claims, including what is now independent claim 4, to recite “reading an extended segment of data.” Ex. 1010, 11–12. Applicants explained that they amended the claims “to incorporate features indicated as allowable by the Examiner.” *Id.* at 18. “In particular,” Applicants stated, those claims “have incorporated the feature of reading extended segments of data.” *Id.* The Examiner allowed the claims without further comment. Ex. 1011.

Although the Examiner considered Lee and Dharap during prosecution, we decline to deny institution of this challenge under § 325(d) due to the particular circumstances of this case. Here, Petitioner asserts that the combination of Lee, Dharap, and Karedla would have rendered claims 4 and 5 obvious. The Examiner did not reject claims based on that particular combination of references. And, as explained below, we find Petitioner presents a reasonable likelihood that it would prevail in showing that Karedla teaches the feature of "reading extended segments of data." *See infra* at § II.C.3. Petitioner's obviousness challenge also relies on the testimony of its witness, Dr. Kubiatowicz; Patent Owner does not argue that this testimony is duplicative of any evidence previously submitted to the Office. Under these circumstances, therefore, we decline to exercise our discretion to deny institution under § 325(d) even though Lee and Dharap were previously considered by the Office.

2. The prior art

a. Lee

Lee is directed to the implementation and evaluation of a caching mechanism utilizing a "Least Recently/Frequently Used" method for eliminating files from the cache ("the LRFU replacement policy"). Ex. 1003, Abstract. As background, Lee teaches that the "speed gap between processors and disks is becoming wider" and that "[o]ne solution to overcome this speed gap is to use a caching technique which keeps disk blocks that are likely to be accessed in the future in DRAM memory." *Id.* at 106. But, Lee explains, "[s]ince the buffer cache size is necessarily limited, an effective scheme is needed for the block replacement policy that decides which block should be replaced." *Id.*

Lee discloses that the "[p]revious research on block replacement policies [for caches] can be divided into two groups." *Id.* The "Least Frequently Used (LFU)" policy is an example of "policies [that] base their replacement decision on the frequency of references." *Id.* And the "Least Recently Used (LRU)" policy is an example of "policies that base their replacement decision" "on the recency of references." *Id.* Lee teaches that, "[r]ecently, a new block replacement policy called the LRFU (Least Recently/Frequently Used) policy was proposed that subsumes both the LRU and LFU policies, and provides a spectrum of replacement policies between them." *Id.*

Lee explains that the LRFU replacement policy "associates a value" "called the CRF (Combined Recency and Frequency) value" "with each block" that "quantifies the likelihood that the block will be referenced in the near future." *Id.* at 107. "Each reference to a block in the past adds its contribution to this [CRF] value and its contribution is determined by [a] *weighing function* . . . ." *Id.* The weighing function includes "a control parameter that determines a trade-off between recency and frequency." *Id.* In other words, the control parameter controls the relative contributions of recency and frequency to the CRF values. *See id.* "The LRFU policy replaces the block whose CRF value is the smallest in the cache." *Id.*

b. Dharap

Dharap teaches "information processing systems that utilize cache memory to minimize latency." Ex. 1004 ¶ 2. Dharap provides as background a description of cache systems. "Traditionally," Dharap explains, "cache memory is filled with copies of information resources that a user receives from a remote source, 'remote' being defined as being further removed from the user than the cache memory, e.g., local main memory or a server in a client-server architecture." *Id.* ¶ 4. Dharap describes the time-saving benefit of this system: "If the user subsequently requests the same resource, the resource's copy is provided from the cache memory, rather than from the original remote source, thereby saving the time required to receive the resource from the remote source for a second time." *Id.*

Dharap explains, that "[w]hen the cache memory becomes full," however, the cache controller must "make room for copies of new resources that the user accesses." *Id.* Dharap then explains that "[a] variety of criteria, commonly termed caching policies, are available to determine which resource copy to remove from cache memory" to make room for new information. *Id.* "Such caching policies can be based on: the duration since the last access, the number of times accessed since originally received, the amount of memory allocated to the resource, the difficulty of retrieving the resource from the remote site, etc." *Id.*

Dharap teaches that "the cache controller in a conventional system also removes copies of resources from the cache memory when it is predicted or determined that the source information has changed, because the copy of the resource in cache memory is outdated, or 'stale.'" *Id.* ¶ 5. Dharap points to Internet content as one example of information that may become outdated. *Id.*

Dharap discloses "a cache system that caches copies of resources based on the *semantic type* of the resource." *Id.* ¶ 9 (emphasis added). Dharap explains:

> The expression "semantic type" as used within this context refers to the different connotative meaning that the information contents of resources can have, as perceived by the user. For example, some information content may be perceived as highly volatile (e.g., being of short term relevance such as web sites dedicated to the results of sports matches, to specific stock market news or currency exchange rates), other information content may be perceived as rather static (e.g., being of long-

> term relevance such as glossaries on the Internet). Semantic types that can be expected to contain dynamic information, such as news Web sites and weather Web sites, need a caching policy wherein the copy in the cache memory is selected for replacement based upon the duration of time that the copy has been in the cache memory. Conversely, semantic types that can be expected to relate to static resources, such as encyclopedic information, glossaries, etc., need a more conservative caching policy, such as least-recently-used (LRU) or least-frequently-used (LFU), that are substantially independent of the time duration that the copy remains in the cache memory.

Id. Dharap also teaches that "a combination of caching policies" may be employed for certain types of resources. *Id.*

c. Karedla

Karedla provides an overview of caching strategies that "increase system response time and improve the data throughput of the disk subsystem." Ex. 1006, 38. In particular, Karedla "examine[s] some popular caching strategies and cache replacement algorithms." *Id.*

Karedla explains that "[t]he chief metric for cache performance is its *miss rate*." *Id.* at 39 (emphasis added). The miss rate represents the percentage of all I/O requests not found in the cache. *Id.* Typically, Karedla explains, "[a]n I/O request to a storage device, especially a read request, searches the cache first." *Id.* "A request is said to be a *cache hit* when all of the requested data is found in the cache," but a "request is said to *miss the cache* when any of its data blocks are not found in the cache." *Id.* One cache-design parameter that "has a profound impact on cache performance in terms of miss rate," Karedla explains, is the "line replacement algorithm." *Id*. Karedla notes that vendors attempt to design "efficient replacement algorithms that can offer higher cache performance for relatively small cache sizes." *Id.*

Karedla also notes that "[c]aches work on the premise that data in the cache will be reused often, thus reducing the number of accesses to the backing store." *Id.* "To achieve this," Karedla explains, "caches exploit the principles of spatial and temporal locality of reference." *Id.* Karedla continues:

> Spatial locality implies that if an object is referenced, then nearby objects will also soon be accessed. Temporal locality implies that a referenced object will tend to be referenced again in the near future.

Id.

Karedla then introduces "[a] read-ahead strategy known as prefetching" that "exploits the principle of spatial locality." *Id.* at 40. Karedla explains that prefetching "minimize[s] latency in data access by anticipating future requests for data and bringing it into the cache." *Id.* "Most disk drives," Karedla notes, "implement prefetch in on on-board cache." *Id.* "However, a large prefetch can have a negative impact on small caches, because it can displace data that would have been useful in the cache." *Id.*

Karedla explains that prefetching relies on the host caches' ability to "predict the type of data most needed in the near future." *Id.* Karedla discloses as an example "files system caches that prefetch the rest of a file that is being read." *Id.* Karedla further teaches that prefetching "is usually done when no I/O request is pending and is typically aborted when a user request arrives." *Id.*

3. Analysis

Having reviewed the record, we determine that Petitioner has shown sufficiently for the purpose of institution that the combination of Lee, Dharap, and Karedla discloses each limitation of claims 4 and 5. *See id.* at 46–52; *see also* Ex. 1002 ¶¶ 143–60. We find that Petitioner reasonably relies on Lee and Dharap for teaching all limitations of claims 4 and 5, except for "reading an extended segment of data in response to a request from an operating system." Pet. 46–52.

For that limitation, Petitioner reasonably points to Karedla, which, Petitioner contends, describes "prefetching, a read-ahead strategy, that exploits spatial locality." *Id.* at 26 (citing Ex. 1006, 40; Ex. 1002 ¶ 91). For example, as noted above, Karedla discloses "files system caches that prefetch the rest of a file that is being read." Ex. 1006, 40.

We also find, at this stage of the proceeding, that Petitioner has set forth sufficient reasoning with rational underpinning for combining the teachings of the references to meet the "reasonable likelihood" standard for instituting trial. We note that Petitioner contends that "[m]odifying Lee to evict a cache entry when the cache memory becomes full or reaches capacity" based on the teachings of Dharap "would have been a simple design choice" to an ordinarily skilled artisan. Pet. 54–55 (citing Ex. 1002 ¶ 165). We also note that Petitioner also asserts that an ordinarily skilled artisan "would have been motivated to combine Lee and Dharap with Karedla to implement prefetching in Lee's cache architecture, to further decrease latency and increase performance in a computer architecture system." *Id.* at 55–56 (citing Ex. 1002 ¶ 167). Petitioner points out that Lee discloses a read-ahead feature and that combining Lee and Karedla "would merely require enabling th[at] existing read-ahead feature." *Id.* at 57 (citing Ex. 1002 ¶ 170).

Patent Owner does not address the merits of these contentions. Based on the record and information before us at this time, we are persuaded to go forward with a trial on claims 4 and 5 as unpatentable under 35 U.S.C. § 103(a) for obviousness over Lee, Dharap, and Karedla. The burden is on Petitioner to demonstrate by a preponderance of the evidence that the claims are unpatentable. 35 U.S.C. § 316(e); 37 C.F.R. § 42.1(d). And that burden of persuasion never shifts to Patent Owner. *See Dynamic Drinkware, LLC v. Nat'l Graphics, Inc.*, 800 F.3d 1375, 1378 (Fed. Cir. 2015); *see also In re Magnum Oil Tools Int'l, Ltd.*, 829 F.3d 1364, 1375–78 (Fed. Cir. 2016) (discussing the burden of proof in *inter partes* review). We emphasize that, at this preliminary stage, the Board has not made a final determination with respect to the patentability of the challenged claims or any underlying factual and legal issues.

4. Summary

For the above reasons, we are satisfied at this stage of the proceeding that the information presented establishes a reasonable likelihood that Petitioner would prevail in showing that claims 4 and 5 are unpatentable under 35 U.S.C. § 103(a) as obvious over Lee in view of Dharap and Karedla.

D. Alleged Obviousness over Lee in view of Dharap, Karedla, and further in view of Robinson

Petitioner contends that claim 6 is unpatentable under 35 U.S.C. § 103(a) as obvious over Lee in view of Dharap, Karedla, and further in view of Robinson. Pet. 57–61. Relying in part on the testimony of Dr.

Kubiatowicz, Petitioner explains how the references allegedly teach or suggest the claim limitations and provides purported reasoning for combining the teachings of the references. *Id.* (citing Ex. 1002 ¶¶ 172–82)

1. 35 U.S.C. § 325(d)

As an initial matter, we decline to exercise our discretion to deny this ground under § 325(d) for the same reasons explained above for the challenges based on Lee, Dharap, and Karedla. *See supra* § II.C.1.

2. The prior art

a. Robinson ’885

Robinson ’885 discloses a method and apparatus “for making cache block replacement decisions based on a combination of least recently used (LRU) stack distance and data reference frequencies.” Ex. 1008, 1:11–14.

Figure 1 of Robinson ’885 is shown below:

FIG. 1

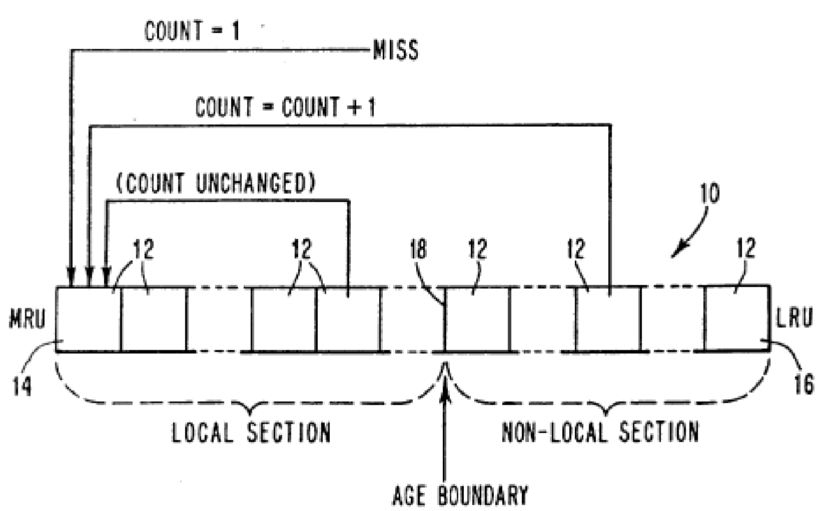

Figure 1 illustrates “how a combination of block aging, boundary condition and reference count techniques may be used.” *Id.* at 3:37–39.

Robinson ’885 teaches that each cache directory entry **12** in cache directory **10** has a “reference count” associated with it. *Id.* at 4:24–26. According to Robinson ’885, when a new block is brought into the cache, the associated cache directory entry is placed into most recently used (MRU) position **14** and the reference count is initialized to one. *Id.* at 4:24–31. Robinson ’885 discloses that “the cache directory essentially works in LRU fashion, with a cache directory entry being put in the MRU position each time a block is referenced.” *Id.* at 4:34–37. However, according to Robinson ’885, “the block associated with the cache directory entry in the LRU position **16** will not necessarily be the one that is replaced when there is a miss.” *Id.* at 4:37–40.

As shown in Figure 1, the cache directory has age boundary **18**. The section of the cache directory “on the MRU side of the boundary” is the “local” section; the section on the other side of the boundary is the “non-local” section. *Id.* at 4:46–47. Robinson ’885 discloses that “when there is a hit on the local section, the count remains the same; when there is a hit on the non-local section, the count is incremented.” *Id.* at 4:50–52. Robinson ’885 discloses ways of using the reference counts to select blocks to replace, including “select[ing] the least recently used block in the non-local section whose count is below a preselected threshold.” *Id.* at 4:55–57.

Figure 2 of Robinson ’885, shown below, illustrates an alternative embodiment in which the cache directory is divided into three sections using two boundaries:

FIG. 2

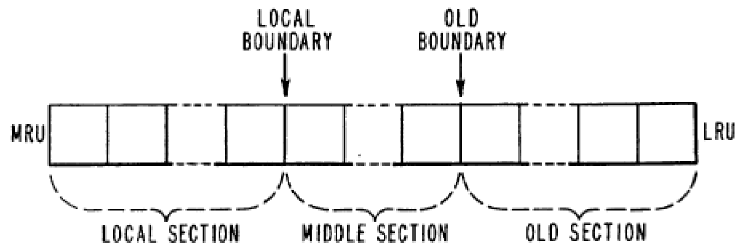

Figure 2, above, shows a cache directory having a local section, a middle section, and an old section. *Id.* at 4:66–5:1.

Robinson ’885 discloses that “[b]locks with reference counts in the old section are the blocks from which replacement selections are made.” *Id.* at 5:1–3. Specifically, Robinson ’885 discloses that “a block is selected to be replaced by finding the block (or blocks) with the smallest count in the old section and then replacing that block (or least recently used such block if there is more than one) if the count is below a predetermined threshold.” *Id.* at 5:4–8. Robinson ’885 further describes maintaining a separate LRU chain for blocks with a count of 1, another separate LRU chain for blocks with a count of 2, and so on. *Id.* at 5:10–14. Robinson ’885 discloses finding a block to replace by “scanning the blocks (from LRU to MRU) in each such LRU chain (in ascending count order) until a block is found in the old section.” *Id.* at 5:15–18.

b. The Robinson article[9]

The Robinson article discloses a cache replacement algorithm based on “a combination of reference frequency and block age.” Ex. 1007, 9. The

[9] Although Petitioner includes Robinson in the patentability challenge, Petitioner does not rely on the Robinson article for teaching any particular limitation of claim 6. Pet. 59.

Robinson article states that "[l]ike previously studied algorithms, reference counts are maintained for each block in the cache; unlike previous algorithms, reference counts are not incremented on every reference." *Id.* at 10. The article discloses that "references are incremented only for those blocks that have aged out of a 'new section' of the cache." *Id.*

Figure 2.2 of the Robinson article, below, illustrates the sections of the cache:

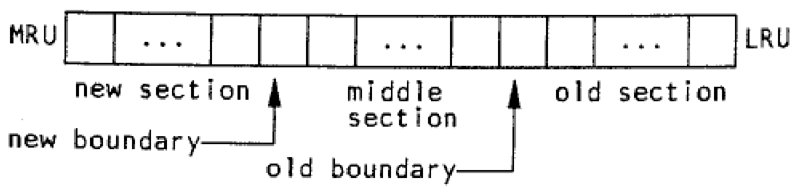

Figure 2.2. Three Sections of the Cache

> Figure 2.2 illustrates the three sections of cache: a new section, a middle section, and an old section. *Id.* at 11.

The Robinson article discloses that the basis of its algorithm is the LRU replacement algorithm, in which the cache consists of a stack of blocks, with the most recently referenced block always pushed to the top of the stack. *Id.* at 10. The Robinson article discloses that a certain portion of the top part of the stack is set as aside as the new section, a bottom part of the stack is defined as the old section, and the remaining part of the stack between the new and old sections is the middle section. *Id.* at 11. According to the Robinson article, "replacement choices are confined to those blocks that have aged into an 'old section' of the cache." *Id.* at 10.

3. Analysis

Claim 6 depends from claim 4 and further recites "scanning from a frequency factor corresponding to a LRU file to a frequency factor corresponding to a MRU file." Ex. 1001, 13:8–11. Petitioner asserts that Robinson '885 teaches the claimed scanning by disclosing that "[f]inding a block to replace . . . consists of scanning the blocks (from LRU to MRU) in each such LRU chain (in ascending count order) until a block is found in the old section." Pet. 59 (quoting Ex. 1008, 5:15–18); *see also* Ex. 1002 ¶ 177 (same).

Having reviewed the record, we determine that the information presented does not establish a reasonable likelihood that Petitioner would prevail in showing that claim 6 is unpatentable under 35 U.S.C. § 103(a) as obvious over Lee in view of Dharap, Karedla, and further in view of Robinson. We find that Petitioner fails to explain sufficiently how the quoted passage from Robinson '885 teaches the scan required by the claims—scanning from a frequency factor corresponding to a LRU file to a frequency factor corresponding to a MRU file, either including or not including the frequency factor for the MRU file—given that Robinson '885 discloses that the scan ends upon finding a block "in the old section."

Robinson '885 discloses that the old section is only part of the entire cache, as shown in Figure 2 below:

FIG. 2

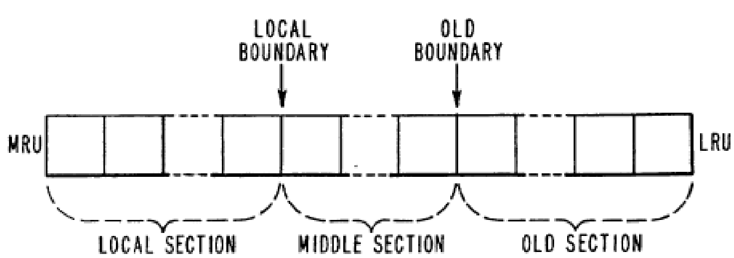

Figure 2, above, illustrates a cache including three sections: a local section, a middle section, and an old section. Ex. 1008, 4:64–5:1. Robinson discloses—in a sentence preceding the one quoted by Petitioner—"maintaining a separate LRU chain for blocks with a count of **1**, another separate LRU chain for blocks with a count of **2**, and so forth up to the preselected threshold count." *Id.* at 5:10–14. Given that, according to Robinson '885, the scan proceeds "from LRU to MRU" in each such LRU chain "until a block is found in the old section," blocks in the middle section and local section are not scanned as part of the cited operation, and, thus, the scan will not proceed to a frequency factor corresponding to "a MRU file," as required by claim 6. *See id.* at 5:10–18; *see also id.* at 5:1–3 ("Blocks with reference counts in the old section are the blocks from which replacement selections are made."). Petitioner fails to provide sufficient explanation why Robinson '885 would have taught this claim requirement to a person of ordinary skill in the art in view of its express disclosure that the scanning is up until a block is found in the old section. *See* Pet. 59. Dr. Kubiatowicz simply repeats the statements from the Petition without providing any further explanation and thus, fails to help Petitioner carry its burden. *See* Ex. 1002 ¶ 177.

c. Summary

For the above reasons, we determine that the information presented does not establish a reasonable likelihood that Petitioner would prevail in showing that the subject matter of claim 6 is unpatentable under 35 U.S.C. § 103(a) as obvious over Lee, Dharap, Karedla, and Robinson.

E. Word Limit

Throughout its Petition, Petitioner uses atypical citation formats. For example, Petitioner cites to exhibits by combining an abbreviation for the word "exhibit" with the particular exhibit number (e.g., "Ex1002") and omits a space between the paragraph symbol and the paragraph number (e.g., "¶24"). Petitioner's citations to the Robinson references are also unusual: an underscore connects "Robinson" with a Roman numeral (i.e., "Robinson_I" and "Robinson_II") to differentiate between Robinson '885 and the Robinson article. By this approach, Petitioner creates one word for what would be counted as two if standard citation formats were used instead. For example, Ex. 2002, ¶ 24, Robinson I, and Robinson II count as nine words using word processing software, whereas Ex2002, ¶24, Robinson_I, and Robinson_II count as five. Nevertheless, Petitioner's lead counsel certified, pursuant to 37 C.F.R. § 42.24(d), that the word count for the Petition is 13,939. Pet. i.

Patent Owner states that, "[b]ecause more than 370 of these non-standard citations appear in the Petition, the Petition is over the word limit by about 300 or more words." Prelim. Resp. 35. Patent Owner also points out that Petitioner did not file a motion seeking to waive the word count. *Id.* at 36–37. Patent Owner argues that "the Board . . . should require all parties to count two-word phrases—regardless of how they are formatted—as two words." *Id.* at 36.

We agree with Patent Owner that Petitioner's use of atypical citations amounts to formatting tricks designed to avoid the word count limit for petitions set forth in our rules. *See* 37 C.F.R. § 42.24(a)(1)(i) (providing a 14,000-word count limit for a petition requesting an *inter partes* review). In our view, the appropriate remedy at this point in the proceeding is to waive the word count limit for Patent Owner's Response. *See* 37 C.F.R. § 42.24(b)(2) (providing a 14,000-word count limit for a petition requesting an *inter partes* review). Specifically, Patent Owner may, in its Response, exceed the 14,000-word count limit by 370 words, the number of non-standard citations Patent Owner identifies in the Petition. *See* 37 C.F.R. § 42.5(b) (stating that the Board "may waive or suspend a requirement of parts 1, 41, and 42").

We expect Petitioner to use common citation formats going forward.

III. CONCLUSION

For the above reasons, we determine that the information presented establishes a reasonable likelihood that Petitioner would prevail in showing that claims 4 and 5 are unpatentable under 35 U.S.C. § 103(a) as obvious over Lee, Dharap, and Karedla. We do not institute trial, however, as to claim 6 on any ground for the reasons explained above.

IV. ORDER

Accordingly, it is:

ORDERED that pursuant to 35 U.S.C. § 314(a), an *inter partes* review is instituted as to claims 4 and 5 of the '745 patent on the ground of obviousness over Lee, Dharap, and Karedla;

FURTHER ORDERED that no other ground set forth in the Petition is authorized for *inter partes* review; and

FURTHER ORDERED that pursuant to 35 U.S.C. § 314(c) and 37 C.F.R. § 42.4, notice is hereby given of the institution of a trial, which commences on the entry date of this decision.

FOR PETITIONER:

Peter Dichiara
Theodoros Konstantakopoulos
WILMERHALE
peter.dichiara@wilmerhale.com
theodoros.konstantakopoulos@wilmerhale.com

Christopher Centurelli
Benjamin Weed
K&L GATES LLP
christopher.centurelli@klgates.com
benjamin.weed.ptab@klgates.com

Thomas Brown
tom.brown@emc.com

FOR PATENT OWNER:

John R. King
Ted M. Cannon
KNOBBE, MARTENS, OLSON & BEAR, LLP
2jrk@knobbe.com
2tmc@knobbe.com

Tim Seeley
James Hietala
INTELLECTUAL VENTURES
tims@intven.com
jhietala@intven.com