

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

APPLE INC.,
Petitioner,

v.

Zentian Limited
Patent Owner.

Case IPR2023-00037
Patent No. 10,971,140

PATENT OWNER'S SUR-REPLY

TABLE OF CONTENTS

I.	Petitioner has failed to show obviousness as to limitations 1(d) and 1(e)	1
II.	The Petition’s combination lacks a reasonable expectation of success	4
A.	Patent Owner’s Response presented detailed evidence directed to the <i>Petition’s combination</i> and the abilities of the <i>Petition’s</i> POSA	4
B.	Petitioner’s Reply fails to rebut Patent Owner’s evidence, and instead relies on incorrect premises	7
1.	Petitioner’s false logic 1: “a POSITA is <i>assumed</i> to have requisite knowledge” to enable the Petition’s combination	8
2.	Petitioner’s false logic 2: A POSA would have known how to achieve <i>Petitioner’s combination</i> through general knowledge of parallel processing architectures and various speech recognition systems	15
3.	Petitioner’s false logic 3: “Zentian does not argue that the combination is inoperable, [or] could not be done. . . .”	18
4.	Petitioner’s false logic 4: “Zentian’s alleged complications . . . would occur regardless of the type of information being processed.” Paper 22 at 14.	19
5.	Petitioner’s false logic 5: “Chen envisions its processors and memories are directly and adjacently accessible.” Paper 22 at 16.	19
III.	The evidence proves no motivation to combine	21
A.	Patent Owner’s Response disproved Petitioner’s motivations	21
B.	Petitioner’s attempts to rehabilitate its motivations to combine are ineffective	23

EXHIBIT LIST

Exhibit No.	Description
2017	Deposition of Christopher Schmandt dated September 6, 2023
2018	<i>Intentionally left blank</i>
2019	<i>Intentionally left blank</i>
2020	Declaration of David Anderson, Ph.D
2021	Binu K. Mathew et al., A Gaussian Probability Accelerator for SPHINX 3, ("Mathew I")

I. Petitioner has failed to show obviousness as to limitations 1(d) and 1(e)

Limitations 1(d) and 1(e) require that the “*speech recognition circuit* is configured to generate an initial score,” 1(d), which is “used to determine whether to continue processing to determine a final score via processing a larger amount of data than that was processed to generate the initial score.” 1(e).

To demonstrate obviousness, Mr. Schmandt and Petitioner were required to explain how Jiang’s tree search 74 *modified in view of Chen*, i.e., the combination as a whole, would allegedly operate to meet limitations 1(d) and 1(e). *ActiveVideo Networks, Inc. v. Verizon*, 694 F.3d 1312, 1327 (Fed. Cir. 2012); Ex. 2017 at 46:12-22 (Schmandt admitting his theory requires *modifying* Jiang’s tree search engine 74 to use Chen’s clustered processor architecture); Ex. 2017 at 68:17-25, 83:10-14 (Schmandt admitting his theory relies on tree search engine 74 for limitations 1(d) and 1(e)).

Mr. Schmandt admitted, however, that his declaration’s discussions of limitations 1(d) and 1(e) contains no such explanation. Ex. 2017 at 79:11-23, 81:8-22, 83:15-18. The Petition fails on that basis alone. *ActiveVideo*, 694 F.3d at 1327.

Moreover, the *only* theory Mr. Schmandt offered at his deposition was that *one* processor in *one* cluster in Chen would have been used to generate the “initial

score” of limitation 1(d) and to determine whether to continue processing as recited in limitation 1(e). Ex. 2017 at 102:5-103:3.

19 Just so I'm clear, your
20 testimony now at your deposition is that the
21 result would be compiling or scored in one
22 processor within one cluster in Jiang -- in
23 Chen -- excuse me -- for instance; is that
24 fair?

25 A. Yes. At that point we have
1 Schmandt
2 finished our parallel operation and we're
3 back to serial operations.

As Patent Owner's Response demonstrated, however, that theory would fail because it would require one processor in one of Chen's clusters to receive necessary information from *all four memories* within Chen's Fig. 4 embodiment, Paper 19 at 32-34, whereas Mr. Schmandt's declaration admits that the four memories are not all "interconnected," and that no one processor in one cluster can directly or adjacently access all of the data stored across the four memories. Paper 19 at 34-35; Ex. 1003 ¶ 95; Ex. 2020 ¶¶ 50-51; Ex. 1005, 9:10-39.

Accordingly, Patent Owner demonstrated that Petitioner has not proven that its *combination* would meet limitations 1(d) and 1(e).

Petitioner's Reply arguments only further prove that reality. Petitioner argues that "Jiang's algorithm of generating an initial score and using the initial score to determine whether to continue processing (the functionality of claims 1(d)-1(e)) would be *replicated* across Chen's processors." Paper 22 at 2 (emphasis Petitioner's). That theory, however, does not solve Petitioner's problems; it makes them worse.

First, Petitioner's Reply theory requires *every processor* in Chen's four clusters to perform limitations 1(d) and 1(e), and is thus not Mr. Schmandt's deposition theory, which required only one processor in one cluster to do so. Petitioner thus abandons its own expert's testimony, discrediting Mr. Schmandt.

Second, Petitioner's Reply theory would require every processor in Chen's four clusters to calculate *the same* initial score and make *the same* determination whether to continue processing towards a final score—a design proposal that makes no sense and for which Petitioner has provided neither a reasonable expectation of success nor a motivation.

Third, Mr. Schmandt's deposition testimony explained that there would be no advantage from using *all* of Chen's processors to calculate the initial score and

determine whether to continue processing because “[a]t that point, we have *finished* our parallel processing operation and we’re back to serial operation.” Ex. 2017 at 102:5-103:3.

Fourth, Petitioner’s Reply theory would *still* require *each processor* to obtain necessary information stored across *all four memories* 104(a)-(d) in Chen, Ex. 2020 ¶ 51, whereas none of Chen’s processors can directly or adjacently access all cluster memories. Ex. 1003 ¶ 95; Ex. 2020 ¶¶ 50-51; Ex. 1005, 9:10-39. Indeed, Petitioner’s Reply admits that “[o]nce the cumulative probability score is calculated, inter-cluster processing of the results for each tree search would be communicated[.]” Paper 22 at 16. Petitioner’s Reply theory is thus no more feasible than Mr. Schmandt’s deposition theory.

Fifth, Petitioner’s Reply theory was not set forth in the Petition. Petitioner’s Reply citations are to aspects of the Petition discussing other limitations, not limitations 1(d) and 1(e). Paper 22 at 2 (citing Pet. at 1, 17-18, 24, 30, Ex. 1003 ¶¶ 74, 76-77, 83). Petitioner’s Reply theory is thus foreclosed.

Accordingly, Petitioner has not proven unpatentability.

II. The Petition’s combination lacks a reasonable expectation of success

A. Patent Owner’s Response presented detailed evidence directed to the *Petition’s combination* and the abilities of the *Petition’s* POSA

Patent Owner's Response demonstrated through extensive evidence that *the Petition's combination* lacks a reasonable expectation of success. Paper 19 at 7-16. Each aspect of Patent Owner's argument was directed to problems and deficiencies arising out of *the Petition's specific combination*, not the '140 Patent. *Id.*

In particular, Patent Owner's Response proved:

- The Petition's combination requires performing the speech recognition techniques of Jiang on the clustered processor and memory architecture of Chen, Ex. 1003 ¶ 68;
- Jiang does not enable performing its speech recognition on Chen's clustered processors, nor does Chen enable using its hardware to perform Jiang's speech recognition, Ex. 2017 at 86:9-17, Ex. 2020 ¶ 24;
- Mr. Schmandt's declaration did not provide any underlying details to support his conclusion as to the POSA's capability, Ex. 2017 at 48:13-21;
- Mr. Schmandt has never supervised anyone involved in mapping a speech recognition model to a clustered processor and memory architecture like Chen's, Ex. 2017 at 145:20-24, nor has he built the processor to memory architecture of any speech recognition system identified in his background experience, *Id.* at 34:24-35:5, 32:3-9, 106:10-23.

- Implementing the Petition's combination of Jiang and Chen would have been highly complex and well beyond the skill level of the POSA prior to the '140 Patent, Ex. 2020 ¶¶ 17-19, 27 (page 16), because:
 - porting Jiang's speech recognition techniques to Chen's clustered parallel processors and memories would have required coordinating multiple caches, avoiding memory conflicts, controlling task sharing in an efficient manner, resolving synchronous bottlenecks, addressing communication bandwidth and latency issues among the various hardware, and developing a messaging strategy to coordinate information sharing between and within the clusters, Ex. 2020 ¶ 27, whereas;
 - the ordinary artisan in the field of speech recognition was a specialist in digital signal processing and speech recognition, not parallel processing architectures or high performance computing, Ex. 2020 ¶ 28, and further because;
 - Jiang teaches the use of the Viterbi decoding algorithm, which requires extensive communication between computational components of a system, whereas Chen's cluster memories are not all directly or adjacently accessible to the processors and memories in

each other cluster, rendering it unlikely that Jiang's Viterbi algorithm could successfully be run on Chen's clustered processors in a practical system. Ex. 2020 ¶ 29;

- It was known in the art that communication between parallel processing nodes can determine the feasibility of a parallel processing architecture for a given problem, and that a memory model in which memory is connected only within a cluster (as in Chen) may be appropriate “for applications that require little or no communication,” whereas Jiang's speech recognition stage requires significant communication, and is thus poorly suited for Chen. Ex. 2020 ¶ 30;
- Even Dr. Anderson, who is a person of *extraordinary skill*, along with a team of engineers and a DSP expert from Texas Instruments, could not successfully transition existing signal processing software to a parallel processing architecture after six months. Ex. 2020 ¶ 31.

B. Petitioner's Reply fails to rebut Patent Owner's evidence, and instead relies on incorrect premises

In the face of the extensive evidence summarized above, Petitioner does not present a supplemental declaration from Mr. Schmandt. Moreover, Petitioner

abandoned its deposition of Dr. Anderson after merely one hour, and thus has nothing substantive to introduce from Dr. Anderson’s testimony. Ex. 1035.

Instead, Petitioner’s Reply presents a series of incorrect conclusions built on inaccurate or inapplicable premises. Each of Petitioner’s logically false arguments is addressed and refuted in turn below.

1. Petitioner’s false logic 1: “a POSITA is *assumed* to have requisite knowledge” to enable the Petition’s combination

Petitioner attempts to conflate *the Petition’s combination* with the embodiments and claims of the ’140 Patent. Petitioner argues that if the *patent* did not discuss the problems with the *Petition’s combination*, it must be assumed that the POSA would have known how to overcome the problems with the *Petition’s combination*. Paper 22 at 4-6. That logic is false.

“Skill in the art does ***not*** act as a bridge over gaps in substantive presentation of an obviousness case, but instead supplies an important guarantee of ***objectivity*** in the process.” *Al-Site Corp. v. VSI Int’l, Inc.*, 174 F.3d 1308, 1324 (Fed. Cir. 1999) (emphasis added).

Here, the Petition’s combination begins with Jiang and Chen, and proposes to combine those references in particular ways that neither reference enables. Section II.A, *supra*. Petitioner has failed to demonstrate through actual evidence

that the POSA, as defined in the Petition, would have the necessary skill or knowledge to in fact achieve *the Petition's combination* with a reasonable expectation of success. Section II.A, *supra*. Patent Owner has demonstrated through detailed evidence that the POSA could not. Section II.A, *supra*.

Petitioner's assertion that the POSA should be “***assumed to have requisite knowledge***” to achieve the Petition's contorted and highly complex combination, Paper 22 at 4, is the essence of hindsight bias. Petitioner cannot weaponize the level of *ordinary* skill in the art, which is a safeguard against hindsight bias, to fill substantive gaps in its obviousness case. *Al-Site*, 174 F.3d at 1324.

Petitioner's premise is that if the '140 Patent did not teach how to solve the specific problems arising out of *Petitioner's combination*, then the Board must conclude that the POSA would know how to solve *Petitioner's problems*. Paper 22 at 2-5. That reasoning is false: the '140 Patent does not teach performing Jiang's speech recognition techniques, which require extensive communication between a system's computational components, within memory and processor clusters that are not all directly or adjacently accessible to each other, as Petitioner's combination requires. Ex. 2020 ¶¶ 29-30. The '140 Patent thus has no reason to teach solutions to problems arising out of *Petitioner's combination*. Paper 22 at 4. The patent's silence on those issues is certainly not evidence that the solutions

were within the skill level of the Petition's POSA. The consequence of the Petition's absurd combination theory lies with Petitioner, not Patent Owner.

Moreover, Dr. Anderson explained at his deposition that the '140 Patent *does* provide important teachings and solutions that Petitioner has not shown were known in the prior art, and that are *distinct* in key respects from Petitioner's combination of Jiang and Chen. Ex. 1035 at 23:6-27:3.

Petitioner may not, of course, demonstrate obviousness based on the '140 Patent's own teachings. "[T]he inventor's own path itself never leads to a conclusion of obviousness; that is hindsight." *Univ. of Strathclyde v. Clear-Vu Lighting LLC*, 17 F.4th 155, 162 (Fed. Cir. 2021) (emphasis added); *TQ Delta, LLC v. Cisco Sys.*, 942 F.3d 1352, 1362 (Fed. Cir. 2019).

Petitioner's case law citations do not establish otherwise. In *Uber Techs., Inc. v. X One, Inc.*, 957 F.3d 1334, 1339 (Fed. Cir. 2020), the issue was whether a POSA would have been motivated to select between one of only two known, finite approaches. Because the challenged patent was directed to *both* solutions, the Federal Circuit held that the appropriate choice was within the level of skill of the POSA. *Id.* Here, the issue is expectation of success, not motivation to combine, there are not finite choices, and the '140 Patent does not teach Petitioner's combination.

In *In re Publicover*, 813 F. App’x 527, 531-32 (Fed. Cir. 2020), the prior art and the challenged patent **both** taught identification of vestibulo-ocular movement, a type of eye movement. Patent Owner presented only “attorney argument,” not evidence, that the prior art’s disclosures failed to adequately explain how to identify that type of eye movement. *Id.* The Federal Circuit rejected that “attorney argument” because the challenged patent likewise **taught** identifying that type of eye movement **but did not** provide a detailed teaching of **how** to do so. *Id.* Here, neither Jiang nor Chen teaches implementing Jiang’s speech recognition techniques on a clustered parallel processing, shared cluster-memory architecture—much less in the manner Petitioner has proposed. Ex. 2017 at 86:9-17, Ex. 2020 ¶ 24. Accordingly, the prior art does not and cannot carry any presumption of enablement **as to Petitioner’s combination**. *In re Antor Media Corp.*, 689 F.3d 1282, 1287-88 (Fed. Cir. 2012) (prior art only presumed enabling for what it teaches). Moreover, the ’140 Patent teaches approaches that are **distinct** in key respects from Petitioner’s combination of Jiang and Chen, and provides solutions that, in any case, have not been shown in the prior art. Ex. 1035 at 23:6-27:3. Thus, the ’140 Patent provides no assumption that Petitioner’s combination was within the POSA’s skill. Finally, Patent Owner’s demonstration of a lack of a

reasonable expectation of success in this proceeding is based on extensive evidence, not “attorney argument.” *Publicover* is thus inapposite.

In re Epstein, 32 F.3d 1559, 1568 (Fed. Cir. 1994), fails to support Petitioner for the same reasons. There, the Federal Circuit simply held that “one of skill in the art would have known how to implement ***the features of the references***” as taught in each reference individually. *Id.* (emphasis added). Here, Petitioner’s combination does not simply require practicing what Jiang or Chen taught individually—it requires a combination that neither reference taught, and that a POSA would not have been able to achieve. Ex. 2020 ¶¶ 23-32 (pages 15-20).

Petitioner’s reliance on the Board’s decisions in *Walmart* and *Samsung* is misplaced because those decisions do not carry legal authority. Nonetheless, both decisions also fail to support Petitioner’s argument.

In *Walmart*, the Board relied on the challenged patent’s teaching that “a plurality of light panels 40 may be disposed end to end,” and the patent’s silence as to how to address challenges that would arise “***when disposing the panels in electrical series.***” IPR2022-00534, Paper 50 at 32 (emphasis added). By contrast, the problems Patent Owner has identified with Petitioner’s combination do ***not*** arise out of practicing what the ’140 patent itself teaches; they arise out of

Petitioner's combination design. The Board's decision in *Walmart* is thus inapposite.

In *Samsung*, the prior art disclosed wireless communication, and thus was entitled to a presumption of enablement ***for its own teachings***. IPR2022-00324, Paper 33 at 36-37. Moreover, the challenged patent itself disclosed the ***same*** wireless communication techniques as the prior art without implementation details. By contrast, the problems with Petitioner's combination here do not arise out of the individual teachings of any reference, but from a combination of teachings that is neither disclosed nor enabled by either reference. Section II.A, *supra*. Petitioner's combination thus carries no presumption of enablement. Nor does the '140 patent disclose Petitioner's combination without implementation details. Ex. 1035 at 23:6-27:3.

Petitioner's reliance on *Keynetik* is further misplaced. Paper 22 at 6. There, the patent owner had "cited no evidence" to contradict petitioner's assertions of a reasonable expectation of success, and had instead relied only on "attorney argument." *Keynetik, Inc. v. Samsung Elecs. Co., Ltd.*, 2023 U.S. App. LEXIS 3567, at *3-4 (Fed. Cir. 2023). The Federal Circuit held that even though petitioner's alleged evidence was "brief, in the absence of any contradictory evidence, it constitutes substantial evidence to support the Board's finding." *Id.*

Here, by contrast, Patent Owner has presented extensive evidence disproving a reasonable expectation of success, and showing that Petitioner’s combination would not have simply required routine and simple software programming within the POSA’s skill. Section II.A, *supra*. In particular, Dr. Anderson (one of extraordinary skill) has testified that he and a team of engineers failed to successfully port far simpler signal processing software to a parallel processing hardware architecture. Ex. 2020 ¶ 31. *Keynetik* is thus inapposite.

Finally, the problems Patent Owner has identified with respect to Petitioner’s combination are not “generic,” as Petitioner contends. Paper 22 at 7. They are highly detailed and specific to Petitioner’s combination, and supported by the testimony of *both* experts. Section II.A, *supra*; Ex. 1035 at 23:6-27:3. For the same reasons, the *challenged claims* do not need to recite either the problems arising from Petitioner’s combination theory or the solutions to those problems, as Petitioner argues. Paper 22 at 7. The question of a reasonable expectation of success *for the combination* must be assessed based on the details Petitioner has proposed for the *combination*, not the words of the challenged claims. *Arctic Cat Inc. v. Bombardier Rec. Prods.*, 876 F.3d 1350, 1360-61 (Fed. Cir. 2017) (requiring proof that “an ordinarily skilled artisan would [] believe that a

particular combination would have a reasonable expectation of ‘anticipated success’) (emphasis added).

2. Petitioner’s false logic 2: A POSA would have known how to achieve *Petitioner’s combination* through general knowledge of parallel processing architectures and various speech recognition systems

Petitioner argues that because Jiang allegedly taught “using multiple processors,” and because references like Mathew taught parallel processing in speech recognition, it follows that the POSA would have known how to achieve *Petitioner’s specific combination* of Jiang and Chen. Paper 22 at 11-15.

That logic is false. Prior art is only presumptively enabled *for what it teaches*. *In re Antor Media Corp.*, 689 F.3d 1282, 1287-88 (Fed. Cir. 2012). Mr. Schmandt has admitted that neither Chen nor Jiang teaches *Petitioner’s combination*, and thus neither reference is entitled to a presumption of enablement with respect to *Petitioner’s combination*, as Petitioner contends. *Compare* Paper 22 at 12-15 *with* Ex. 2017 at 86:9-17, Ex. 2020 ¶¶ 23-32 (p. 15-19); *Raytheon Techs. Corp. v. GE*, 993 F.3d 1374, 1380-81 (Fed. Cir. 2021). Even assuming Chen teaches an “operable clustered processor architecture,” Paper 22 at 15, Chen does *not* teach or enable using its architecture to run Jiang’s speech recognition

techniques (much less in the manner the Petition has proposed), or to run *any* speech recognition. Ex. 2017 at 86:9-17.

And Petitioner itself states that “Mathew is *dissimilar architecture*” to the combination, so Mathew could not have enabled the combination. Paper 22 at 21 (emphasis added). The same is true for Hon and Ravishankar. Indeed, the evidence is uncontroverted that Petitioner’s combination appears *nowhere* in the prior art, but is instead a novel fabrication concocted from multiple modifications to the teachings of two entirely distinct references, one of which (Chen) is not even in the field of speech recognition, in a manner not taught or enabled by any reference. Section II.A, *supra*. The law does not permit *assuming* a combination to be enabled when the prior art of record does not enable the combination. *Raytheon*, 993 F.3d at 1380-81.

Moreover, Dr. Anderson’s detailed testimony refutes Petitioner’s conclusory assumptions of a reasonable expectation of success. Ex. 2020 ¶¶ 23-32. Indeed, Dr. Anderson has explained that, despite his *extraordinary* level of skill, and the help of multiple other engineers, he could not successfully transition existing signal processing software to a parallel processing architecture after many months. Ex. 2020 ¶ 31.

Petitioner argues that Dr. Anderson's testimony merely "impugns Dr. Anderson's credibility," but "does not prove a POSITA would lack the knowledge to reasonably succeed in the combination." Paper 22 at 10-11. That is incorrect. Dr. Anderson's testimony demonstrates that even a person of *extraordinary skill* in the art "could not necessarily expect success with respect to far simpler combinations than the one the Petition has proposed," Ex. 2020 ¶ 31, and thus supports the credibility of Dr. Anderson's testimony that the Petition's combination was beyond the level of *ordinary skill*. Ex. 2020 ¶¶ 23-32 (pages 15-20). Petitioner hypothesizes that "[p]erhaps Dr. Anderson simply did not have the necessary skills," Paper 22 at 11, but Dr. Anderson's qualifications and experience far exceed Mr. Schmandt's, and Petitioner did not even attempt to question Dr. Anderson regarding the challenges he faced in porting existing signal processing software to a parallel processing architecture. Ex. 1035.

By contrast, Mr. Schmandt's lack of experience in building the processor to memory architecture of any speech recognition system, Ex. 2017 at 34:24-35:5, 32:3-9, 106:10-23, or supervising anyone involved in mapping a speech recognition model to a clustered memory and processor architecture, *id.* at 145:20-24, greatly undermines his opinions regarding a reasonable expectation of success. And Mr. Schmandt's opinion is further discredited by his testimony that his

declaration does not explain *how* a POSA could have achieved his combination, Ex. 2017 at 48:13-21.

Petitioner argues that Mr. Schmandt’s testimony is *admissible* despite those flaws, Paper 22 at 9-10, but the issue here is not admissibility but rather credibility. Mr. Schmandt’s conclusory assertions of a reasonable expectation of success should be given no weight in view of the evidence of record, including his own admissions.

3. Petitioner’s false logic 3: “Zentian does not argue that the combination is inoperable, [or] could not be done. . . .”

Petitioner appears to suggest that Patent Owner must prove that Petitioner’s combination could never be achieved even with unlimited knowledge, skill, and resources. Paper 22 at 18. That is false. Patent Owner need only demonstrate that *Petitioner has not proven* that a person of *ordinary* skill in the art at the time of the invention would have reasonably expected to succeed in combining the prior art elements in the way *Petitioner has proposed*. *KEYnetik, Inc. v. Samsung Elecs. Co.*, 841 F. App’x 219, 227 (Fed. Cir. 2021); *Arctic Cat Inc. v. Bombardier Rec. Prods.*, 876 F.3d 1350, 1360-61 (Fed. Cir. 2017); *In re Cyclobenzaprine Hydrochloride Extended-Release Capsule Patent Litig.*, 676 F.3d 1063, 1069 (Fed. Cir. 2012). Patent Owner has met and exceeded that requirement. Section II.A,

supra. Petitioner’s arguments fundamentally ignore the crucial distinction between persons of *ordinary* skill and those of *extraordinary* skill.

4. Petitioner’s false logic 4: “Zentian’s alleged complications . . . would occur regardless of the type of information being processed.” Paper 22 at 14.

Petitioner contends that the problems arising out of its combination would occur in *any* clustered processing architecture. Paper 22 at 14. But Petitioner cites *no evidence* in support of that assertion. *Id.*

Petitioner later attempts to flip its burden onto Patent Owner, arguing that Zentian never explains why certain problems “would be present in the combination but not” in the system of Claim 1. Paper 22 at 20. That argument also fails. Petitioner cannot shift its burden of proof to Patent Owner. Nonetheless, Patent Owner has demonstrated in detail that Petitioner’s problems arise out of the requirements Petitioner and its expert have identified for Petitioner’s combination, and the teachings of the prior art references they have relied upon, Ex. 2020 ¶¶ 23-32, not the ’140 Patent. Ex. 1035 at 23:6-27:3.

5. Petitioner’s false logic 5: “Chen envisions its processors and memories are directly and adjacently accessible.” Paper 22 at 16.

Petitioner admits that its combination *must* communicate the probability scores calculated by the various processors and stored within the various memories

of Chen's clusters *across* the different clusters. Paper 22 at 16 ("Once the cumulative probability score is calculated, inter-cluster processing of the results of each tree search would be communicated").

Petitioner instead contends that Chen's "processors and memories are directly and adjacently accessible." *Id.* But Petitioner cites no testimony from its expert witness. Instead, it cites to Chen's "Prior Art" discussion, where Chen theorizes an idealized parallel processing system in general terms. *Id.* (citing Ex. 1005 at 1:56-2:1). That passage of Chen does not describe the embodiment Petitioner relies upon. Rather, Mr. Schmandt's own declaration stated that *in the relevant embodiment of Chen*, "not all[] the clusters are interconnected," and "the processors 102a in cluster 100a . . . **cannot directly or adjacently access** the memory 104d of cluster 100d, while processors 102d in cluster 100d . . . **cannot directly or adjacently access** the memory 104a." Ex. 1003 ¶ 95. That fact is further set forth in Chen, Ex. 1005 at 9:10-39, and is demonstrated by Dr. Anderson. Ex. 2020 ¶¶ 50-51. Petitioner's Reply contention is thus false.

Petitioner next argues that Chen's "node switching mechanism" permits the necessary communications *between* different cluster processors and memories. Paper 22 at 17. Again, Petitioner does not cite to any testimony from Mr. Schmandt. And once again, Petitioner's assertions are false. Petitioner cites to

Chen at 5:54-6:3, which explains that the “two or more processors, and one or more input/output ports” *within a single cluster* are “symmetrically connected by a node switching mechanism to a homogenous *cluster shared memory*.” *Id.*

(emphasis added). Thus, Chen’s “node switching mechanism” allows processors *within* one cluster to access the “cluster shared memory” of *that same cluster*, and thus does not support Petitioner or disprove Patent Owner. Petitioner’s citation to Chen at 10:36-55 contains the same teaching, describing access by processors 102a within cluster 104a to the memory 100a *of that same cluster*. Ex. 1005 at 10:36-55. Petitioner’s citation to Chen at 8:52-64 is likewise off the mark. That passage does not address or negate Chen’s express teachings at 9:34-39 that the processors in cluster 100a cannot directly or adjacently access the memory of cluster 100d, and the processors of cluster 100d cannot directly or access the memory of cluster 100a. Ex. 1005 at 9:34-39, Ex. 1003 ¶ 95, Ex. 2020 ¶¶ 50-51.

III. The evidence proves no motivation to combine

A. Patent Owner’s Response disproved Petitioner’s motivations

Patent Owner’s Response refuted each of the Petition’s motivations through the testimony of both experts. Paper 19 at 17-28.

With respect to increasing processing power and speed, Mr. Schmandt’s declaration did not provide any analysis to explain *why* Chen would allegedly

speed up Jiang or relax its pruning threshold, or by *how much*. Ex. 2020 ¶ 35. To the contrary, Mathew I taught that a five-processor system parallel processing speech recognition system was *slower* than a two-processor system due to high synchronization overhead. Ex. 2021 at 11; Ex. 2020 ¶¶ 36-37. Thus, a POSA would not have assumed that the combination of Jiang and Chen would improve processing speed or relax Jiang's pruning threshold. Ex. 2020 ¶ 37.

As for Petitioner's "cost" motivations, Mr. Schmandt admitted he did not specify any costs that would be saved, nor did he identify *what* the combination would be cheaper than. Ex. 2017 at 61:14-62:9. Further, he did not compare the cost of Chen's clustered processing architecture to any other clustered processing architecture. *Id.* at 63:8-15. Nor did he compare the cost of Jiang in unmodified form with the cost of combining Jiang with Chen. *Id.* at 40:23-41:4. There is thus no basis for Petitioner's "cost" motivation. Ex. 2020 ¶¶ 38-39.

Regarding "flexibility and scalability," Mr. Schmandt's deposition and Dr. Anderson's testimony established that there is no such advantage to the combination compared to Jiang alone. Ex. 2020 ¶¶ 40-41; Ex. 2017 at 105:8-106:4.

In addition, Dr. Anderson demonstrated *against* a motivation to combine because (1) converting Jiang's shared memory architecture to Chen's memory architecture would have required significant reprogramming effort, (2) Chen's

memory architecture would have required onerous messaging overhead to pass information across its memory clusters, as shown by Mathew I, (3) Chen does not allow direct communication between all of the clusters, which would be critical for Jiang’s Viterbi decoding technique, and (4) the POSA would not have expected to successfully manage the combination’s complications in order to achieve worthwhile benefits. Ex. 2002 ¶¶ 42-45.

B. Petitioner’s attempts to rehabilitate its motivations to combine are ineffective

Petitioner argues that Mr. Schmandt did not need to “build” Petitioner’s combination to prove obviousness, Paper 22 at 18, but that argument attacks a strawman. Petitioner’s motivations rest on *allegations* of benefits that Petitioner does not prove, and that Patent Owner’s evidence refutes. While a combination may not need to be “the *most* optimal, desirable, or preferred,” Paper 22 at 19 (emphasis added), Petitioner must prove that the POSA would have been motivated to pursue the combination. The evidence here proves otherwise.

Petitioner argues that its cost motivation is premised on comparing multiple commercially-available processors to one solitary, single core processor. Paper 22 at 19. But as Mr. Schmandt admitted, he did not perform any analysis to conclude that Petitioner’s combination would cost less than any other solution. Ex. 2017 at

62:2-9. Indeed, Mr. Schmandt admitted he had no idea what different types of processors would cost at the time of the invention. *Id.* Critically, Mr. Schmandt provided no evidence that the Petition's combination would have cost **less** than the alleged "high cost" solitary, single-core processors with the "necessary" computer power.

Petitioner next contends that Patent Owner did not prove that *Petitioner's combination* would slow down the system *of the challenged claims*. Paper 22 at 20. Petitioner's misdirection is unavailing. Patent Owner's evidence correctly focused on *Petitioner's allegations*, and showed that the POSA **would not** have expected to achieve more speed in a parallel processing speech recognition system by simply adding more processors to that system. Ex. 2021 at 11 (five processor system slower than two processor system); Ex. 2020 ¶¶ 36-37. Petitioner argues that Mathew I did not discuss a clustered architecture with a partitioned acoustic model like Petitioner's combination of Jiang and Chen. Paper 22 at 20-21. But *Petitioner* has provided no evidence of an improvement that is specific to the combination's "clustered architecture" or its "partitioned acoustic model." Instead, Petitioner rested on the general assertion that more processors are better. Ex. 2017 at 49:15-50:3 (Mr. Schmandt alleging that "more processors is going to be faster"), Ex. 2020 ¶ 35, Ex. 1003 ¶¶ 67, 74. Mathew I proves that in the context of parallel

processing speech recognition systems, that is not necessarily true. Ex. 2020 ¶¶ 36-37; Ex. 2021 at 11. Dr. Anderson has further shown that a POSA would not have expected a speed benefit from the combination. Ex. 2020 ¶¶ 42-45. Petitioner's reliance on Dr. Wang's "Formula One engine" analogy is likewise misplaced. Paper 22 at 20. Dr. Wang opined that the POSA would know that Petitioner's combination "makes no sense," and would not have been motivated to pursue it. Ex. 2002 ¶ 30. There is thus no basis for Petitioner's assertion that its combination system is "higher-speed than Jiang alone." Paper 22 at 24.

Petitioner alleges, in passing, that Mr. Schmandt's declaration testimony regarding striking "a balance between cost and processing power" is "sufficient to support a motivation of the combination." Paper 22 at 21-22. But as shown, Petitioner's evidence neither shows a cost benefit nor a processing improvement, much less *both*. Moreover, Petitioner simply ignores Dr. Anderson's proof that the POSA would *not* have desired the combination given the significant reprogramming effort, onerous messaging overhead, lack of direct communication across all clusters, and the POSA's ultimate lack of an expectation of net benefits. Ex. 2002 ¶¶ 42-45. The evidence thus weighs against a motivation to combine regardless of Petitioner's "cost" motivation.

Petitioner falls back on the incantation of a “known technique to improve a similar device in the same way.” Paper 22 at 22. But Chen never teaches that its hardware would improve any speech recognition system, much less Jiang’s. Indeed, Petitioner has no evidence that Chen’s hardware architecture was *ever* applied to a speech recognition system in the prior art, much less that its use improved any speech recognition system. Moreover, the evidence from Mathew I and Dr. Anderson proves that the POSA would *not* have expected a motivating improvement from combining Jiang with Chen. Ex. 2020 ¶¶ 33-45. Thus, the combination is *not* the application of a “known technique” (Chen’s specific clustered processor architecture) and would *not* have improved Jiang’s system in any way.

Petitioner next resurrects its “flexibility and scalability” theory. Paper 22 at 23. But Hon’s “flexibility and scalability” teaching is entirely tied to the benefits of a programmable pipeline (which the combination does not propose) and extendable processor architecture compared to a system that did not have that capability (which Jiang already had). Ex. 2020 ¶¶ 40-41. There was thus no “flexibility and scalability” advantage, and certainly no motivation in view of Dr. Anderson’s evidence of the combination’s many prohibitive disadvantages. Ex. 2002 ¶¶ 42-45.

Finally, Petitioner argues that: (a) the combination does not use a “distributed memory model among the clusters,” Paper 22 at 26, and (b) all of Chen’s processors allegedly *can* access the cluster memory of every cluster. *Id.* at 27. Both assertions are incorrect.

Petitioner’s no “distributed memory model” argument rests on its allegation that its combination does not provide access from one cluster’s memory to adjacent clusters. *Id.* at 26. The Petition and Chen prove that argument to be false. The Petition relies on the embodiment of Chen’s Fig. 4. Pet. at 15, 16, 22, 23. Chen teaches that in the embodiment of Fig. 4, “the processors 102a in cluster **100a** . . . can **adjacently access** the cluster shared memory 104b and 104c of clusters **100b and 100c**, respectively.” Ex. 1005 at 9:24-28 (emphasis added). Thus, the Petition’s combination system would require significant reprogramming effort, as Dr. Anderson opined. Ex. 2020 ¶¶ 42-44.

Petitioner’s argument that each processor in Chen can access every cluster memory relies on Chen at 10:40-43. But as explained, that portion of Chen describes access by processors 102a within cluster 104a to the memory 100a *of that same cluster*, not to every other cluster memory, as Petitioner contends. Ex. 1005 at 10:36-55. To the contrary, the testimony of both experts and Chen itself proves that no processor in any of Chen’s clusters can directly or adjacently access

the memory of every other cluster. Ex. 1005 at 9:34-39, Ex. 1003 ¶ 95, Ex. 2020 ¶¶ 50-51. Accordingly, “Chen presents a particularly poor choice” for combination with Jiang, as Jiang’s search stage requires extensive communication that Chen cannot support. Ex. 2020 ¶ 43.

Respectfully submitted,

/s/ Kayvan B. Noroozi

Peter C. Knops, Reg. No. 37,659
Kayvan B. Noroozi, *Pro Hac Vice* (To be Filed)
NOROOZI PC
11601 Wilshire Blvd., Suite 2170
Los Angeles, CA 90025
Attorneys for the Patent Owner

Date: February 20, 2024

CERTIFICATION OF WORD COUNT

The undersigned hereby certifies that the portions of the above-captioned **PATENT OWNER'S SUR-REPLY** specified in 37 C.F.R. § 42.24 have 5,594 words, in compliance with the 5,600 word limit set forth in 37 C.F.R. § 42.24(b)(1). This word count was prepared using Microsoft Word 365.

/Peter C Knops/

Peter C Knops

Counsel for Patent Owner

Date: February 20, 2024

CERTIFICATION OF SERVICE

The undersigned hereby certifies that the foregoing **PATENT OWNER'S SUR-REPLY** was served electronically via e-mail on February 20, 2024, on the following counsel of record for Petitioner:

Adam P. Seitz
Adam.Seitz@eriseip.com
PTAB@eriseip.com

Jennifer C. Bailey
Jennifer.Bailey@eriseip.com

Adam M. Sandwell
Adam.Sandwell@eriseip.com

Erise IP, P.A.
7015 College Blvd., Ste. 700
Overland Park, Kansas 66211

Christina Canino
Christina.Canino@eriseip.com

Erise IP, P.A.
5299 DTC Blvd., Suite 1340
Greenwood Village, CO 80111

/Peter C. Knops/
Peter C. Knops
Counsel for Patent Owner

Date: February 20, 2024