

Demonstratives of Patent Owner Vervain LLC

Case Nos.: IPR2021-01547, -01548, -01549, -01550
USPTO Patent Trial and Appeal Board

Oral Hearing: January 12, 2023

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

1

The Challenged '298, '385, '240, and '300 Patents

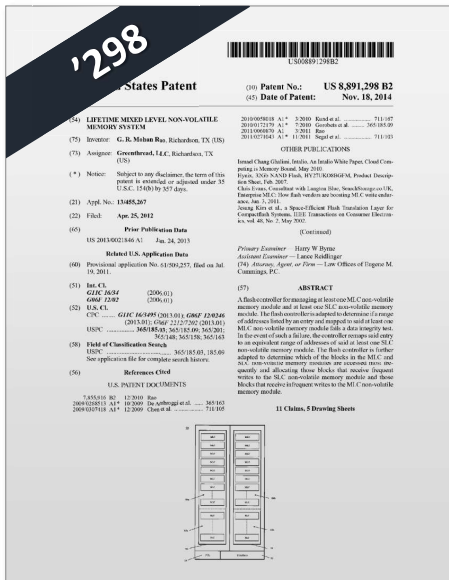
VERVAIN

'298 Patent (-01547)

'385 Patent (DIV of '298) (-01548)

'240 Patent (CON of '385) (-01549)

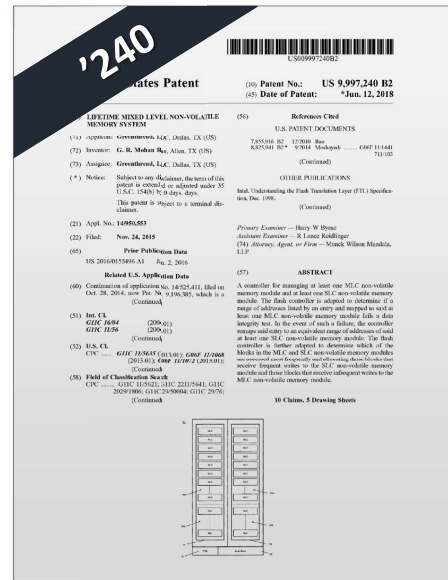
'300 Patent (CON of '240) (-01550)



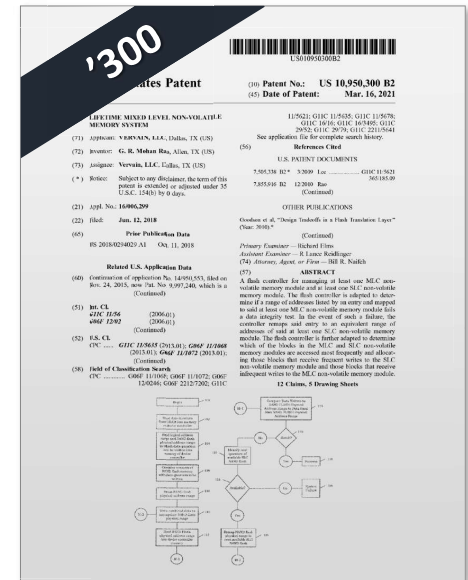
-01547: Ex. 1001 at Cover



-01548: Ex. 1003 at Cover



-1549: Ex. 1005 at Cover

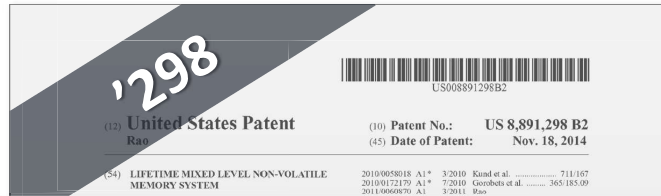


-01550: Ex. 1007 at Cover

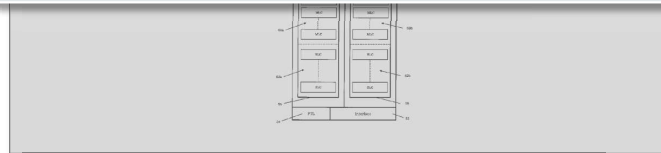
-01547 Pet. at 1; -01548 Pet. at 1; -01549 Pet. at 1; -01550 Pet. at 1

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

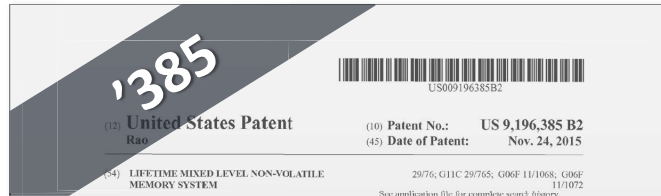
'298 Patent (IPR2021-01547): Instituted Grounds



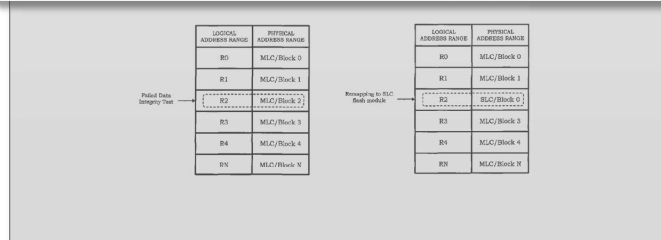
Ground 1	Claims 1-5 and 11 are obvious over Dusija and Sutardja in view of knowledge of POSA
Ground 2	Claims 8-9 are obvious over Dusija, Sutardja, and Li in view of knowledge of POSA
Ground 3	Claims 1-5 and 11 are obvious over Moshayedi and Dusija in view of knowledge of POSA
Ground 4	Claim 11 is obvious over Moshayedi, Dusija, and Sutardja in view of knowledge of POSA
Ground 5	Claims 8-9 are obvious over Moshayedi, Dusija, and Li in view of knowledge of POSA



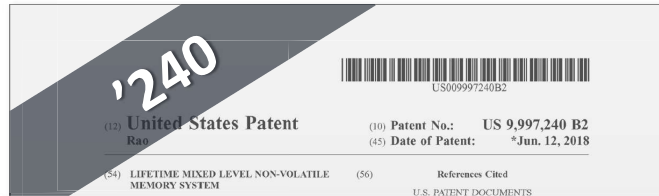
'385 Patent (IPR2021-01548): Instituted Grounds



Ground 1	Claims 1-5 and 11-13 are obvious over Dusija and Sutardja in view of knowledge of POSA
Ground 2	Claims 1-5 and 11-13 are obvious over Moshayedi and Dusija in view of knowledge of POSA
Ground 3	Claim 11 is obvious over Moshayedi, Dusija, and Sutardja in view of knowledge of POSA



'240 Patent (IPR2021-01549): Instituted Grounds

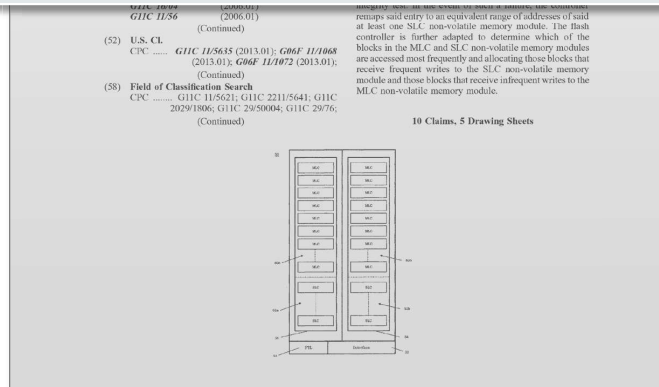


Ground 1

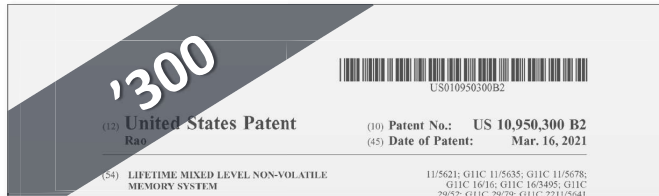
Claims 1-2 and 6-7 are obvious over Dusija and Sutardja in view of knowledge of POSA

Ground 2

Claims 1-2 and 6-7 are obvious over Dusija, Sutardja, and Chin in view of knowledge of POSA



'300 Patent (IPR2021-01550): Instituted Grounds

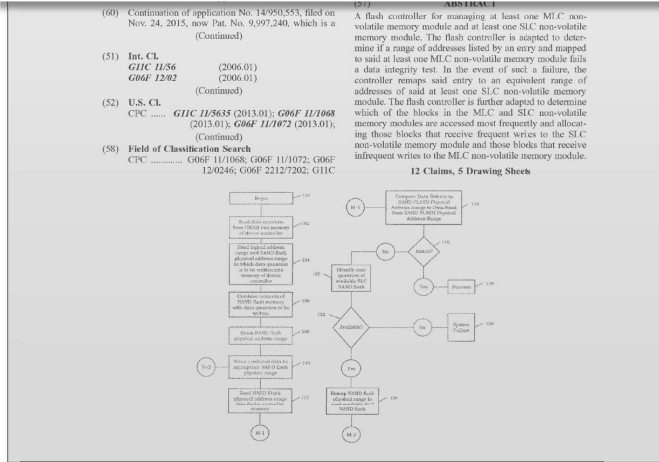


Ground 1

Ground 2

Claims 1-9 and 11-12 are obvious over Dusija in view of knowledge of POSA

Claim 10 is obvious over Dusija and Sutardja in view of knowledge of POSA

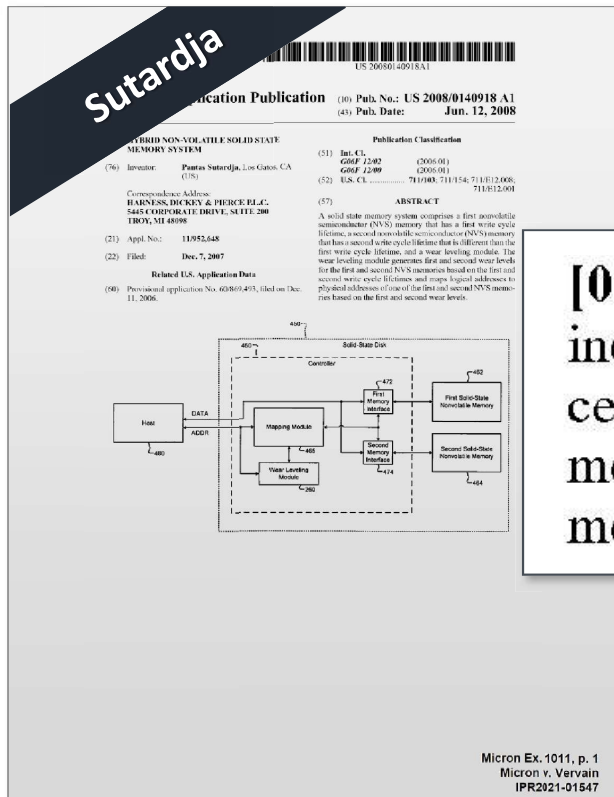


	1. Micron's Expert is not credible on several issues
○	1.1 Sutardja's First and Second Memories
	1.2 Dr. Rao's Controller
	1.3 Dusija's Preferred Embodiment
	2. '298 patent (IPR2021-01547)
	3. '385 patent (IPR2021-01548)
	4. '240 patent (IPR2021-01549)
	5. '300 patent (IPR2021-01550)



“If the Board finds he gave inconsistent testimony, the Board shall consider the impact on the specific patents at issue in the trial testimony *as well as* on his credibility as a whole.”

Ultratec, Inc. v. CaptionCall, LLC, 872 F.3d 1267, 1275 (Fed. Cir. 2017) (emphasis in original)



[0108] The first solid-state nonvolatile memory **204** may include single-level cell (SLC) flash memory or multi-level cell (MLC) flash memory. The second solid-state nonvolatile memory **206** may include single-level cell (SLC) flash memory or multi-level cell (MLC) flash memory.

– 01547: Ex. 1011 (Sutardja) at ¶ 0108

1.1 Dr. Liu's Testimony Regarding Sutardja's First and Second NVS Memories VERVAIN

Dr. Liu

Q. I understand that's what you're trying to establish, but you haven't provided support in Sutardja, and you haven't explained any such support, as to the first or the second memory being exclusively MLC or SLC; correct?

A. I believe I stated very clear, and I believe that a POSA looking at it will find it obvious. And so, to the extent -- to the extent -- to be quite honest, it is my opinion that the POSA will find it obvious, and **that's the only way**, that the first NVS is MLC and second NVS SLC.

To try to say it other- -- it's not obvious to a POSA -- **to try to say otherwise I don't believe is credible**. I believe it's not -- it's not genuine. Yeah.

– 01547: Ex. 2020 at 99:1-17

1.1 Dr. Liu's Testimony Regarding Sutardja's First and Second NVS Memories VERVAIN

Dr. Liu

Q. In your previous response, you said "that's the only way," regarding first MLC, second SLC, and you also said "to try to say otherwise I don't believe is credible."

Do you stand by those statements?

A. What I meant -- I didn't say --

I did not use the way it worked. I'm just saying that for the conclusion, given all the information in the conclusion, that is the most obvious. **Maybe I will take -- take back the "only."** That's the most obvious. First --

– 01547: Ex. 2020 at 101:3-16

1.1 Dr. Liu's Testimony Regarding Sutardja's First and Second NVS Memories VERVAIN

Dr. Liu

- Q. And is an implementation in Sutardja of first memory being SLC and second memory being MLC a credible implementation, in your opinion?
- A. I would not use the word "credible." **I should not have used the word "credible."** But I will say that – because the statement 108 clearly says "may," but that statement has to be taken in context with other things. We will actually run into certain -- I wouldn't even use the word "contradiction," but I would say that it may not jibe with the general knowledge of the characteristic of SLC and MLC.

– 01547: Ex. 2020 at 128:6-19

1.1 Dr. Liu's Testimony Regarding Sutardja's First and Second NVS Memories VERVAIN

Dr. Liu

Q. Putting aside what embodiment you consider most obvious, I'm asking what Sutardja discloses. Do you understand there is a difference between something being disclosed and something being obvious?

A. So, that's fine, but I'm just saying that Sutardja also described -- **described the first memory being MLC and second memory being SLC.**

Q. And Sutardja describes the reverse; correct?

A. **Sutardja never explicitly described the reverse.** Sutardja say "may," "may," but never explicitly, whereas Sutardja does explicitly mention that first NVS being SLC and second NVS being SLC.

– 01547: Ex. 2020 at 112:16-113:8

1.1 Dr. Liu's Testimony Regarding Sutardja's First and Second NVS Memories VERVAIN

Dr. Liu

- A. So, to make sure I explain clearly, so if I make the first one -- let's say for Vervain the sake of argument, just render my -- if I make the first one SLC, and the second one necessarily will have to be S -- or **if I make the first one SLC, the second one necessarily would have to be MLC.** And vice-versa, I can also crisscross -- if the first one is MLC, the second one is SLC.

So, **this is almost like a permutation you can pick,** but you can only pick one for the first one, and the other, the second one will have to be what -- it cannot be the same as the first one. Otherwise, it will not be a hybrid system. The second one would have to be the other memory type.

– 01547: Ex. 2020 at 198:24-199:15

1.1 Dr. Liu's Testimony Regarding Sutardja's First and Second NVS Memories VERVAIN

Dr. Liu

154. Sutardja discloses a “data shift analysis.” Sutardja at [0148]-[0149]. When time to perform the analysis has arrived, Sutardja discloses determining if “a number of write operations to a first block of the first NVS memory [*i.e.*, MLC] during a predetermined time is greater than or equal to a predetermined threshold” (“*blocks that receive the most frequent writes*”). *Id.* at [0149], Figure 7C, [0106], claim 37 (first memory is MLC; second memory is SLC). Next, if the number of writes is greater than the threshold, the “control maps the logical addresses that correspond to the first block to a second block of the second NVS memory [*i.e.*, SLC memory module] in step 522” (“*allocate*” and “*by transferring the respective contents of those blocks to the at least one SLC non-volatile memory module*”). *Id.* at [0149].

– 01547: Ex. 1009 (Liu Decl.) at ¶ 154

Dr. Liu

- Q. You talked about going around a lot of these clues that Sutardja gave. You didn't discuss any of that in your original declaration for limitation 1.G; correct?
- A. I believe we went through this, and in the deposition I asked for opportunity to explain, and I believe in this case, I -- again, let me just state simply, I described the first memory is MLC, second memory is SLC, and as a POSA, **this is the obvious implementation.** Okay?

– 01547: Ex. 2020 at 106:19-107:5

1.1 Dr. Liu's Testimony Regarding Sutardja's First and Second NVS Memories VERVAIN

Dr. Liu

- A. And I will be happy to walk through the logical deduction, but -- so, **my statement in the first declaration simply is of that, that is the most obvious to** -- from Sutardja's teaching, that the first memory is MLC and second memory is SLC. I stated clearly.

– 01547: Ex. 2020 at 100:21-101:2

Dr. Liu

Q. In your original declaration, you did not describe anything being obvious about exclusively having one MLC and another memory SLC; correct?

A. **Perhaps I give credit too much to other people thinking the same** -- thinking of the most obvious way, but I disclose clearly first memory is MLC, second memory is SLC.

– 01547: Ex. 2020 at 107:20-108:3

	1. Micron's Expert is not credible on several issues
	1.1 Sutardja's First and Second Memories
	1.2 Dr. Rao's Controller
	1.3 Dusija's Preferred Embodiment
	2. '298 patent (IPR2021-01547)
	3. '385 patent (IPR2021-01548)
	4. '240 patent (IPR2021-01549)
	5. '300 patent (IPR2021-01550)

1.2 Controller in the '298 Patent



block address (LBA) used by the host system. The controller's wear-leveling algorithm determines which physical block to use each time data is programmed, eliminating the relevance of the physical location of data and enabling data to be stored anywhere within the memory array and thus prolonging the service life of the flash memory. Depending on the wear-leveling method used, the controller typically either writes to the available erased block with the lowest erase count (dynamic wear leveling); or it selects an available target block with the lowest overall erase count, erases the block if necessary, writes new data to the block, and ensures that blocks of static data are moved when their block erase count is below a certain threshold (static wear leveling).

-01547: Ex. 1001 at 3:1-13

Dr. Liu

- Q. The sentence which you just read aloud describes the controller using a physical block; correct?
- A. It's -- to a POSA, it conveys much more, but the controller's wear leveling algorithm, the wear leveling algorithm is based on a mapping of logical block to physical block, and knowing -- and then also keeping track of -- of the counts to a logical block as well as to a physical block, and then based on that information, the controller will decide the logical block mapping to the physical block and which physical block to use when the data is programmed.

– 01547: Ex. 2020 at 41:8-22

1.2 Dr. Liu's Testimony Regarding the '298 Specification

VERVAIN

Dr. Liu

- Q. So, to summarize, the controller does determine which physical block to use; correct?
- A. My -- my point is that if you read the sentence, "eliminating the relevance of the physical location of data," so controller sees -- controller use logical block address, and based on the logical block address, you have different wear leveling algorithm, and that's taught quite well by the prior art, and you count both the access to the logical block, and you also count the access to the physical block, and you make a determination, first, this logical block, that's where I am going to map to the physical block, such that -- so that the physical location of data is - is -- the relevance of the physical location of data is eliminated is purely based on the wear leveling, and based on the connection between the logical block and physical block.

- 01547: Ex. 2020 at 41:23-42:18

Dr. Liu

Q. I don't think you answered my question, Dr. Liu. I am basically noting that the sentence starting at column three, line one, literally says that the controller determines which physical block to use, and I am asking you if you agree with that.

Do you agree with the sentence saying that the controller determines which physical block to use? Yes or no, please.

A. **I disagree.** The sentence says "the controller's wear leveling algorithm," and I am trying to provide the context of the wear leveling algorithm.

- 01547: Ex. 2020 at 42:19-43:6

Dr. Liu

Q. In that context, yes, or in that context, no?

A. In the context that it is the controller's wear leveling algorithm that determines, therefore if -- see, so you can -- under that context, you can say that **the controller determines.**

Q. Determines what?

A. **Which physical block to use.** So, it's -- it's the wear leveling algorithm of the controller that determines which physical block to use. How about that?

– 01547: Ex. 2020 at 45:13-24

Dr. Liu

- Q. Yes. I see that's what the sentence says. And because the wear leveling algorithm is the controller's wear leveling algorithm, the controller is determining which physical block to use; correct?
- A. **That's not what the specification says**, and I -- I think I answered the question very clear. The controller's wear leveling algorithm determines, so the wear leveling algorithm of the controller determines. So, that's the extent, and if you want to add something to it, I stand by my testimony.

– 01547: Ex. 2020 at 45:25-46:12

Dr. Liu

- Q. Would a person of ordinary skill have understood the controller of the '298 patent to be determining which physical block to use? Yes or no, please. Would they have that understanding or would they not?
- A. **It's not a yes-or-no question**, because you are qualifying -- in that case, if you put a lot of qualifying context, you would have to say that the controller would have to have the proper wear leveling algorithm.

– 01547: Ex. 2020 at 47:14-25

Dr. Liu

Q. You still haven't answered the question, which seems to be apparent from the words on the page. So, I will give you another chance to answer it, and then I'm going to move on.

Are you disagreeing with the controller determining which physical block to use each time data is programmed?

A. Again, let me say that the wear leveling algorithm of the controller determines which physical block to use.

– 01547: Ex. 2020 at 48:12-23

	1. Micron's Expert is not credible on several issues
	1.1 Sutardja's First and Second Memories
	1.2 Dr. Rao's Controller
	● 1.3 Dusija's Preferred Embodiment
	2. '298 patent (IPR2021-01547)
	3. '385 patent (IPR2021-01548)
	4. '240 patent (IPR2021-01549)
	5. '300 patent (IPR2021-01550)

1.3 Dr. Liu's Testimony Regarding Dusija ('300 IPR)

VERVAIN

Dr. Liu

Q. And you have not stated in either of your declarations, your original declaration or your reply declaration, that Dusija's preferred embodiment discloses caching data in the flash memory?

Correct?

A. I testified that the **preferred embodiment, based on the language, is under the alternative embodiment.** When Dusija introduced figure to illustrate first memory being used to store the data, Dusija used the word as an "alternative embodiment." I'm simply referring to that. And then after that, there are statements to the effect that **preferred embodiment under the context of the alternative embodiment.**

– 01550: Ex. 2020 at 12:4-20

1.3 Dr. Liu's Testimony Regarding Dusija ('300 IPR)

VERVAIN

Dr. Liu

Q. Can there be an alternative embodiment under a preferred embodiment?

A. If it is alternative embodiment under a preferred embodiment, then I would think that it would be specified as the preferred embodiment under preferred embodiment, or the first embodiment under the preferred embodiment.

I am not sure the context of "alternative." It seems like "alternative" is a -- is -- the word "alternative" means it's -- it has connotation of being a replacement, being a **replacement for the primary embodiment.**

– 01550: Ex. 2020 at 13:8-22

1.3 Dr. Liu's Testimony Regarding Dusija ('300 IPR)

VERVAIN

Dr. Liu

- Q. I'm trying to understand your terminology, Dr. Liu. Is primary embodiment the same as preferred embodiment in your testimony?
- A. As I testified in my first deposition, there are primary embodiments under the context of alternative embodiment -- there are preferred embodiments under the context of alternative embodiment. Let me just make sure the record is clean. I'm sorry for messing up.

To the extent **I'm not here trying to -- to define "primary."** I'm simply trying to -- at the time, trying to help us avoid the confusion, because I was using the word and you were using the word "preferred." I was just trying to categorize which preferred embodiment was under the context of alternative embodiment.

– 01550: Ex. 2020 at 18:1-11; 18:21-19:3

1.3 Dr. Liu's Testimony Regarding Dusija ('300 IPR)

VERVAIN

Dr. Liu

Q. You spoke a few minutes ago about a replacement for the primary embodiment as being an alternative embodiment; correct?

A. I'm not here to define "alternative." I'm sorry.

I'm not here to define "alternative." I'm simply saying that Dusija used "alternative." That means that it is not the primary ...

- 01550: Ex. 2020 at 14:21-15:7

1.3 Dr. Liu's Testimony Regarding Dusija ('300 IPR)

VERVAIN

Dr. Liu

Q. Is there in Dusija a preferred embodiment which is not under an alternative embodiment?

A. Let's go to Dusija.

If you read paragraph 18, there is talk about the general aspect of the invention. It talks about data is written to the second portion. Afterward, the data is read back. Okay. It's read back to check for errors. Reading back. So, that is mentioned, and it is also mentioned in the context of ECC. Okay.

So, this is mentioned first and foremost with a read back, and then the paragraph after that, "in an alternative embodiment" that is mentioned. So, there is a general - for the lack of better word, general embodiment or general scope of the invention, which involves read back, and then there is this "in an alternative embodiment" of first memory used to store incoming data.

That is my testimony, and that is the basis of my declaration and my deposition -- and my testimony in my deposition.

- 01550: Ex. 2020 at 29:6-30:7

1.3 Dr. Liu's Testimony Regarding Dusija ('300 IPR)

Dusija

US 201009460A1

Application Publication (10) Pub. No.: US 2011/0099460 A1
(43) Pub. Date: Apr. 28, 2011

NONVOLATILE MEMORY AND METHOD WITH POST-WRITE READ AND ALTERNATE RE-WRITE TO MANAGE ERRORS

(75) Inventors: **Gautam Ashik Dusija**, Milpitas, CA (US); **Frank Chang**, Menlo Park, CA (US); **Chirukunuru Sumanvulu**, CA (US); **Jianming Huang**, Sunnyvale, CA (US); **Lee M. Gavron**, Milpitas, CA (US)

(21) Appl. No.: 12/942,728
(22) Filed: Dec. 18, 2009

Related U.S. Application Data
(63) Continuation-in-part of application No. 12/607,522, filed on Oct. 26, 2009.

Publication Classification
(51) Int. Cl. G11C 29/32 (2006.01); G06F 13/10 (2006.01); G06F 12/16 (2006.01); G06F 12/00 (2006.01)
(52) U.S. CL. 714773; 714105; 714113; 714112; 105; 714112; 001; 714111; 04

ABSTRACT
Data errors in non-volatile memory inevitably increase with usage and with higher density of bits stored per cell. The memory is configured to have a first portion operating with low error rate of low density storage, and a second portion operating with a higher density but less robust storage. Input data is written and staged in the first portion before being copied to the second portion. An error management module provides checking the quality of the copied data for excessive error bits. The copying and checking are repeated on a different location in the second portion until either a predetermined quality is attained or the number of repeats exceeds a predetermined limit. The error management is not started when a memory is new with little or no errors, but started after the memory has aged to a predetermined amount as determined by the number of erase-program cycling it has experienced.

Micron Ex. 1010, p. 1
Micron v. Vervain
IPR2021-01550

[0019] In one preferred embodiment, the first portion has each memory cell storing one bit of data and the second portion has each memory cell storing more than one bit of data.

[0020] In an alternative embodiment, the first portion serves as a cache for incoming data, so a cache copy of the input data is programmed into the cache. Then a first copy of data is programmed into the second portion. If the post-write read has not detected an excessive amount of error in the first copy, the first copy will be deemed valid and subsequent access will be directed to access the first copy. On the other hand, if the post-write read has detected an excessive amount of error in the first copy, the cached copy in the first portion will replace the first copy in the second portion as valid data. The first copy will become obsolete and subsequent access will be directed to the cached copy.

[0021] In the preferred embodiment, the first portion is further provided with a first section and a second section. The incoming data is cached in the first section of the first portion and a first copy of the data is written to the second portion. Afterwards, the first copy in the second portion is read back to check for excessive error bits. If the error bits exceeded a predetermined amount, a second copy of the incoming data is written to the second section of the first portion.

– 01550: Ex. 1010 (Dusija) at ¶¶ 0019-0021

1.3 Dr. Liu's Testimony Regarding Dusija ('300 IPR)

VERVAIN

Dr. Liu

- Q. Dr. Liu, you agree that when "the preferred embodiment" is mentioned in paragraph 21, there are only two possibilities. It either means the same thing as "one preferred embodiment" of paragraph 19, or it means a different thing. You agree those are the only two possibilities; correct?
- A. The reason -- Counsel, I'm trying to -- it's my job as an expert not only to answer questions truthfully, but also to educate and provide the proper context. I'm -- the reason some of the questions are a little difficult is because they may be taken out of context.

– 01550: Ex. 2020 at 38:18-39:11

Dr. Liu

Q. Is paragraph 21 further narrowing details regarding paragraph 19?

A. Under the context of paragraph -- you have paragraph 19, and you have an alternative embodiment, and -- and **within the alternative embodiment, you have the preferred embodiment**, which is paragraph 21.

– 01550: Ex. 2020 at 42:13-19

1.3 Dr. Liu's Testimony Regarding Dusija ('300 IPR)

VERVAIN

Dr. Liu

Q. Is there in Dusija a preferred embodiment which is not under an alternative embodiment?

A. Let's go to Dusija.

So, this is mentioned first and foremost with a read back, and then the paragraph after that, "in an alternative embodiment" that is mentioned. So, there is a general -- for the lack of better word, general embodiment or general scope of the invention, which involves read back, and then there is this "in an alternative embodiment" of first memory used to store incoming data. That is my testimony, and **that is the basis of my declaration and my deposition -- and my testimony in my deposition.**

– 01550: Ex. 2020 at 29:6-11, 29:20-30:7

1.3 Dr. Liu's Testimony Regarding Dusija ('300 IPR)

Dr. Liu

64. I understand Patent Owner to argue that “an off-chip comparison would have been contrary to well-known design principles regarding fast cache operation.” POR, 55. I note that although Patent Owner cites Dr. Khatri’s declaration for support, Dr. Khatri makes no such statement. *Id.*, 55 (citing Khatri Decl., ¶ 97). Specifically, Dr. Khatri states that “such an approach,” namely, performing an *on-chip comparison*, “would have been understood by a POSA to be *consistent* with well-known design criteria regarding cache design and chip operation/design.” Khatri Decl., ¶ 97. But Dr. Khatri does not say the converse, namely, that an *off-chip comparison* would have been *contrary* to well-known design principles. And for good reason: a POSA would have known that off-chip data verification was typical in flash memory devices.

– 01550: Ex. 1057 (Liu Reply Decl.) at ¶ 64

1.3 Dr. Khatri's Testimony Regarding Dusija ('300 IPR)

Dr. Khatri

98. As a POSA would have well understood at the time, a primary concern for caching operations is speed. *See, e.g.*, Ex. 2017, 1:63-2:1. A POSA would also know that moving data between components takes time, introducing unwanted delays. Ex. 2017, 2:33-35 (“the use of external cache memory ... compromises the main storage access speed”). Because relying on external chips for caching can “lead[] to a complicated and slower system” (*id.*, 5:20-23), one solution known to a POSA at the time would be to integrate the cache and main memory on a single device. Ex. 2017, 4:42-43, 6:14-16.

– 01550: Ex. 2014 (Khatri Decl.) at ¶ 98

1.3 Dr. Liu's Testimony Regarding Dusija ('300 IPR)

VERVAIN

Dr. Liu

Q. You don't think a primary concern for caching operations is speed?

A. In the context of flash memory cache. Also --

Q. Is that a yes or a no, please? I can't tell if you are agreeing or disagreeing.

A. I do not -- **I do not agree necessarily the primary concern. I say one concern.** You have to present it in the whole picture with everything under consideration.

– 01550: Ex. 2020 at 169:8-20

1.3 Dr. Liu's Testimony Regarding Dusija ('300 IPR)

Dr. Liu

Q. So, is it your opinion that whether the cache should be fast depends on the type of memory, in terms of whether it is flash memory or not flash memory?

A. Let me answer this way. If compared with MLC and SLC, if I use SLC to cache for MLC, yes, SLC is faster in terms of caching for MLC. That's correct.

Now, if I am using RAM as caching for non-volatile memory, then RAM is faster -- RAM has faster caching speed than non-volatile memory.

So, everything has context. So, in the -- in the paragraph 98, Dr. Khatri mentioned caching speed, and it's not in the context of flash memory caching.

Now, I can take that caching --

Q. Dr. Khatri's -- I'm sorry. Go ahead. I didn't know you weren't done.

A. I can take the caching in many, many comparison. I just said it. Between SLC/MLC, SLC could be used as MLC's cache, because SLC is faster than MLC. Now, in terms of between RAM and non-volatile memory, in that case RAM would be better caching speed than a non-volatile memory. That's why we -- I said it already. That's why we use RAM as buffer in the controller for the data that comes in from the host very fast, and cache memory cannot take it right away, and that's why we cache it.

So, I agree a primary concern for caching operation is speed, but it has to have context.

...

-- 01550: Ex. 2020 at 178:2-179:12

1.3 Dr. Liu's Testimony Regarding Dusija ('300 IPR)

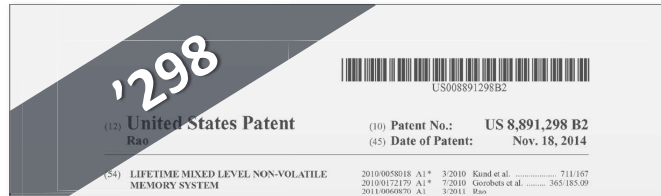
VERVAIN

Dr. Liu

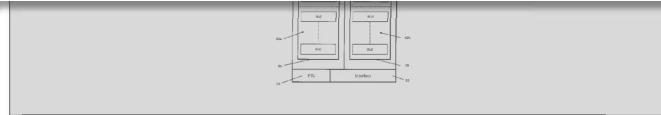
- Q. You stated, "I agree a primary concern for caching operation is speed," in your previous response; correct?
- A. I'm saying **the statement of a primary concern for caching operation is speed, that statement is true**, but it has to have a context of relative -- relativity, in terms of comparison.

– 01550: Ex. 2020 at 180:1-8

'298 Patent (IPR2021-01547): Instituted Grounds



Ground 1	Claims 1-5 and 11 are obvious over Dusija and Sutardja in view of knowledge of POSA
Ground 2	Claims 8-9 are obvious over Dusija, Sutardja, and Li in view of knowledge of POSA
Ground 3	Claims 1-5 and 11 are obvious over Moshayedi and Dusija in view of knowledge of POSA
Ground 4	Claim 11 is obvious over Moshayedi, Dusija, and Sutardja in view of knowledge of POSA
Ground 5	Claims 8-9 are obvious over Moshayedi, Dusija, and Li in view of knowledge of POSA



1. Micron's Expert is not credible on several issues
2. '298 patent (IPR2021-01547)
● 2.1 Claim construction for "blocks"
2.2 Dusija-Sutardja does not disclose or suggest determining which of the blocks are accessed most frequently (limitation [1.F])
2.3 Dusija-Sutardja does not disclose or suggest transferring contents of blocks to SLC (limitation [1.G])
2.4 Moshayedi-Dusija does not disclose or suggest transferring contents of blocks to SLC (limitation [1.G])
2.5 Moshayedi's logical block addresses do not disclose or suggest "blocks" (limitations [1.F]-[1.G])
3. '385 patent (IPR2021-01548)
4. '240 patent (IPR2021-01549)
5. '300 patent (IPR2021-01550)

'298

1. A system for storing data comprising:

at least one MLC non-volatile memory module comprising a plurality of individually erasable blocks;

at least one SLC non-volatile memory module comprising a plurality of individually erasable blocks; and

a controller coupled to the at least one MLC non-volatile memory module and the at least one SLC non-volatile memory module wherein the controller is adapted to:

- a) maintain an address map of at least one of the MLC and SLC non-volatile memory modules, the address map comprising a list of logical address ranges accessible by a computer system, the list of logical address ranges having a minimum quanta of addresses, wherein each entry in the list of logical address ranges maps to a similar range of physical addresses within either the at least one SLC non-volatile memory module or within the at least one MLC non-volatile memory module;

- b) determine if a range of addresses listed by an entry and mapped to a similar range of physical addresses within the at least one MLC non-volatile memory module, fails a data integrity test, and, in the event of such a failure, the controller remaps the entry to the next available equivalent range of physical addresses within the at least one SLC non-volatile memory module;
- c) determine which of the blocks of the plurality of the blocks in the MLC and SLC non-volatile memory modules are accessed most frequently by maintaining a count of the number of times each one of the blocks is accessed; and
- d) allocate those blocks that receive the most frequent writes by transferring the respective contents of those blocks to the at least one SLC non-volatile memory module.

-01547: Ex. 1001 at Claim 1

2.1 PO's Construction of "Blocks" Should Be Adopted

"Blocks"

Patent Owner

"In the context of the '298 Patent, a 'block' (singular form of the plural 'blocks' recited in claim 1) should be construed as 'in a non-volatile memory, a physical group of memory cells that must be erased together.'"

-01547: Resp. at 23

Petitioner

"For good reason, the Board already rejected PO's attempt to limit 'blocks' to 'physical blocks.'"

-01547: Reply at 2

Dr. Khatri

31. The primary difference between SLC and MLC is what data each threshold voltage is interpreted to represent. With SLC flash, the transistor stores only a 1 or 0, so a wide range of threshold voltages can be allotted to a single bit. This allows for faster and more reliable memory access. On the other hand, MLC flash must be slowly and carefully programmed using a narrow, precise range of threshold voltages, with each threshold voltage range representing a specific sequence of bits (see the figure above, which shows four sequences of bits—11, 10, 01, and 00—corresponding to different ranges of threshold voltages). This results in a less reliable memory as well. MLC stores more bits per transistor, and is hence more dense.

– 01547: Ex. 2014 (Khatri Decl.) at ¶ 31

2.1 Controller in Claim 1 of '298 Patent

'298

1. A system for storing data comprising:
- at least one MLC non-volatile memory module comprising a plurality of individually erasable blocks;
 - at least one SLC non-volatile memory module comprising a plurality of individually erasable blocks; and
 - a controller coupled to the at least one MLC non-volatile memory module and the at least one SLC non-volatile memory module wherein the controller is adapted to:
 - a) maintain an address map of at least one of the MLC and SLC non-volatile memory modules, the address map comprising a list of logical address ranges accessible by a computer system, the list of logical address ranges having a minimum quanta of addresses, wherein each entry in the list of logical address ranges maps to a similar range of physical addresses within either the at least one SLC non-volatile memory module or within the at least one MLC non-volatile memory module;
 - b) determine if a range of addresses listed by an entry and mapped to a similar range of physical addresses within the at least one MLC non-volatile memory module, fails a data integrity test, and, in the event of such a failure, the controller remaps the entry to the next available equivalent range of physical addresses within the at least one SLC non-volatile memory module;
 - c) determine which of the blocks of the plurality of the blocks in the MLC and SLC non-volatile memory modules are accessed most frequently by maintaining a count of the number of times each one of the blocks is accessed; and
 - d) allocate those blocks that receive the most frequent writes by transferring the respective contents of those blocks to the at least one SLC non-volatile memory module.

-01547: Ex. 1001 at Claim 1

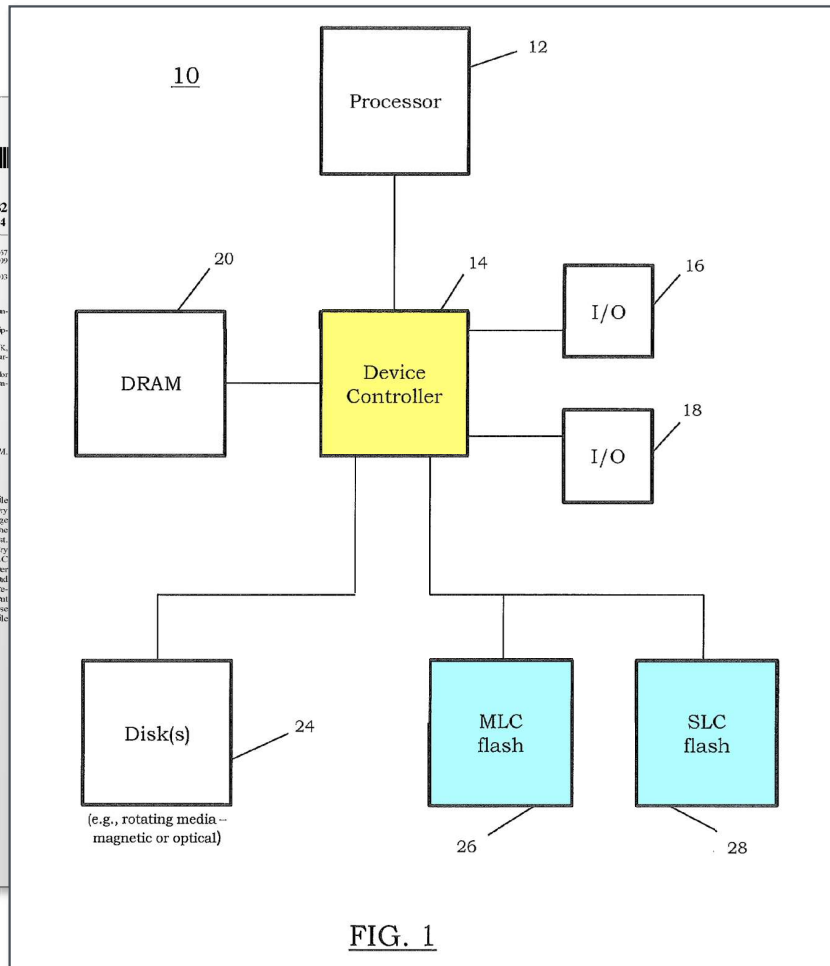
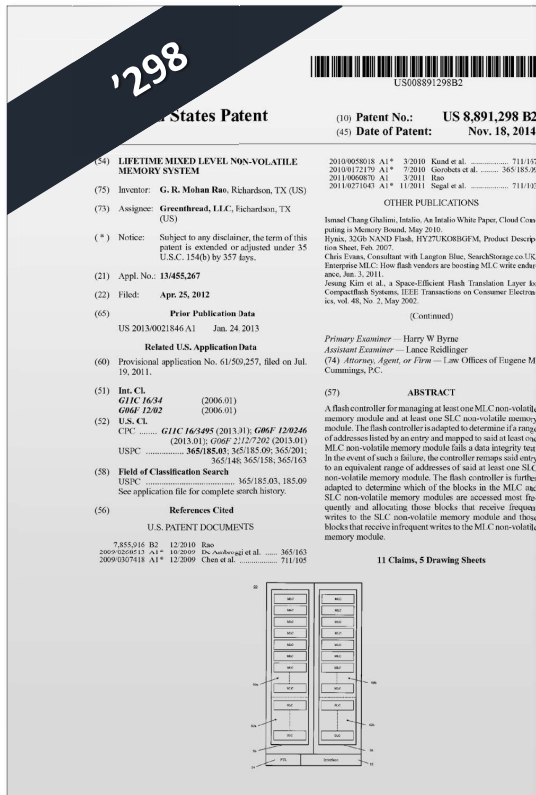
2.1 PO's Construction of "Blocks" Should be Adopted

'298

usually performed within a flash translation layer. In most cases, the controller maintains a lookup table to translate the memory array physical block address (PBA) to the logical block address (LBA) used by the host system. The controller's wear-leveling algorithm determines which physical block to use each time data is programmed, eliminating the relevance of the physical location of data and enabling data to be stored anywhere within the memory array and thus prolonging the service life of the flash memory. Depending on the wear-leveling method used, the controller typically either writes to the available erased block with the lowest erase count (dynamic wear leveling); or it selects an available target block with the lowest overall erase count, erases the block if necessary, writes new data to the block, and ensures that blocks of static data are moved when their block erase count is below a certain threshold (static wear leveling).

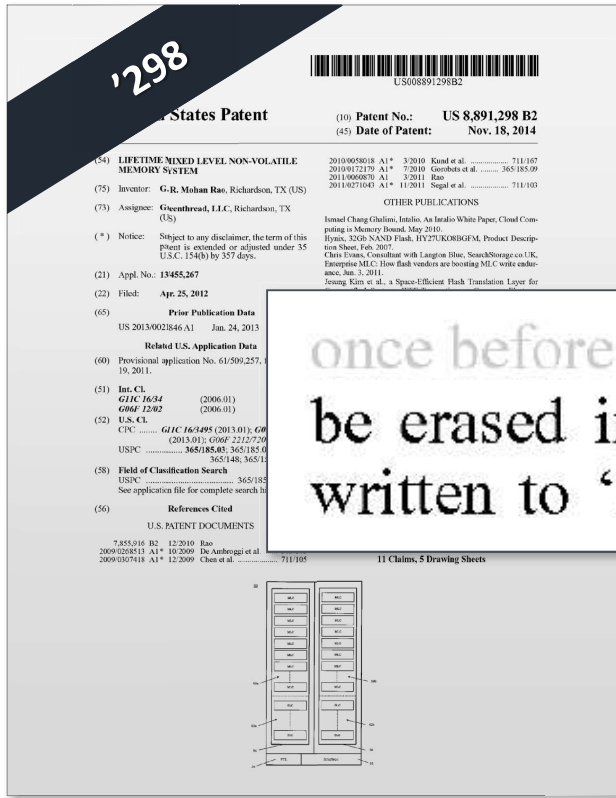
– 01547: Ex. 1001 at 2:65-3:13

2.1 Host Processor 12 and Controller 14 of '298 Patent



- 01547: Ex. 1001 at FIG. 1

2.1 PO's Construction of "Blocks" Should be Adopted



once before the entire block must be erased. Blocks can only be erased in their entirety, and, when erased, are usually written to '1' bits. However, if an erased block is already

- 01547: Ex. 1001 at 2:43-45

Dr. Khatri

46. Based on such disclosure of determining which *physical* block to use in the context of *erasable* blocks, a POSA would have understood that the "blocks" of claim 1, which are "erasable" according to limitations [1.A] and [1.B], are physical (and not logical) groups of memory cells.

– 01547: Ex. 2014 (Khatri Decl.) at ¶ 46

2.1 Dr. Khatri's Testimony Regarding "Blocks"

Dr. Khatri

entirety, and when erased, are usually written to '1' bits." Ex. 1001, 2:43-45. Such disclosure confirms that the "blocks" are physical groups of memory cells, because a POSA would have understood that erasing or writing bits in memory cells is functionality applicable to physical groups of memory cells, whereas logical blocks are mapped to physical blocks. Further, the statement that erasing results in the writing of a '1' bit agrees with the well known behavior of a floating gate transistor, which is a physical circuit that is used to implement blocks in a SLC or MLC flash memory. Dr. Liu attested to this in his deposition. Ex. 2015, 11:8-12:3, 30:4-31:3. Additionally, based on such disclosure, a POSA would have understood that "blocks" as claimed in the '298 patent (unless explicitly referred to as "logical blocks") refer to physical groups of memory cells *that must be erased all at once*, as reflected in Patent Owner's construction. Indeed, Sutardja explains that a block is a "group of memory cells *that is erased together*." Ex. 1011, [0157].

- 01547: Ex. 2014 (Khatri Decl.) at ¶ 47

2.1 '298 Specification and Dr. Liu's Testimony Regarding "Blocks"

'298

once before the entire block must be erased. Blocks can only be erased in their entirety, and, when erased, are usually written to '1' bits. However, if an erased block is already

– 01547: Ex. 1001 at 2:43-45

Q. You said in your response that an erase is simply having a smaller amount of electrons on the floating gate; right?

Dr. Liu

A. In the context what we just described in paragraph 46, where, again, to the host is really the convention of logical state that the host will see. So, in the context of a floating-gate NAND structure, the way the host can see a logical one for erase is having a small amount of charge on the floating gate.

– 01547: Ex. 2015 (Liu Depo.) at 30:4-18

2.1 Dr. Khatri's Testimony Regarding Erasing Flash Memory Cells

Dr. Khatri

25. Flash memory uses a special type of transistor (called a “floating gate transistor”) that, unlike a regular transistor, has two gates – a “floating gate” and a “control gate.” In a floating gate transistor, charge is stored on an electrically isolated conductor, called the floating gate. This charge has no path to dissipate (or discharge). The charge on the floating gate effectively changes the threshold voltage of the floating gate transistor, and thereby controls the current flowing between the source and drain. This change in current allows the user to determine the value stored in the cell.

26. To erase the cell, a voltage is applied between the drain and the control gate. Charge is then dissipated through a mechanism called Fowler-Nordheim (FN) tunneling.

– 01547: Ex. 2014 (Khatri Decl.) at ¶¶ 25-26

2.1 Patent Owner's Argument

PO's Sur-Reply

Claim 1 recites the “erasable blocks” are within MLC and SLC memory modules. *Id.*, 7:9-12. MLC and SLC refer to memory cells in flash memory (Ex. 2014, ¶30), e.g., implemented with transistors—and thus *physical* (as opposed to logical) groups of memory cells. *Id.*, ¶¶31-32. Thus, erasable “blocks” in the context of claim 1 are erasable physical groups of MLC or SLC memory cells that the controller (as opposed to host) erases.

As Dr. Khatri mentions (Ex. 1059, 112:14-117:2), the '298 patent discloses that physical blocks are programmed (Ex. 1001, 3:2-3) and that writing and programming are the same operation (*id.*, 3:28). Therefore, because writes are to erased blocks (*id.*, 3:8), erased blocks are physical groups of memory cells.

-01557 Sur-Reply at 5

1. Micron's Expert is not credible on several issues
2. '298 patent (IPR2021-01547)
2.1 Claim construction for "blocks"
● 2.2 Dusija-Sutardja does not disclose or suggest determining which of the blocks are accessed most frequently (limitation [1.F])
2.3 Dusija-Sutardja does not disclose or suggest transferring contents of blocks to SLC (limitation [1.G])
2.4 Moshayedi-Dusija does not disclose or suggest transferring contents of blocks to SLC (limitation [1.G])
2.5 Moshayedi's logical block addresses do not disclose or suggest "blocks" (limitations [1.F]-[1.G])
3. '385 patent (IPR2021-01548)
4. '240 patent (IPR2021-01549)
5. '300 patent (IPR2021-01550)

2.2 Claim 1 of the '298 Patent

'298

1. A system for storing data comprising:

at least one MLC non-volatile memory module comprising a plurality of individually erasable blocks;

at least one SLC non-volatile memory module comprising a plurality of individually erasable blocks; and

a controller coupled to the at least one MLC non-volatile memory module and the at least one SLC non-volatile memory module wherein the controller is adapted to:

- a) maintain an address map of at least one of the MLC and SLC non-volatile memory modules, the address map comprising a list of logical address ranges accessible by a computer system, the list of logical address ranges having a minimum quanta of addresses, wherein each entry in the list of logical address ranges maps to a similar range of physical addresses within either the at least one SLC non-volatile memory module or within the at least one MLC non-volatile memory module;

- b) determine if a range of addresses listed by an entry and mapped to a similar range of physical addresses within the at least one MLC non-volatile memory module, fails a data integrity test, and, in the event of such a failure, the controller remaps the entry to the next available equivalent range of physical addresses within the at least one SLC non-volatile memory module;

- c) determine which of the blocks of the plurality of the blocks in the MLC and SLC non-volatile memory modules are accessed most frequently by maintaining a count of the number of times each one of the blocks is accessed; and

- d) allocate those blocks that receive the most frequent writes by transferring the respective contents of those blocks to the at least one SLC non-volatile memory module.

← Limitation [1.F]

← Limitation [1.G]

-01547: Ex. 1001 at Claim 1

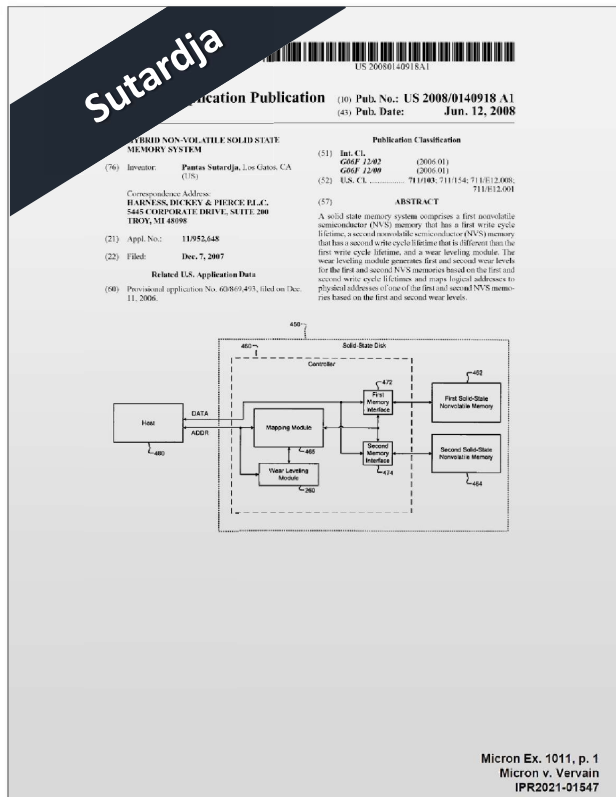
Claim 1

**Limitation
[1.F]**

'298

c) determine which of the blocks of the plurality of the blocks in the MLC and SLC non-volatile memory modules are accessed most frequently by maintaining a count of the number of times each one of the blocks is accessed: and

- 01547: Ex. 1001 at Claim 1



[0112] FIGS. 4A and 4B include additional modules that help to control wear leveling. In FIG. 4A, the wear leveling module determines how frequently data is written to each of the logical addresses. Logical addresses that are the target of relatively frequent writes or erases should be mapped to physical addresses that have not experienced as much wear.

[0113] In FIG. 4B, a write mapping module receives write frequency information from the host 220. The write frequency information identifies the logical addresses that correspond to data that is expected to change relatively frequently and/or the logical addresses that correspond to data that is expected to change relatively infrequently. In addition, the write mapping module may determine how frequently data is actually written to the logical addresses, as in FIG. 4A. FIG. 5 shows a solid-state disk where degradation of the memory and resulting remaining life is determined empirically, in addition to or instead of estimating remaining life based on the number of writes or erases.

- 01547: Ex. 1011 (Sutardja) at ¶¶ 0112-0113

2.2 Petitioner's Arguments Regarding Sutardja

Petition

of times each one of the blocks is accessed"). In other words, “the wear leveling module determines how frequently data is written to each of the logical addresses.” *Id.*, [0112]-[0113]. And Sutardja uses this count to determine the most frequently written blocks to transfer to SLC memory. *Id.*, [0149] (“if a number of write

– 01547: Pet. at 42

Dr. Liu

Q. You also cited paragraphs 112 and 113 of Sutardja; correct?

A. That's correct.

Q. And those paragraphs, in your words in the declaration, describe determining how frequently data is written to each of the logical addresses; correct?

A. I cite 112 and 113 as example that -- some of the feature that wear leveling module do. In this case, in this example I cited, it would determine how frequently data is written to each of the **logical addresses**.

– 01547: Ex. 2015 at 33:15-34:3

2.2 Dr. Khatri's Testimony

Dr. Khatri

73. Moreover, Sutardja's system determines how frequently data is written to each of the logical addresses. Ex. 2015, 43:20-44:6. But the write frequency of a logical address does not necessarily correspond with the access frequency of a physical block (as correctly construed, *see supra* Section VII.A). A logical address may be remapped to a new physical address due to, for example, wear leveling, garbage collection, or bad block management. When this occurs, the frequency of writes to the logical address may be different from the frequency of writes to the physical block that it is currently mapped to. If, for example, "LogicalAddressA" is remapped from Block1 to Block2, there may be 100 writes to LogicalAddressA, but only 70 writes to Block1 and 30 writes to Block2.

– 01547: Ex. 2014 (Khatri Decl.) at ¶ 73

Sutardja

US 2008/014918 A1
 Publication Publication (06) Pub. No.: US 2008/014918 A1
 (43) Pub. Date: Jun. 12, 2008

THIRD NON-VOLATILE SOLID STATE MEMORY SYSTEM

(76) Inventor: **Petrus Sutardja**, Los Gatos, CA (US)

Correspondence Address: **HARNS, DICKIN & PIERCE PLLC**, 5445 CORPORATE DRIVE, SUITE 200 TROY, MI 48068

(21) Appl. No.: **11,952,648**
 (22) Filed: **Dec. 7, 2007**

Related U.S. Application Data
 (60) Provisional application No. 60369493, filed on Dec. 11, 2006.

Publication Classification
 (51) Int. Cl. **G06F 12/02** (2006.01)
G06F 12/09 (2006.01)
 (52) U.S. Cl. **710/903**; **710/934**; **710/912**; **710/912.09**

ABSTRACT
 A solid state memory system comprises a first nonvolatile semiconductor (NVS) memory that has a first write cycle lifetime, a second nonvolatile semiconductor (NVS) memory that has a second write cycle lifetime that is different than the first write cycle lifetime, and a wear leveling module. The wear leveling module generates first and second wear level for the first and second NVS memories based on the first and second write cycle lifetimes and maps logical addresses to physical addresses of one of the first and second NVS memories based on the first and second wear levels.

Micron Ex. 1011, p. 1
 Micron v. Vervain
 IPR2021-01547

[0111] With various nonvolatile memories, writing data to a block may require erasing or writing to the entire block. In such a block-centric memory, the wear leveling module may track the number of times that each block has been erased or written. When a write request arrives from the host, the wear leveling module may select the block of memory that has been written to the least from among the available blocks. The wear leveling module then maps the incoming logical address to the physical address of this block. Over time, this may produce a nearly uniform distribution of write operations across memory blocks.

– 01547: Ex. 1011 (Sutardja) at ¶ 0111

Petition

In addition, and unlike the "data shift," Sutardja discloses a second way to use the frequency count to transfer data from MLC memory to SLC memory. Specifically, upon receiving a host write to an MLC block, if the frequency count to that **logical address** is high, it may direct the write to a physical SLC block, thereby transferring the data from MLC memory modules to SLC memory modules. *Id.*, [0146]-[0147]; Liu Decl., ¶ 157.

– 01547: Pet. at 44

1. Micron's Expert is not credible on several issues
2. '298 patent (IPR2021-01547)
2.1 Claim construction for "blocks"
2.2 Dusija-Sutardja does not disclose or suggest determining which of the blocks are accessed most frequently (limitation [1.F])
● 2.3 Dusija-Sutardja does not disclose or suggest transferring contents of blocks to SLC (limitation [1.G])
2.4 Moshayedi-Dusija does not disclose or suggest transferring contents of blocks to SLC (limitation [1.G])
2.5 Moshayedi's logical block addresses do not disclose or suggest "blocks" (limitations [1.F]-[1.G])
3. '385 patent (IPR2021-01548)
4. '240 patent (IPR2021-01549)
5. '300 patent (IPR2021-01550)

Claim 1

**Limitation
[1.G]**

'298

d) allocate those blocks that receive the most frequent writes by transferring the respective contents of those blocks to the at least one SLC non-volatile memory module.

- 01547: Ex. 1001 at Claim 1

2.3 Petitioner's Argument Regarding Sutardja for Limitation [1.G]

VERVAIN

Petition

to a second block of the second NVS memory [*i.e.*, SLC memory module] in step 522” (“*allocate. . .by transferring the respective contents of those blocks to the at least one SLC non-volatile memory module*”). *Id.*, [0149]; Liu Decl., ¶ 154. Note that by Sutardja disclosing remapping the logical address to SLC, it would have been understood that it must copy the data to SLC as well, which Sutardja describes in paragraph 167 (“swapping data”). Liu Decl., ¶ 156.

– 01547: Pet. at 44

2.3 Sutardja's Wear Leveling

Sutardja

US 2008/0140918 A1

Publication Publication (06) Pub. No.: US 2008/0140918 A1
(43) Pub. Date: Jun. 12, 2008

NON-VOLATILE SOLID STATE MEMORY SYSTEM

(76) Inventor: Paras Sutardja, Los Gatos, CA (US)

Correspondence Address: HARVENS, DICKNEY & PIERCE PLLC, 5445 CORPORATE DRIVE, SUITE 200 TROY, MI 48068

(21) Appl. No.: 11/952,648
(22) Filed: Dec. 7, 2007

Related U.S. Application Data
(60) Provisional application No. 60/869,491, filed on Dec. 11, 2006.

Publication Classification
(51) Int. Cl. G06F 12/02 (2006.01)
G06F 12/09 (2006.01)
(52) U.S. Cl. 719/983, 711/154, 711/12,008, 711/192,001

ABSTRACT
A solid state memory system comprises a first nonvolatile semiconductor (NVS) memory that has a first write cycle lifetime, a second nonvolatile semiconductor (NVS) memory that has a second write cycle lifetime that is different than the first write cycle lifetime, and a wear leveling module. The wear leveling module generates first and second wear levels for the first and second NVS memories based on the first and second write cycle lifetimes and maps logical addresses to physical addresses of one of the first and second NVS memories based on the first and second wear levels.

Micron Ex. 1011, p. 1
Micron v. Vervain
IPR2021-01547

[0167] At various times, such as periodically, the wear leveling module may analyze the wear levels of the blocks, and remap relatively frequently rewritten logical addresses to blocks with low wear levels. In addition, the wear leveling module may remap relatively infrequently rewritten logical addresses to blocks with high wear levels, which is known as static data shifting. Remapping may involve swapping data in two blocks. During the swap, the data from one of the blocks may be stored in an unused block, or in temporary storage.

– 01547: Ex. 1011 (Sutardja) at ¶ 0167

2.3 Sutardja's Wear Leveling

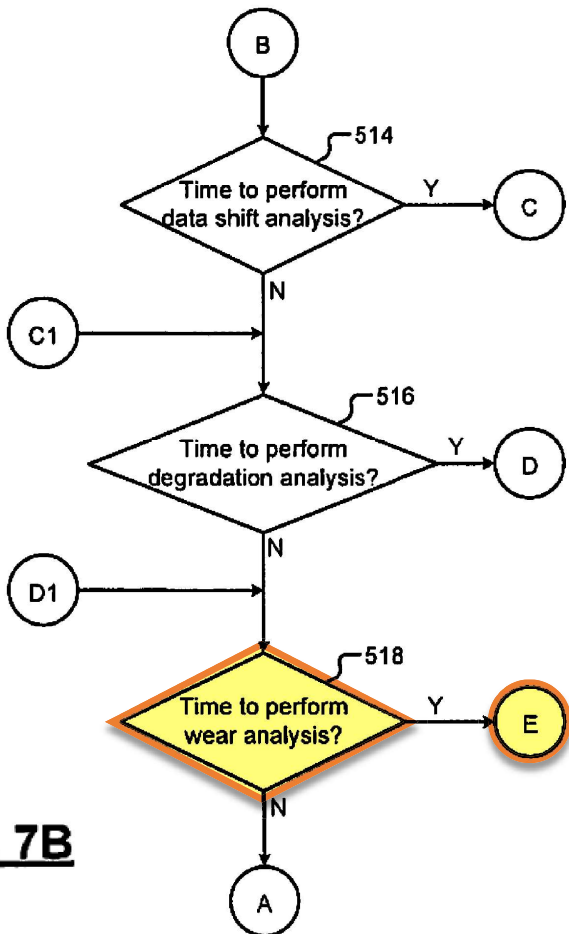


FIG. 7B

Sutardja

[0148] In FIG. 7B, control determines whether time to perform data shift analysis has arrived in step 514. If the result of step 514 is false, control determines whether time to perform degradation analysis has arrived in step 516. If the result of step 516 is false, control determines whether time to perform wear level analysis has arrived in step 518. If the result of step 514 is false, control returns to step 510.

- 01547: Ex. 1011 (Sutardja) at ¶ 0148

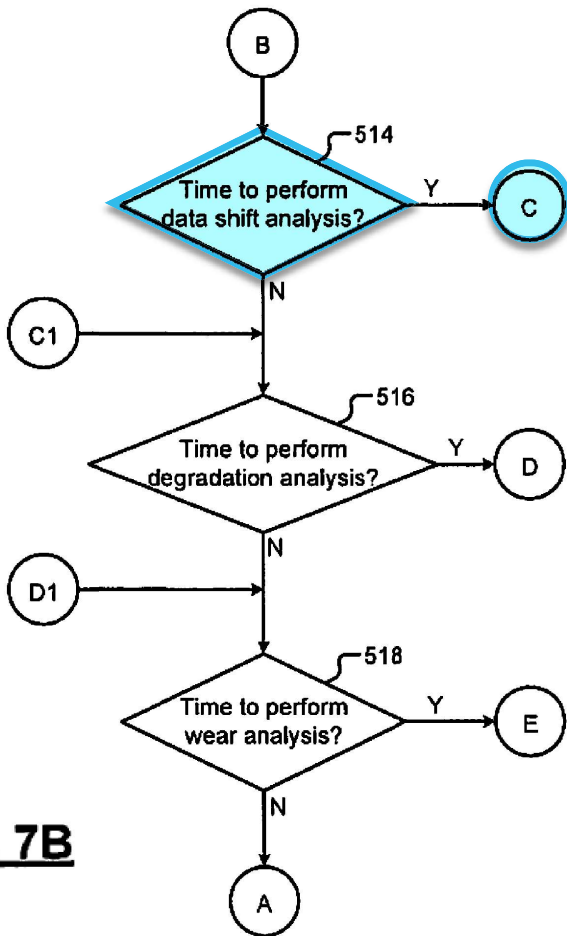


FIG. 7B

Sutardja

[0149] In FIG. 7C, when the result of step 514 is true, control determines in step 520 if a number of write operations to a first block of the first NVS memory during a predetermined time is greater than or equal to a predetermined threshold. If the result of step 520 is false, control returns to step 516. If the result of step 520 is true, control maps the logical addresses that correspond to the first block to a second block of the second NVS memory in step 522.

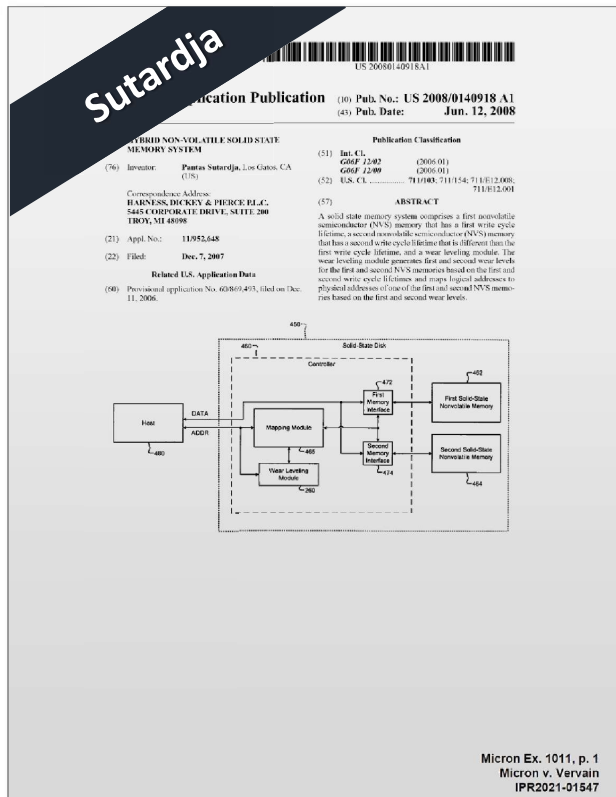
– 01547: Ex. 1011 (Sutardja) at ¶ 0149

2.3 Petitioner’s “Second Way” Argument for Limitation [1.G]

Petition

In addition, and unlike the “data shift,” Sutardja discloses a second way to use the frequency count to transfer data from MLC memory to SLC memory. Specifically, upon receiving a host write to an MLC block, if the frequency count to that logical address is high, it may direct the write to a physical SLC block, thereby transferring the data from MLC memory modules to SLC memory modules. *Id.*, [0146]-[0147]; Liu Decl., ¶ 157.

– 01547: Pet. at 44



[0146] In FIG. 7A, the method 500 begins at step 502. Control receives write frequencies for logical addresses where data is to be written from the host in step 504. Control maps the logical addresses having low write frequencies (e.g., having write frequencies less than a predetermined threshold) to the first NVS memory in step 506. Control maps the logical addresses having high write frequencies (e.g., having write frequencies greater than a predetermined threshold) to the second NVS memory in step 508.

[0147] Control writes data to the first and/or second NVS memories in step 510 according to the mapping generated in steps 506 and 508. Control measures actual write frequencies at which data is in fact written to the logical addresses and updates the mapping in step 512.

- 01547: Ex. 1011 (Sutardja) at ¶¶ 0146-147

2.3 Petitioner's Argument Regarding the Second NVS Memory

Petition

⁵ This Petition relies on the scenario in which the first memory is MLC and second memory is SLC. Sutardja, [0106], Claim 37; Liu Decl., ¶ 103-104.

– 01547: Pet. at 25 n.5

2.3 Sutardja's First and Second NVS Memories

Sutardja

US 2008/0140918 A1

Publication Publication (06) Pub. No.: US 2008/0140918 A1
(43) Pub. Date: Jun. 12, 2008

NON-VOLATILE SOLID STATE MEMORY SYSTEM

(76) Inventor: Paras Sutardja, Los Gatos, CA (US)

Correspondence Address: HARVINS DICKNEY & PIERCE PLLC, 5445 CORPORATE DRIVE, SUITE 200 TROY, MI 48068

(21) Appl. No.: 11/952,648
(22) Filed: Dec. 7, 2007

Related U.S. Application Data
(60) Provisional application No. 60/869,491, filed on Dec. 11, 2006.

Publication Classification
(51) Int. Cl. G06F 12/02 (2006.01)
G06F 12/09 (2006.01)
(52) U.S. Cl. 719/983, 711/154, 711/157, 711/158

ABSTRACT
A solid state memory system comprises a first non-volatile semiconductor (NVS) memory that has a first write lifetime, a second non-volatile semiconductor (NVS) memory that has a second write cycle lifetime that is different from the first write cycle lifetime, and a wear leveling module that generates first and second wear levels for the first and second NVS memories based on the first and second write cycle lifetimes and maps logical addresses to physical addresses of one of the first and second NVS memories based on the first and second wear levels.

Micron Ex. 1011, p. 1
Micron v. Vervain
IPR2021-01547

[0149] In FIG. 7C, when the result of step 514 is true, control determines in step 520 if a number of write operations to a first block of the first NVS memory during a predetermined time is greater than or equal to a predetermined threshold. If the result of step 520 is false, control returns to step 516. If the result of step 520 is true, control maps the logical addresses that correspond to the first block to a second block of the **second NVS memory** in step 522.

– 01547: Ex. 1011 (Sutardja) at ¶ 0149

2.3 Sutardja's Second NVS Memory

Sutardja

US 2008/014918 A1

Publication Publication (06) Pub. No.: US 2008/014918 A1
(43) Pub. Date: Jun. 12, 2008

MEMBRANE NON-VOLATILE SOLID STATE MEMORY SYSTEM

(76) Inventor: Paras Sutardja, Los Gatos, CA (US)

Correspondence Address: HARDEK, DICKNEY & PIERCE PLLC, 5445 CORPORATE DRIVE, SUITE 200 TROY, MI 48068

(21) Appl. No.: 11,952,648
(22) Filed: Dec. 7, 2007

Related U.S. Application Data
(60) Provisional application No. 60/869,493, filed on Dec. 11, 2006.

Publication Classification
(51) Int. Cl. G06F 12/02 (2006.01)
G06F 12/09 (2006.01)
(52) U.S. Cl. 719/003, 711/154, 711/12,008, 711/12,001

(57) ABSTRACT
A solid state memory system comprises a first nonvolatile semiconductor (NVS) memory that has a first write cycle lifetime, a second nonvolatile semiconductor (NVS) memory that has a second write cycle lifetime that is different than the first write cycle lifetime, and a wear leveling module. The wear leveling module generates first and second wear levels for the first and second NVS memories based on the first and second write cycle lifetimes and maps logical addresses to physical addresses of one of the first and second NVS memories based on the first and second wear levels.

Micron Ex. 1011, p. 1
Micron v. Vervain
IPR2021-01547

[0105] In FIG. 2, a functional block diagram of an exemplary solid-state disk 200 according to the present disclosure is presented. The solid-state disk 200 includes a controller 202 and first and second solid-state nonvolatile memories 204 and 206. Throughout the remainder of this disclosure, solid-state nonvolatile memories may be implemented as integrated circuits (IC). The controller 202 receives access requests from a host 220. The controller 202 directs the access requests to the first solid-state nonvolatile memory 204 or the second solid-state nonvolatile memory 206, as will be described below.

- 01547: Ex. 1011 (Sutardja) at ¶ 0105

2.3 Sutardja's Second NVS Memory

Sutardja

US 2008/0140918 A1

Publication Publication (16) Pub. No.: US 2008/0140918 A1
(45) Pub. Date: Jun. 12, 2008

THIRD NON-VOLATILE SOLID STATE MEMORY SYSTEM

(76) Inventor: **Pantaz Sutardja**, Los Gatos, CA (US)

Correspondence Address: **HARNESS, DICKEY & PIERCE PLLC**, 5445 CORPORATE DRIVE, SUITE 200 TROY, MI 48068

(21) Appl. No.: 11,952,648
(22) Filed: Dec. 7, 2007

Related U.S. Application Data
(60) Provisional application No. 603869491, filed on Dec. 11, 2006.

Publication Classification
(51) Int. Cl. G06F 12/02 (2006.01)
G06F 12/09 (2006.01)
(52) U.S. Cl. 711/013; 711/014; 711/012; 711/012.008; 711/012.009

ABSTRACT
A solid state memory system comprises a first nonvolatile semiconductor (NVS) memory that has a first write cycle lifetime, a second nonvolatile semiconductor (NVS) memory that has a second write cycle lifetime that is different than the first write cycle lifetime, and a wear leveling module generated for the first and second NVS memory to manage write cycle lifetimes and physical addresses of one of the memories based on the first and second

Micron Ex. 1011, p. 1
Micron v. Vervain
IPR2021-01547

37. The solid state memory system of claim 1 wherein said second NVS memory includes single-level cell (SLC) flash memory and said first NVS memory include multi-level cell (MLC) flash memory.

– 01547: Ex. 1011 (Sutardja) at Claim 37

2.3 Sutardja's Second NVS Memory

Sutardja

US 2008/040918 A1

Publication Publication (06) Pub. No.: US 2008/0140918 A1
(43) Pub. Date: Jun. 12, 2008

MEMORY SYSTEM

Publication Classification
(51) Int. Cl. G06F 12/02 (2006.01)
G06F 12/09 (2006.01)
(52) U.S. Cl. 710/003, 711/054, 711/012, 008, 711/012, 009

(76) Inventor: Paras Sutardja, Los Gatos, CA (US)

Correspondence Address: HARDEK, DICKIN & PIERCE PLLC, 5445 CORPORATE DRIVE, SUITE 200 TROY, MI 48068

(21) Appl. No.: 11,952,648
(22) Filed: Dec. 7, 2007

Related U.S. Application Data
(60) Provisional application No. 60/869,491, filed on Dec. 11, 2006.

ABSTRACT
A solid state memory system comprises a first nonvolatile semiconductor (NVS) memory that has a first write lifetime, a second nonvolatile semiconductor (NVS) memory that has a second write cycle lifetime that is different than the first write cycle lifetime, and a wear leveling module. The wear leveling module generates first and second wear for the first and second NVS memories based on the first and second write cycle lifetimes and maps logical address physical addresses of one of the first and second NVS memories based on the first and second wear levels.

Micron Ex. 1011, p. 1
Micron v. Vervain
IPR2021-01547

[0108] The first solid-state nonvolatile memory 204 may include single-level cell (SLC) flash memory or multi-level cell (MLC) flash memory. The second solid-state nonvolatile memory 206 may include single-level cell (SLC) flash memory or multi-level cell (MLC) flash memory.

– 01547: Ex. 1011 (Sutardja) at ¶ 0108

2.3 Petitioner's Argument for Limitation [1.G]

PO Response

vs. MLC in both the first and second memories. Indeed, Dr. Liu did not even cite paragraph [0108] of Sutardja for his analysis of limitation [1.G] (Ex. 1009, ¶¶153-157), and where Dr. Liu did cite paragraph [0108], he stated that Sutardja's first memory may be MLC and the second memory may be SLC but failed to mention that the same paragraph of Sutardja (i.e., [0108]) also discloses the opposite. In other words, Dr. Liu's analysis does not consider that Sutardja paragraph [0108] reveals that Sutardja does not care which type of memory cell (SLC vs. MLC) is present at the first and second memories, nor does it consider Sutardja's emphasis on normalized wear levels instead of targeting SLC specifically for transferring data. Dr. Liu's failure to address the foregoing disclosures of paragraph [0108] of Sutardja is particularly noteworthy given that paragraph [0108] is the only place throughout Sutardja's specification to mention SLC.

--01547: PO Resp. at 54-55

2.3 Sutardja's Normalized Wear Levels

Sutardja

US 2008/014091A1

Publication Publication (16) Pub. No.: US 2008/014091A1
(45) Pub. Date: Jun. 12, 2008

NON-VOLATILE SOLID STATE MEMORY SYSTEM

Publication Classification

(76) Inventor: **Pantaz Sutardja**, Los Gatos, CA (US)

(51) Int. Cl. G06F 12/02 (2006.01)
G06F 12/09 (2006.01)

(52) U.S. Cl. 71/043, 71/054; 71/012,008; 71/012,009

(57) **ABSTRACT**
A solid state memory system comprises a first nonvolatile semiconductor (NVS) memory that has a first write cycle lifetime; a second nonvolatile semiconductor (NVS) memory that has a second write cycle lifetime that is different than the first write cycle lifetime; and a wear leveling module. The wear leveling module generates first and second wear levels for the first and second NVS memories based on the first and second write cycle lifetimes and maps logical addresses to physical addresses of one of the first and second NVS memories based on the first and second wear levels.

(21) Appl. No.: 11/952,648
(22) Filed: Dec. 7, 2007

Related U.S. Application Data
(60) Provisional application No. 60/869,491, filed on Dec. 11, 2006.

Micron Ex. 1011, p. 1
Micron v. Vervain
IPR2021-01547

Sutardja

[0162] A block in the first memory that has been erased 1,000 times would then have a **normalized wear level** of $1/10$, while a block in the second memory that has been erased 1,000 times would then have a normalized wear level of $1/100$. Once the wear levels have been normalized, a wear leveling algorithm can be employed across all the blocks of both the first and second memories as if all the blocks formed a single memory having a single write cycle lifetime. Wear levels as used herein, unless otherwise noted, are **normalized wear levels**.

--01547: Ex. 1011 (Sutardja) at ¶ 0162

2.3 Dr. Khatri's Testimony about Sutardja

Dr. Khatri

Accordingly, Sutardja is agnostic as to whether to write to SLC or instead MLC, and Sutardja instead focuses on the fraction of the overall lifetime that has been reached so far. Ex. 1011, [0161]-[0162]. This is consistent with Sutardja's teaching in paragraph [0108] where both the first and second NVS memories are taught to possibly include SLC or MLC flash memory.

--01547: Ex. 2014 (Khatri Decl.) at ¶ 97

2.3 Dr. Khatri's Testimony about Sutardja

Dr. Khatri

44. Indeed, Sutardja's emphasis on the normalized wear level (i.e., fraction of lifetime that has been reached), instead of the type of memory (MLC vs. SLC), discourages Petitioner's proposed obviousness combination of Dusija and Sutardja. Sutardja describes a fundamentally different approach than the one claimed in claim 1 and limitation [1.G], because transferring the contents of blocks to a specific type of memory (where the type of memory is the basis for the transfer) is qualitatively different than writing to a memory based on an attribute (namely, normalized wear level) that is *independent of* memory type.

– 01547: Ex. 2014 (Khatri Decl.) at ¶ 98

2.3 Petitioner’s “Second Way” Argument for Limitation [1.G]

Petitioner’s Reply

After remapping and writing the data, Sutardja has “*transferr[ed] the respective contents of those [frequently written logical] blocks to the at least one SLC non-volatile memory module,*” because the data for that logical block (“*contents*”) has been transferred to the SLC module. Thus, the write operation and remapping “*transfer[s]*” the “*contents*” of those frequently written logical blocks to “*at least one SLC non-volatile memory module.*” Liu Reply, ¶ 37. This is all that the limitation requires, and the Board should reject PO’s implicit constructions for the reasons set forth in Section III.A.1.

– 01547: Reply at 18-19

2.3 PO is Analyzing the Plain Language of “Transferring” and “Contents Of”

VERVAIN

PO’s Sur-Reply

Regarding limitation [1.G], the Reply relies on Petitioner’s claim construction argument for “blocks,” but as noted above that argument is incorrect. Reply, 18. Like with its Moshayedi-based argument, Petitioner again presents in its Reply an argument for “transferring...” that simultaneously reads “*contents of* those blocks” in a manner at odds with plain meaning while incorrectly ascribing to PO an interpretation of “transferring” (PO did not construe that term). *Id.* It is Petitioner that is again incorrect regarding the “transferring...” limitation, for the reasons discussed above regarding Moshayedi (*supra* Section III.A.1). Contrary to Petitioner’s characterization of “PO’s implicit constructions” (Reply, 19), PO is analyzing the plain language of “transferring” and “contents of.” Petitioner is attempting to escape the plain language of “contents of.”

-01557 Sur-Reply at 13

1. Micron's Expert is not credible on several issues
2. '298 patent (IPR2021-01547)
2.1 Claim construction for "blocks"
2.2 Dusija-Sutardja does not disclose or suggest determining which of the blocks are accessed most frequently (limitation [1.F])
2.3 Dusija-Sutardja does not disclose or suggest transferring contents of blocks to SLC (limitation [1.G])
● 2.4 Moshayedi-Dusija does not disclose or suggest transferring contents of blocks to SLC (limitation [1.G])
2.5 Moshayedi's logical block addresses do not disclose or suggest "blocks" (limitations [1.F]-[1.G])
3. '385 patent (IPR2021-01548)
4. '240 patent (IPR2021-01549)
5. '300 patent (IPR2021-01550)

Moshayedi

Application Publication (109) Pub. No.: US 2009/0327591 A1
 (45) Pub. Date: Dec. 31, 2009

US 2009027591A1

SLC-MLC COMBINATION FLASH STORAGE DEVICE

(75) Inventor: Mark Moshayedi, Newport Coast, CA (US)

Correspondence Address: MICRORIMOTT WILL & EMERY LLP
 JEFF VON KARMAN AVE., SUITE 500
 IRVINE, CA 92614-7108 (US)

(73) Assignee: STEC, INC., Santa Ana, CA (US)

(21) Appl. No.: 12,492,113

(22) Filed: Jun. 25, 2009

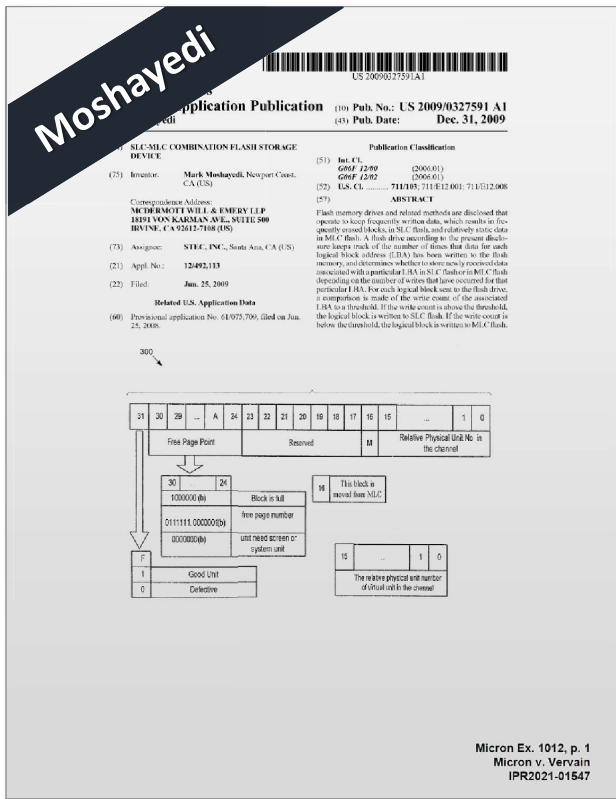
Related U.S. Application Data
 (60) Provisional application No. 61/055,790, filed on Jun. 25, 2008.

Publication Classification
 (51) Int. Cl. G06F 2/00 (2006.01)
 G06F 2/02 (2006.01)
 (52) U.S. Cl. 711/105; 711/811; 801; 711/012; 808
 (57) ABSTRACT
 Flash memory drives and related methods are disclosed that operate to keep frequently written data, which tends to frequently erase blocks in SLC flash, and relatively static data in MLC flash. A flash drive according to the present disclosure keeps track of the number of times that data for each logical block address (LBA) has been written to the flash memory, and determines whether to move new or previously associated with a particular LBA in SLC flash or in MLC flash depending on the number of writes that have occurred for that particular LBA. For each logical block kept in the flash drive, a comparison is made of the write count of the associated LBA to a threshold. If the write count is above the threshold, the logical block is written to SLC. If the write count is below the threshold, the logical block is written to MLC flash.

Micron Ex. 1012, p. 1
 Micron v. Vervain
 IPR2021-01547

[0032] The flash drive may use the erase count lists to move data between MLC flash and SLC flash. For example, once a block in MLC flash reaches a threshold erase count (e.g., 500), the next write operation to that block triggers a swap where the data from the MLC flash block is written to a block in SLC flash. In this manner, data for an LBA that is frequently written and causes frequent erasures is moved to SLC flash which can perform more erase cycles than MLC flash.

- 01547: Ex. 1012 at ¶ 0032



(57)

ABSTRACT

Flash memory drives and related methods are disclosed that operate to keep frequently written data, which results in frequently erased blocks, in SLC flash, and relatively static data in MLC flash. A flash drive according to the present disclosure keeps track of the number of times that data for each logical block address (LBA) has been written to the flash memory, and determines whether to store newly received data associated with a particular LBA in SLC flash or in MLC flash depending on the number of writes that have occurred for that particular LBA. For each logical block sent to the flash drive, a comparison is made of the write count of the associated LBA to a threshold. If the write count is above the threshold, the logical block is written to SLC flash. If the write count is below the threshold, the logical block is written to MLC flash.

– 01547: Ex. 1012 at Abstract

Dr. Liu

Q. So, in short, the abstract of Moshayedi, which states a criterion, as you mentioned, is describing if-then logic to write data into SLC or instead into MLC; correct?

A. In that context, that is fair, yes.

– 01547: Ex. 2015 at 50:7-12

Moshayedi

Application Publication (10) Pub. No.: US 2009/0327591 A1
 (45) Pub. Date: Dec. 31, 2009

Publication Classification
 Int. Cl. G06F 2/00 (2006.01)
 G06F 2/02 (2006.01)
 U.S. Cl. 711/105; 711/111; 711/106; 711/112; 700

SLC-MLC COMBINATION FLASH STORAGE DEVICE
 Inventor: Mark Moshayedi, Newport Coast, CA, US

Correspondence Address: MICHELMOTT WILLIAMS & EMERY LLP
 JEFFI VON KARMAN, AUC, SUITE 900
 BOSTON, CA 95013-7108 (US)

Assignee: STEC, INC., Santa Ana, CA (US)

Appl. No.: 12,492,113
 Filed: Jun. 25, 2009

Related U.S. Application Data
 Provisional application No. 61,055,790, filed on Jun. 25, 2008.

Flash memory drives and related methods are disclosed that operate to keep frequently written data, which tends to frequently exceed blocks in SLC flash, and relatively static data in MLC flash. A flash drive according to the present disclosure keeps track of the number of times that data for each logical block address (LBA) has been written to the flash memory, and determines whether to move or re-write data associated with a particular LBA in SLC flash or in MLC flash depending on the number of writes that have occurred for that particular LBA. For each logical block kept in the flash drive, a comparison is made of the write count of the associated LBA to a threshold. If the write count is above the threshold, the logical block is written to SLC. If the write count is below the threshold, the logical block is written to MLC flash.

Micron Ex. 1012, p. 1
 Micron v. Vervain
 IPR2021-01547

[0073] As the level of user blocks within the SLC goes above the configured limit, the logical blocks with the lowest write count are moved from SLC to MLC, therefore freeing up space in the SLC. No converse operation is necessary (copy from MLC to SLC) as logical blocks are directed towards SLC when they are written by the host if the write count is at a sufficient level to warrant this. There are many different configurations regarding proportion of SLC to MLC, user blocks to total NAND capacity (for SLC and MLC) and what algorithm to use to calculate the threshold at

- 01547: Ex. 1012 at ¶ 0073

Dr. Liu

Q. You did not address paragraph 73, did you?

A. **I did not cite paragraph 73**, but I did address in the sense that I did review thoroughly Moshayedi and every paragraph of Moshayedi in rendering my opinion. Again, I need to say that there is a clause given -- it has a context. There is a when, there's a limited context of when, and it's stated in 73.

– 01547: Ex. 2015 at 56:22-57:6

2.4 Petitioner's Argument Regarding "Transferring"

Petitioner's Reply

PO characterizes the "swap" operation as only writing "a single block of data" received with the write operation to SLC. POR, 57. Even if this were correct (it's not), it satisfies the claims under the correct claim interpretation. Note that for this write operation, the "newly received data [is] associated with a particular LBA." Moshayedi, ¶24.⁵ Thus, upon receiving the write operation with new "data," that "data" are "*the respective contents of those blocks*," because they are "associated with a particular LBA" (*i.e.*, the LBA for the MLC block). *Id.* The act of writing that data to SLC "*transfer[s] the respective contents of those blocks to the at least one SLC non-volatile memory module.*" Liu Reply, ¶ 18. The limitation requires nothing more.

-01557: Reply at 8

2.4 Patent Owner's Argument Regarding "Transferring"

PO's Sur-Reply

Petitioner states "PO [r]elies on an [i]ncorrect [c]onstruction of '[t]ransferring.'" Reply, 8. But PO has not sought to construe "transferring" (Response, 22-30)—and it is *Petitioner* that is making an argument based on a peculiar reading of this term. By stating Moshayedi's to-be-written data is somehow "the respective contents of those blocks" (Reply, 8), it is Petitioner that is trying to construe the phrase "contents of." But it is apparent from the plain meaning of "contents of" that what is intended in the claim language is something actually contained in those blocks (not something that will *eventually* be in those blocks, after the write operation finishes). Similarly, the Petition states that "remapping the logical address from MLC to SLC [] transfer[s] the contents for that *logical address* from MLC to SLC" and "the write operation causes the data associated with the *logical address* to be transferred from MLC to SLC memory space" (Petition, 63)², but such statements in the Petition and Reply are incorrect because Petitioner incorrectly relies on contents of logical addresses, not contents of "blocks" as properly construed; and Petitioner's strained argument that "contents of" blocks includes something not even contained in the blocks yet is incorrect. Petitioner did not pursue any construction of "contents" in its Petition.

-01557 Sur-Reply at 6-7

1. Micron's Expert is not credible on several issues
2. '298 patent (IPR2021-01547)
2.1 Claim construction for "blocks"
2.2 Dusija-Sutardja does not disclose or suggest determining which of the blocks are accessed most frequently (limitation [1.F])
2.3 Dusija-Sutardja does not disclose or suggest transferring contents of blocks to SLC (limitation [1.G])
2.4 Moshayedi-Dusija does not disclose or suggest transferring contents of blocks to SLC (limitation [1.G])
● 2.5 Moshayedi's logical block addresses do not disclose or suggest "blocks" (limitations [1.F]-[1.G])
3. '385 patent (IPR2021-01548)
4. '240 patent (IPR2021-01549)
5. '300 patent (IPR2021-01550)

2.5 Moshayedi's Logical Block Address

Moshayedi

US 2009/027591A1

Application Publication (109) Pub. No.: US 2009/0327591 A1
(43) Pub. Date: Dec. 31, 2009

SLC-MLC COMBINATION FLASH STORAGE DEVICE

Publication Classification

(51) Int. Cl. G06F 2/00 (2006.01)
G06F 2/02 (2006.01)
(52) U.S. Cl. 711/105; 711/111; 711/101; 711/121; 708
(57) **ABSTRACT**

Flash memory drives and related methods are disclosed that operate to keep frequently written data, which results in frequently erased blocks, in SLC flash, and relatively static data in MLC flash. A flash drive according to the present disclosure keeps track of the number of times that data for each logical block address (LBA) has been written to the flash memory, and determines whether to store newly received data associated with a particular LBA in SLC flash or in MLC flash depending on the number of writes that have occurred for that particular LBA. For each logical block sent to the flash drive, a comparison is made of the write count of the associated LBA to a threshold. If the write count is above the threshold, the logical block is written to SLC flash. If the write count is below the threshold, the logical block is written to MLC flash.

300

Micron Ex. 1012, p. 1
Micron v. Vervain
IPR2021-01547

[0024] A flash drive according to the present disclosure keeps track of the number of times that data for each logical block address (LBA) has been written to the flash memory, and determines whether to store newly received data associated with a particular LBA in SLC flash or in MLC flash, depending on the number of writes that have occurred for that particular LBA. For each logical block sent to the flash drive by the host, the host compares the write count of the associated LBA against a threshold. If the write count is above the threshold, the logical block is written to SLC flash. If the write count is below the threshold, the logical block is written to MLC flash. The threshold may be set at 0 initially, resulting in all data being written to SLC flash, and then increased as needed. When SLC flash has reached a designated capacity, blocks with the lowest write counts are moved down to MLC flash. The flash drive uses sets of link lists to keep track of a number of metrics used to manage the flash memory. A set is kept for both the SLC flash and the MLC flash in each channel of the flash drive.

- 01547: Ex. 1012 at ¶ 0024

Dr. Liu

- Q. So, you agree that the abstract is describing that, depending on the **write count of the logical block address** being either higher or lower than a threshold, the write occurs to one type of memory or to another type of memory.
- A. I would say that abstract describes a way of determining to MLC or SLC, the abstract cites one way is to look at the **write count to an LBA** as the criterion, as one criterion, a criterion to do that.

– 01547: Ex. 2015 at 49:8-18

1. Micron's Expert is not credible on several issues
2. '298 patent (IPR2021-01547)
3. '385 patent (IPR2021-01548)
4. '240 patent (IPR2021-01549)
5. '300 patent (IPR2021-01550)

3. Claim 1 of the '385 Patent

'385

1. A system for storing data comprising:
at least one MLC non-volatile memory module comprising a plurality of individually erasable blocks;
at least one SLC non-volatile memory module comprising a plurality of individually erasable blocks; and
a flash translation layer (FTL); wherein the FTL is adapted to:
a) maintain an address map of at least one of the MLC and SLC non-volatile memory modules, the address map comprising a list of logical address ranges accessible by a computer system, the list of logical address ranges having a minimum quanta of addresses, wherein each entry in the list of logical address ranges maps to a similar range of physical addresses within either the at least one SLC non-volatile memory module or within the at least one MLC non-volatile memory module;

- b) determine if a range of addresses listed by an entry and mapped to a similar range of physical addresses within the at least one MLC non-volatile memory module, fails a data integrity test, and, in the event of such a failure, the controller remaps the entry to the next available equivalent range of physical addresses within the at least one SLC non-volatile memory module;
- c) determine which of the blocks of the plurality of the blocks in the MLC and SLC non-volatile memory modules are accessed most frequently by maintaining a count of the number of times each one of the blocks is accessed; and
- d) allocate those blocks that receive the most frequent writes by transferring the respective contents of those blocks to the at least one SLC non-volatile memory module.

← Limitation [1.F]

← Limitation [1.G]

-01548: Ex. 1003 at Claim 1

3. Limitations [1.F] and [1.G] of the '385 Patent

'298

c) determine which of the blocks of the plurality of the blocks in the MLC and SLC non-volatile memory modules are accessed most frequently by maintaining a count of the number of times each one of the blocks is accessed; and

d) allocate those blocks that receive the most frequent writes by transferring the respective contents of those blocks to the at least one SLC non-volatile memory module.

-01547: Ex. 1001 at Claim 1

Limitation [1.F]

Limitation [1.G]

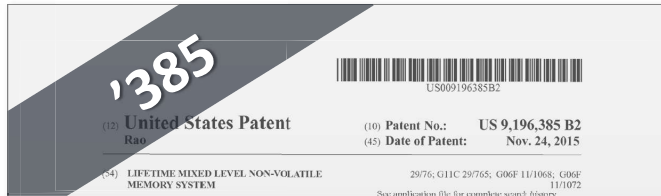
'385

c) determine which of the blocks of the plurality of the blocks in the MLC and SLC non-volatile memory modules are accessed most frequently by maintaining a count of the number of times each one of the blocks is accessed; and

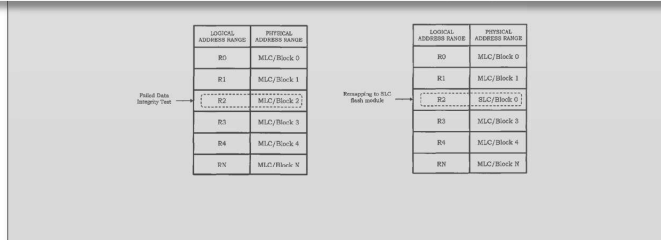
d) allocate those blocks that receive the most frequent writes by transferring the respective contents of those blocks to the at least one SLC non-volatile memory module.

-01548: Ex. 1003 at Claim 1

3. '385 Patent (IPR2021-01548): Instituted Grounds



Ground 1	Claims 1-5 and 11-13 are obvious over Dusija and Sutardja in view of knowledge of POSA
Ground 2	Claims 1-5 and 11-13 are obvious over Moshayedi and Dusija in view of knowledge of POSA
Ground 3	Claim 11 is obvious over Moshayedi, Dusija, and Sutardja in view of knowledge of POSA



3. Flash Translation Layer (FTL)

'385

FIG. 4 depicts another embodiment of the present disclosure. The embodiment is entirely resident within a NAND flash module 50. In particular, a standard NAND flash interface 52 is managed by flash translation layer (FTL) logic 54. The flash translation layer (FTL) 54 manages two NAND flash memory banks 56 and 58, whereby memory bank 56 comprises a plurality of MLC NAND flash memory modules 60a and a plurality of SLC NAND flash memory modules 62a. Memory bank 58 comprises a plurality of MLC NAND flash memory modules 60b and a plurality of SLC NAND flash memory modules 62b.

This embodiment of the present disclosure could function similarly to the system level embodiment discussed earlier with reference to FIGS. 1-3B, but the control functions, such as maintenance of the translation table/address map (FIGS. 2A and 2B); could be conducted within the flash translation layer (FTL) 54 instead of in a device controller 14.

-01548: Ex. 1003 at 6:40-56

3. Flash Translation Layer (FTL)

PO's Sur-Reply

Reply, 6. As explained above, the specification makes clear that the FTL, like the controller, deals with blocks that are *physical* groups of memory cells. Ex. 1003, 3:4-16.

Petitioner argues logical blocks were known to be erasable. Reply, 5. But, Petitioner cites references such as Exhibit 1062, which disclose a *host* (not controller) issuing a sector erase command to erase a logical sector. *Id.* (citing Ex. 1062, 2:10-34). As noted above, claim 1 recites functionality that the FTL, like the controller, performs regarding the claimed “blocks,” and the specification explains the controller deals with *physical* groups of memory cells. The portion of Exhibit 1062 cited by Petitioner (Reply, 5) does not indicate whether a flash translation layer (as required by claim 1 of the '385 patent) performs the functionality cited by Petitioner. As explained by the '385 patent, a host processor deals with logical blocks (Ex. 1003, 2:18-24), i.e., is aware of the blocks at a higher level of abstraction than the FTL (or controller) is aware. *Id.*, FIG. 1, 3:4-16, 3:30-33.

– 01548: Sur-Reply at 4-5

3. Claim 12 of the '385 Patent

'385

12. The system of claim 1, wherein the FTL is adapted to determine which of the blocks of the plurality of the blocks in the MLC and SLC non-volatile memory modules are accessed less frequently by maintaining a count of the number of times each one of the blocks is accessed.

-01548: Ex. 1003 at Claim 12

Petition

As discussed above, Moshayedi's FTL tracks the number of writes and erases to each LBA. Moshayedi, [0024], [0032], [0049]; Section VII.D.1.g. Moshayedi uses this count to "keep ... relatively static data in MLC flash," *i.e.*, it determines the MLC and SLC "modules [that] are accessed less frequently" with this count. *Id.*, Abstract, [0049] (explaining this includes transferring lesser written blocks to MLC); Liu Decl., ¶ 254.

– 01548: Pet. at 66

3. Claim 13 of the '385 Patent

'385

13. The system of claim **12**, wherein the FTL is further adapted to allocate those blocks that receive less frequent writes by transferring the respective contents of those blocks to the at least one MLC non-volatile memory module.

-01548: Ex. 1003 at Claim 13

3. Petitioner’s Argument for Claim 13 of the ‘385 Patent (Dusija-Sutardja Ground)

Petition

First, as discussed above, Sutardja discloses a “data shift analysis,” which occurs as a background operation, in which Sutardja’s FTL copies frequently written data from MLC to SLC. Section VII.C.1.h [1.G]. As part of this analysis, if more space in SLC is required, “control identifies a block of the second NVS memory [*i.e.*, SLC] is a LUB in step 526. Control maps the logical addresses that correspond to the LUB to a block of the first NVS memory [*i.e.*, MLC] in step 528.” Sutardja, [0150]; *see also id.* [0127] (LUB means least used block), [0106], Claim 37 (first memory is MLC; second memory is SLC). Thus, as it goes through this process, it will repeatedly move LUBs to MLC as spaced is necessary. *Id.*, [0145]-[0153]; Liu Decl., ¶ 187. Note that by Sutardja disclosing remapping the logical address to SLC, it would have been understood that it must copy the data to SLC as well, which Sutardja describes in paragraph 167 (“swapping data”). Liu Decl., ¶ 188.

– 01548: Pet. at 49

Petition

Second, as discussed above, Sutardja also discloses determining that a frequency count to a logical address is high and directing an incoming write to a physical SLC block, thereby transferring the data from MLC memory modules to SLC modules. Section VII.C.1.h [1.G]. It also does this in reverse. Specifically, Sutardja discloses “[c]ontrol maps the logical addresses having low write frequencies (e.g., having write frequencies less than a predetermined threshold) to the first NVS memory [*i.e.*, SLC] in step 506.” Sutardja, [0146]-[0147]; Liu Decl., ¶ 189.

– 01548: Pet. at 50

3. Dr. Khatri's Testimony Regarding Claim 13

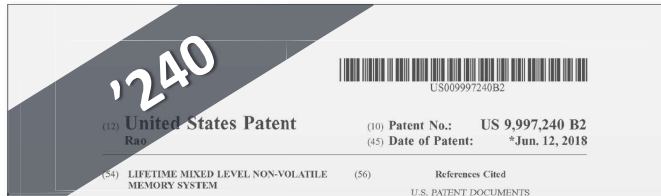
Dr. Khatri

110. Indeed, Sutardja only mentions SLC (or MLC) at three places: claim 37, claim 39, and paragraph [0108]. By disclosing that both the first and second memories may include SLC or MLC flash memory (Ex. 1011, [0108]), Sutardja emphasizes that it is indifferent (or agnostic) to the usage of SLC vs. MLC in either of the first or second NVS memories. *Supra* Section IX.A.2(d). Sutardja's disclosure of paragraph [0162] (discussed above for claim 1, *supra* Section IX.A.2(d)) makes clear to a POSA that Sutardja does **not** seek to transfer contents of blocks to **MLC specifically**, unlike the approach recited in claim 13.

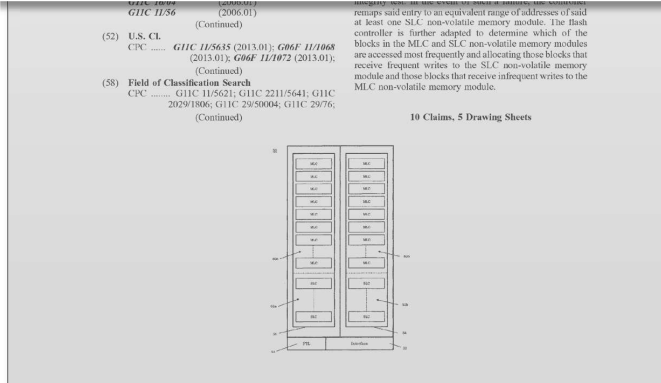
– 01548: Ex. 2014 (Khatri Decl.) at ¶ 110

1. Micron's Expert is not credible on several issues
2. '298 patent (IPR2021-01547)
3. '385 patent (IPR2021-01548)
4. '240 patent (IPR2021-01549)
5. '300 patent (IPR2021-01550)

4. '240 Patent (IPR2021-01549): Instituted Grounds



Ground 1	Claims 1-2 and 6-7 are obvious over Dusija and Sutardja in view of knowledge of POSA
Ground 2	Claims 1-2 and 6-7 are obvious over Dusija, Sutardja, and Chin in view of knowledge of POSA



4. Claim 1 of the '240 Patent

'240

1. A system for storing data comprising:
at least one MLC non-volatile memory module comprising a plurality of individually erasable blocks;
at least one SLC non-volatile memory module comprising a plurality of individually erasable blocks; and
a controller coupled to the at least one MLC non-volatile memory module and the at least one SLC non-volatile memory module, the controller maintaining an address map of at least one of the MLC and SLC non-volatile memory modules, the address map comprising a list of logical address ranges accessible by a computer system, the list of logical address ranges having a minimum quanta of addresses, wherein each entry in the list of logical address ranges maps to a similar range of physical addresses within either the at least one SLC non-volatile memory module or within the at least one MLC non-volatile memory module;
wherein the controller is adapted to determine if a range of addresses listed by an entry and mapped to a similar range of physical addresses within the at least one MLC non-volatile memory module, fails a data integrity test, and, in the event of such a failure, the controller remaps the entry to the next available equivalent range of physical addresses within the at least one SLC non-volatile memory module; and

wherein the controller is further adapted to determine which of the blocks of the plurality of the blocks in the MLC and SLC non-volatile memory modules are accessed most frequently and wherein the controller segregates those blocks that receive frequent writes into the at least one SLC non-volatile memory module and those blocks that receive infrequent writes into the at least one MLC nonvolatile module, and maintain a count value of the blocks in the MLC non-volatile memory module determined to have received frequent writes and that are accessed most frequently on a periodic basis when the count value is a predetermined count value, transfer the contents of the counted blocks in the MLC non-volatile memory module determined to have received frequent writes after reaching the predetermined count value to the SLC non-volatile memory module and which determined blocks in the SLC are determined in accordance with the next equivalent range of physical addresses determined by the controller.

← Limitation
[1.F]

← Limitation
[1.G]

-01549: Ex. 1005 at Claim 1

4. Petitioner's Argument Regarding Dusija for Limitation [1.F]

VERVAIN

Petition

As to [1.F.i], Dusija discloses a “hot count” employed in the memory that may “track[] the endurance or the number of times the erase block has been cycled through erase and program operations” as part of the “enhanced post-write-read error management” process. Dusija, [0153], Figure 19 (block 720). A POSA would have understood that this operation is implemented by the controller because Dusija’s controller “controls and manages higher level memory operations.” *Id.*, [0060]; Liu Decl., ¶ 157.

To the extent Patent Owner might argue that Dusija does not satisfy the limitation, Sutardja discloses a “wear leveling module” as part of its controller that manages a memory that includes MLC and SLC blocks of memory. *Sutardja*,

– 01549: Pet. at 44

4. Dr. Khatri's Testimony Regarding Limitation [1.F]

Dr. Khatri

81. The Petition relies on various portions of Sutardja relating to its “wear leveling” function. Petition, 43-44 (citing Ex. 1011, [0111], [0121], [0146], [0147]). As I discussed above, these paragraphs of Sutardja all relate to logical addresses that do not meet the correct construction of “blocks” (*see supra* Sections VII.A, IX.A.1(a)). But even setting aside this deficiency, the Petition’s analysis for the above-identified portion of limitation [1.F] is flawed. For example, paragraph [0111], cited by the Petition, explicitly discloses a block which has been written to *the least*. Petition, 44; Ex. 1011, [0111]. None of the cited paragraphs which disclose “write frequencies” teach or suggest determining the blocks which are accessed *most frequently*; rather, the Petition simply states in a conclusory manner that the claim language is satisfied.

– 01549: Ex. 2014 (Khatri Decl.) at ¶ 81

4. Dr. Khatri's Testimony Regarding Limitation [1.F]

VERVAIN

Dr. Khatri

82. Petitioner has not explained why a POSA would have sought to modify Dusija to “segregate those blocks that receive frequent writes into the at least one SLC ... module and those blocks that receive infrequent writes into the at least one MLC ... module,” as claimed in limitation [1.F]. For limitation [1.F], the Petition cites to both Sutardja’s “data shift” method (which I discuss in more detail below) and its “wear leveling” function. Petition, 45-46. However, there is no mention in the cited paragraphs or elsewhere in Sutardja of selectively segregating data to SLC or MLC in the manner recited in limitation [1.F].

– 01549: Ex. 2014 (Khatri Decl.) at ¶ 82

4. Dr. Khatri's Testimony Regarding Limitation [1.G]

VERVAIN

Dr. Khatri

92. As discussed above in Section IX.A.1(a), the Petition does not analyze limitation [1.F] under the proper construction of “blocks” (*see supra* Section VII.A). These same deficiencies carry over to both of the Petition’s interpretations of limitation [1.G]. Thus, as explained above in Section IX.A.1(a), the Petition has not established that Sutardja discloses determining the blocks that are accessed, or written to, most frequently. Therefore, the Petition similarly has not established that Sutardja discloses transferring those blocks, much less transferring them in the manner recited in limitation [1.G].

– 01549: Ex. 2014 (Khatri Decl.) at ¶ 92

4. Dr. Khatri's Testimony Regarding Limitation [1.G]

VERVAIN

Dr. Khatri

94. Limitation [1.G] recites in part “transferring the respective contents of those blocks [that are accessed most frequently] ... to the SLC non-volatile memory module.” For the claimed “transferring...” the Petition relies on two disclosures of Sutardja in connection with one another: 1) Sutardja’s disclosure at paragraphs [0148-0149] regarding mapping logical addresses; and 2) Sutardja’s disclosure at paragraph [0167] regarding swapping data. Petition, 48-49. I disagree with the Petition’s analysis, as explained in the following paragraphs.

– 01549: Ex. 2014 (Khatri Decl.) at ¶ 94

4. Dr. Khatri's Testimony Regarding Sutardja's First/Second Memories

VERVAIN

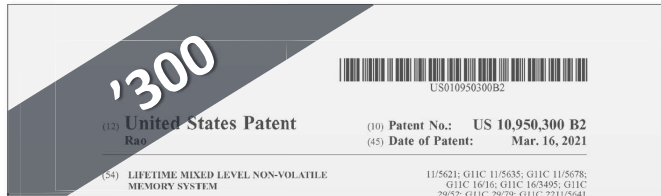
Dr. Khatri

87. The above-identified disclosure from paragraph [0162] of Sutardja makes clear to a POSA that Sutardja does *not* seek to segregate contents of blocks to *SLC or MLC in particular*, unlike the approach claimed in limitation [1.F]. Indeed, Sutardja's emphasis on the normalized wear level (i.e., fraction of lifetime that has been reached), instead of the type of memory (MLC vs. SLC), discourages a POSA from implementing Petitioner's proposed obviousness combination of Dusija and Sutardja. Sutardja describes a fundamentally different approach than the one claimed in claim 1 and limitation [1.F], because segregating the contents of blocks to a specific type of memory (where the type of memory is the basis for the segregation) is qualitatively different than writing to a memory based on an attribute (namely, normalized wear level) that is *independent of* memory type.

– 01549: Ex. 2014 (Khatri Decl.) at ¶ 87

1. Micron's Expert is not credible on several issues
2. '298 patent (IPR2021-01547)
3. '385 patent (IPR2021-01548)
4. '240 patent (IPR2021-01549)
5. '300 patent (IPR2021-01550)
○ 5.1 Petitioner's Mapping of the Claimed "Random Access Volatile Memory" to an Implementation of Dusija's Cache is Deficient (limitation [1.E])
5.2 Dusija Does Not Disclose or Suggest "Retain[ing] Such Stored Data in the Random Access Volatile Memory" (limitation [1.G.2])
5.3 Dusija Does Not Disclose or Suggest "Comparing the Stored Data to the Retained Data in the Random Access Volatile Memory" (limitation [1.H])

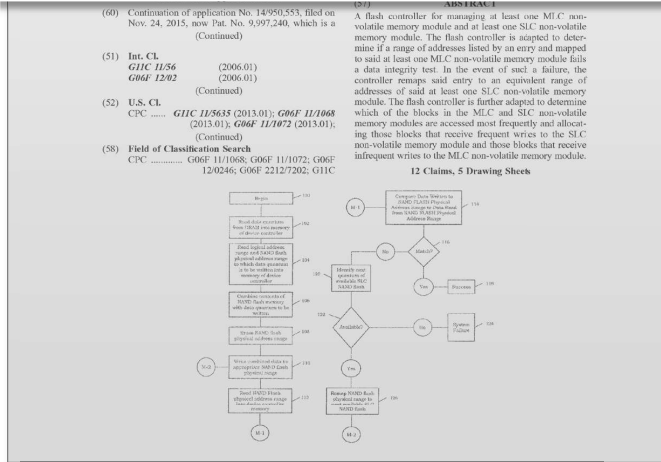
5. '300 Patent (IPR2021-01550): Instituted Grounds



- Ground 1**
- Ground 2**

Claims 1-9 and 11-12 are obvious over Dusija in view of knowledge of POSA

Claim 10 is obvious over Dusija and Sutardja in view of knowledge of POSA



5. Claim 1 of the '300 Patent

'300

1. A system for storing data comprising:
memory space containing volatile memory space and nonvolatile memory space, wherein the nonvolatile memory space includes both multilevel cell (MLC) memory space and single level cell (SLC) memory space;
at least one controller to operate memory elements and associated memory space;
at least one MLC nonvolatile memory element that can be mapped into the MLC memory space;
at least one SLC nonvolatile memory element that can be mapped into the SLC memory space;
▶ at least one random access volatile memory;
an FTL flash translation layer, wherein the at least one controller, or FTL, or a combination of both maintain an address table in one or more of the memory elements and random access volatile memory;
▶ the controller controlling access of the MLC and SLC nonvolatile memory elements and the random access volatile memory for storage of data therein, the controller, in at least a Write access operation to the MLC nonvolatile memory element, operable to store data in the MLC nonvolatile memory element and retain such stored data in the random access volatile memory;

Limitation [1.E]

Limitation [1.G.2]

the controller performing a data integrity test on stored data in the MLC nonvolatile memory element after at least a Write access operation performed thereon by comparing the stored data to the retained data in the random access volatile memory;
wherein the address table maps logical and physical addresses adaptable to the system, wherein the mapping is performed as necessitated by the system to maximize lifetime, and wherein the mapping maps blocks, pages, or bytes of data in either volatile or nonvolatile, or both, memories; and
wherein a failure of the data integrity test performed by the controller results in a remapping of the address space to a different physical range of addresses and transfer of data corresponding to the stored data to those remapped physical addresses from those determined to have failed the data integrity test to achieve enhanced endurance.

Limitation [1.H]

- 01547: Ex. 1007 at claim 1

Petition

g. [1.E] “at least one random access volatile memory;”

As discussed above, Dusija discloses or renders obvious “[a] *system comprising . . . at least one random access volatile memory.*” See limitation [1.A.1] (describing that Dusija would have been understood, or it would have been obvious, for its controller to include RAM and for its cache to be RAM). Liu Decl., ¶ 158.

h. [1.F] “an FTL flash translation layer, wherein the at least one controller, or FTL, or a combination of both maintain an address table in one or more of the memory elements and random access volatile memory”

– 01550: Pet. at 42

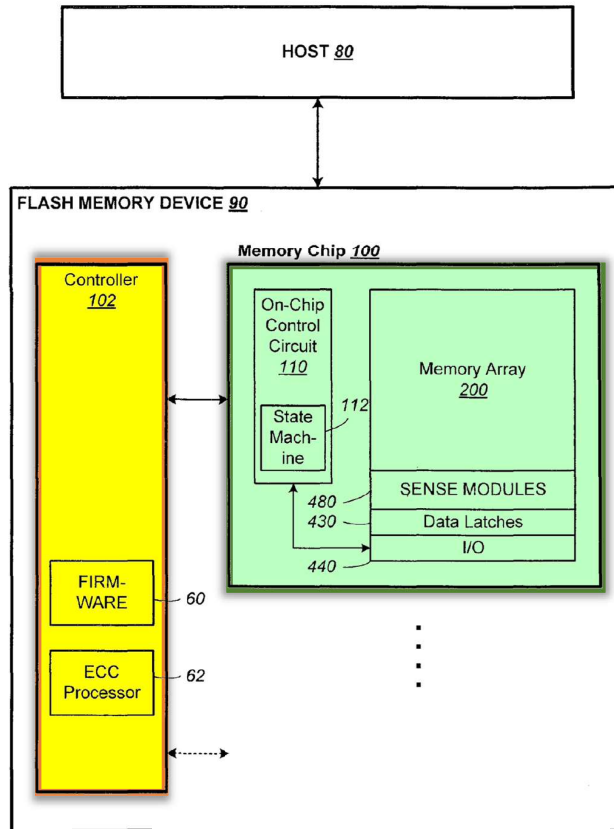
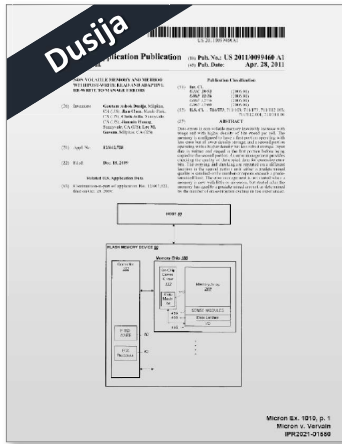


FIG. 1

- 01550: Ex. 1010 (Dusija) at Fig. 1

5.1 Dusija's Cache

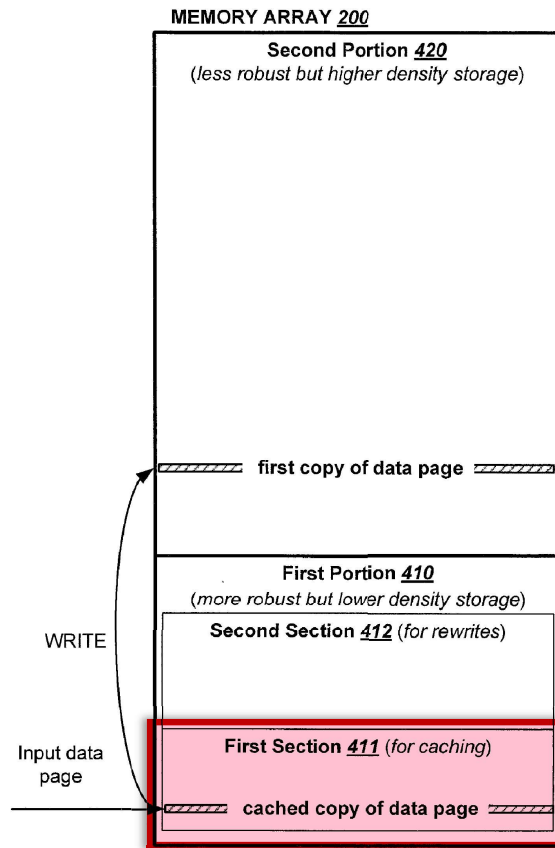
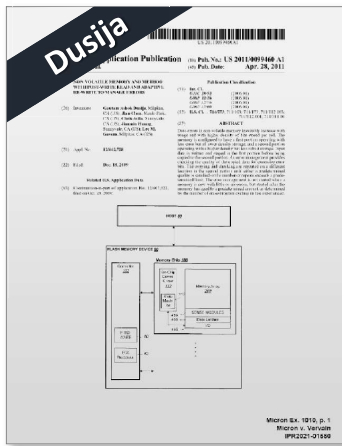
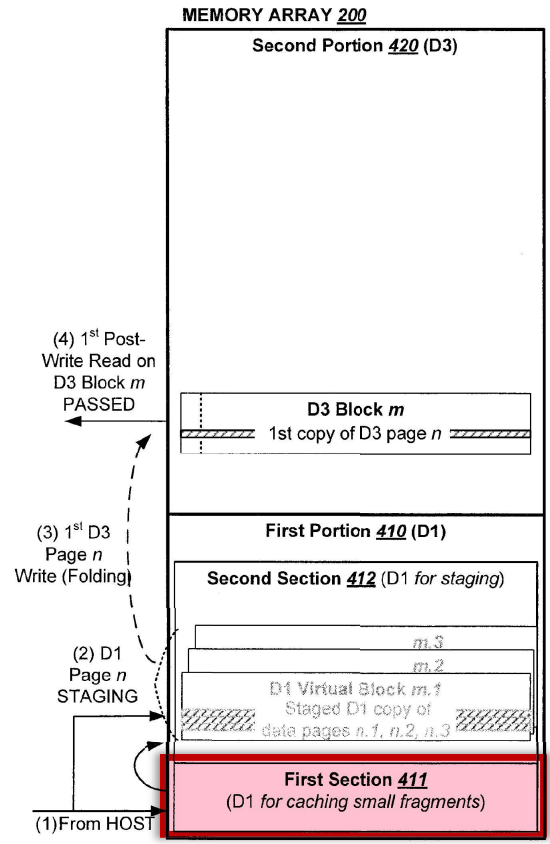
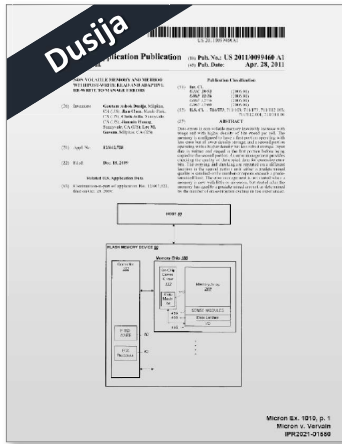


FIG. 16A

- 01550: Ex. 1010 (Dusija) at Fig. 16A

5.1 Dusija's Cache



Example of successful D1 to D3 Folding

FIG. 20A

- 01550: Ex. 1010 (Dusija) at Fig. 20A

5.1 Dr. Liu's Testimony Regarding the Cache

Dr. Liu

Q. I will ask the question again. I know that the question has been asked, but it has not been answered.

Is the cache at the controller, or is it at the memory array, when your implementation has been done?

A. The cache will be used in conjunction with Dusija's controller. That would be obvious.

Q. And when that is done, as you propose, the cache is then located at the controller; is that right?

A. Again, cache would be used in conjunction with the controller and that cache would be RAM.

– 01550: Ex. 2015 at 87:3-22

5.1 Dr. Liu's Testimony Regarding the Cache

Dr. Liu

Q. Dr. Liu, you still haven't stated where the cache is located in your proposed obviousness argument. It has to be somewhere, unless you tell me that somehow it's not anywhere.

Where is it located in your proposed obviousness implementation?

A. I am stating the obvious – it would be obvious to use a cache in conjunction with the controller, and that cache will be implemented with RAM. And I think it's very clear.

– 01550: Ex. 2015 at 87:23-88:11

5.1 Dr. Khatri's Testimony Regarding Dusija's Cache

69. In my opinion, a POSA would have understood that Dusija's cache is not a "random access volatile memory" as required by limitation [1.E]. Instead, Dusija's cache (annotated below with an orange oval in FIG. 16A of Dusija) is within flash memory (specifically, at a first section 411 of a first portion 410 of flash memory array 200, which is annotated with a green oval in FIGS. 1 and 16A), and the flash memory array is non-volatile. Therefore, a POSA would have understood that Dusija's cache is **non-volatile** memory.

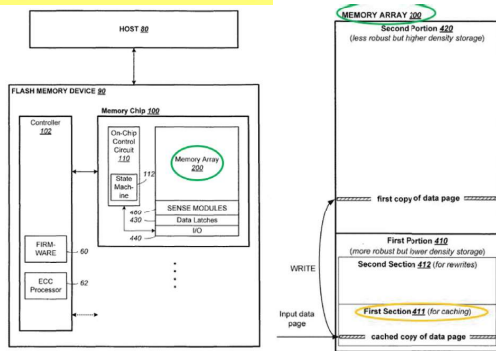


FIG. 1

FIG. 16A

Ex. 1010, FIGS. 1, 16A (annotated), [0003] ("Unlike RAM ..., flash memory is **non-volatile**...") (emphasis added); *see also id.*, [0020] ("the first portion serves as a cache for incoming data"), [0130] ("The first portion acts as a low-density write cache"), [0131] ("The incoming data is cached in the first section of the first portion"), [0133] ("The first portion 410 is further provided with a first section 411 for caching incoming data"), [0134] ("a cached copy is cached in the first section 411 of the first portion 410").

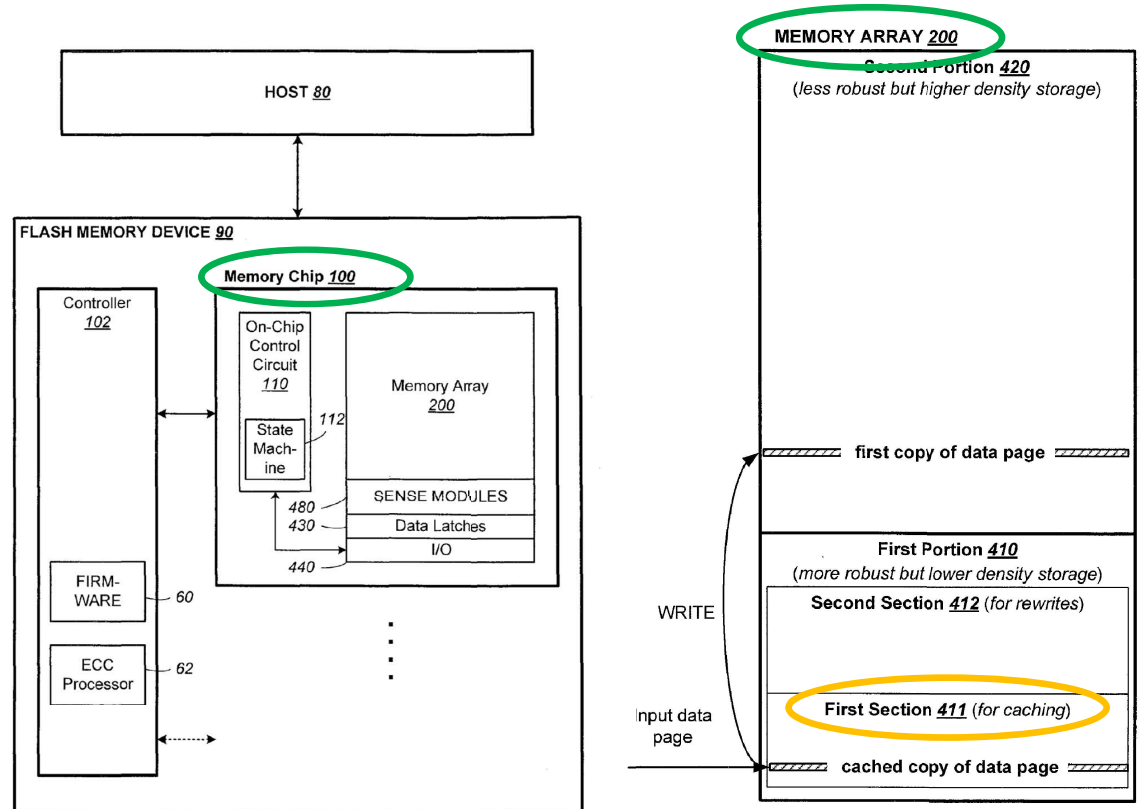


FIG. 1

FIG. 16A

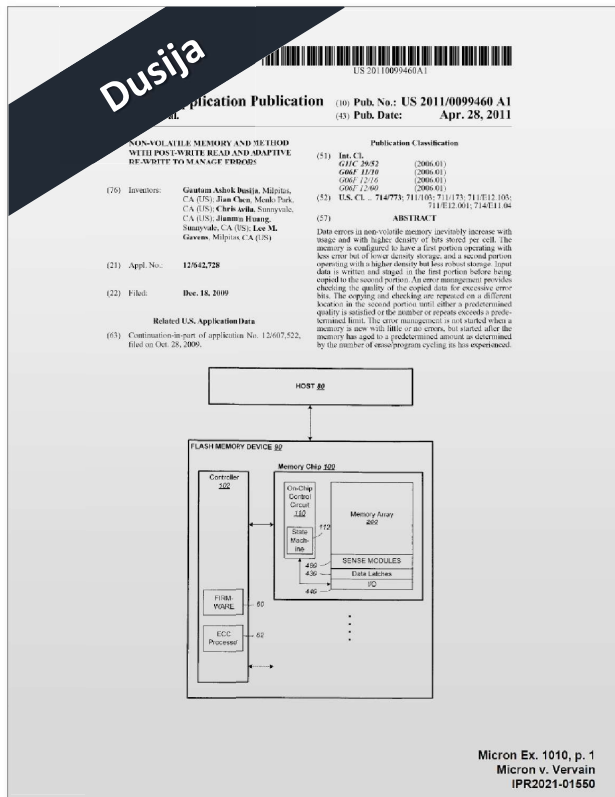
– 01550: Ex. 2014 (Khatri Decl.) at ¶ 69; Ex. 1010 (Dusija) at Figs. 1 and 16A

Petitioner's Reply

⁶ Dusija states that its “cache” could be, “[i]n an alternative embodiment,” flash memory (what Dusija calls the “first portion). Dusija, [0020]. This only reinforces that the primary embodiments (not “alternative”) would be understood to employ a RAM cache and certainly does not act to limit the claim scope. Liu Decl., n. 6.

– 01550: Pet. at 33 n.6

5.1 Dusija's Preferred Embodiment



[0021] In the preferred embodiment, the first portion is further provided with a first section and a second section. The incoming data is cached in the first section of the first portion and a first copy of the data is written to the second portion. Afterwards, the first copy in the second portion is read back to check for excessive error bits. If the error bits exceeded a predetermined amount, a second copy of the incoming data is written to the second section of the first portion.

[0131] In the preferred embodiment, the first portion is further provided with a first section and a second section. The incoming data is cached in the first section of the first portion and a first copy of the data is written to the second portion. Afterwards, the first copy in the second portion is read back to check for excessive error bits. If the error bits exceeded a predetermined amount, a second copy of the incoming data is written to the second section of the first portion.

– 01550: Ex. 1010 (Dusija) at ¶¶ 0131, 0021

5.1 Dr. Liu's Testimony About Dusija

Dr. Liu

- Q. Is caching described in the first portion of Dusija's memory in Dusija's preferred embodiment?
- A. We need to be very careful, because Dusija sometimes in the context in his description refer to a preferred embodiment within the alternative embodiment. So, when you use the word "preferred embodiment," it could be in the context of **already being an alternative embodiment**. So, I do not want to be confusing as somehow this preferred embodiment that was used in the context of alternative embodiment all of a sudden supersedes this alternative embodiment. No.

Dusija described in a few instances the **preferred embodiment within the alternative embodiment**, but that preferred embodiment in that context is still in the context of being an alternative embodiment, and so, I would prefer to use the word, the embodiment that is mentioned first and foremost in the conventional embodiment as what I meant perhaps earlier, and which I would not use anymore as a preferred embodiment. I will just say that as an embodiment of choice, as the primary embodiment, simply because it was mentioned first and was not mentioned in the context of being an alternative embodiment.

– 01550: Ex. 2015 at 61:21-63:1

5.1 Dusija's Disclosure of RAM

Dusija

US 2010099460A1

Application Publication (10) Pub. No.: US 2011/0099460 A1
(43) Pub. Date: Apr. 28, 2011

NONVOLATILE MEMORY AND METHOD WITH PERSISTENT READ AND WRITABLE RE-WRITE TO MANAGE ERRORS

(75) Inventors: **Guatam Ashik Dusija**, Milpitas, CA (US); **Frank Chao**, Menlo Park, CA (US); **Chiraj Gupta**, Sunnyvale, CA (US); **Jianming Huang**, Sunnyvale, CA (US); **Lee M. Gavron**, Milpitas, CA (US)

(21) Appl. No.: 12/942,728
(22) Filed: Dec. 18, 2009

Related U.S. Application Data
(63) Continuation-in-part of application No. 12/607,522, filed on Oct. 26, 2009.

Publication Classification
(51) Int. Cl. G1C 28/32 (2006.01)
G06F 13/10 (2006.01)
G06F 12/16 (2006.01)
G06F 12/00 (2006.01)
(52) U.S. CL. 714/773; 714/105; 714/113; 714/112; 105; 714/112; 001; 714/111; 04

ABSTRACT
Data errors in non-volatile memory inevitably increase with usage and with higher density of bits stored per cell. The memory is configured to have a first portion operating with low error rate of low density storage, and a second portion operating with a higher density but less robust storage. Input data is written and staged in the first portion before being copied to the second portion. An error management unit provides checking the quality of the copied data for excessive error bits. The copying and checking are repeated on a different location in the second portion until either a predetermined quality is attained or the number of repeats exceeds a predetermined limit. The error management is not started when a memory is new with little or no errors, but started after the memory has aged to a predetermined amount as determined by the number of erase-program cycling it has experienced.

Micron Ex. 1010, p. 1
Micron v. Vervain
IPR2021-01550

[0003] Solid-state memory capable of nonvolatile storage of charge, particularly in the form of EEPROM and flash EEPROM packaged as a small form factor card, has recently become the storage of choice in a variety of mobile and handheld devices, notably information appliances and consumer electronics products. Unlike RAM (random access memory) that is also solid-state memory, flash memory is non-volatile, and retaining its stored data even after power is turned off. Also, unlike ROM (read only memory), flash memory is rewritable similar to a disk storage device. In spite of the higher cost, flash memory is increasingly being used in mass storage applications. Conventional mass storage, based on rotating magnetic medium such as hard drives and floppy disks, is unsuitable for the mobile and handheld environment. This is because disk drives tend to be bulky, are prone to mechanical failure and have high latency and high power requirements. These undesirable attributes make disk-based storage impractical in most mobile and portable applications. On the other hand, flash memory, both embedded and in the form of a removable card are ideally suited in the mobile and handheld environment because of its small size, low power consumption, high speed and high reliability features.

- 01550: Ex. 1010 (Dusija) at ¶ 0003

5.1 Patent Owner's Argument Regarding Dusija's Cache

PO's Sur-Reply

In any event, to the extent Petitioner is now streamlining its obviousness theory to be about “add[ing]” a RAM cache in a manner that leaves in place Dusija’s existing cache, Petitioner has not adequately explained why a person of ordinary skill in the art (POSITA) would have implemented *two* caches, particularly, when (as discussed below) Dusija explicitly states that locating its cache outside of flash memory is undesirable.² To the extent Petitioner contends that the RAM cache is added as a *replacement* for Dusija’s flash-based cache (i.e., with Petitioner’s obviousness implementation moving the cache to be outside of the flash memory), Petitioner has not adequately established why a POSITA would have sought to do so against the guidance of Dusija. To the extent Petitioner contends that flash-based caching is only disclosed in the context of an “alternative” embodiment Dusija (Petition, 33 n.6), that is incorrect—paragraph [0131] of Dusija clearly discloses that

– 01550: Sur-Reply at 7

5.1 Dr. Khatri's Testimony Regarding Limitation [1.E]

Dr. Khatri

84. In my opinion, Petitioner's proposed obviousness implementation for limitation [1.E] would have been recognized as undesirable and disadvantageous for several reasons. For example, Petitioner's proposed "implement[ation] [of] Dusija's cache as RAM" (Petition, 36) would have a) entailed replacing Dusija's intrinsic disclosure of a cache with a fundamentally different component, and b) it would also have changed the fundamental principle of Dusija's operation with respect to using the cache. Indeed, such an implementation would have unnecessarily moved the cache outside the very flash memory in which Dusija places it, which a POSA would have found to be an efficient and technically well-motivated placement). Additionally, Petitioner's proposed implementation would have placed the cache outside the flash memory in a manner that detracts from performance. One reason for this is that is because the cached data will no longer be in the flash memory, which a POSA would have found to be advantageous in order for the write to flash memory to be completed efficiently. For all of these reasons, it is my opinion that a POSA would not have sought to modify Dusija's cache in the manner proposed in the Petition for limitation [1.A.1] (and incorporated for the Petition's analysis of limitation [1.E]).

– 01550: Ex. 2014 (Khatri Decl.) at ¶ 84

5.1 Dusija's Data Comparison

Dusija

US 20 1009460A1

Application Publication (10) Pub. No.: US 2011/0099460 A1
(43) Pub. Date: Apr. 28, 2011

NONVOLATILE MEMORY AND METHOD WITH PSEUDO-RANDOM WRITE AND READ-AND-WRITE TO MANAGE ERRORS

(75) Inventors: **Geetan Ashik Dusija**, Milpitas, CA (US); **Frank Chao**, Menlo Park, CA (US); **Chris LaBella**, Sunnyvale, CA (US); **Shamir Huang**, Sunnyvale, CA (US); **Lee M. Garrow**, Milpitas, CA (US)

(21) Appl. No.: 12/942,728
(22) Filed: Dec. 18, 2009

Related U.S. Application Data
(63) Continuation-in-part of application No. 12/607,522, filed on Oct. 26, 2009.

Publication Classification
(51) Int. Cl. G1C 29/32 (2006.01); G06F 13/10 (2006.01); G06F 12/16 (2006.01); G06F 12/00 (2006.01); U.S. CL. 714/773; 714/105; 714/113; 714/112; 105; 714/112.001; 714/111.04

ABSTRACT
Data errors in non-volatile memory inevitably increase with usage and with higher density of bits stored per cell. The memory is configured to have a first portion operating with low error bit of lower density storage, and a second portion operating with a higher density but less robust storage. Input data is written and staged in the first portion before being copied to the second portion. An error management module provides checking the quality of the copied data for excessive error bits. The copying and checking are repeated on a different location in the second portion until either a predetermined quality is attained or the number of repeats exceeds a predetermined limit. The error management is not started when a memory is new with little or no errors, but started after the memory has aged to a predetermined amount as determined by the number of erase-program cycling it has experienced.

Micron Ex. 1010, p. 1
Micron v. Vervain
IPR2021-01550

[0135] According to another preferred embodiment, the memory array is provided with a set of data latches on an integrated circuit chip, the checking of the error bits in the first copy is accomplished by loading the first copy and the cached copy into the set of data latches and making a comparison at the set of data latches.

[0136] By not making the comparison at the controller, the data does not have to be toggled out to the controller, much time can be saved. FIG. 1 shows the data latches 430, which is on-chip, for the data comparison to take place.

– 01550: Ex. 1010 (Dusija) at ¶¶ 0135-0136

5.1 Dr. Liu's Testimony Regarding Dusija's Comparison

VERVAIN

Dr. Liu

Q. And Dusija's disclosure of doing the comparison at the flash memory itself is advantageous compared to doing the comparison outside the flash memory; right?

A. I would not say it's advantageous. I would say that a certain aspect, it may -- you may have a different -- for lack of better words, you may have different perspective. But taken in totality, it's not -- I would be very cautious to use the word "advantageous," because again, I want to offer many, many perspective where it's disadvantageous to do it Dusija's way.

So, I would be very cautious when you use the word "advantageous" to characterize, because it's -- it's debatable, and it's only maybe in one narrow sense, in one particular perspective, but taken overall, it's not advantageous.

– 01550: Ex. 2015 at 110:25-111:14, 111:22-112:2

5.1 Dr. Liu's Testimony Regarding Dusija's Controller

Dr. Liu

130. Second, Dusija discloses controller 102, which includes firmware 60, that manages memory operations. Dusija at Fig. 1, [0059], [0062]. A POSA would have understood, and at least found it obvious, that the controller includes RAM in order to buffer data and commands, execute the firmware, and manage the memory operations. Ex. 1019, Micheloni at 132 (showing "Data SRAM" as being

– 01550: Ex. 1009 at ¶ 130

Dusija's controller lacks RAM. POR, 45 (citing Khatri Decl., ¶ 83). And, contrary to Patent Owner's assertion, I have already explained why a POSA would have understood Dusija's controller to have RAM. Liu Decl., ¶ 130.

– 01550: Ex. 1057 at ¶ 23

1. Micron's Expert is not credible on several issues
2. '298 patent (IPR2021-01547)
3. '385 patent (IPR2021-01548)
4. '240 patent (IPR2021-01549)
5. '300 patent (IPR2021-01550)
5.1 Petitioner's Mapping of the Claimed "Random Access Volatile Memory" to an Implementation of Dusija's Cache is Deficient (limitation [1.E])
● 5.2 Dusija Does Not Disclose or Suggest "Retain[ing] Such Stored Data in the Random Access Volatile Memory" (limitation [1.G.2])
5.3 Dusija Does Not Disclose or Suggest "Comparing the Stored Data to the Retained Data in the Random Access Volatile Memory" (limitation [1.H])

5.2 Petitioner's Argument for Limitation [1.G.2]

Petition

Second, Dusija discloses “*retain[ing] such stored data in the random access volatile memory.*” Dusija explains that once data is written to MLC, the data is read back and compared to the original copy that had been retained in a “cache.” Dusija, [0112]-[0115]. As Section VII.C.1.b, limitation [1.A.1] demonstrates, a POSA would have understood Dusija’s teaching of retaining data in a “cache” to disclose retaining the data in a RAM, and thus to “*retain such stored data in the random access volatile memory.*” Sections VII.C.1.b and VII.C.1.g demonstrate that at a minimum, it would have been obvious to implement Dusija’s “cache” as random access volatile memory. Liu Decl., ¶¶ 173-74.

Dusija discloses this same process as part of a “post-write read and adaptive rewrite” which it depicts in flow-chart form in Figure 15. Dusija, [0119]-[0124].

- k. [1.H] “the controller performing a data integrity test on stored data in the MLC nonvolatile memory element after at least a Write access operation performed thereon by comparing the stored data to the retained data in the random access volatile memory;”

Dusija discloses, and at the minimum renders obvious, this limitation.

– 01550: Pet. at 47-48

Dr. Khatri

89. Even though the Petition makes no argument for the first mapping, I have considered Petitioner's first mapping (discussed above for limitation [1.E]) (i.e., the mapping involving Dusija's controller) for limitation [1.G.2], and it is my opinion that that mapping is deficient for the reasons discussed above for limitation [1.E]. Dusija teaches that there are speed and efficiency advantages to storing a copy of data in a cache on-chip within Dusija's memory chip 100 (and specifically, within memory array 200 thereof) for two reasons. First, this approach saves time during the data compare steps—e.g., by not having to “toggle” the data out to a controller (Ex. 1010, [0136]). Secondly, in response to excessive errors, the cached copy of the embodiment cited by Petitioner (the embodiment comprising FIGS. 14A-14B of Dusija, *see* Petition, 33) simply becomes the valid copy without requiring another write operation from an external controller, Ex. 1010, [0129]. In my opinion, a POSA would have readily recognized that a mapping which requires retaining data in an off-chip controller goes against Dusija's explicit teachings and guidance, and indeed sacrifices Dusija's explicitly claimed advantages. A POSA would not reasonably have considered such a modification desirable in the context of Dusija's disclosure.

– 01550: Ex. 2014 (Khatri Decl.) at ¶ 89

Dr. Khatri

90. As I noted above in this section (Section IX.A.2), for limitation [1.G.2] the Petition only relies on the second mapping introduced for limitation [1.E] (where antecedent basis for “the random access volatile memory” of limitation [1.G.2] is found). That second mapping—involving a proposed modification of Dusija’s cache to include RAM to allegedly arrive at the “at least one random access volatile memory” of limitation [1.E]—is deficient for at least the reasons that I discussed above regarding limitation [1.E]. The Petition’s main argument for limitation [1.G.2] that “a POSA would have understood Dusija’s teaching of retaining in a ‘cache’ to disclose retaining the data in a RAM, and thus to ‘retain such stored data in the random access volatile memory’” (Petition, 48) is incorrect. As I discussed above for limitation [1.E] (*see supra* Section IX.A.1), Dusija’s cache is within flash memory (Ex. 1010, FIGS. 16A-16B, FIGS. 20A-20C, [0131], [0133], [0135], [0137], [0163-0178]), which is different than RAM, as stated in the background section of Dusija. Ex. 1010, [0003] (“Unlike RAM ... flash memory is non-volatile”).

– 01550: Ex. 2014 (Khatri Decl.) at ¶ 90

Dr. Khatri

91. The Petition's backup argument for limitation [1.G.2] that "it would have been obvious to implement Dusija's 'cache' as random access volatile memory" (Petition, 48) is also erroneous, as explained above for limitation [1.E] (*see supra* Section IX.A.1). Petitioner's proposed "implement[ation] [of] Dusija's 'cache' as random access volatile memory" (Petition, 48) would have entailed replacing Dusija's cache (which is intrinsically described as part of flash memory in Dusija) with a RAM that is necessarily outside of the flash memory array. This is so because flash memory is non-volatile, and RAM is volatile, *see* Ex. 1010, [0003]. A POSA would not have been motivated to retain data in such a RAM *outside* of the flash memory array, because, as explained above in Section IX.A.1 for limitation [1.E], Dusija's caching of data within the flash memory, in the embodiment cited by the Petitioner (the embodiment comprising FIGS. 14A-14B of Dusija), Petition, 33,

– 01550: Ex. 2014 (Khatri Decl.) at ¶ 91

Dr. Khatri

92. In other words, retaining the data in a RAM *outside* of the flash memory array would require additional processing to be performed (requiring additional time and/or resources) in order to transfer the data into the first portion of the memory (which would be in RAM, according to this approach of the Petitioner) in the event of too many errors detected at Dusija's post-write read, and indeed is counseled against by Dusija itself. Ex. 1010, [0136].

93. The Petition's analysis for limitation [1.G.2] does not address these issues at all, and the Petition's analysis for limitation [1.A.1] (cited at the Petition's analysis of limitation [1.G.2]) similarly discusses caches and RAM generally without considering the specific details of Dusija's system/process relevant to Petitioner's proposed modification. Thus, in my opinion the Petition's discussion of limitation [1.G.2] is flawed.

3. *Limitation [1.H] ("the controller performing a data integrity test on stored data in the MLC nonvolatile memory element after at least a Write access operation performed thereon by comparing the stored data to the retained data in the random access volatile memory")*

– 01550: Ex. 2014 (Khatri Decl.) at ¶¶ 92-93

Petitioner's Reply

cache as RAM.” Petition, 36 (emphasis added). **Third**, the discussion of limitation [1.G.2] cites the paragraph of Dr. Liu’s declaration explaining that “[Dusija’s] ‘cache’ would have been understood to be *Dusija’s controller RAM*.” Petition, 48 (citing Liu Decl., ¶ 173). The Petition’s obviousness theory is sufficiently clear.

As to the substance of both so-called “mappings,” the POR merely repeats the arguments made for limitation [1.E] that are refuted above in Section III.A.3. PO’s arguments as to limitation [1.G.2] should be rejected for the same reasons.

– 01550: Reply at 23-24

1. Micron's Expert is not credible on several issues
2. '298 patent (IPR2021-01547)
3. '385 patent (IPR2021-01548)
4. '240 patent (IPR2021-01549)
5. '300 patent (IPR2021-01550)
5.1 Petitioner's Mapping of the Claimed "Random Access Volatile Memory" to an Implementation of Dusija's Cache is Deficient (limitation [1.E])
5.2 Dusija Does Not Disclose or Suggest "Retain[ing] Such Stored Data in the Random Access Volatile Memory" (limitation [1.G.2])
● 5.3 Dusija Does Not Disclose or Suggest "Comparing the Stored Data to the Retained Data in the Random Access Volatile Memory" (limitation [1.H])

5.3 Petitioner's Arguments for Limitation [1.H]

Petition

Third, Dusija would have been understood to disclose that the copy of data read back from MLC is stored with the original copy in volatile memory (or at a minimum it would have been obvious). Recall that data is retained in Dusija's "cache," which would have been understood, and at least obvious, to be the controller's RAM. See Section VII.C.1.b. As Paley, Micheloni, and Harari illustrate, a controller typically has a RAM as a cache and to support other processing. *Id.* Indeed, it was well known for a controller to use such RAM for such operations, i.e., Dusija's data integrity test. *Id.*; Section V.A.4, VII.C.1.b, VII.C.1.g; Liu Decl. ¶ 178.

Dusija's data integrity test may be performed after each write. Dusija, [0203]; Liu Decl. ¶ 179.

Thus, Dusija discloses performing the data integrity test "*by comparing the stored data to the retained data in the random access volatile memory.*" See Dusija, [0112] ("post write read" and "comparison with the original copy which may be

– 01550: Pet. at 49

Petitioner's Reply

PO's only additional argument is its meritless contention that "an off-chip comparison would have been contrary to well-known design principles regarding fast cache operation." POR, 55. Though the POR cites Dr. Khatri's declaration, he makes no such statement and the evidence on which he relies does not support PO's contention. Ex. 2014, ¶ 97; Liu Reply, ¶ 61. In fact, PO's assertion is contradicted by the common and well-understood use of RAM caches in SSDs. Liu Decl., ¶ 130; Liu Reply, ¶¶ 64-71.

Otherwise, in substance, PO's argument for both "mappings" merely repeats the flawed arguments made above for limitations [1.E] and [1.G.2], namely, that Dusija teaches away from the use of a controller RAM cache. Those arguments should be rejected for the reasons explained in Section III.A.3.

– 01550: Reply at 24-25

5.3 Dusija's Data Comparison

Dusija

US 201009466A1

Application Publication (10) Pub. No.: US 2011/0099460 A1
 (43) Pub. Date: Apr. 28, 2011

NONVOLATILE MEMORY AND METHOD WITH PERSISTENT READ AND WRITABLE RE-WRITE TO MANAGE ERRORS

(75) Inventors: **Gautam Ashok Dusija**, Milpitas, CA (US); **Frank Chua**, Menlo Park, CA (US); **Chirukunthu Sumanvile**, CA (US); **Jianming Huang**, Sunnyvale, CA (US); **Lee M. Gavens**, Milpitas, CA (US)

(21) Appl. No.: 12/942,728
 (22) Filed: Dec. 18, 2009

Related U.S. Application Data
 (63) Continuation-in-part of application No. 12/607,522, filed on Oct. 26, 2009.

Publication Classification
 (51) Int. Cl. G1C 29/32 (2006.01)
 G06F 11/09 (2006.01)
 G06F 12/16 (2006.01)
 G06F 12/60 (2006.01)
 (52) U.S. CL. 714/773; 711/105; 711/173; 711/174; 711/112; 091; 714

ABSTRACT
 Data errors in non-volatile memory inevitably increase with higher density of bits stored per memory cell. A memory is configured to have a first portion (read) with a low error rate of lower density storage, and a second portion (write) with a higher density but a higher error rate. Data is written and staged in the first portion before being copied to the second portion. An error management circuit checks the quality of the copied data for excess bits. The copying and checking are repeated on a location in the second portion until either a preset quality is satisfied or the number or excess bits exceeds a predetermined limit. The error management is not started if the memory is new, write-able, or an error, but started if the memory has aged to a predetermined amount as set by the number of erase-program cycling in its life span.

MICRON EX. 1010, p. 1
 Micron v. Vervain
 IPR2021-01550

[0135] According to another preferred embodiment, the memory array is provided with a set of data latches on an integrated circuit chip, the checking of the error bits in the first copy is accomplished by loading the first copy and the cached copy into the set of data latches and making a comparison at the set of data latches.

[0136] By not making the comparison at the controller, the data does not have to be toggled out to the controller, much time can be saved. FIG. 1 shows the data latches 430, which is on-chip, for the data comparison to take place.

– 01550: Ex. 1010 (Dusija) at ¶¶ 0135-0136