

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC,
Petitioner,

v.

SANDPIPER CDN, LLC,
Patent Owner.

Case IPR2025-00860
Patent No. 10,924,573

PETITIONER'S REPLY TO PATENT OWNER'S RESPONSE

TABLE OF CONTENTS

I.	Introduction.....	1
II.	Ground 1 Renders Claims 1-20 Unpatentable.....	2
	A. Newton-471 Anticipates Claims 1-20.....	2
	B. Patent Owner’s Contrary Arguments Should Be Rejected.....	3
III.	Ground 2 Renders Claims 1-20 Unpatentable.....	6
	A. A Skilled Artisan Would Have Been Motivated to Combine Dilley, Pai, and Wang	6
	1. Patent Owner’s Arguments Misread Dilley and Pai.....	7
	2. The Petition Aligns with Petitioner’s Litigation Construction	8
	3. The Combination Does Not Render Dilley or Pai Inoperable.....	9
	B. Dilley, Pai and Wang Render Obvious the Claimed “popularity designation” (claims 1, 11, 19).....	10
	1. Patent Owner Improperly Imports a Limitation	11
	C. Dilley, Pai and Wang Render Obvious Claim 19’s Second Server Request “based at least on the popularity designation”.....	13
	1. The Art Discloses This Feature; There Is No Gap-Filling	14
	2. A Skilled Artisan Would Have Combined Dilley and Pai	16
	D. Dilley, Pai, and Wang Render Obvious the Claimed “redirect instruction” (claims 1, 11).....	16
	1. The Petition Presents One Combination.....	16
	2. Patent Owner Argues the References Individually, Ignoring That Wang Supplies a Known Alternative	18
	3. Wang Does Not Inject Redundant Architecture	19

4.	Wang’s Redirect Mechanism Offers Predictable Benefits	20
E.	Dilley, Pai, and Wang Render Obvious the Claimed “peer server” (claims 5, 15)	22
F.	Dilley, Pai, and Wang Render Obvious the Claimed “receiving an instruction” (claim 3)	24
IV.	Conclusion	25

TABLE OF AUTHORITIES

	Page(s)
Federal Cases	
<i>Am. Airlines, Inc. v. Intellectual Ventures I LLC</i> , IPR2025-01055, Paper 11 (P.T.A.B. Nov. 21, 2025).....	25
<i>Ariad Pharms., Inc. v. Eli Lilly & Co.</i> , 598 F.3d 1336 (Fed. Cir. 2010)	5
<i>Axonics, Inc. v. Medtronic, Inc.</i> , 73 F.4th 950 (Fed. Cir. 2023)	10
<i>Ford Motor Co. v. AutoConnect Holdings LLC</i> , IPR2025-01342, Paper 27 (P.T.A.B. May 12, 2026)	25
<i>Intel Corp. v. Qualcomm Inc.</i> , 21 F.4th 801 (Fed. Cir. 2021)	26
<i>Koninklijke Philips N.V. v. Google LLC</i> , 948 F.3d 1330 (Fed. Cir. 2020)	16
<i>KSR Int’l Co. v. Teleflex Inc.</i> , 550 U.S. 398 (2007).....	10, 17
<i>Lockwood v. Am. Airlines, Inc.</i> , 107 F.3d 1565 (Fed. Cir. 1997)	2
<i>Qualcomm Inc. v. Intel Corp.</i> , No. 2022-1824, 2024 WL 4258364 (Fed. Cir. Sept. 23, 2024).....	10
<i>Randall Mfg. v. Rea</i> , 733 F.3d 1355 (Fed. Cir. 2013)	16
<i>Rembrandt Diagnostics, LP v. Alere, Inc.</i> , 76 F.4th 1376 (Fed. Cir. 2023)	16
<i>Revvo Techs., Inc. v. Cerebrum Sensor Techs., Inc.</i> , IPR2025-00632, Paper 20 (P.T.A.B. Nov. 3, 2025).....	25

<i>Rivera v. Int’l Trade Comm’n</i> , 857 F.3d 1315 (Fed. Cir. 2017)	3
<i>Tesla, Inc. v. IV II Ltd.</i> , IPR2025-00340, Paper 18, 3-4 (P.T.A.B. Nov. 5, 2025)	25, 26
<i>United Servs. Auto. Ass’n v. PNC Bank N.A.</i> , No. 2023-2171, 2025 WL 339662 (Fed. Cir. Jan. 30, 2025).....	20
<i>Unwired Planet, LLC v. Google Inc.</i> , 841 F.3d 995 (Fed. Cir. 2016)	20, 23
Other Authorities	
M.P.E.P. § 2173.05(i)	4

I. Introduction

Patent Owner's Response largely recycles arguments that the Board rejected at institution. The Board should again reject Patent Owner's efforts to treat known cache-hierarchy behavior as patentably distinct and find claims 1-20 unpatentable.

First, Newton-491 anticipates claims 1-20 under the proper 2019 priority date in Ground 1. There is no dispute the "instruction to not cache" limitation was first introduced in 2019 in the '791 application leading to the '573 patent, rather than in any priority application. Patent Owner's use of gap-filling to equate the *absence* of an instruction to cache (e.g., when there is no option to cache or not cache) with an *affirmative* no-cache instruction is a non sequitur and legally insufficient.

Second, Patent Owner's Ground 2 refrain that the prior art cannot be combined because "Dilley is content delivery" and "Pai is content placement" ignores both the proposed combination and the realities of CDNs. In a tiered CDN, *where* content is placed is directly related to *where* content is retrieved. This supposed "distinction" does not undermine the combination.

Patent Owner's alternative Ground 2 arguments ignore the combination, treat the references in isolation, and add unclaimed features that distort the claims. They should be rejected. The Petition presents a logical combination of related systems to produce "an improved and dynamic tiered CDN with a popularity service," offering

a predictable integration of known building blocks. Google respectfully requests the Board find claims 1-20 unpatentable.

II. Ground 1 Renders Claims 1-20 Unpatentable

A. Newton-471 Anticipates Claims 1-20

There is no dispute Newton-471 is anticipatory. POR at 13-17. The dispute is whether Newton-471 qualifies as prior art. It does.¹

No application in the priority chain discloses an “instruction to not cache.” As explained in the Petition, the ’573 specification, all intervening applications, and the provisional application lack adequate support for an “instruction to not cache.” Pet. at 8-11. As the Board correctly found, the provisional does not “explicitly disclose” an “instruction to not cache.” ID at 13-15. This finding alone should be dispositive, since Patent Owner cannot gap-fill using obviousness to provide written description support. *Lockwood v. Am. Airlines, Inc.*, 107 F.3d 1565, 1571-1572 (Fed. Cir. 1997) (rejecting written description support based on an “obvious variant of that which is disclosed in the specification”); *Rivera v. Int’l Trade Comm’n*, 857 F.3d 1315, 1322-

¹ This is a live issue. Patent Owner mischaracterizes Petitioner’s expert, Dr. Mowry, as agreeing with the Institution Decision. POR at 17. Dr. Mowry acknowledged the Board’s authority; he did not retract his analysis that the priority chain lacks adequate support for the “instruction to not cache” limitation. EX-2017 13:11-15:6.

23 (Fed. Cir. 2017) (same). Because the '573 patent can claim priority to no earlier than February 25, 2019, Newton-471 is prior art.

B. Patent Owner's Contrary Arguments Should Be Rejected

Patent Owner argues it is irrelevant “instruction to not cache” is not “literally recited.” POR at 16-17. It instead contends the provisional discloses that limitation because (i) not-caching is an alternative to caching; (ii) the provisional discusses “reasons for not caching unpopular resources”; and (iii) the provisional discloses a follow-me tag, the absence of which purportedly leads to not caching. *Id.* at 13-17. These points do not provide adequate written description for an “instruction to not cache.”

First, that caching and not caching are alleged “alternatives” does not obviate the “instruction to not cache” requirement. The prior art discusses ways to cache or not cache, including, for example, setting policies for when/whether to cache. *E.g.*, EX-1019, 43 (selectively preventing caching to improve performance and conserve resources); EX-1013, ¶11 (implementing policies to avoid storing on edge servers); EX-1006, ¶63; EX-1037, 14:35-45 (instructions for “hot” content items); EX-1036, 5:31-33 (preventing automatic caching and requiring thresholds), 12:31-50. But the claims do not cover these no-cache mechanisms. They require *an explicit instruction* that says do not cache this resource.

Indeed, that understanding of the “instruction to not cache” is why the claims were allowed during prosecution. The Applicant amended the independent claims to include an “instruction to not cache” to get around prior art of record. EX-1007, 17. While the prior art disclosed policies to manage edge caching (Raju) and popularity-based caching (Swildens)—both of which inherently involved caching some content and not others—the Examiner found the prior art did not disclose “an instruction to not cache,” which the Examiner explained is *more than* “[d]etermining whether [to] cache.” *Id.*; Pet. at 8-11; EX-1006, ¶¶47-53; EX-1036, 5:31-33, 12:31-50; EX-1038, ¶¶68, 126-127.

Contrary to Patent Owner’s suggestion, the “instruction to not cache” is not a negative limitation or a necessary alternative to an “instruction to cache.” POR at 16 (citing M.P.E.P. §2173.05(i)). It requires something affirmative (send an instruction) and cannot be met by the mere absence of an “instruction to cache”—indeed, mere absence of an instruction to file a document (i.e., silence as to whether or not to take the action) does not amount to providing an instruction to *not* file the document.

Second, Patent Owner argues that there are “technical reasons” why unpopular content should not be cached. POR at 15-16. But this is immaterial. As Dr. Mowry explained, all these benefits (namely, storage and cost savings) reasonably conveyed is a *generalized* “determination of whether to cache content,” *not a specific no-cache instruction*, as claimed. EX-1006, ¶¶61-65; *e.g.*, *Ariad Pharms., Inc. v. Eli Lilly &*

Co., 598 F.3d 1336, 1351 (Fed. Cir. 2010). And as discussed, even with benefits to not caching certain content, there are still many different ways to achieve that result.

Last, even if Patent Owner’s points were true, the provisional’s disclosure of a redirect without the “follow me” tag does not provide adequate written description support for the “instruction to not cache.” POR at 15. The specification explains if the edge server receives a redirect with a follow me tag, it “obtains and caches the resource” and then serves it to the client. EX-1011, ¶44. If, however, the edge server receives a redirect *without* a follow me tag, it “forwards the redirect to the client.” *Id.* Importantly, when the follow-me tag is missing, the edge server *does not obtain the resource*, so whether it should cache the content is irrelevant; instead, the client is redirected elsewhere to obtain the content. And when the follow-me tag is present (only branch where the edge server obtains a resource), the resource is obtained and cached. *Id.*, ¶44, Fig. 5. Thus, the ’573 specification provides: (1) forward-only, with no edge retrieval, and (2) retrieve-and-cache. There is no disclosure of any retrieve-and-do-not-cache action, much less one responding to an “instruction to not cache.” *Id.*, ¶44, Fig. 5.

Notably, Patent Owner knew how to provide disclosure for an “instruction to not cache,” since it did so in Newton-471. For example, it explained in Newton-471: “[f]or unpopular content, the mid-tier 708 returns a no-cache instruction,” and “edge server 706 does not cache the content in response to the no cache instruction.” EX-

1002, ¶¶99, claims 2, 19. Such a disclosure is missing from the '573 patent's priority chain, which at most discloses silence (lack of "follow me" tag) in situations where there is no relevant caching option. Because no application in the chain discloses an affirmative "instruction to not cache," the '573 patent claims priority no earlier than February 25, 2019, and Newton-471 is anticipatory prior art.

III. Ground 2 Renders Claims 1-20 Unpatentable

A. A Skilled Artisan Would Have Been Motivated to Combine Dilley, Pai, and Wang

The Petition more than adequately explained why a skilled artisan would have been motivated to combine Dilley, Pai, and Wang to achieve the claimed invention. Each reference operates in the same technological field as the challenged claims and addresses the same problems as the '573 patent—efficient and reliable management of network resources. Pet. at 36.

The Petition first explained Dilley discloses using a tiered cache hierarchy and distributing popular content to the edge to reduce origin-server load. *Id.* at 37. Pai supplies popularity-based distribution, dynamically updating popularity values and distributing content across tiers in view of those values. *Id.* at 37-38. A skilled artisan would have used Pai's popularity-based distribution in Dilley's tiered architecture to provide an effective placement mechanism. *Id.* This would predictably yield an efficient, scalable, cost-effective CDN, and a skilled artisan would reasonably expect success given Dilley and Pai's shared framework. *Id.* at 38-39.

The Petition further demonstrated a skilled artisan would have been motivated to incorporate Wang into the combination to introduce redirects and an “instruction not to cache.” *Id.* at 39-40. Like Dilley and Pai, Wang discloses a three-tiered CDN having different levels storing content of different popularities. *Id.* Wang improves performance by identifying resources to not cache using cache-control attributes and avoiding unnecessary resource consumption. *Id.* A skilled artisan would have used this in the combination to avoid the costs and resource usage of caching content that should not be cached. *Id.* In view of Dilley teaching un-cacheable content should be served from the origin, Wang’s no-cache instruction would have been a predictable complement to Dilley’s combined-tiered approach. *Id.* Combining Dilley-Pai-Wang applies known mechanisms, within a tiered CDN, to yield predictable results. *Id.*

1. Patent Owner’s Arguments Misread Dilley and Pai

Patent Owner argues the skilled artisan would not combine Dilley and Pai, for Dilley uses metadata to determine where to retrieve content, and Pai uses popularity designations to determine where to store content. POR at 22-23. According to Patent Owner, neither teaches “content *delivery* (external routing)” based on popularity. *Id.* (emphasis in original). These arguments should be rejected.

Patent Owner’s argued “gap” between content placement and content delivery ignores that, in hierarchical caching systems, retrieval and external routing follows placement. Where content is distributed in a hierarchy determines where the content

is retrieved. EX-2017, 60:16-61:3 (“where you decide to place data within a cache hierarchy” is “closely related” to “where you fetch data”). To say Pai only speaks to internal placement and cannot inform a request to a second server ignores how tiered CDNs operate. *See* EX-1006, ¶¶185-187.

Further, Dilley discloses a general policy for fetching, caching, and delivering content in a hierarchy, using metadata broadly as “the set of all control options and parameters.” EX-1003, 4:27-30, 4:63-5:21, 5:27-36. That broad disclosure matters because Dilley identifies a problem Pai solves. Dilley recognizes the importance of handling flash-popular content differently from other content and explains its tiered distribution hubs accelerates content to end users and insulates the origin server from overwhelming requests for popular content and large files. *Id.* at 2:67-3:2, 4:48-62. Pai provides a known, concrete way to achieve those goals—determining popularity values (one type of metadata) and distributing content amongst server levels based on popularity thresholds. EX-1004, ¶20. A skilled artisan would have wanted to use this dynamic popularity-based framework to implement Dilley’s general tiering and caching objectives. Pet. at 38-39.

2. The Petition Aligns with Petitioner’s Litigation Construction

The Petition has not used “popularity service” inconsistent with Google’s use of the term in litigation. POR at 23-24. The Petition consistently explained that the

“popularity service” determines content popularity and provides information used to determine where to retrieve content. EX-2027 at 3. Pai’s use of popularity values to distribute/redistribute content determines where content will be served from. ID at 21 (citing Pet. at 44-46). And Pai’s determination of where to store content based on popularity works with its goal of improving content delivery to users by determining where to fetch content based upon popularity. EX-2017, 60:16-61:3. Patent Owner’s effort to divorce placement from delivery location ignores how CDN’s operate. Pet. at 35, 38, 56-57, 83-84.

3. The Combination Does Not Render Dilley or Pai Inoperable

Patent Owner argues the references “are incompatible,” for Dilley uses parent hubs for popular content, and Pai stores popular content at level one servers, and less popular content at level two servers. *Id.* at 24-26. This argument should be rejected.

Dilley does not disclose a rigid popularity-threshold architecture. It teaches a tiered system that serves content through peer, parent, *or* origin paths, and identifies popular/large objects as particularly well-suited for hierarchical handling. EX-1003, 4:63-5:21; EX-1006, ¶232; EX-2017, 46:3-16 (“advantageous to put more popular objects ... close to users”). Pai then supplies specific thresholds for deciding which popularity bands belong at which tier. That a skilled artisan would calibrate the tiers in different ways does not show incompatibility; it shows obviousness. Claims that

combine familiar elements using known methods to yield predictable results would have been obvious. *KSR Int’l Co. v. Teleflex Inc.*, 550 U.S. 398, 416 (2007).

Moreover, Patent Owner’s arguments improperly suggest the Petition is trying to bodily incorporate Pai into Dilley. Not so. The Board rejected these arguments at institution, noting Petitioner was not relying on bodily incorporation (it is not), and Pai’s popularity service would supplement Dilley’s existing tiered architecture in a predictable way (it would have). ID at 24. Nothing in the Response undermines that determination.

Nonetheless, obviousness turns on what the teachings would have suggested to a skilled artisan, not whether a reference can be inserted unchanged into another. *Axonics, Inc. v. Medtronic, Inc.*, 73 F.4th 950, 957 (Fed. Cir. 2023); *Qualcomm Inc. v. Intel Corp.*, No. 2022-1824, 2024 WL 4258364, at *4 (Fed. Cir. Sept. 23, 2024). Here, the combined teachings disclose a CDN hierarchy (Dilley), where popularity (Pai) determines if content should stay at the edge, migrate to a higher tier, or remain farther from the client. Pet. at 38, 44-45.

B. Dilley, Pai and Wang Render Obvious the Claimed “popularity designation” (claims 1, 11, 19)

The Petition explained that Dilley discloses distributing popular content closer to the edge, and Pai supplies implementation details using a popularity-based media distribution system. Pet. at 44. This combination predictably determines a popularity

designation by characterizing content with respect to popularity thresholds, and uses that designation to control content placement. *Id.* at 44-46, 55-60; *see also id.* at 39-40 (storing content in tiered network using popularity was known). This combination yields a more efficient, scalable, cost-effective CDN. *Id.* at 38-40.

1. Patent Owner Improperly Imports a Limitation

The Board correctly found at institution that Pai’s popularity service monitors content popularity, updates popularity values based upon requests, and dynamically redistributes content using popularity thresholds. ID at 22-23. To rebut this, Patent Owner reads into the claims an “eligibility” determination—determining if content is “eligible” for popularity-based handling before popularity analysis occurs. POR at 26 (arguing Pai does not assess eligibility *before* assigning popularity value). This claim rewriting should be rejected.

First, the plain claim language does not recite a determination of “eligibility” for popularity-based handling. EX-1001, 13:35-54, 14:30-49, 15:13-16:14. Claim 1 recites “accessing a popularity service ... to determine a popularity designation.” *Id.* at 13:41-43. Claim 11 recites “a popularity service tracking a popularity designation” *Id.* at 14:45-46. And claim 19 has similar concepts. *Id.* at 16:1-4. Each requires determining the resource’s popularity and using the determination to route and cache resources. Nothing in this plain claim language requires assessing eligibility for the popularity analysis *before* assessing what the popularity designation may be.

The dependent claims confirm this understanding, explaining that “popularity designation” simply requires making the popularity threshold determination. Claims 2 and 3, for example, provide the popularity designation “indicates that the requested resource is [or is not] popular.” *Id.* at 13:55-64. And claim 6 states that the popularity designation “comprises an indication whether a current popularity value ... exceeds popularity threshold.” *Id.* at 14:4-8. Claims 12, 13, and 16 recite similar limitations. *Id.* at 14:50-60, 15:1-5. If “popularity designation” required some binary eligibility determination distinct from popularity threshold determinations, it would not make sense for the dependent claims to then define “popularity designation” as “popular,” “not popular,” or whether the threshold is exceeded. Instead, the claims would define the popularity designation as indicating whether content was labeled for a popularity service. They do no such thing.

Second, the '573 specification does not establish that “popularity designation” requires any eligibility assessment. Patent Owner cites the yes/no “Designated?” box in Figure 5, but that does not support its position. POR at 27. Again, the claims say: “accessing a popularity service ... to determine a popularity designation.” EX-1001, 13:41-43. The “Designated?” box (506) appears in the edge server *prior to* accessing a popularity service and, thus, cannot be the claimed “popularity designation.” That occurs after accessing the popularity service. Patent Owner conflates two Figure 5 concepts: (i) an *unclaimed* routing decision performed at the edge (box 506), with

(ii) the *claimed* popularity designation performed by the popularity service. The '573 specification describes the former as an optional step (“[w]hen only certain content uses the popularity service,” EX-1001, 5:48-50), and it is not required by any of the challenged claims. As the Board correctly explained, the claimed popularity service shown in Figure 5 does the latter—making “determinations on the popularity of the object, and [outputting] three possible responses.” ID at 5. Those “three possible responses” reflect the popularity designations in the dependent claims, not a yes/no eligibility tag.

This aligns with Dr. Mowry’s testimony showing the “popularity designation” “indicates popularity” and allows the system “to determine where [the resource falls] relative to some thresholds.” EX-2017, 36:20-37:10; *id.* at 30:10-31:11. And that is what Pai discloses. EX-1004, ¶20; Pet. at 44 (citing EX-1004, ¶¶13, 20, 39, 47); EX-1006, ¶¶231-236. When Pai determines a resource’s popularity value and classifies it relative to thresholds, that is “determining a popularity designation associated with the requested resource,” as claimed. ID at 22-23.

C. Dilley, Pai and Wang Render Obvious Claim 19’s Second Server Request “based at least on the popularity designation”

Patent Owner first argues the prior art fails to disclose “us[ing] the popularity designation to request a resource from a particular server,” and that Petitioner relies “exclusively on the testimony of Dr. Mowry to fill this gap.” POR at 29-33. Patent

Owner also argues a skilled artisan would not have combined Dilley and Pai for this feature because they “are incompatible.” *Id.* at 30, 34-35. Neither point has merit.

1. The Art Discloses This Feature; There Is No Gap-Filling

Dilley and Pai render obvious “requesting the resource from a second server ... based at least on the popularity designation.” Pet. at 83-84. The Petition explains Dilley satisfies requests using second-tier servers (Pet. at 37, 83 (citing EX-1003, 5:6-21); *see* ID at 17), and its routing mechanism could incorporate popularity-based inputs because it recognizes the importance of delivering flash-popular content. Pet. at 37. The Petition nonetheless uses Pai’s popularity-based distribution system to yield a combination where server selection for retrieval is based on popularity driven placement. *Id.* at 83-85; EX-1003, 5:6-21; EX-1004, ¶¶38-45; EX-1006, ¶¶354-355. A skilled artisan would have understood that, when a request is received, the content is retrieved from the tier in which it is stored, which would be a “second server” in a higher tier depending on popularity. Pet. at 84 (citing EX-1006, ¶354). This is the behavior described in claim 19.

Instead of addressing the combination, Patent Owner repeats its flawed refrain that Dilley uses its metadata to determine where to fetch content from, and Pai cannot influence fetching decisions because it is a storing mechanism. POR at 31-33. But as explained, this argument ignores that, in hierarchical caching systems, retrieval follows placement. EX-2017, 60:16-61:3. Indeed, the Board found “content in Pai is

appropriately distributed ... based on popularity such that when [the] user requests a resource, the system retrieves that resource from the relevant media source based on the content's current storage location.” ID at 21 (citing Pet. at 45-46). To argue Pai speaks only to internal placement and cannot inform a request to a second server ignores how tiered CDNs operate. EX-1006, ¶¶185-187.

Patent Owner also argues the Petition relies on “conclusory expert testimony” to fill in alleged gaps in the prior art. POR at 30-32. Not so. As explained, the prior art renders obvious claim 19, and expert testimony simply explained how the skilled artisan would have understood the teachings and combinability. *Koninklijke Philips N.V. v. Google LLC*, 948 F.3d 1330, 1337 (Fed. Cir. 2020) (an obviousness inquiry “necessarily depends on such artisan’s knowledge”); *Randall Mfg. v. Rea*, 733 F.3d 1355, 1362 (Fed. Cir. 2013) (*KSR* “required an analysis that reads the prior art in context”). Indeed, Dr. Mowry explained, based upon Dilley’s disclosure of handling flash popular content to insulate the origin server, the skilled artisan “would have understood that content designated as unpopular would be served from the parent or origin servers, while popular content would be served from the edge,” tying request routing to popularity. Pet. at 83-84 (citing EX-1006, ¶¶352-353). This is appropriate expert testimony. *See Rembrandt Diagnostics, LP v. Alere, Inc.*, 76 F.4th 1376, 1386 (Fed. Cir. 2023) (properly relying on expert’s “conclusions regarding how [a skilled artisan] would have interpreted the identified disclosures”).

2. A Skilled Artisan Would Have Combined Dilley and Pai

The Petition’s combination also relies on well-established principles of CDN design and predictable integration of known techniques to establish the obviousness of the claimed invention. Pet. at 37-39. This is a classic application of *KSR*, where a purported invention would have been obvious if it offered no more than combining known elements according to established functions to yield predictable results (here dynamic popularity-based routing within a hierarchical CDN). *Id.*; *see also KSR*, 550 U.S. at 416-418.

Patent Owner’s attempt to undermine this obviousness showing by arguing it conflates “internal distribution” and “external delivery” ignores that such systems in fact are “closely related” (EX-2017, 60:16-61:3) and focused on improving content management of network resources. Pet. at 36-39; POR at 34-36. Neither Dilley nor Pai draws the rigid distinction Patent Owner asserts. Moreover, as explained above and in the Petition, it is not a “leap” to use Pai’s popularity distribution together with Dilley’s retrieval of content based on popularity. *See supra* Sections III.A.1-III.A.2, III.B.1; Pet. at 37-39.

D. Dilley, Pai, and Wang Render Obvious the Claimed “redirect instruction” (claims 1, 11)

1. The Petition Presents One Combination

Patent Owner contends the Petition “switches” between combinations to map each claim element (POR at 36), and ignores Pai for limitation 1d, presenting only a

Dilley-Wang combination. POR at 38. Not so. The Petition combines Dilley, Pai, and Wang, and explains how this single combination teaches each limitation. Pet. at 36-40. Indeed, the Petition explains why the skilled artisan would “Combin[e] Dilley and Pai” to create an improved, popularity-based tiered CDN, and then explains why a skilled artisan would “Further Combin[e] Dilley and Pai with Wang” to incorporate Wang’s redirect and no-cache teachings into that same architecture. *Id.* The Petition explains the motivations to combine separately rather than restating Pai’s role under every later sub-limitation, as is common practice in IPRs. *See id.*

Patent Owner improperly treats the prior art as isolated sub-combinations and faults Petitioner for not re-describing the combination under every limitation. POR at 39. But the Petition did not ignore Pai for 1d. Pai’s role is unchanged: providing dynamic popularity-based functionality in a tiered system. Pet. at 44-45. A skilled artisan would have combined Pai’s teachings with Dilley’s and Wang’s to produce “an improved and dynamic tiered CDN with a popularity service,” which includes a redirect and a cache/no-cache instruction. Pet. at 37-40.

Patent Owner argues not addressing Pai is “problematic,” repeating its flawed content “storage” versus “retrieval” dichotomy advanced elsewhere. POR at 38, 40; Section III.A.1. Patent Owner’s reliance on *Intelligent Bio-Systems* is also misplaced since Petitioner is not advancing a new rationale. *Id.* at 40. The Petition presented a three-reference combination, explained the motivation to combine all three, and used

Wang's teachings as the source for the redirect limitation. Pet. at 36-40, 50-51. Any further explanation here is merely responding to Patent Owner's mischaracterization of that disclosed theory.

2. Patent Owner Argues the References Individually, Ignoring That Wang Supplies a Known Alternative

Patent Owner argues Dilley does not teach a second server sending a redirect instruction to a first server because its parent servers request content from the origin server. POR at 41. This ignores the combination. Dilley discloses a tiered CDN in which a first server requests content from a second server and, whenever that second server cannot satisfy a request locally, Wang supplies a known technique returning redirect and cache/no-cache instructions to the lower-level server. Pet. at 50-51; EX-1003, 7:31-40, 8:53-58; EX-1005, ¶¶123-124.

Patent Owner further argues the combination is based upon a "false premise" that Dilley "does not provide specifics" regarding origin retrieval. POR at 41-42. In Patent Owner's view, Dilley has "clear, specific disclosure of how to retrieve content from an origin server" and does not need to be combined with Wang. *Id.* at 42-43. Tellingly, Patent Owner relies upon instances where Dilley has essentially "turn[ed] the cache hierarchy off." EX-1003, 6:30-40. But these passages do not address the relevant situation: when content *is* a candidate for cache hierarchy and is not located at the edge. In that circumstance, Wang supplies the necessary details regarding how

a second server resolves a cache-miss scenario by causing the first server to retrieve content from the origin. EX-1006 ¶252, Pet. at 50-51.

But even if Patent Owner were correct about the level of detail in Dilley, that would not defeat obviousness. Petitioner does not need Dilley to be “silent” as to the retrieval mechanisms before relying on Wang. *See United Servs. Auto. Ass’n v. PNC Bank N.A.*, No. 2023-2171, 2025 WL 339662, at *3 (Fed. Cir. Jan. 30, 2025). Nor is Patent Owner correct that motivation to combine cannot exist if Dilley performs a complete origin-retrieval path (POR at 42). *See Unwired Planet, LLC v. Google Inc.*, 841 F.3d 995, 1002-03 (Fed. Cir. 2016). In other words, whether Dilley provides one origin-retrieval path is of no moment, as Wang teaches another well-known path that would be predictably used with Dilley’s hierarchy. EX-1006, ¶¶253-257; EX-1005, ¶¶123-124. Further, Wang operates in the same way as the “follow me” flag in the ’573 patent. EX-2017, 83:12-84:11 (explaining if follow me flag “could possibly be an instruction not to cache,” it “is identical to what’s in Wang”). The skilled artisan would have thus understood Wang to supply predictable implementation details for Dilley’s system.

3. Wang Does Not Inject Redundant Architecture

Patent Owner argues Dilley does not suggest any redirects because it discloses its own retrieval process and, thus, a skilled artisan would not have reason to turn to Wang for the redirect instruction. POR at 43-45. But that argument simply highlights

Dilley's exemplary routing method; it does not show the skilled artisan would have been discouraged from using Wang's redirect as another known way to implement hierarchical escalation upon a cache miss. Pet. at 50-51; EX-1003, 6:10-15, 6:31-36; EX-1006, ¶¶263-264.

Nonetheless, Dilley's routing method and Wang's redirect are not mutually exclusive. A skilled artisan would advantageously use Dilley's metadata and guide process to determine whether the request proceeds upward in the CDN, while using Wang's redirect for the specific handoff in which the second server causes the first server to obtain a resource from the origin server. Pet. at 50-51, 53 (Dilley's direct-to-origin path would still be used for objects that should not be cached); EX-1006, ¶¶255-256. And Patent Owner's contention that a skilled artisan would not look for a redirect instruction since Dilley does not suggest using one is meritless.

4. Wang's Redirect Mechanism Offers Predictable Benefits

Patent Owner contends the Petition's reason for combining Dilley and Wang is "hindsight" because Dilley discusses that having a persistent connection to parent servers can reduce latency for poorly connected edge servers. POR at 46-47. But that does not establish hindsight. Dilley does not require this persistent connection. EX-1003, 5:10-12. It is optional. And Patent Owner's argument conflicts with Dilley's explanation that most content is fetched using a policy that bypasses the hierarchy, having the allegedly "poorly connected" edge servers request content directly from

the origin without issue. EX-1003, 6:10-15, 31-36. This direct retrieval disclosure also defeats Patent Owner's further argument that the combination would allegedly eliminate the connection to Dilley's allegedly "poorly connected regions." POR at 49; EX-1003, 6:10-15. Patent Owner also contends the combination does not provide latency benefit because its diagrams count "four steps" in both systems. POR at 47-48. This ignores the data being transferred. Under Patent Owner's reading of Dilley, once the parent misses, the full resource (i.e., popular object) travels from origin, to parent, to edge. POR at 48. Under the Wang-inspired modification, the second server need only transmit a small instructional message; the resource travels directly from origin to edge, preventing the second server from obtaining and processing large or popular content unnecessarily. EX-1005, ¶¶123-124. As Dr. Mowry explained, the skilled artisan would find this beneficial as a small redirect message is not equivalent in cost, bandwidth, and/or processing burden to fetching and relaying large resources through the parent tier. EX-1006, ¶¶256-257.

Patent Owner's parent-server-availability argument is also misplaced. POR at 49. Just because servers can handle multiple requests concurrently does not mean no benefits exist when preventing a parent server from processing and relaying large, popular objects unnecessarily. Freeing parent resources from that role leaves them available for other requests and reduces the amount of data that they must manage.

POR at 51. That is important in a CDN designed to protect upstream infrastructure from requests for popular content. EX-1003, 2:26-31, 2:67-3:2.

Finally, Patent Owner argues Wang is superfluous because Dilley discloses a guide process to retrieve content. POR at 49-50. The references are not redundant. The guide-process relates to how the edge server determines whether an object is a candidate for the cache hierarchy and, if so, which parent region to contact in the first instance. EX-1003, 7:14-18, 8:22-38. This is an initial routing decision, not a description of how ultimate retrieval occurs. The Petition explains Wang provides a well-known technique by which an upstream server can instruct a downstream server to retrieve content directly from an origin server with or without caching. *See* EX-1005, ¶¶123-124. And it explains the benefits of improving such a mechanism in the combination. EX-1006, ¶¶253-257. Where a secondary reference contributes a well-known mechanism having predictable benefits, an obviousness determination is not defeated simply because the primary reference has another way of accomplishing a related result. *Unwired Planet*, 841 F.3d at 1002-03.

E. Dilley, Pai, and Wang Render Obvious the Claimed “peer server” (claims 5, 15)

Patent Owner argues the Petition failed to establish the obviousness of claims 5 and 15 because it purportedly only mapped parent-server scenarios to the redirect mechanism. POR at 51-52. Not so.

First, Patent Owner’s recycled argument regarding the redirect instruction in claim 1d (POR at 51) fails for the same reasons above. Section III.D.2. Next, Dilley expressly teaches peer-server queries when the first edge server does not have the requested object. EX-1003, Fig. 6, 8:11-21. Wang teaches the well-known technique of returning redirect and cache/no-cache instructions in that circumstance. EX-1005, ¶¶123-124. The Petition cites as much when discussing claims 5 and 15. Pet. at 50-52, 69, 82.

The skilled artisan would have readily applied the same redirect technique to any “second server” (peer server) in Dilley’s hierarchy when that peer cannot satisfy the request locally. Pet. at 69; EX-1006, ¶258. It is not that a peer is a “middleman” in the same sense as a parent; rather, the peer is an intermediate node in the request path, and redirecting a first server toward a content source avoids futile peer-to-peer handoffs and resource expenditure. Pet. at 69; EX-1006, ¶¶258-260.

Patent Owner’s attempt to cabin the obviousness rationale to parent servers (exclusive of peer servers) improperly narrows the analysis. POR at 52. Claims 5 and 15 recite the “second server is a peer server of the first tier of servers.” EX-1001, 14:1-3, 14:64-67. Dilley supplies the peer-server scenario and Wang supplies a well-known redirect technique; applying Wang’s technique to Dilley’s peer configuration would have been an obvious design choice. Pet. at 50-51, 67-69; EX-1006, ¶¶251-260; EX-1003, 5:4-21, 7:31-40; EX-1005, ¶¶123-124.

F. Dilley, Pai, and Wang Render Obvious the Claimed “receiving an instruction” (claim 3)

Patent Owner argues that Petitioner’s preliminary district court indefiniteness position on claim 3 precludes Petitioner’s obviousness showing here. POR at 53-54. Not so.

The Petition identified an internal inconsistency because claim 1 recites a no-cache instruction, and claim 3 adds a cache instruction; claim 3 accordingly requires competing cache/no-cache instructions. This is not a situation where a petitioner is presenting “inconsistent positions to gain an upper hand,” but rather identifying an incompatibility, which the Board has noted may be proper. *Tesla, Inc. v. IV II Ltd.*, IPR2025-00340, Paper 18, 3-4 (P.T.A.B. Nov. 5, 2025); *compare Ford Motor Co. v. AutoConnect Holdings LLC*, IPR2025-01342, Paper 27 (P.T.A.B. May 12, 2026). Patent Owner cites inapposite caselaw (POR at 53-56), with a petitioner refusing to explain genuinely inconsistent positions. *Am. Airlines, Inc. v. Intellectual Ventures I LLC*, IPR2025-01055, Paper 11 (P.T.A.B. Nov. 21, 2025); *Tesla*, IPR2025-00340, Paper 18; *Revvo Techs., Inc. v. Cerebrum Sensor Techs., Inc.*, IPR2025-00632, Paper 20 at 4-5 (P.T.A.B. Nov. 3, 2025). That is not the case here. Still, “notwithstanding the alleged indefiniteness,” the Petition showed “an ordinarily skilled artisan would understand that the asserted art satisfies” claim 3. *Tesla*, IPR2025-00340, Paper 18,

3-4; *Intel Corp. v. Qualcomm Inc.*, 21 F.4th 801, 813-14 (Fed. Cir. 2021); Pet. at 25, n.4, 61, n.5; EX-1003, 8:15-18; EX-1004, ¶¶3, 30; EX-1005, ¶¶123-124.

IV. Conclusion

For the foregoing reasons, and as explained in the Petition, Google requests the Board find claims 1-20 unpatentable.

Dated: June 8, 2026

Respectfully submitted,

/Erika H. Arner/

Erika H. Arner, Lead Counsel
Reg. No. 57,540

CERTIFICATE OF COMPLIANCE

Pursuant to 37 C.F.R. § 42.24(c)(1), the undersigned hereby certifies that the foregoing **Petitioner's Reply to Patent Owner's Response** contains 5,577 words, excluding the parts exempted under 37 C.F.R. § 42.24(c), as measured by the word-processing system used to prepare this paper.

/Erika H. Arner/
Erika H. Arner, Lead Counsel
Reg. No. 57,540

CERTIFICATE OF SERVICE

The undersigned certifies that a copy of the foregoing **Petitioner's Reply to Patent Owner's Response** was served on June 8, 2026, via email directed to counsel of record for the Patent Owner at the following:

Jason D. Eisenberg
jasone-PTAB@sternekessler.com

Daniel S. Block
dblock-PTAB@sternekessler.com

Kristina Caggiano Kelly
kckelly-PTAB@sternekessler.com

Xiaowei Jin
xjin-PTAB@sternekessler.com

PTAB@sternekessler.com

Dated: June 8, 2026

/Daniel E. Doku/

Daniel E. Doku
Senior Litigation Paralegal
Finnegan, Henderson, Farabow,
Garrett & Dunner, LLP