

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

APPLE INC.,
Petitioner

v.

ZENTIAN LIMITED,
Patent Owner

Inter Partes Review Case No. IPR2023-00037
U.S. Patent No. 10,971,140

PETITIONER'S REPLY TO PATENT OWNER'S RESPONSE

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	THE JIANG-CHEN COMBINATION RENDERS OBVIOUS CLAIMS 1(D) AND 1(E).....	1
III.	ZENTIAN PROVIDES NO EVIDENCE OF A LACK OF REASONABLE EXPECTATION OF SUCCESS.....	3
A.	ZENTIAN’S ARTIFICIAL COMPLEXITY REQUIREMENTS ARE UNCLAIMED AND UNDESCRIBED IMPLEMENTATION DETAILS	4
B.	LEVEL OF SKILL.....	8
1.	<i>Zentian Does Not Dispute the Level of Skill</i>	<i>8</i>
2.	<i>Mr. Schmandt’s Level of Skill</i>	<i>9</i>
3.	<i>Dr. Anderson’s Level of Skill</i>	<i>10</i>
4.	<i>Zentian’s “Specialized” Requirements Are Found in the Cited References</i>	<i>11</i>
C.	SKILLED PERSONS KNEW OF VARIOUS SPEECH RECOGNITION HARDWARE ARCHITECTURES	12
D.	ZENTIAN’S “INACCESSIBLE MEMORY” THEORY	15
IV.	ZENTIAN DOES NOT IDENTIFY REASONS FOR NON-OBVIOUSNESS.....	18
A.	ZENTIAN’S “QUANTITATIVE ANALYSIS” THEORY	18
1.	<i>Cost Reduction</i>	<i>19</i>
2.	<i>Improved Processing Speed.....</i>	<i>19</i>
3.	<i>Obviousness Includes Evaluating Tradeoffs and Sufficiency of Known Techniques</i>	<i>21</i>
B.	FLEXIBILITY AND SCALABILITY	23
C.	NO DESIRE TO ADD ANOTHER PROCESSOR.....	24
D.	REPROGRAMMING EFFORT	25
V.	CONCLUSION	27

TABLE OF AUTHORITIES

Cases

<i>Allied Erecting and Dismantling Co. v. Genesis Attachments, LLC</i> , 825 F.3d 1373 (Fed. Cir. 2016)	21
<i>Antor Media Corp.</i> , 689 F.3d 1282 (Fed. Cir. 2012)	15
<i>Corephotonics, LTD. v. Apple Inc., et al.</i> , No. 2020-1961, 2021 WL 4944471 (Fed. Cir. Oct. 25, 2021)	21
<i>Elekta Limited v. ZAP Surgical Systems, Inc.</i> , 81 F.4th 1368 (Fed. Cir. 2023)	19
<i>Facebook, Inc. v. Windy City Innovations, LLC</i> , 973 F.3d 1321 (Fed. Cir. 2020)	7
<i>Home Depot U.S.A., Inc. v. Lynk Labs, Inc.</i> , No. IPR2021-01368, Paper 49 (PTAB Jan. 27, 2023)	3
<i>In re Epstein</i> , 32 F.3d 1559 (Fed. Cir. 1994)	5
<i>In re Mouttet</i> , 686 F.3d 1322 (Fed. Cir. 2012)	18
<i>In re Publicover</i> , 813 F. App'x. 527 (Fed. Cir. 2020)	5
<i>In re Wands</i> , 858 F.2d 731 (Fed. Cir. 1988)	15
<i>Intel Corp. v. PACT XPP Schweiz AG</i> , 61 F.4th 1373 (Fed. Cir. 2023)	18, 22
<i>Intel Corp. v. Qualcomm Inc.</i> , 21 F.4th 784 (Fed. Cir. 2021)	22
<i>Keep Truckin, Inc., v. Innovative Glob. Sys., LLC</i> , No. IPR2020-00694, Paper 33 (PTAB July 21, 2021)	9
<i>Keynetik, Inc. v. Samsung Electronics Co., Ltd.</i> , No. 2022-1127, 2023 WL 2003932 (Fed. Cir. 2023)	6
<i>MCM Portfolio LLC v. Hewlett-Packard Co.</i> , 812, F.3d 1284 (Fed. Cir. 2015)	3
<i>Samsung Electronics Co., Ltd. et al. v. Staton Techiya, LLC</i> , IPR2022-00324, Paper 33 (PTAB July 10, 2023)	5
<i>SEB S.A. v. Montgomery Ward & Co.</i> , 594 F.3d 1360 (Fed. Cir. 2010)	9
<i>Sundance, Inc. v. DeMonte Fabricating Ltd.</i> , 550 F.3d 1356 (Fed. Cir. 2008)	9

<i>TQ Delta, LLC v. CISCO Systems, Inc.</i> , 942 F.3d 1352 (Fed. Cir. 2019)	15
<i>Uber Techs., Inc. v. X One, Inc.</i> , 957 F.3d 1334 (Fed. Cir. 2020).....	5
<i>Walmart Inc., et al. v. Power Concepts, LLC</i> , No. IPR2022-00534, Paper 50 (PTAB Sept. 15, 2023)	5

I. INTRODUCTION

Zentian does not dispute the Petition’s combination teaches each of the claim elements. Nor does Zentian dispute the Petition’s level of skill for a POSITA or that Mr. Schmandt (Apple’s expert) qualifies as a POSITA. Zentian instead argues a lack of motivation to combine for various reasons that are either technically incorrect, overly generalized and not responsive to the Petition’s mapping, or are not supported by Federal Circuit caselaw.

II. THE JIANG-CHEN COMBINATION RENDERS OBVIOUS CLAIMS 1(D) AND 1(E)

Zentian incorrectly argues the Petition and Declaration do not “explain how Jiang’s tree search engine 74 *modified in view of Chen* would operate to meet limitation 1(d) and 1(e).” (Paper 19, 29) (emphasis original). The Petition specifies “each of the one or more clusters in the modified *Jiang-Chen* circuit would have been configured by a POSITA to include a plurality of processors described by *Chen* performing the techniques described regarding the tree search engine 74 described by *Jiang*, **replicating the tree search engine’s functionality amongst each of the plurality of *Chen*’s processors.**” (Paper 1, 17-18 (emphasis added), 24 (the combination “**would have stored at least a portion of the acoustic model data** (HMM acoustic model from *Jiang*’s model memory 72) in the shared cluster memories 104a-d” of the *Chen* clusters), 30 (“*each* of the plurality of processors” of the *Chen* cluster “would have been utilized to perform the probability computations

expressly taught by *Jiang* for its single-processor embodiment”); Ex. 1003, 74, 76-77, 83).

Mr. Schmandt confirmed at length during deposition how the *Jiang-Chen* combination operated, repeatedly pushing back on Zentian’s counsel’s question that Ex. 1003, ¶¶ 83-87 did not explain how the *Jiang-Chen* combination operated. (Ex. 2017, 71:6-23, 73:12-17, 44:2-11, 45:10–46:10, 86:25–87:23, 93:7–94:10, **94:16-23, 95:19–96:10**, 97:8-17, 99:17–102:12) (particularly relevant sections bolded). Zentian argues Mr. Schmandt allegedly admitted in deposition his declaration contains no explanation for the combination. (Paper 19, 30-31). Zentian’s counsel confined their questions to only Ex. 1003, ¶¶ 83-87 and whether those paragraphs discussed certain aspects of the combination. *Id.*, citing Ex. 2017, 79:11-23, 81:8-22, 83:15-18. Mr. Schmandt repeatedly answered that he did not mention *Chen* in ¶¶ 83-87, but did discuss the *Chen* combination elsewhere. *Id.*

As mapped and undisputed by Zentian, *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 1, 17-18, 24, 30; Ex. 1003, 74, 76-77, 83). Claim 1 does not require any particular processor from Claim 1(c)’s “plurality of processors” to perform any particular portion or sub-step of the speech recognition steps of Claims

1(d)-(e). Claim 1(d) only specifies the “speech recognition circuit” recited in the preamble performs the steps of Claims 1(d)-(e).

Zentian’s argument turns on bodily incorporation of *Chen’s* architecture. (Paper 19, 29 (arguing Jiang’s tree search engine “would have been entirely modified to use Chen’s clusters of processors, distributed memories, and parallel processing architecture”)). The ““test for obviousness is not whether the features of a secondary reference may be bodily incorporated into the structure of the primary reference.... Rather, the test is what the combined teachings of the references would have suggested to those of ordinary skill in the art.”” *Home Depot U.S.A., Inc. v. Lynk Labs, Inc.*, No. IPR2021-01368, Paper 49, at 16 (PTAB Jan. 27, 2023) (rejecting patent owner’s “attack” on the petition for “not explaining how to bodily incorporate” the references), *citing MCM Portfolio LLC v. Hewlett-Packard Co.*, 812, F.3d 1284, 1294 (Fed. Cir. 2015). Neither the Challenged Claims nor caselaw requires Apple to explain how the combination utilizes Chen’s clusters of processors, memories, etc.

III. ZENTIAN PROVIDES NO EVIDENCE OF A LACK OF REASONABLE EXPECTATION OF SUCCESS

Zentian incorrectly contends a POSITA could never make the proposed combinations and there would be no reasonable expectation of success because the ’140 Patent combines speech recognition and computer architecture technologies that a speech recognition artisan allegedly “would not have been capable of

addressing.” (Paper 19, 13, 12 (contending the combination “would have been highly complex, and far outside the skill of the ordinary artisan” without the “benefit of the ’140 Patent”), 7 (arguing a person meeting Petitioner’s level for a POSITA would be a skilled artisan in either speech recognition or “high performance computing and parallel processing” but “would not be both”))).

A. Zentian’s Artificial Complexity Requirements Are Unclaimed and Undescribed Implementation Details

Zentian identifies several issues a POSITA would need to allegedly address in making the combination, such as “coordinating multiple caches, avoiding memory conflicts, controlling task sharing in an efficient manner, resolving synchronous bottlenecks, and addressing communication bandwidth and latency issues....” (Paper 19, 12-13). These alleged issues are absent from the ’140 Patent Specification and claims, indicating a POSITA would know how to address each.

Both the Federal Circuit and Board have consistently held a POSITA is assumed to have requisite knowledge to address complications arising from a combination or from how to achieve a claimed limitation, especially when the challenged patent’s specification does not correspondingly address such complications or how to achieve the claimed limitation. For example, the Federal Circuit held a specification “entirely silent” on how to achieve a claimed step or function “suggest[s] that a [POSITA] was more than capable of selecting between the known methods of accomplishing” the claimed step or function, noting an

argument that the prior art is “too sparse to adequately explain to a skilled artisan how to modify [the primary reference’s] system” to achieve the claimed limitation fails when the application is “just as sparse on how a system would” achieve the claimed limitation. *Uber Techs., Inc. v. X One, Inc.*, 957 F.3d 1334, 1339 (Fed. Cir. 2020), *cited by In re Publicover*, 813 F. App’x. 527, 531-32 (Fed. Cir. 2020); *In re Epstein*, 32 F.3d 1559, 1568 (Fed. Cir. 1994) (“Rather, the Board’s observation that appellant did not provide the type of detail in his specification that he now argues is necessary in prior art references” supports a finding that a POSITA would have known how to implement the references and would have concluded they were enabling.); *Walmart Inc., et al. v. Power Concepts, LLC*, No. IPR2022-00534, Paper 50, at 31-32, 20-21, 55, 60-61, 115-16, 126 (PTAB Sept. 15, 2023) (rejecting patent owner’s contention the combination “would require modifications that would introduce complications discouraging a [POSITA] from making the modifications[,]” noting the challenged patent “does not describe any techniques to address the complications” and finding “[i]n the absence of a specific description of such techniques in the [challenged patent],” addressing the complications was “within the knowledge of a person of ordinary skill”); *Samsung Electronics Co., Ltd. et al. v. Staton Techiya, LLC*, IPR2022-00324, Paper 33, at 36-38, 29, 42, 63 (PTAB July 10, 2023) (noting prior art references are presumed enabling and include “no

implementation details,” undermining patent owner’s argument the combination would have been beyond a POSITA).

Federal Circuit caselaw also does not require the combination to describe specific software modifications. *Keynetik, Inc. v. Samsung Electronics Co., Ltd.*, No. 2022-1127, Slip Op. at 4, 2023 WL 2003932 (Fed. Cir. 2023) (“Normally, once the function to be performed by software has been identified, writing code to achieve that function is within the skill of the art.”) (not reported in Fed. Rptr.); *cf.* Paper 19, 10-11 (complaining Mr. Schmandt’s software programming opinions were conclusory).

The ’140 Patent does not discuss the alleged complications identified by Zentian. (Paper 19, 12-13). The ’140 Patent has a single sentence regarding a solution to potential “bandwidth” limitations, stating “[b]y providing a plurality of processors in a group with a common memory, flexibility in the processing is provided without being bandwidth limited by the interface to the memory that would occur if only a single memory were used for all processors.” *’140 Patent*, 4:1-9. Zentian notes this citation (Paper 19, 5), yet does not point to any other discussion alleging techniques to address complexities in the claimed clustered processors. The ’140 Patent, 4:1-9 describes no more than what is expressly claimed. Details describing overcoming bandwidth complications are absent from the ’140 Patent.

Dr. Anderson admitted specific details regarding memory structures and cache consistency are not required by Claim 1, stating “implementation details are discussed in the specification. The claim...is not ever expected to have all the implementation details.” (Ex. 2017, *Anderson Depo. Tr.*, 23:6-25 (discussing alleged “difficulty” of coordinating multiple caches), 26:2–27:3 (describing implementation details regarding memory and caches), 27:9-28:2 (admitting such implementation details are not recited)).

Zentian’s alleged complications are generic, not identifying specifically *how* or *why* memory conflicts, latency issues, or other impediments to the combination would arise. A petitioner is not required to predict and proactively explain every unrecited and generic implementation detail. *Facebook, Inc. v. Windy City Innovations, LLC*, 973 F.3d 1321, 1343 (Fed. Cir. 2020) (affirming Board’s rejection of patent owner’s argument “about the difficulty of physically creating the combination,” as the petitioner was not arguing bodily incorporation). Given the lack of any description in the ’140 Patent of challenges (or solutions thereto) for performing known speech recognition techniques on clustered processors, the Board should determine Zentian’s generic complications would have been within a POSITA’s skill to address.

B. Level of Skill

1. Zentian Does Not Dispute the Level of Skill

The Parties agree to the Petition's POSITA skill level. (Paper 10, 8, *citing* Paper 1, 5; Paper 19, 8-9). Zentian does not dispute the level of skill, nor provide an alternative level of skill. (Ex. 2017, 9:20–10:12). Zentian's declarant, Dr. Anderson, did not dispute the Petition was correct:

Q. Okay. Do you dispute that this is a correct level of skill for the 140 and the 319 patents?

A. I do not.

Id. at 10:9-12. Dr. Anderson opined “the Petitioner put forth a person of ordinary skill in the art definition. **It was reasonable.** Somebody with experience in speech recognition. **I believe somebody with those particular backgrounds could implement the patents at issue, and so those are the ones that I used.**” *Id.* at 9:20-13:4 (pinpoint at 12:1-6), 10:10:5-8 (Dr. Anderson agreeing the level of skill includes education and experience), 11:12-17, 13:2-4 (opining the level is reasonable, he adopted it, and is “not proposing anything different”)).

The Board considered the level of a POSITA at Institution, determining the combination of *Jiang-Chen* did not exceed the qualifications of a POSITA. (Paper 10, 8).

2. *Mr. Schmandt's Level of Skill*

Though Zentian does not dispute Mr. Schmandt meets the Petition's level of skill, Zentian complains Mr. Schmandt "lacks an adequate basis from his personal knowledge and experience to opine" on the proposed combination because he has not personally built (nor supervised the building of) the combination's particular architecture. (Paper 19, 9-12, pinpoint at 12). Whether Mr. Schmandt has built or supervised building the claimed invention does not determine whether Mr. Schmandt meets the skill level of a POSITA, whether his opinions are credible, or whether he has personal knowledge.

The Board previously held "with respect to the issue of whether an expert has sufficient qualifications to offer admissible testimony, there is no requirement that the expert's experience be a perfect match to the field of art at issue" or that the expert has experience working on specific projects providing first-hand knowledge of claimed details. *Keep Truckin, Inc., v. Innovative Glob. Sys., LLC*, No. IPR2020-00694, Paper 33, at 12 (PTAB July 21, 2021), *citing* *SEB S.A. v. Montgomery Ward & Co.*, 594 F.3d 1360, 1373 (Fed. Cir. 2010). "To testify as an expert, an individual does not need to be a person of ordinary skill in the art, but rather must be 'qualified in the pertinent art.'" *Id.*, *citing* *Sundance, Inc. v. DeMonte Fabricating Ltd.*, 550 F.3d 1356, 1363-64 (Fed. Cir. 2008). It is sufficient that the expert has the requisite advanced degree and "considerable technical experience related to the "[t]echnical

[f]ield” of the [] patent[.]” *Id.* at 13. Mr. Schmandt having not built the specifically-claimed speech recognizer does not determine his skill level analysis or qualification to testify.

In fact, Mr. Schmandt was educated in multiprocessor architectures, parallelism, concurrency, and memory contention at MIT. (Ex. 1003, 113 (Appendix A – Education and Professional Experience at MIT)). Mr. Schmandt also co-authored a paper in 1993 on hardware architectures for speech recognition entitled *Speech Recognition Architectures for Multimedia Environments*. (Ex. 1003, 113 (Appendix A – Publications)). Mr. Schmandt is also the author of a textbook that includes a chapter entitled “Architectures” discussing various computer speech recognition architectures. (Ex. 1012, 9).

Further, Zentian did not timely object to Mr. Schmandt’s Declaration and therefore cannot move to exclude the Declaration. *Keep Truckin*, No. IPR2020-00694, Paper 33, at 11.

3. Dr. Anderson’s Level of Skill

Zentian’s arguments recite Dr. Anderson’s experience, or lack thereof. Zentian states Dr. Anderson has “never taught a course that would have equipped a masters-level student in speech recognition to apply a parallel processing architecture like Chen’s to known speech recognition techniques.” (Paper 19, 7). Zentian also provides an anecdotal story of Dr. Anderson’s failed attempts in

constructing a parallel computing system for a computer vision project (unrelated to speech recognition) while employed at Texas Instruments. (Paper 19, 1-2, 15-16).

Notwithstanding that Zentian denigrates its own declarant, whether Dr. Anderson has taught a specific course or failed at a parallel processing project is irrelevant to whether a POSITA would have had a reasonable expectation of success. Perhaps Dr. Anderson simply did not have the necessary skills to lead a parallel processing project in an unrelated field. Though Zentian's argument impugns Dr. Anderson's credibility, it does not prove a POSITA would lack the knowledge to reasonably succeed in the combination.

4. *Zentian's "Specialized" Requirements Are Found in the Cited References*

Zentian contends the skilled person would not "have specialized in parallel processing architectures and methods or [sic] high performance computing[.]" (Paper 19, 6-7). "Specializing" in parallel processing architectures or high performance computing is not required of the POSITA definition nor necessary for a reasonable expectation of success. It is unclear from Zentian's argument what experience would encompass *specializing*, and Zentian purposefully provides no alternative POSITA definition for consideration. (Ex. 2017, 11:12-17 (Dr. Anderson opining he provides no alternative skill level definition, the Petition's level of skill "was a reasonable place to start," and he "adopted it")). Regardless, the cited art

establishes the skilled person would have been knowledgeable in both high performance computing and parallel processing, as discussed in Section III.C.

Chen expressly states its disclosed embodiments are “**equally applicable to any type of parallel processing computer system**, regardless of the performance levels of the individual processors and memory components that make up such a system.” *Chen*, 11:33-41, 10:14-35 (disclosing use of commercially-available microprocessors and memory chips); Paper 1, 16-17, 23-24; Ex. 1003, 73-74, 76. *Chen* is thus not only directed to high performance computing. Because *Chen* is analogous art (undisputed by Zentian), a skilled person seeking to utilize parallel processing techniques for speech recognition would have recognized *Chen*’s applicability to “any type of parallel processing computer system.” *Id.*

Chen and *Jiang* (see 4:60-65, 6:39-42) are also directed to parallel processing architectures (using multiple processors). Given *Jiang* comprises multiple processing architectures for speech recognition and is enabling for its teachings, a POSITA would have certainly had knowledge of parallel processing architectures for speech recognition, contrary to Zentian’s assertions. (Paper 19, 6-7).

C. Skilled Persons Knew of Various Speech Recognition Hardware Architectures

Zentian asserts a POSITA “would not have been capable of addressing the complex challenges of combining *Chen* with *Jiang* in the manner the Petition has

proposed[.]” (Paper 19, 13). Evidence shows the exact issues raised by Zentian were being addressed by skilled artisans.

Evaluating hardware architectures for speech recognition was of significant interest to POSITAs as of the 1990s. Speech recognition researchers recognized issues and tradeoffs of speed, accuracy, power, scalability, monetary cost, ease of building, flexibility, expandability, miniaturization, etc. *See, e.g.,* Ex. 1006, *Hon*, 6 (¶ 2), 17-19; Ex. 2021, *Mathew*, 17 (discussing energy consumption and performance tradeoffs).

For example, Ex. 1006 (*Hon*) from 1991 is entitled *A Survey of Hardware Architectures Designed for Speech Recognition*. *Hon* discusses speech recognition architectures, including pipelined and shared memory multiprocessor machines. (Ex. 1006, 6, ¶ 2). *Hon* notes “[o]f course, every architecture has its own pros and cons.” *Id.* *Hon*’s stated purpose is to assist future speech recognition researchers: “We would like this document to serve as a future reference for people who may contemplate building new speech architectures or for those who would like to choose among the architecture alternatives that are available.” *Id.*

In other examples, *Ravishankar* explains a “memory bottleneck” impedes speech recognition parallelization, which may be overcome through techniques such as load balancing and/or processing state selection. (Ex. 1007, 92, 111). Mr. Schmandt’s own textbook provides a brief discussion of several known hardware

computing platforms for speech recognition, including the parallel-processing Sphinx system. (Ex. 1012, *Schmandt*, 151-152).

A POSITA would have been aware of survey articles like *Hon* discussing available hardware and programming tools for resolving complications such as memory conflicts in speech recognizers. The “hypothetical ‘person having ordinary skill in the art’ to which the claimed subject matter pertains would, of necessity have the capability of understanding the scientific and engineering principles applicable to the pertinent art[.]” *Ex parte Hiyamizu*, 10 USPQ2d 1393, 1394 (Bd. Pat. App. & Inter. 1988). A speech recognition POSITA would have knowledge to address “trivial” issues arising in the combination, at least because the references expressly disclose skilled persons were building speech recognizers utilizing parallel processing architectures. (Ex. 1003, 56-60; Ex. 2017, 137:21–143:10 (*Schmandt* opining making the combination is “trivial” and within the POSITA’s knowledge); Ex. 2021, 11 (discussing parallelization techniques for the Sphinx recognizer)).

Zentian does not directly dispute any of Mr. Schmandt’s motivations for the combination, generically arguing Mr. Schmandt did not identify “tasks” or “challenges” in the combination. (Paper 19, 13). Zentian’s assertions are so amorphous as to be meaningless.

Dr. Anderson provides no supporting factual bases, instead simply repeating attorney argument from the Patent Owner Response. *See, e.g.*, Ex. 2020, 24, 27-30,

33-35, 38, 50-53. Dr. Anderson’s conclusory, verbatim opinions should be given no weight. *TQ Delta, LLC v. CISCO Systems, Inc.*, 942 F.3d 1352, 1358-59 (Fed. Cir. 2019) (“This court’s opinions have repeatedly recognized that conclusory expert testimony is inadequate to support an obviousness determination on substantial evidence review.”).

Finally, Zentian does not raise a non-enablement argument nor provide a *Wands* factors analysis. *In re Wands*, 858 F.2d 731, 737 (Fed. Cir. 1988). The references are presumed enabling. *Antor Media Corp.*, 689 F.3d 1282, 1288 (Fed. Cir. 2012) (discussing reference presumed enabling). *Chen* already teaches an operable clustered processor architecture; the combination simply requires utilizing the architecture for speech recognition according to *Jiang’s* multiprocessor technique. Zentian’s alleged complications are specific to a clustered processor architecture and would occur regardless of the information being processed (e.g., for speech recognition or other uses). *Chen* further directly envisions accounting for memory latency. *See, e.g., Chen*, 11:25-32, 12:1-13, 18:24-42. Therefore, addressing memory conflicts or bandwidth latency, for example, are issues *Chen* is already operable to address because *Chen* is enabling.

D. Zentian’s “Inaccessible Memory” Theory

Zentian’s final argument regarding an alleged lack of a reasonable expectation of success regards memory access. (Paper 19, 14-15). Zentian argues some

processors in *Chen's* clusters may not directly or adjacently access memories in other clusters. (Paper 19, 14-15, 1). Per Zentian, because *Jiang* teaches Viterbi decoding, which requires alleged “extensive communication between computational components[,]” *Chen's* clustered processor architecture “is not likely to be successful in a practical speech recognition system.” *Id.* at 14.

In the combination, a portion of the acoustic model is processed by each processor, i.e., each processor operates on a portion of the lexical trees. *See* Section II, *citing* Ex. 1003, 74, 76-77, 83. This processing does not require inter-cluster communication. Once the cumulative probability score is calculated, inter-cluster processing of the results for each tree search would be communicated, but Zentian provides no evidence that such communication would impede processing or require “extensive communication[.]” (Paper 19, 14).

Zentian's argument that information needed by one cluster “**may** be in a cluster that is not directly accessible” misunderstands *Chen's* architecture. *Id.* at 14; *see also* 1 (arguing “Chen does not allow communication between **some** of its processors and memories”) (emphases added). *Chen* envisions its processors and memories **are** directly and adjacently accessible. *Chen*, 1:56–2:1. Zentian's generalized theory that “extensive communication” is required does not address *Chen's* disclosed communication amongst processors. *Id.* Dr. Anderson's opinions

are unhelpful, providing essentially verbatim copy of Zentian's Response without factual basis. (Ex. 2020, 29-30).

Additionally, *Chen's* operable clustered-computing platform already provides infrastructure by which its clusters may communicate, such as the switching node (or "node switching mechanism") of Fig. 4 that "allows for direct and symmetrical access by" processors of one cluster "to the cluster shared memory" of another cluster.

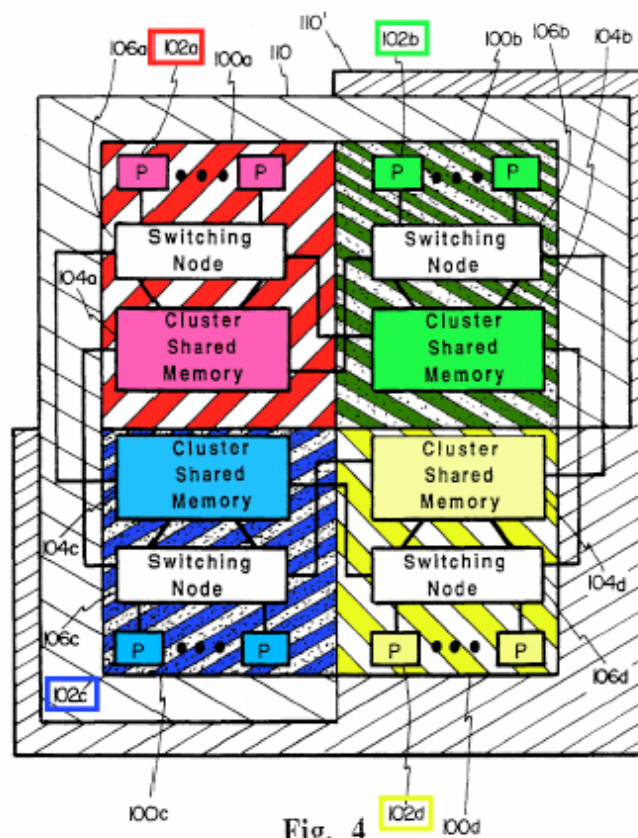


Fig. 4

(Paper 1, 15-18, 39; *Chen*, Fig. 4 (annotated), 10:36-55, 5:54-6:3; *Jiang*, 8:52-64).

Thus, to the extent tree search results of *Jiang's* speech recognition algorithm must

be communicated between clusters, the Petition’s *Jiang-Chen* system provides necessary framework for inter-cluster communication.

IV. ZENTIAN DOES NOT IDENTIFY REASONS FOR NON-OBVIOUSNESS

Zentian does not argue the combination is inoperable, could not be done, or renders the art unsatisfactory for its intended purpose. Instead, Zentian argues a variety of unsupported, generic arguments that contradict Federal Circuit obviousness law.

A. Zentian’s “Quantitative Analysis” Theory

Zentian argues Mr. Schmandt did not perform “quantitative analysis” of real processors and speech recognition systems available at the time of the ’140 Patent. (Paper 19, 17, 19-22). Obviousness does not require building the combination and testing its efficacy: “It is well-established that a determination of obviousness based on teachings from multiple references does not require an actual, physical substitution of elements.” *In re Mouttet*, 686 F.3d 1322, 1332 (Fed. Cir. 2012); *Intel Corp. v. PACT XPP Schweiz AG*, 61 F.4th 1373, 1381 (Fed. Cir. 2023) (finding that petitioner need not show an improvement but rather a “suitable option”). Additionally, “it is not necessary that the inventions of the references be physically combinable to render obvious the invention under review.” *Id.*, citing *In re Keller*, 642 F.2d 413, 425 (CCPA 1981); see also *In re Nievelt*, 482 F.2d 965, 968 (CCPA 1973) (“Combining the teachings of references does not involve an ability to

combine their specific structures.”). Nor is it necessary that the combination be the most optimal, desirable, or preferred. *Elektro Limited v. ZAP Surgical Systems, Inc.*, 81 F.4th 1368, 1374 (Fed. Cir. 2023).

1. Cost Reduction

Mr. Schmandt explained a POSITA would have been motivated to utilize *Chen*’s clusters of simple (cheaper) **commercially-available** processors to improve (reduce) the cost of computing resources, thereby enabling a larger amount of computing resources for a given cost. (Paper 1, 9-10; Ex. 1003, 67, 74). Mr. Schmandt opined “solitary, single-core processors sophisticated enough (for the time) to provide the necessary computing power would incur a high financial cost....” (Ex. 1003, 67). Zentian does not rebut that special-purpose, single-core processors would have been more expensive than multiple, commercially-available processors.

2. Improved Processing Speed

Zentian argues Mr. Schmandt did not “undertake any quantitative or other particularized analysis to demonstrate *why* Chen would allegedly speed up Jiang and relax its pruning threshold, or *by how much*.” (Paper 19, 17-19, 26 (emphasis original); Paper 1, 9-10, 17-19, 24, 31 (providing motivations to combine), *citing* Ex. 1003, 67-68, 71-72, 74, 76; *Hon*, 24-25). Again, Zentian expects “actual, physical substitution of elements[,]” which is not required to establish obviousness.

In re Mouttet, 686 F.3d at 1332. Establishing obviousness does not require the challenger to build the system and then quantify “by how much” improvements in speed or cost are realized. (Paper 19, 17). This is true at least because the combination need not be the most optimal. *Elekta*, 81 F.4th at 1374.

Regardless, Zentian never explains why such alleged slowdown would be present in the combination but not the clustered-processor speech recognition system of Claim 1. (Paper 19, 2, 17-19, 26). Zentian’s new “slower” theory is in direct contradiction to Zentian’s prior declarant, Dr. Wang, who opined using *Chen’s* processing platform was analogous to employing “the engine from a Formula One car” to provide “more speed” to the system. (Paper 6, 13 *citing* Ex. 2002, 30).

Zentian also argues *Mathew* refutes an improvement in speed in the combination. (Paper 19, 18-19). Zentian cites to *Mathew* to contend a 4-processor HMM phase was slower due to high synchronization overhead. *Id.*, *citing* Ex. 2021, 11. Zentian argues “the ordinary artisan could not have assumed” the combination “would have improved the speed” of speech recognition. (Paper 19, 18-19).

Zentian is not actually arguing the combination would not have improved the speed of speech recognition. Instead, Zentian uses *Mathew’s* disclosure of 4-processor high synchronization to argue the POSITA cannot *assume* improved speed with additional processors. Zentian does not respond to the Petition’s mapping of *Chen’s* clustered processors performing *Jiang’s* algorithm with the acoustic model

partitioned amongst the processors. Instead, Zentian takes a single sentence from *Mathew* to assume the combination would have “high synchronization overhead.” (Paper 19, 19). Zentian provides no evidence other than the *Mathew* sentence and Dr. Anderson’s conclusory verbatim opinions. (Ex. 2020, 36-37; *TQ Delta*, 942 F.3d at 1358-59). *Mathew* is a dissimilar architecture, as there is no evidence the *Mathew* multi-processor system was clustered or how its acoustic model was partitioned amongst processors.

3. *Obviousness Includes Evaluating Tradeoffs and Sufficiency of Known Techniques*

Zentian alleges no cost benefit or improved speed, failing to recognize that design decisions often entail tradeoffs between simultaneous advantages and disadvantages, which does not obviate a motivation to combine. *Corephotonics, LTD. v. Apple Inc., et al.*, No. 2020-1961, 2021 WL 4944471, Slip. Op., at 6 (Fed. Cir. Oct. 25, 2021) (“it is a commonplace fact that design decisions entail making tradeoffs among multiple objectives”) (not reported in Fed. Rptr.), *citing Allied Erecting and Dismantling Co. v. Genesis Attachments, LLC*, 825 F.3d 1373, 1381 (Fed. Cir. 2016) (“[a] given course of action often has simultaneous advantages and disadvantages, and this does not necessarily obviate motivation to combine”). The Petition details such a tradeoff analysis between the cost and speed of a speech recognition system. (Paper 1, 9-10). Mr. Schmandt’s declaration testimony that *Chen’s* clustered processors utilizing commercially-available processors strikes a

balance between cost and processing power motivations known to a POSITA is sufficient to support the combination's motivation. (Ex. 1003, 74).

Mr. Schmandt also opined the combination is use of a known technique (e.g., *Chen's* clustering processors with memory) to improve a similar device (*Jiang's* multiprocessor speech recognition circuit) in the same way....” (Ex. 1003, 67-68). Zentian does not rebut this motivation or otherwise respond.

The Federal Circuit has instructed the “known-technique” rationale from *KSR* alone provides motivation: “And if there’s a known technique to address a known problem using ‘prior art elements according to their established functions,’ then there is a motivation to combine.” *Intel Corp. v. PACT XPP Schweiz AG*, 61 F.4th 1373, 1380 (Fed. Cir. 2023), *quoting Intel Corp. v. Qualcomm Inc.*, 21 F.4th 784, 799-800 (Fed. Cir. 2021). A petitioner need only show the combination is a “suitable option.” *Id.*, *quoting Intel*, 21 F.4th at 800.

Here, Zentian does not dispute that clustered architectures were known and advantageous for increased speed at lower cost than a special-purpose processor. Claim 1 of the '140 Patent recites basic, well-known speech recognition steps performed on clustered architectures. Because Zentian's critiques demand hard, quantitative numbers for the improvements of cost and speed—which is not required—and because Zentian does not dispute the known technique rationale, the

motivations provided in the Petition and Declaration provide ample evidence of combination's obviousness.

B. Flexibility and Scalability

Zentian requires the obviousness analysis prove a negative by showing *Jiang* “itself, without modification, was not already both flexible and scalable in the same manner” as the cited art. (Paper 19, 23; *see* Paper 1, 17-19; Ex. 1003, 59, 71-72, 74, *citing Hon*, 24-25). This argument misrepresents an appropriate obviousness analysis. Applying Zentian's logic that *Jiang* already includes a flexible and scalable architecture actually supports the *Chen* combination because the combination allows further scaling of *Jiang's* architecture using clustered processors. Because *Jiang* already teaches a multiprocessor architecture, *Jiang* does in fact teach scalability. (Ex. 1003, 72, *citing Jiang*, 4:51–5:3).

Zentian further incorrectly argues *Hon* does not teach flexibility and scalability for clustered processors. *Hon* discusses the MARS recognizer, expressly stating: “...MARS is scalable because it can be expanded by adding more clusters.” (Ex. 1006, 24, ¶ 3). This expressly states *scalability* is a result of adding more clusters. Zentian once again ignores the Petition's mapping and motivations.

Finally, Zentian does not respond to the other multiple background references provided in discussing the motivations to utilize clustered architectures in speech recognition. (Ex. 1003, 71-72, *citing Tremblay, Wulf, Garde, and Deo*).

C. No Desire to Add Another Processor

Zentian uses an out-of-context excerpt from an answer by Mr. Schmandt to assert incorrectly that Mr. Schmandt testified “it is *never* desirable to add another processor to a speech recognition system when that system can already perform adequately.” (Paper 19, 23-24 (emphasis original), *citing* Ex. 2017, 163:4-12). Mr. Schmandt’s *entire* answer includes an initial discussion regarding undesirability of overcomplicating a **small vocabulary recognizer**, such as a voice dialer with limited address book. (Ex. 2017, 163:4–164:22). Mr. Schmandt goes on to explain in **large vocabulary systems** (such as *Jiang’s*, *see* Paper 1, 18, *citing Jiang*, 2:19-46), one would employ “multiple processors to get real-time performance because there’s a heck of a lot of names that I’d have to search for to compare against the acoustic data....” *Id.*

The Petition established a known problem in the art regarding the need for high-speed processing at an affordable cost, and the combination achieves a low-cost large vocabulary speech recognition system that is higher-speed than *Jiang* alone. (Paper 1, 17-19, 24-25). Thus, the Board should maintain its finding at Institution that “using *Chen’s* clusters of processors to perform *Jiang’s* method of speech recognition provides the speed and processing power needed to relax the pruning threshold while reducing the expense of grouping commercially available processors into clusters.” (Paper 10, 16).

D. Reprogramming Effort

Zentian argues no motivation for the combination because of an alleged “significant reprogramming effort” of porting a software program from a shared memory model to a distributed or private memory model. (Paper 19, 24-26, *citing Chen*, 2:6-9). Zentian’s analysis misunderstands *Chen* and the mapped architecture model, misapplying *Chen*’s teachings. When considered in context, *Chen*’s “significant reprogramming effort” discussion explains Zentian’s cited challenges regarding distributed memory models are **solved** by *Chen*’s architecture, not **caused** by *Chen*’s architecture. Even if the combination included reprogramming efforts, such does not negate obviousness nor is finding the optimal approach required. *Elekta*, 81 F.4th 1368 at 1374. Depending on the goals and tradeoffs of the POSITA, the clustered processor architecture provides increased benefits over a single-processor architecture, which Zentian does not rebut. *Corephotonics*, No. 2020-1961, Slip. Op., at 6.

Additionally, Zentian incorrectly applies *Chen*’s teachings. *Chen* defines both shared and distributed memory models, stating: “In a distributed memory model, data is stored in the private local memory of each processor and is communicated among processors by some type of message passing scheme. In a shared memory model, data is stored in a common shared memory that is equally accessible to all processors in the parallel processing system.” *Chen*, 1:56-62. At *Chen*, 2:6-9, a prior

art architecture is discussed that is a parallel processing system with a *shared* memory model. *Chen*, 1:66–2:1. *Chen* explains shared memory models are easier to program than distributed memory models in some respects because a distributed memory model, the program is divided into separate tasks that are small enough to fit on each processor’s local memory. *Id.* at 2:1-29.

Zentian states *Jiang* is a shared memory model, whereas the *Chen* embodiment mapped in the Petition relies on “multiple private [distributed] memories, with one memory per cluster, in which access to those private memories is extended also to directly adjacent clusters.” (Paper 19, 24). Here, Zentian is incorrect that for each cluster’s memory, access is extended to adjacent clusters (which would then form a distributed memory model among the clusters). In the mapped combination, each cluster’s memory stores a particular portion (context) of the entire lexicon according to *Jiang*’s description, e.g., each cluster’s memory stores acoustic model data in its respective cluster shared memory. (Ex. 1003, 95). To perform speech recognition, each cluster does not need to continually access the other clusters’ memories. Thus, in the mapped combination, there are multiple, clustered processors, with each cluster of processors sharing a single memory. This architecture is a shared memory model per *Chen*’s definition and does not require any significant reprogramming effort. *Chen*, 5:20-25 (“[a] shared memory model

can be used with programs that can be executed in the cluster shared memory of a single cluster”).

Further, *Chen* explicitly teaches its connection nodes 106 “allow for direct and symmetrical access **by all of the processors 102a** to the cluster shared memory 104a[.]” *Chen*, 10:40-43. Therefore, contrary to Zentian’s assertions, the software program would not need to be divided into separate tasks across each cluster’s memory and would not require significant reprogramming effort.

V. CONCLUSION

For the reasons provided above and in the Petition and supporting evidence, Petitioner Apple Inc. requests cancellation of the Challenged Claims.

Respectfully submitted,

ERISE IP, P.A.

BY: /s/ Jennifer C. Bailey
Jennifer C. Bailey Reg. No. 52,583
Adam P. Seitz, Reg. No. 52,206
7015 College Blvd., Suite 700
Overland Park, KS 66211
P: (913) 777-5600
F: (913) 777-5601
jennifer.bailey@eriseip.com
adam.seitz@eriseip.com

ATTORNEYS FOR PETITIONER
APPLE INC.

Exhibit List

Exhibit 1001	U.S. Patent No. 10,971,140 (“’140 Patent”)
Exhibit 1002	Prosecution History for the 10,971,140 Patent (“’140 File History”)
Exhibit 1003	Declaration of Christopher Schmandt (“Dec.”)
Exhibit 1004	U.S. Patent No. 6,374,219 to Jiang (“Jiang”)
Exhibit 1005	U.S. Patent No. 5,428,803 to Chen et al. (“Chen”)
Exhibit 1006	Hsiao-Wuen Hon, <i>A survey of hardware architectures designed for speech recognition</i> , Carnegie Mellon University, 1991 (“Hon”)
Exhibit 1007	Ph.D. Thesis of Mosur Ravishankar (“Ravishankar”)
Exhibit 1008	U.S. Patent Application Publication No. 2001/0053974 to Lucke et al. (“Lucke”)
Exhibit 1009	U.S. Patent No. 5,983,180 to Robinson (“Robinson”)
Exhibit 1010	U.S. Patent No. 5,036,539 to Wrench Jr., et al. (“Wrench”)
Exhibit 1011	Frederick Jelinek, <i>Statistical Methods for Speech Recognition</i> , The MIT Press, 1997 (“Jelinek”)
Exhibit 1012	Christopher Schmandt, <i>Voice Communication with Computers</i> , Van Nostrand Reinhold, 1994 (“Schmandt”)
Exhibit 1013	Lawrence Rabiner and Biing-Hwang Juang, <i>Fundamentals of Speech Recognition</i> , Prentice Hall PTR, 1993 (“Rabiner”)
Exhibit 1014	Richard Klevans and Robert Rodman, <i>Voice Recognition</i> , Artech House, 1997 (“Klevans”)
Exhibit 1015	U.S. Patent No. 7,120,582 to Young et al. (“Young”)
Exhibit 1016	John Holmes and Wendy Holmes, <i>Speech Synthesis and Recognition</i> , 2nd Edition, Taylor & Francis, 2001 (“Holmes”)
Exhibit 1017	U.S. Patent No. 4,926,488 to Nadas et al. (“Nadas”)
Exhibit 1018	U.S. Patent No. 6,182,037 to Maes (“Maes”)
Exhibit 1019	U.S. Patent Application Publication No. 2003/0220796 to Aoyama et al. (“Aoyama”)
Exhibit 1020	U.S. Patent No. 6,092,045 to Stubley et al. (“Stubley”)
Exhibit 1021	U.S. Patent No. 6,151,574 to Lee et al. (“Lee”)
Exhibit 1022	U.S. Patent No. 6,580,814 to Ittycheriah et al. (“Ittycheriah”)
Exhibit 1023	U.S. Patent No. 6,452,348 to Toyoda (“Toyoda”)
Exhibit 1024	William A. Wulf and C.G. Bell, <i>C.mmp—A multi-mini-processor</i> , Carnegie-Mellon University, 1972 (“Wulf”)
Exhibit 1025	Lee D. Erman, Richard D. Fennell, Victor R. Lesser, and D. Raj Reddy, <i>System Organizations for Speech Understanding</i> :

	<i>Implications of Network and Multiprocessor Computer Architectures for AI</i> , IEEE Transactions on Computers, Vol. C-25, No. 4, April 1976 (“ <i>Erman</i> ”)
Exhibit 1026	U.S. Patent No. 6,393,481 to Deo et al. (“ <i>Deo</i> ”)
Exhibit 1027	U.S. Patent No. 6,615,338 to Tremblay et al. (“ <i>Tremblay</i> ”)
Exhibit 1028	U.S. Patent No. 5,922,076 to Garde (“ <i>Garde</i> ”)
Exhibit 1029	Lawrence R. Rabiner, <i>A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition</i> , Proceedings of the IEEE, Vol. 77, No. 2, February 1989 (“ <i>Rabiner 89</i> ”)
Exhibit 1030	Intentionally Left Blank
Exhibit 1031	Intentionally Left Blank
Exhibit 1032	Zentian, Ltd. v Apple Inc. WDTX Waco (6:22-cv-122) Scheduling Order “ <i>Scheduling Order</i> ”
Exhibit 1033	Declaration of June Ann Munford (“ <i>Munford</i> ”)
Exhibit 1034	Declaration of Christina Canino
Exhibit 1035	Deposition Transcript of David V. Anderson, Ph.D. of November 17, 2023

CERTIFICATION OF WORD COUNT

The undersigned certifies pursuant to 37 C.F.R. § 42.24 that the foregoing Petitioner's Reply to Zentian's Response, excluding any table of contents, mandatory notices under 37 C.F.R. § 42.8, certificates of service or word count, or appendix of exhibits, contains 5,598 words according to the word-processing program used to prepare this document (Microsoft Word).

Dated: January 9, 2024

/s/ Jennifer C. Bailey

Jennifer C. Bailey, Reg. No. 52,583

ATTORNEY FOR PETITIONER

CERTIFICATE OF SERVICE ON ZENTIAN
UNDER 37 C.F.R. § 42.6

Pursuant to 37 C.F.R. § 42.6(e), the undersigned certifies that on January 9, 2024, the foregoing *Petitioner Apple Inc.'s Reply to Patent Owner's Response* was served via electronic filing with the Board and via Electronic Mail on the following practitioners of record for Zentian:

Peter C. Knops (peter@noroozipc.com)
Kayvan B. Noroozi (kayvan@noroozipc.com)
Katherine Rhoades (katherine.rhoades@bartlitbeck.com)
Nevin Gewertz (nevin.gewertz@bartlitbeck.com)

/s/ Jennifer C. Bailey

Jennifer C. Bailey, Reg. No. 52,583

ATTORNEY FOR PETITIONER