Xiaohua Huang, *pro se*
P.O. Box 1639, Los Gatos CA95031
Tel: 669 273 5633
Email: paul-huang03@outlook.com

IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF CALIFORNIA

C 24 07181 VKD

Xiaohua Huang *Pro Se,*

Plaintiff,

v.

TetraMem, Inc.

Defendant.

Civil Action No.:

TRIAL BY JURY DEMANDED

PLAINTIFF XIAOHUA HUANG'S COMPLAINT FOR INFRINGEMENT OF PATENT

Plaintiff, Xiaohua Huang ("Plaintiff" or "Huang"), hereby files his Complaint for Patent Infringement against Defendants, THAT'S TetraMem, Inc. ("TetraMem" or "Defendant and respectfully alleges as follows:

NATURE OF THE ACTION

1. This is a civil action for patent infringement under the Patent Laws of the United States, 35 U.S.C. §101, et seq., to prevent and enjoin Defendant TetraMem, Inc. (hereinafter "TetraMem" or "Defendant") from infringing and profiting, in an illegal and unauthorized manner and without authorization and/or consent from Plaintiff, from U.S. Patent No. RE45259 (the "'RE259patent" which is attached hereto as Exhibit A and incorporated herein by reference, and pursuant to 35 U.S.C.

§271, and to recover damages and costs.

THE PARTIES

2. Xiaohua Huang is an individual, he currently resides at Campbell, CA95008. Huang has developed the state of the art high speed and low power U.S. patented circuit and logic design to build high speed and low power IC chip since the year of 2000, one of those patents is US Patent RE45259.

3. TetraMem is or purports to be a company having its head quarter office address in: 4027 Clipper Ct, Fremont, CA 94538 with its website: https://www.tetramem.com/. TetraMem was founded by Mr.Glenn (Ning)Ge, Mr. Joshua Yang and Qiangfei Xia based on the ReRAM research initiated from HP Lab. and continued in University of Massachusetts and University of Southern California.

JURISDICTION AND VENUE

4. This action arises under the patent laws of the United States, 35 U.S.C. § 101, et seq. This Court has jurisdiction over the subject matter of this action pursuant to 28 U.S.C. §§ 1331 and 1338(a). Venue is proper in this District pursuant to 28 U.S.C. §§1391(b) - (c) and 28 U.S.C. § 1400(b) in that Defendant which has its regular business and main office in 4027 Clipper Ct, Fremont, CA 94538 can be sued in this judicial district.

BACKGROUND FACTUAL ALLEGATION

5. A true and correct copy of the 'RE259patent is attached hereto as Exhibit A.

The 'RE259patent is valid and owned by Plaintiff Mr. Huang as the inventor.

6. In Nov. 2000 "Huang" found CMOS Micro Device Inc. ("CMOS") in Campbell, California. Huang developed advanced TCAM design in CMOS and invented many patented high speed and low power IC design circuit and logic such as the circuit and logic design claimed in the claim 29 of US patent RE45259. At least the claim 29 of US patent RE45259 read the circuit and logic design found in the ReRAM IP and chip which have been designed and made by TetraMem in its office in 4027 Clipper Ct, Fremont, CA 94538.

7. In previous case 5:23-cv-04936-SVK, under Defendant's instruction Defendant's Counsel used all the erroneous information and the erroneous decision of the previous cases to defame Plaintiff and cheat the Court. The Court was biased by defendant's conduct, Plaintiff were prejudiced by the Court, plaintiff withdrew that complaint. In case 5:23-cv-04936-SVK defendant Counsels uses case No. 2:15-cv-1413 Mr. Xiaohua Huang v. Huawei Technologies Ltd. to prove that Plaintiff is a frivolous litigant because Plaintiff was sanctioned by Judge Judge Rodney Gilstrap/ Roy Payne. Case No. 2:15-cv-1413 was my first case filed I have no any experience and knowledge of the Court and litigation. I collected and produced evidence that many chips used in Huawei's Internet router and switches infringes my asserted patents. Judge Roy Payne was very kind and nice to me and gave me two month time to retain the counsels who are very experienced in the local court. I have been grateful and felt in debt to Judge Payne although I failed to retain the counsels. After that Huawei used six its employees, one internal counsel and the

counsels to make false statement and perjured declaration. Although I produced the evidence to prove all the Huawei's declaration are contradictive to the public material evidence the Court took Huawei's false statement and perjured declaration to judge me loose the case and have me sanctioned. Defendant Counsels uses case No. 3:18-cv-6654 Mr. Xiaohua Huang v. Nephos Inc. to defame me. Case No. 3:18-cv6654 Mr. Xiaohua Huang v. Nephos Inc. was the third case I filed. In 2014 I taught the engineers of MediaTek how to use my patents to design the low power TCAM for their networking chips, then the power consumption of MediaTek's chip reduced from over 200watt to much less than 100 watt, and sold to Amazon.com, Inc., but MediaTek refused to pay anything to me, I filed complaint in 2018. Hon. William Alsup assigned Magistrate Jude Kim to mediate the case. Defendant Counsel Brandon Stroy cheated me and said to me that he can help me to have the defendant to pay me the particular amount to settle the case, that amount is many times more than the amount which Judge Kim mediated for me, so I did not accept the mediation result. After that Brandon Story told me that his client only want to pay one-tenth of the amount that Judge Kim mediated. The case was not settled, then Brandon Story imitated Huawei's case No. 2:15-cv-1413 and have the employees of MediaTek to make perjured declaration to cheat the Judge and win the case. I should follow Judge Kim. Although losing the case but I am grateful and in debt to Hon. Judge Alsup and Kim. Everyone knows that my case with Huawei No. 2:15-cv-1413 was erroneously decided by the Courts. The court system tolerates the errors made by Judge. Some defendant counsels used what the Court taking the

false statements and perjured declaration in Xiaohua Huang v. Huawei No. 2:15-cv-1413 made erroneous judgment as reference example to attack me, make false statements and perjured declaration and successfully induced the trial court, as Serpent seduces Eve, to imitate case No. 2:15-cv-1413 rather than focus on the case objectively to make erroneous decision.

In California, before obtaining the permission, the lawyer solemnly declares to GOD by putting his/her hand on bible "I solemnly swear (or affirm) that I will support the Constitution of the United States and the Constitution of the State of California, and that I will faithfully discharge the duties of an attorney and counselor at law to the best of my knowledge and ability. As an officer of the court, I will strive to conduct myself at all times with dignity, courtesy, and integrity." Defendant Counsels such as PAUL J. ANDRE, LISA KOBIALKA, VIOLAINE have violated their oath to the God, their conducts may have himself, his family and descendants suffer from their wrongful conduct. State Bar Announces 23 Attorneys Disbarred in the Third Quarter of 2023 BY STATE BAR OF CALIFORNIA ON NOVEMBER 20, 2023. Some attorneys even violates the law and commits crime for their personal interests , do not mention their conduct would have their descendants suffered such as the Jewish have suffered from King David's wrongful conduct.

Defendant's conduct of instructing their Counsels to lie, cheat and use false information to defame Plaintiff and cheat the Court are somehow consistent with their other unethical illegal conducts. Defendant and their founders filed many

patents to protect their own technology, but use Plaintiff's patent without licensing and refuse to license. Defendants have used Plaintiff's patent to have built several sample chips, measured some data from those chips and published the papers on science journal then obtained tens of million USD income (investment). TetraMem has offered to sell its ReRAM designed with using US patent RE45259 without licensing.

8. The new complaint is also based on the new further information provided by Defendant in Exhibit 3 that defendant's product is design with "1R 1T" structure which is consistent with the structure of Figure 2 and Figure 3 in plaintiff's analysis on how defendant's products infringe the us patent RE45259 Exhibit X1. Defendant sent Plaintiff an email in Exhibit 3 and tried to narrow the claim 29 of US patent RE45259 to one embodiment, but in the same time disclosed that defendant's product use " 1T,1R" ReRAM structure. Plaintiff pointed out defendant counsel's cheating conduct with Exhibit 4 and Exhibit 5.

THE INFRINGING PRODUCTS WHICH DEFENDNAT MAY HAVE MADE AND SOLD

9. Based on the information obtained that the products made by Defendant TetraMem, including but not limited to embedded RRAM IP and chips contains the IC with the function which read the claim 29 of US patent RE45259. In page 5 and page 6 of Exhibit 1 the joint news release made by TetraMem and Synopsys offer to

sell the RRAM of TetraMem with its marketing name " TetraMem's analog-RRAMbased in-memory computing technology .....". Exhibit 3 also claims that TetraMem's ReRAM employ "1T, 1R" (1 transistor, t resistor) which is consistent with Figure 2 and Figure 3 in Exhibit X1. How and why the RRAM of TetraMem have infringed claim 29 of US patent RE45259 is analyzed and explained in Exhibit X1.

COUNT I: INFRINGEMENT OF U.S. PATENT NO. RE45259

10. Plaintiff refers to and incorporates herein the allegations of Paragraphs 1-9 above.

11. On November 25, 2014 U.S. Patent No. RE45259 (the"'RE259Patent") was duly and legally issued for a "Hit ahead hierarchical scalable priority encoding logic and circuits." A true and correct copy of the 'RE259patent is attached hereto as Exhibit A. Xiaohua Huang as inventor is the owner of all rights, title, and interest in and to the 'RE259 patent.

12. On information and belief, Defendant TetraMem has infringed and continue to infringe directly, indirectly, literally, on Doctrine of Equivalent one or more of the claims of the'RE259 patent through making ReMAM IP and chips containing circuit and logic function which have infringed at least claim 29 of the 'RE259 patent under 35 U.S.C. § 271(a), (b) and(c).

13. On information and belief, Defendant TetraMem has induced its Customers to have infringed and continue to infringe directly, indirectly, literally, on Doctrine of Equivalent the claim 29 of the 'RE259 patent by using the devices which infringes the claim 29 of 'RE259 patent. The Customers of the Defendant uses the function of reading and writing data and information which contains the circuit and logic reading the claim 29 of the 'RE259 patent. The using of function of reading and writing data and information in accused devices are completely not a staple article or commodity of commerce suitable for substantial non-infringing use.

14. Defendant TetraMem's acts of infringement, inducing infringement have caused damage to Xiaohua Huang, and Xiaohua Huang is entitled to recover from Defendant for the damages sustained by Xiaohua Huang as a result of Defendant's wrongful acts in an amount subject to proof at trial. Defendant's infringement of Xiaohua Huang exclusive rights under the 'RE259 patent will continue to damage Xiaohua Huang, causing irreparable harm for which there is no adequate remedy at law, unless enjoined by this Court. Defendant's infringement entitle Xiaohua Huang to recover damages under 35 U.S.C.§284 and to legal fees and costs incurred in prosecuting this action under35 U.S.C. § 285.

JURY DEMAND

15. Pursuant to Fed. R. Civ. P. 38(b), Plaintiff Xiaohua Huang requests a trial by jury on all issues.

PRAYER FOR RELIEF

WHEREFORE, Xiaohua Huang prays for the following relief:

(a). A judgment in favor of Xiaohua Huang that Defendant TetraMem has infringed and is infringing U.S. Patent No RE45259;

(b). A judgment that the 'RE259 patent is valid and enforceable;

(c). An order preliminarily and permanently enjoining Defendant and its subsidiaries, parents, officers, directors, agents, servants, employees, affiliates, attorneys and all others in active concert or participation with any of the foregoing, from further acts of infringement of the 'RE259;

(d). An accounting for damages resulting from Defendant's infringement of the 'RE259 patent under 35 U.S.C. § 284;

(e). An assessment of interest on damages;

(f). A judgment awarding damages to Xiaohua Huang for its costs, disbursements, expert witness fees, and legal fees and costs incurred in prosecuting this action, with interest pursuant to 35 U.S.C. § 285 and as otherwise provided by law;

(g). Such other and further relief as this Court may deem just and equitable.

Dated: October 5, 2024

Respectfully submitted

Xiaohua Huang

P.O. Box 1639, Los Gatos CA95031

Tel: 669 273 5633

Email: paul-huang03@outlook.com

Exhibit A US patent RE45259

Exhibit 1 pre-suit background

Exhibit 2 Tetramem-paper

Exhibit 3 TetraMem letter

Exhibit 4 Response to TetraMem letter

Exhibit 5 TetraMem claimed info

Exhibit X1 Pre-suit analysis

Exhibit A

US patent RE45259

US00RE45259E

(19) **United States**

(12) **Reissued Patent**
Huang

(10) **Patent Number: US RE45,259 E**

(45) **Date of Reissued Patent: Nov. 25, 2014**

(54) **HIT AHEAD HIERARCHICAL SCALABLE PRIORITY ENCODING LOGIC AND CIRCUITS**

(76) Inventor: **Xiaohua Huang**, San Jose, CA (US)

(21) Appl. No.: **13/355,449**

(22) Filed: **Jan. 20, 2012**

Related U.S. Patent Documents

Reissue of:

(64) Patent No.: **7,652,903**
Issued: **Jan. 26, 2010**
Appl. No.: **11/073,116**
Filed: **Mar. 4, 2005**

U.S. Applications:

(60) Provisional application No. 60/550,537, filed on Mar. 4, 2004.

(51) **Int. Cl.**
G11C 15/00 (2006.01)

(52) **U.S. Cl.**
USPC **365/49.1**; 365/230.01; 365/230.05; 365/230.06

(58) **Field of Classification Search**
USPC 365/49.1, 230.01, 230.05, 230.06
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,249,449	B1 *	6/2001	Yoneda et al.	365/49.18
6,307,767	B1 *	10/2001	Fuh	365/49.17
6,392,910	B1 *	5/2002	Podaima et al.	365/49
6,505,271	B1 *	1/2003	Lien et al.	711/108
7,043,601	B2 *	5/2006	McKenzie et al.	711/108
7,464,217	B2 *	12/2008	Braceras et al.	711/108

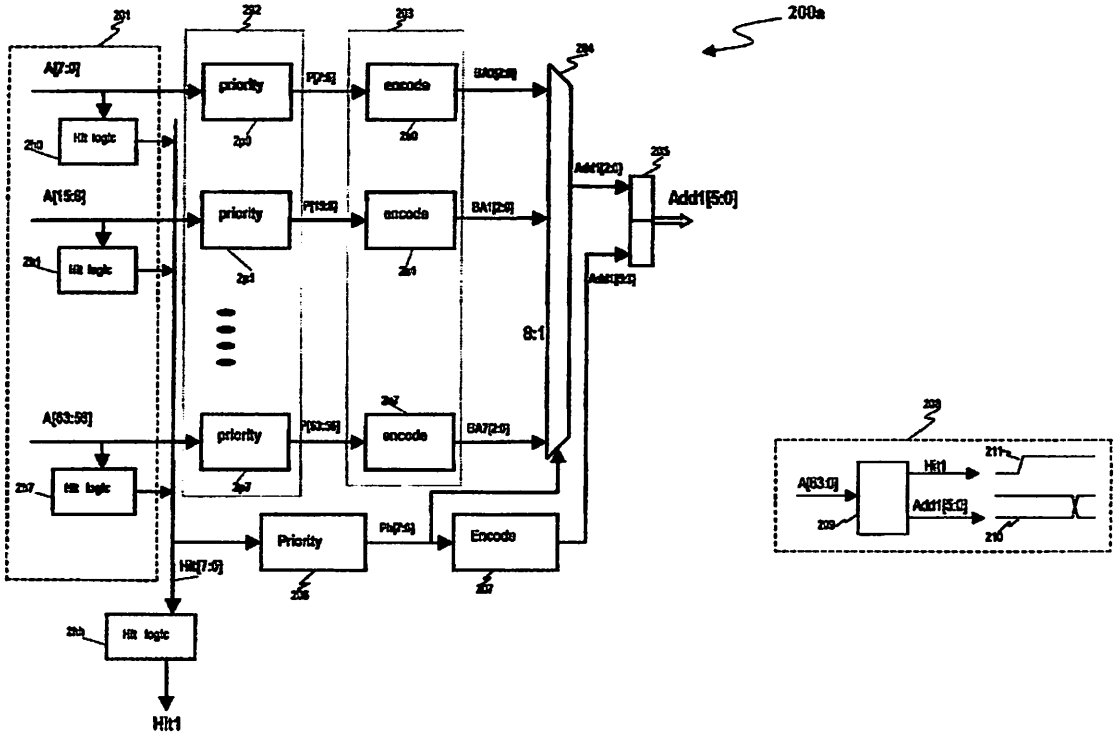
* cited by examiner

Primary Examiner — Han Yang

(57) **ABSTRACT**

In this invention a hit ahead multi-level hierarchical scalable priority encoding logic and circuits are disclosed. The advantage of hierarchical priority encoding is to improve the speed and simplify the circuit implementation and make circuit design flexible and scalable. To reduce the time of waiting for previous level priority encoding result, hit signal is generated first in each level to participate next level priority encoding, and it is called Hit Ahead Priority Encoding (HAPE) encoding. The hierarchical priority encoding can be applied to the scalable architecture among the different sub-blocks and can also be applied with in one sub-block. The priority encoding and hit are processed completely parallel without correlation, and the priority encoding, hit generation, address encoding and MUX selection of the address to next level all share same structure of circuits.

36 Claims, 8 Drawing Sheets



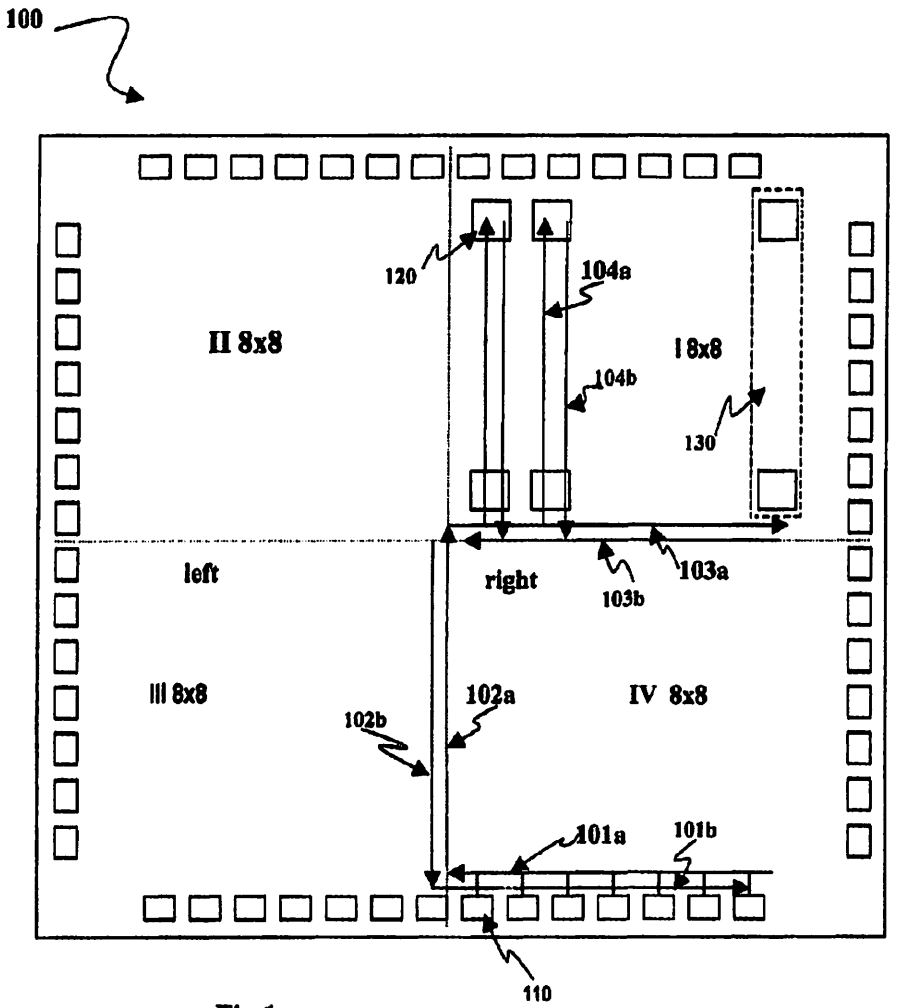

100
120
104a
104b
130
II 8x8
I 8x8
left
right
103a
103b
III 8x8
102a
IV 8x8
102b
101a
101b
110

Fig.1

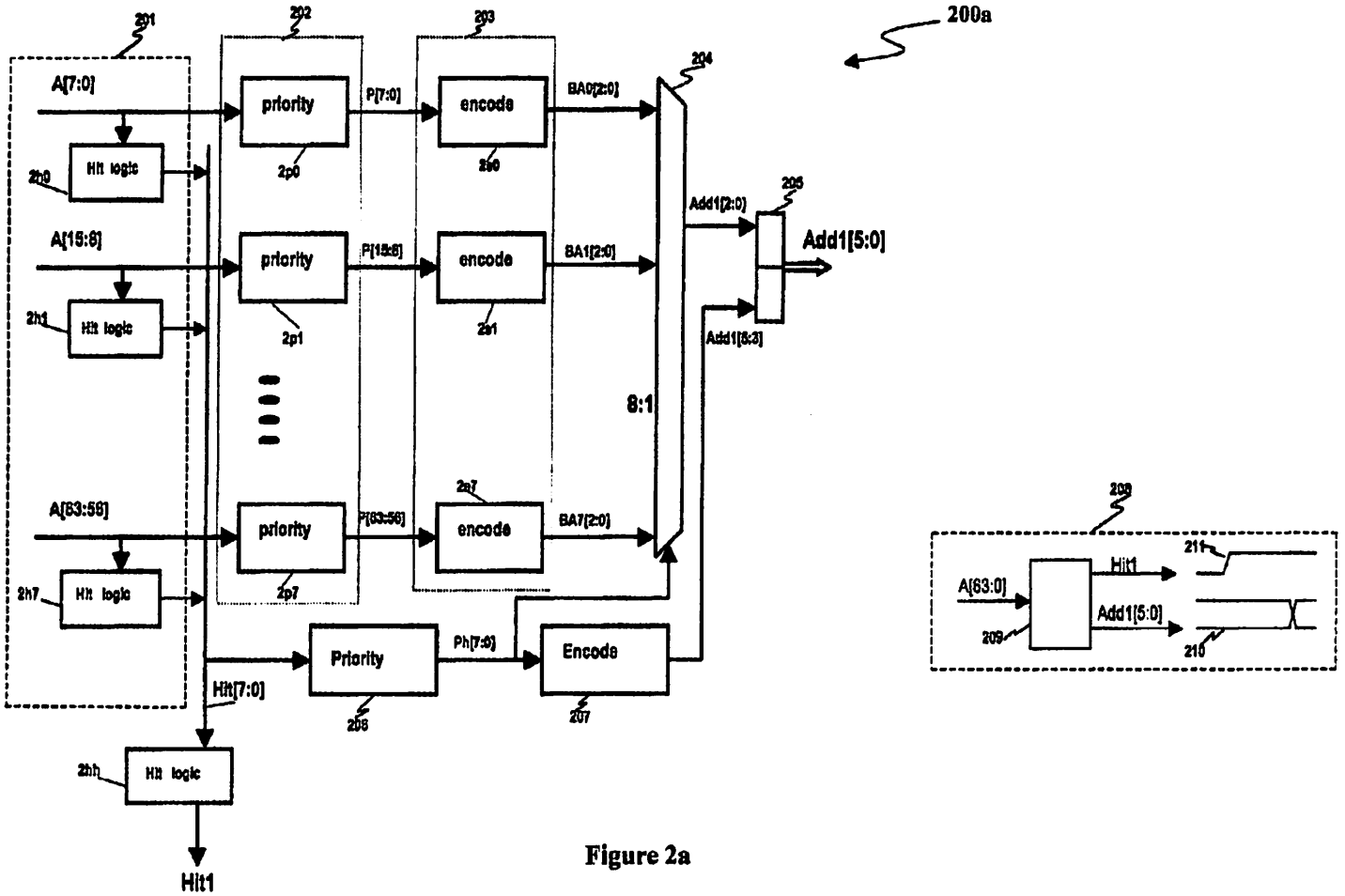
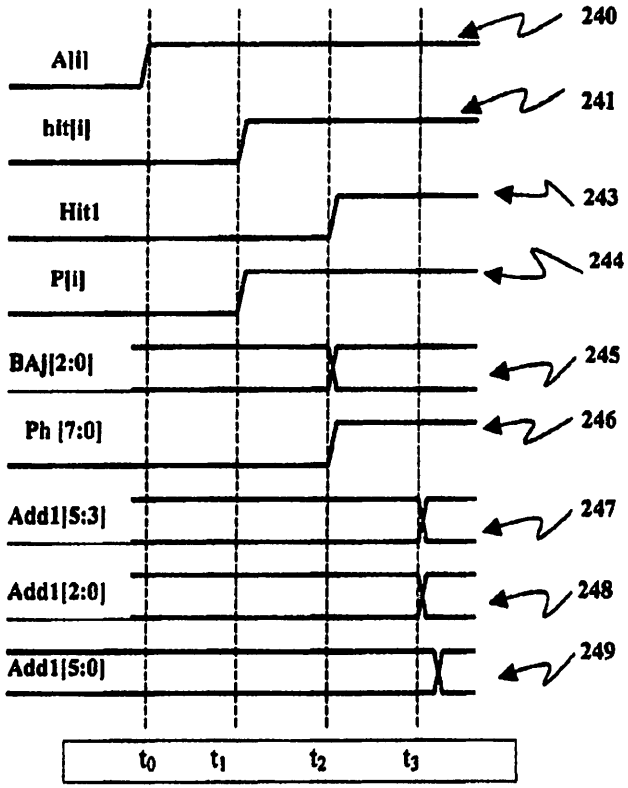
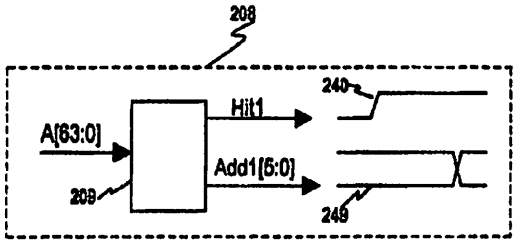

200a
201
202
203
204
205
A[7:0]
priority
P[7:0]
encode
BA0[2:0]
Hit logic
2p0
2e0
2h0
A[15:8]
priority
P[15:8]
encode
BA1[2:0]
Add1[2:0]
Add1[5:0]
Hit logic
2p1
2e1
2h1
Add1[3:3]
8:1
2e7
A[63:56]
priority
P[63:56]
encode
BA7[2:0]
Hit logic
2p7
2h7
Priority
Ph[7:0]
Encode
Hit[7:0]
206
207
2h
Hit logic
Hit1
208
A[63:0]
Hit1
211
Add1[5:0]
209
210

Figure 2a





200b

Figure 2b.

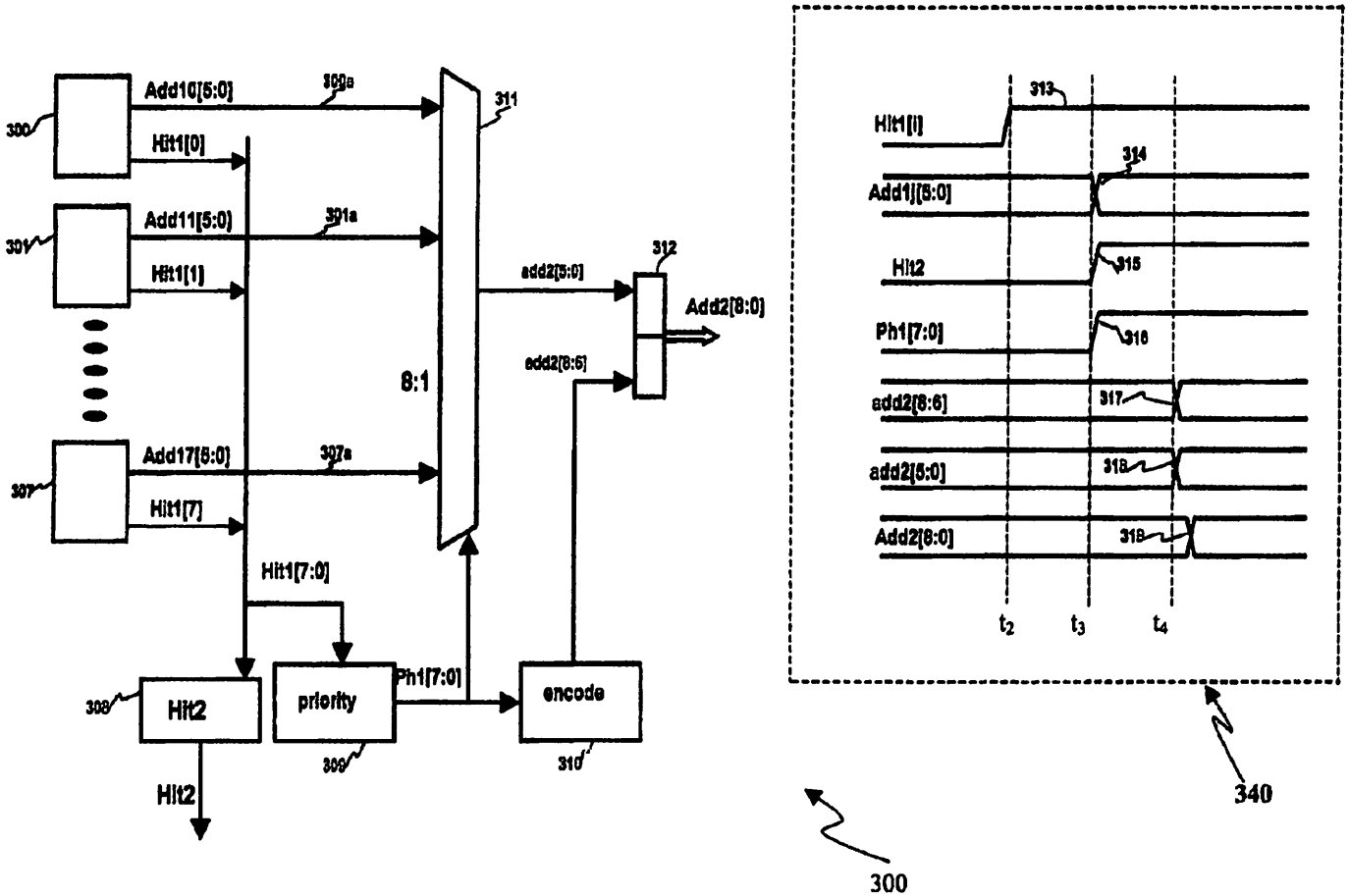

Add10[5:0]
Hit1[0]
Add11[5:0]
Hit1[1]
Add17[5:0]
Hit1[7]
Hit1[7:0]
8:1
add2[5:0]
add2[8:6]
Add2[8:0]
Hit2
priority
Ph1[7:0]
encode
Hit2
Hit1[i]
Add1j[5:0]
Hit2
Ph1[7:0]
add2[8:6]
add2[5:0]
Add2[8:0]
t2
t3
t4
340
300

Figure 3

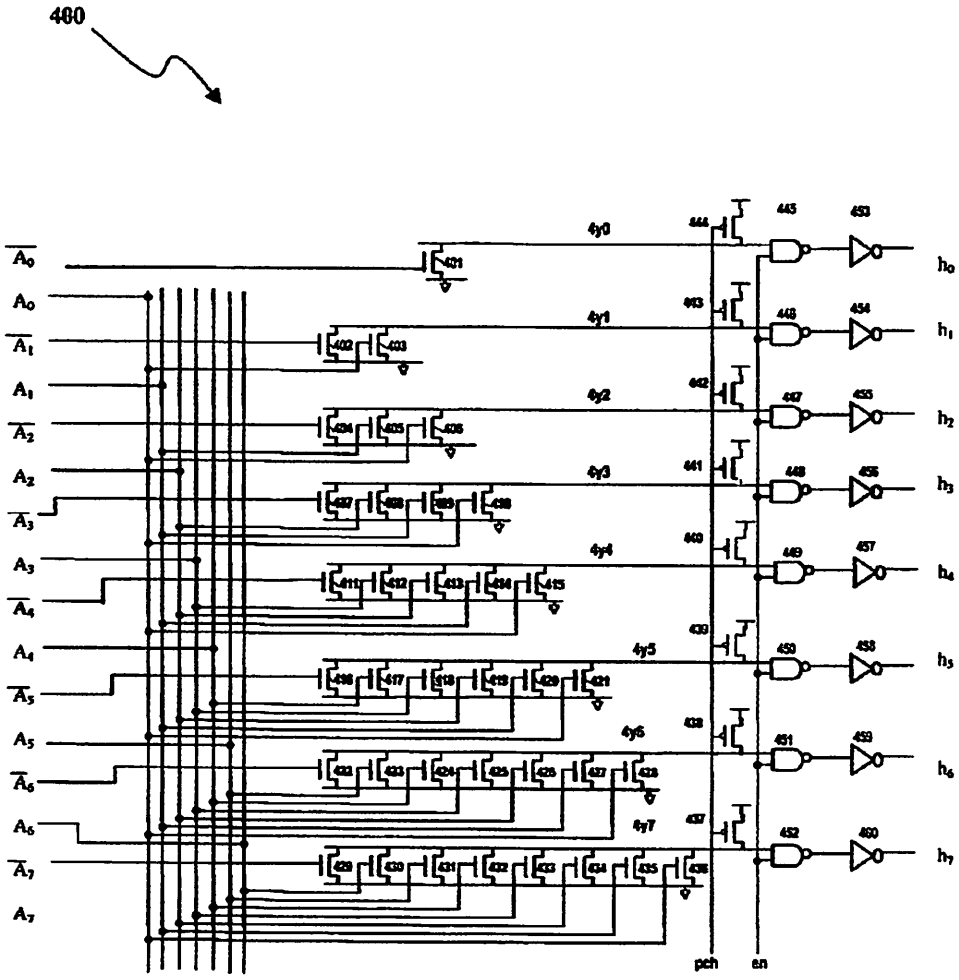
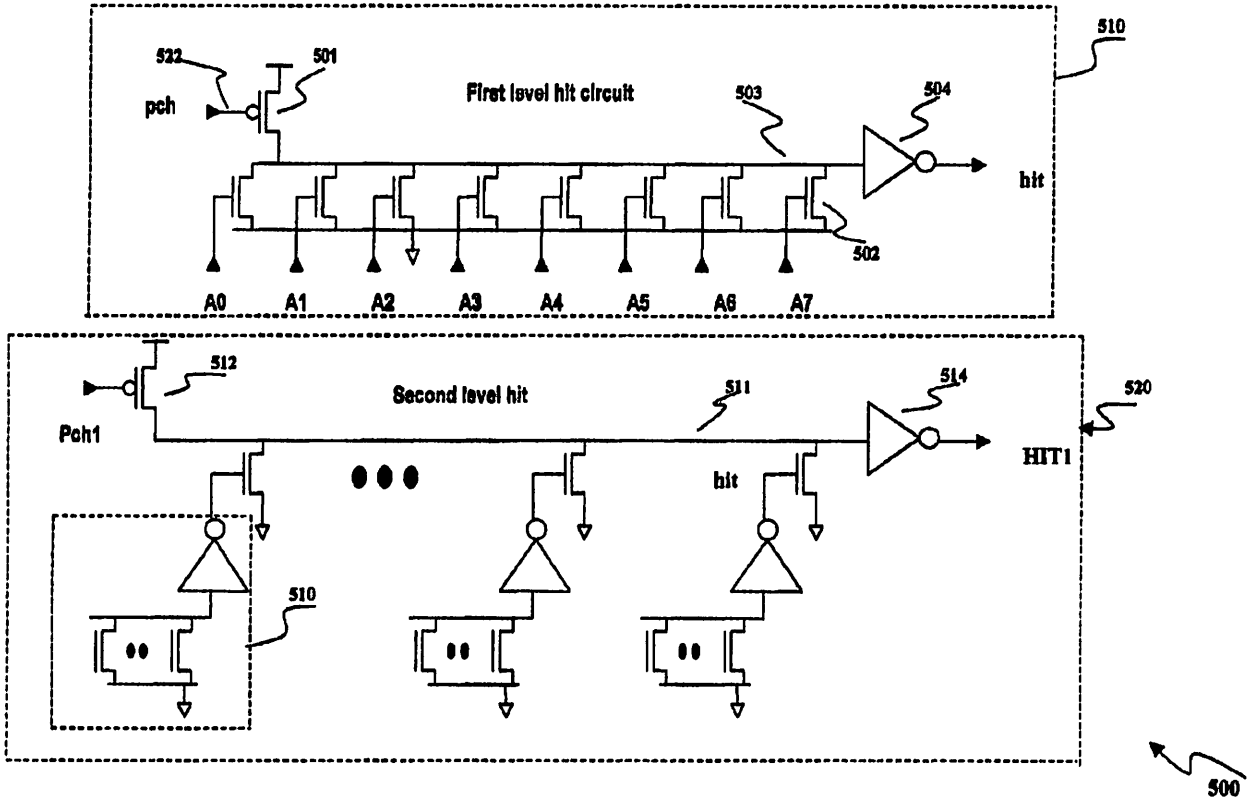


Figure 4



$$Hit=(A0.A1.A2...A7)(A8.A9...)(A16.A17.\ ).....=\overline{\overline{(\overline{A0}+\overline{A1}+....)}+\overline{(\overline{A8}+\overline{A9}+...)}+\overline{(\overline{A16}+\overline{A17}+....)}+.....}$$

Figure 5

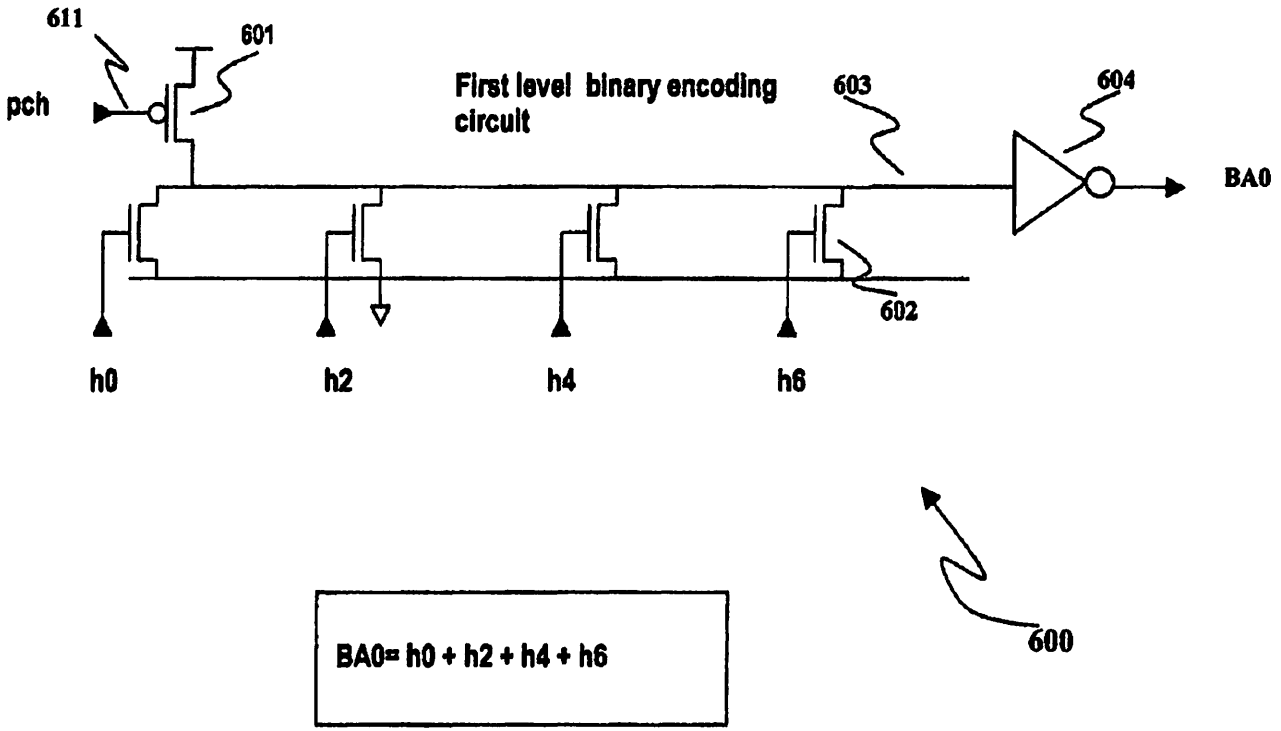

611
601
pch
First level binary encoding circuit
603
604
BA0
602
h0
h2
h4
h6
BA0= h0 + h2 + h4 + h6
600

Figure 6

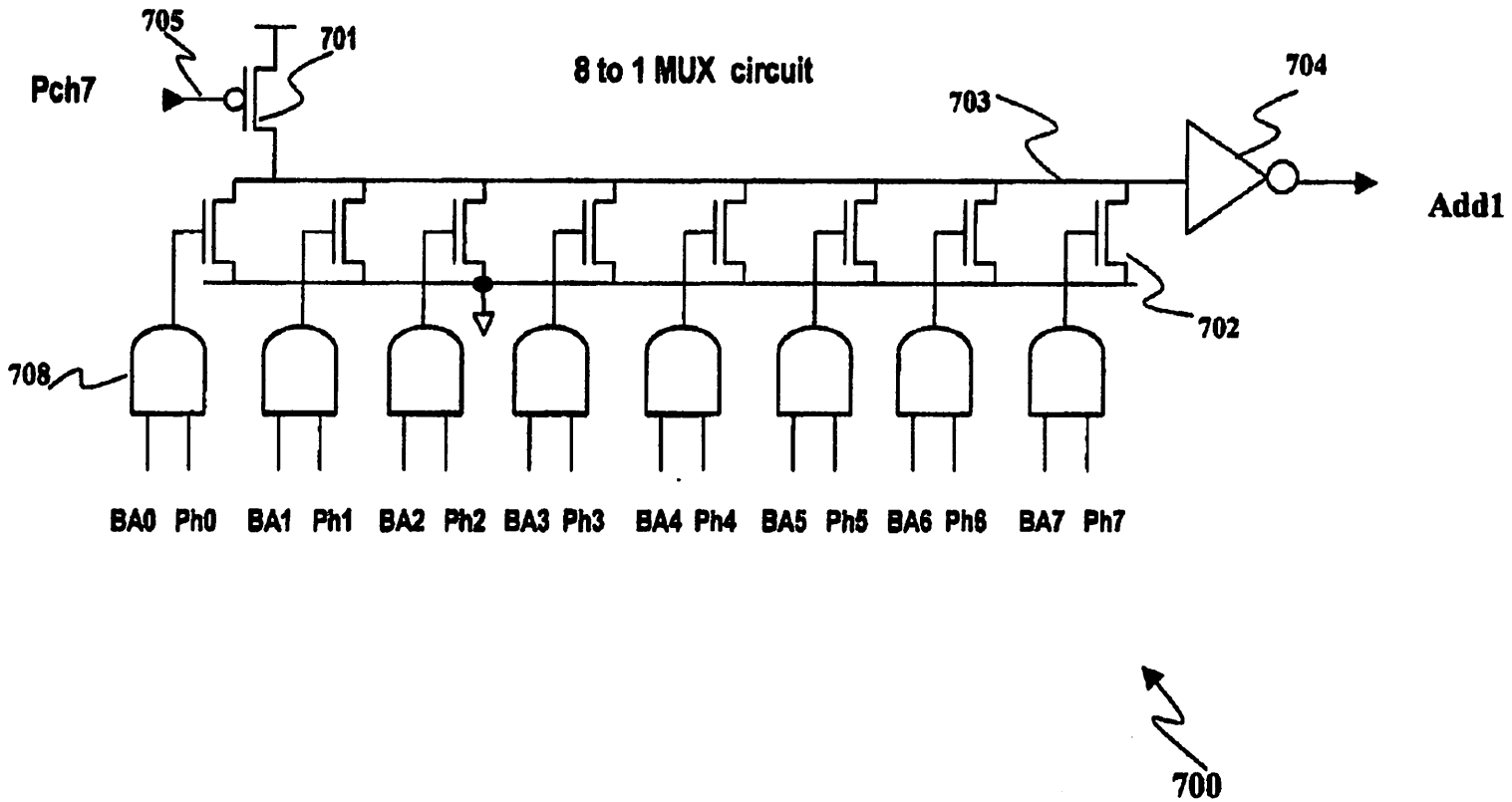

705
701
Pch7
8 to 1 MUX circuit
703
704
Add1
702
708
BA0 Ph0 BA1 Ph1 BA2 Ph2 BA3 Ph3 BA4 Ph4 BA5 Ph5 BA6 Ph6 BA7 Ph7
700

Figure 7

HIT AHEAD HIERARCHICAL SCALABLE PRIORITY ENCODING LOGIC AND CIRCUITS

Matter enclosed in heavy brackets [ ] appears in the original patent but forms no part of this reissue specification; matter printed in italics indicates the additions made by reissue.

This application claims the benefit of provisional U.S. Application Ser. No. 60/550,537, entitled "Priority encoding logic and Circuits," filed Mar. 4, 2004, which is incorporated herein by reference in its entirety for all purposes.

FIELD OF THE INVENTION

The presentation relates to content addressable memory. In particular, the present invention relates to logic and circuits of priority encoding of match or hit address.

BACKGROUND OF THE INVENTION

In ternary content addressable memory, not every bit in each row are compared in the searching or comparing process, so some time in one comparison, there are more than one row matching the input content, it is called multi-hit or match. In multi-hit case, one protocol was made to select the highest priority address. The logic of selecting the highest priority address is called priority encoding.

Assume we have $\{A_0, A_1, \ldots A_{n-1}\}$ hit signals from the corresponding addresses and define A_0 has the highest priority and A_n has the lowest priority. Assume some of $\{A_0, A_1, \ldots A_{n-1}, A_n\}$ are logic "1" and all of the others are logic "0", the priority encoding keep the highest priority "1" as "1" and convert all the other "1" into "0". The logic operation of this transform:

$$\{A_0, A_1, \ldots A_{n-1}, A_n\} \Rightarrow \{h_0, h_1, \ldots h_{n-1}, h_n\} \quad (1)$$

can logically be expressed as:

$$h_0 = A_0 \quad (2)$$

$$h_1 = \overline{A}_0 * A_1$$

$$h_2 = \overline{A}_0 * \overline{A}_1 * A_2$$

$$\ldots$$

$$h_n = \overline{A}_0 * \overline{A}_1 * \overline{A}_2 \ldots A_{n-1} * \overline{A}_n$$

Which means only when A_0 to A_{i-1}, are all zero, $h_i = A_i$, otherwise no matter $A_i = 0$ or 1, $h_i = 0$.

After the priority encoding, the hit address with the highest priority will be encoded to the binary address.

If the entry N are large, say 1K to 128K or even 1M, the calculation of priority logic (2) will take long time if we use serial logic. So we come out the inventions which will be described in the following.

SUMMERY OF THE INVENTION

In this invention, we propose a multi-level hierarchical scalable priority encoding. For example we make 8 entry as one group as first level and 8 first level as a second level, total 64 entry. Then we can make 8 second level as third level, total 512 entry, and so on. The advantage to make hierarchical priority encoding is to improve the speed, and simplify the circuit implementation and make circuit design flexible and scalable.

To reduce the time of waiting for previous level priority encoding result, we generate the hit signal first in each level to participate next level priority encoding, and we call it Hit Ahead Priority Encoding (HAPE) encoding.

The hierarchical priority encoding can be applied to the scalable architecture among the different sub-blocks and can also be applied with in one sub-block.

BRIEF DESCRIPTION OF THE DRAWINGS

Preferred embodiments of the invention will now be described, by way of example only, with reference to the attached Figures, wherein:

FIG. **1** is a block diagram of scalable architecture of CAM with many sub-block in accordance with one embodiment of the present invention.

FIG. **2a** is a logic block diagram of hierarchical priority encoding and match address binary encoding within one sub-block in accordance with one embodiment of present invention.

FIG. **2b** is the and timing diagram in accordance with FIG. **2b** of present invention.

FIG. **3** is a logic block diagram of hierarchical priority encoding and match address binary encoding in higher level or among the different sub-block and timing diagram in accordance with one embodiment of present invention.

FIG. **4** is the circuit implementation of priority encoding with **8** input address in accordance with one embodiment of present invention.

FIG. **5** is the circuit implementation of the HIT generation logic address in accordance with one embodiment of present invention.

FIG. **6** is the circuit implementation of binary encoding logic in accordance with one embodiment of present invention.

FIG. **7** is the circuit implementation of 8 to 1 mux in accordance with one embodiment of present invention.

DETAILED DESCRIPTION OF THE INVENTIONS

To make the priority encoding logic calculation quicker, the entire CAM block can be divided into 256 block and divided into four quadruple, each quadruple has 8×8=64 block and each block has 8×8=64 entry as shown in FIG. **1** with embodiment **100**.

This is just to explain the principle, the entry number of each sub-block and the number of sub-block can be different. Assume the data pad **110** are equally distributed in four side of the chip. If all of the data pad **110** are in one side or less than four side, the principle is same.

First step, route all the data signal in each side (only one side are drawn in the FIG. **1**) to the middle point of that side, which is shown as route **101a** in FIG. **1**. Second step, route all the data signal to the center of the chip shown as route **102a** in FIG. **1**. Third step, in the center point send the data to be compared to both left and right side (only right side path **103a** is shown in FIG. **1**. Fourth step, send data to each one of the 8 column both upper part and down part shown as **104a** in FIG. **1**. Fifth step, the data to be compared are then sent to each sub-block **120** in each column to perform the compari-

son with each entry in every sub-block **120**. In embedded application, the entry number of TCAM is not very large. In that case, the data path start from path **104a**. If only some selected sub-block are searched or compared, the data to be compared will only be sent into those sub-block to save power consumption. After comparison with each entry inside each sub-block **120**, the first level and second level priority encoding and binary encoding are performed which will be explained in details in FIG. **2**, then the priority encoding in each column **130** among 8 sub-block will be performed as third level priority encoding and the hit address are sent out through path **104b**. Next step fourth level priority encoding will be performed among 8 column **130** in each quadruple and the Hit address are sent out through path **103b**. Next step the priority encoding will be performed in the center of chip among four quadruple and the hit address will be sent through path **102b**. Last step the hit address are sent to the output pad **110** through path **101b**. The priority encoding among upper quadruple and lower part quadruple can be performed together in path **103b**.

The priority encoding logic calculation block diagram for each 8×8=64 entry sub-block **120** are shown in FIG. 2a with embodiment **200a**. Each 8 entry of 64 entry are grouped together to do hit logic function from **2h0** to **2h7** and generate Hit[0] to Hit[7] in block **201**. In the same time each 8 entry of 64 entry are performed priority encoding logic calculation in each block from **2p0** to **2p7** of embodiment block **202** to generate P[63:0], then proceed binary encoding from **2e0** to **2e7** in embodiment block **203** to generate any three bit BA0 [2:0] to BA7[2:0] binary address if there is a hit in any 8 bit group. The eight signal of Hit[0] to Hit[7] from block **201** will perform priority encoding in block **206** which is logically exact same as the priority encoding in each 8 entry group from **2p0** to **2p7**. The Priority Hit Ph[7:0] from Hit[0] to Hit[7] will select the 8 to 1 mux **204** and select one three bit binary address from BA0[2:0] to BA7[2:0] and become Add1[2:0]. The priority bit of Hit[0] to Hit[7] is binary encoded in block **207** which is logically same as the binary encoding block from **2e0** to **2e7** to generate the address: Add1[5:3]. Add1[5:3] and Add1[2:0] make Add1[5:0]. Hit[0] to Hit[7] further perform the logic function in block **2hh** which is logically same as any block **2h0** to **2h7** and generate the next level Hit1. Both Add1[5:0] and Hit1 will be passed to the next level.

The timing diagram of embodiment **200a** is shown in FIG. **2b** with embodiment **200b**. Assume all the Hit or miss signal from TCAM comparison A[i] (A[63:0])which is drawn as signal **240** are available in time t_0, the first level hit signal Hit[7:0] generated by block **2h0** to **2h7** are drawn as **241** which is available at time t_1. In the same time A[63:0] are divided into eight group and priority encoded by block **2p0** to **2p7**, generating P[0] to P[63] which are drawn as **244** and available at time t_1. The time delay of generating Ph[7:0] which are drawn as **246** and the time delay of generating.

BA0[2:0] to BA7[2:0] which are drawn as **245** are roughly same and they are generated in time t_2. So the Binary address Add1[2:0] which are drawn as **248** are selected by Ph[7:0] from the 8 group address BA0[2:0] to BA7[2:0] through an eight to one MUX **204** without any further delay except the delay of MUX itself which is (t_3-t_2), and the address Add1 [5:3] which are drawn as **247**, Add1[2:0] and Add[5:0] which are drawn as **249** are available at time t_3.

So the total delay from A[63:0] available to the output of binary hit address Add1[5:0] is about three stage delay(priority **2p0**, binary encoding **2e0** and 8 to 1 MUX **204**), where we call each block(**2p0**, **2e0** and **204** etc) as one stage. The delay of Hit1 **243** is two stage delay. So the output of Hit1 which is available at t_2 which is one stage earlier than the output of binary Hit address Add1[5:0] **249** which is available at t_3. Only Hit1 and Add1[5:0] are sent to the next level priority encoding. The entire sub-block are abstracted as symbol **208**. The timing delay of hit, priority encoding, binary encoding and 8 to 1 mux will be analyzed in details.

FIG. **3** is the logic block diagram of priority encoding of higher level among the eight group of 64 entry sub-block or among the 8 sub-block in every column **130** in FIG. **1**. The Hit signal Hit1[7:0] which is marked as **313** in FIG. **3** are one stage earlier than the binary hit address Add10[5:0] to Add17 [5:0] which are marked as **314**. Eight bit HIT signal of Hit1 [7:0] perform priority encoding in block **309**, then the priority hit signal Ph1[7:0] will select Add2[5:0] from the eight input MUX **311**.

In the same time Ph1[7:0] are encoded into binary address Add2[8:6] in block **310**. Add2[8:6] and Add2[5:0] make Add2[8:0]. In block **308** eight input Hit1[7:0] generate Hit2 at time t_3 which is one stage earlier than Binary hit address Add2[8:0]. From the timing diagram **340** in FIG. **3**, the delay of binary hit address Add1i[5:0] which is signal **314** to Add2 [8:0] which is marked as **319** is an 8 to 1 MUX delay which is (t_4-t_3), where i=0 to 7. In this hierarchical priority design, the delay on each level is an 8 to 1 MUX delay because the selection signal from the priority encoding among the hit signals is available one stage earlier and there is no extra delay to wait for the selection signal.

Another advantage of this hierarchical priority encoding is that the simplicity of circuit design. We already see that each level shares the same logic and circuit design. Say, the priority function block **206**, **309** in each level are same in logic and circuit, which is shown in FIG. **4**, embodiment **400**.

Embodiment **400** in FIG. **4** is a sample implementation of the priority logic equation (2) which can be deduced to equation (3), where n=7.

$$h_0 = A_0 \tag{3}$$

$$h_1 = \overline{A_0} * A_1 = \overline{A_0 + \overline{A_1}}$$

$$h_2 = \overline{A_0} * \overline{A_1} * A_2 = \overline{A_0 + A_1 + \overline{A_2}}$$

...

$$h_n = \overline{A_0} * \overline{A_1} * \overline{A_2} \ldots \overline{A_{n-1}} * A_n = \overline{A_0 + A_1 \ldots + A_{n-1} + \overline{A_n}}$$

The equation (3) is implemented as embodiment **400** in FIG. **4**. Each line from **4y0** to **4y7** connect the drains of a few N transistors and each line **4y0** to **4y7** is the output of dynamic NOR logic of N transistor connected to that line. At the beginning of each cycle, the gate input signals $\overline{A_0}$ to $\overline{A_7}$ and A_0 to A_7 of all the N transistor from **401** to **436** are set to logic zero which turn off all the N transistors and the enable signal en is set to logic zero which makes all the output of NAND gate **445** to **452** to logic one and then turn all the output of inverter **453** to **460** into logic zero. The input pch of the P transistors **437** to **444** are set to logic zero and the P transistor **437** to **444** are turned on, which make the line **4y0** to **4y7** connecting to Vdd with low impedance and pre-charge the potential level of line **4y0** to **4y7** up to Vdd, then the signal pch is turned into Vdd and turn off the P transistors **437** to **444** before the TCAM comparison results A_0 to A_7 and their complementary $\overline{A_0}$ and $\overline{A_7}$ arrive. The Hit signal among A0 to A7 will be logical "one" at potential Vdd and the missed signal among A0 to A7 will be logical zero at potential ground. Only the highest priority hit, the output of the NOR

gates are logically high. For example, A_0=0, A_1=0, A_2=Vdd and A_3=Vdd, the highest priority hit is A_2. The input of N transistor **401** is Vdd and N transistor **401** is turned on and the node **4y0** is discharged to ground. The input of transistor **402** which is the complementary of A_1 is also Vdd and the transistor **402** is ON, the node of **4y1** is also discharged to ground.

Since A_0=0, A_1=0, A_2=Vdd, $\overline{A_2}$=0, so the inputs of transistors **404**, **405**, **406** are all zero and the transistor **404**, **405**, **406** are all OFF and the node **4y2** will not be discharged and will be kept logically "one" at potential Vdd. Since A_2=Vdd, the inputs of transistors **408**, **413**, **419**, **426** and **434** will be Vdd and all the node **4y3**, **4y4**, **4y5**, **4y6** and **4y7** will be pulled to ground no matter if A_3, A_4, A_5, A_6 and A_7 are logically one or zero. The slowest path or worst case is only one input among eight N transistor **429**, **430**, **431**, **432**, **433**, **434**, **435** and **436** connected to node **4y7** is Vdd and all the others are zero, in that case one transistor need to discharge the drain parasitic capacitance of eight transistor and the metal wire capacitance connected to node **4y7**. The signal en is characterized to turned to Vdd later then node **4y7** is discharged in worst case. The worst case delay of eight input priority encoding is that one N transistor discharging the drain parasitic capacitance of eight same size N transistor down to ground plus the delay of one NAND gate and one inverter.

The logic of Hit function block **2h0**, **2h1**, . . . **2hh** and **308** in each level is also same and its logic and circuit are shown in FIG. **5**. The embodiment **510** is the circuit implementation of one block **2h0** and the embodiment **520** is the circuit implementation of both block **201** and block **2hh** in FIG. **2a** together. The operation principle of **510** is: 1) all the input A0 to A7 are set to zero as in embodiment **400** in FIG. **4**. 2) Set the gate input **522** of P transistor **501** to zero to pre-charge the node **503** to Vdd, then turn **522** to Vdd and turn off the P transistor **501** before the signal A0 to A7 arrive. If all the input A0 to A7 are zero, the input of N transistors are zero and all the N transistors **502** are OFF and the node **503** is kept in Vdd and the output signal of inverter **504** is zero. If only one input among A0 to A7 is Vdd and all the others are zero, which is the worst case, the delay of **510** is that one N transistor discharge the drain parasitic capacitance of the eight same size N transistor down to ground plus the delay of one inverter.

The binary encoding logic and circuit is shown as embodiment **600** in FIG. **6**. The operation principle of **600** is: 1) all the input h_0, h_2, h_4 and h_6 are set zero. 2) Set the gate input **611** of P transistor **601** to zero to pre-charge the node **603** to Vdd, then turn **611** to Vdd and turn off the P transistor **601** before the signal h_0, h_2, h_4 and h_6 arrive. If all the input signal h_0, h_2, h_4 and h_6 are zero, the input of N transistors are zero and all the N transistors **602** are OFF and the node **603** is kept in Vdd and the output signal of inverter **604** is zero. If only one input among h_0, h_2, h_4 and h_6 is Vdd and all the others are zero, which is the worst case, the delay of **600** is that one N transistor discharging the drain parasitic capacitance of the four same size N transistor down to ground plus the delay of one inverter.

The MUX logic and circuit is shown in FIG. **7** as embodiment **700**. The operation principle of **700** is: 1) the input signal Ph_0, Ph_1, Ph_2, Ph_3, Ph_4, Ph_5, Ph_6 and Ph_7 are set zero. 2) Set the gate input **705** of P transistor **701** to zero to pre-charge the node **703** to Vdd, then turn **705** to Vdd and turn off the P transistor **701** before the signal Ph_0, Ph_1, Ph_2, Ph_3, Ph_4, Ph_5, Ph_6 and P_7 arrive. Since Ph_0, Ph_1, Ph_2, Ph_3, Ph_4, Ph_5, Ph_6 and Ph_7 are from Priority encoding, only one signal among them is Vdd and all the other are zero if there is hit. After AND logic, only one output of the seven AND gate **708** is equal to the input value which is the selected bit from Ba_0 to Ba_7. If the selected bit from Ba_0 to Ba_7 is zero, the node **703** is kept Vdd and the output of inverter **704** is zero and the selected bit value zero is passed out. If the selected bit from Ba_0 to Ba_7 is Vdd, one N transistor among eight N transistor **702** is turned ON and the node **703** is discharged down to ground and the output of inverter **704** is Vdd(logical one) and the selected bit value Vdd is passed out, which is the worst case, the delay of **700** is one N transistor discharging the drain parasitic capacitance of the eight same size N transistor down to ground plus the delay of one inverter and one AND gate. Usually one AND gate includes one inverter and one NAND gate, so the delay of **700** is one N transistor discharging the drain parasitic capacitance of the eight same size N transistor down to the ground plus the delay of two inverter and one NAND gate.

The entire Priority encoding logic and circuit are simplified as a four basic building block of **400**, **510**, **600** and **700** in FIGS. **4**, **5**, **6** and **7**. The delay of each block **400**, **510**, **600** and **700** are comparable and we call the time of delay of each block **400**, **510**, **600** and **700** one stage. If we define the delay of hit logic block **510** as T_h, one inverter delay is T_i and one NAND gate delay is T_n. The delay of priority encoding block **400** is (T_h+T_n) since the delay of block **400** is one more NAND gate delay comparing with block **510**. The delay of block **600** is roughly T_h. The delay of MUX block **700** is $(T_h+T_n+T_i)$. The extra delay of each higher level priority encoding is a MUX **700** selection delay because that the Hit signal in each priority encoding level is generated one stage earlier than the binary hit address and the selection signal of the MUX is already available when the binary address to be selected arrive and will not suffer extra delay.

The previous description of the disclosed embodiments is provided to enable any person skilled in the art to make or use the present invention. Various modifications to these embodiments will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other embodiments without departing from the spirit or scope of the invention. Thus, the present invention is not intended to be limited to the embodiments shown herein but is to be accorded the widest scope consistent with the principles and novel features disclosed herein.

What is claimed is:

1. A content address able memory(CAM) and hit ahead priority encoding(HAPE) logic, comprising:

a group of blocks which is arranged in column and row, each block has equal number of CAM match signals which are the input signals of priority encoding logic, each block has same priority encoding logic of CAM match signals within the block, the CAM match signals or input signals are arranged from lower priority to higher priority or from higher priority to lower priority, each CAM match signals or input signal has either high logic level "one" which is called hit or low logic level "zero" which is called miss, each block generates block hit when there is at least one CAM match signal is high logic "one" within the block or block miss signal when all the CAM match signals are in low logic level "zero" within the block and block binary address signal corresponding to the CAM match signals of highest priority within the block, a priority encoding logic of block hit or miss signals of each column, each column generates a column hit signal when there is at least one block hit signal within the column or column miss signal when there is only block miss signals within the column and column binary address corresponding to the CAM match signals of highest priority within the column, a priority encoding logic of column hit or miss signals of a group column, a group of column generates a hit signal when there is at least one column hit signal within the

group column or a miss signal when there is only column miss signals within the group column and a group column binary address corresponding to the CAM match signals of highest priority within the group column.

2. A content address able memory(CAM) and hit ahead priority encoding(HAPE) logic of claim 1, further comprising:

a block multiplexer to select the binary address from the block of highest priority hit within the column as less significant portion of the column binary address; and

a priority encoding logic of block hit signals to generate the block multiplexer control signal which select the block of highest priority hit within the column, and a binary address encoding logic of block hit signals to generate the more significant portion of the column highest priority binary address.

3. A content address able memory(CAM) and hit ahead priority encoding(HAPE) logic of claim 1, wherein each block comprises:

a group of sub-blocks, each sub-block has equal number of input signals, each sub-block has priority encoding and binary address encoding logic to generate sub-block highest priority binary address as well as hit or miss generating logic to generate sub-block hit or miss signal, and the sub-block hit or miss signal is generated independently before sub-block binary address;

a block hit or miss generating logic to generate block hit or miss signal and block hit or miss signal is generated independently before the block binary address is generated;

a sub-block multiplexer to select the binary address from the highest priority sub-block within the block as less significant portion of block binary address; and

a priority encoding logic of each sub-block hit signals to generate the control signal of sub-block multiplexer, and a binary address encoding logic of each sub-block hit signals to generate the more significant portion of block binary address.

4. A content addressable memory(CAM) and hit ahead priority encoding(HAPE) logic of claim 3, wherein priority encoding logic, address encoding logic and multiplexer have the logic circuit of same structure.

5. A content address able memory(CAM) and hit ahead priority encoding(HAPE) logic of claim 4, wherein the hit generating logic, priority encoding logic, address encoding logic and multiplexer have dynamic NOR logic.

6. A content address able memory(CAM) and hit ahead priority encoding(HAPE) logic of claim 2, wherein the signal of controlling the multiplexer is generated before or in the same time that the less significant portion of the highest priority local address is generated.

7. A content addressable memory (CAM) and hit ahead priority encoding (HAPE) logic, comprising:

a group of blocks which are arranged in columns and rows, each block having an equal number of CAM match signals which are the input signals of priority encoding logic, each block having a same priority encoding logic of CAM match signals within the block, the CAM match signals or input signals arranged from lower priority to higher priority or from higher priority to lower priority, each CAM match signal or input signal being either a high logic level "one which is called hit or a low logic level "zero" which is called miss, each block configured to generate a block hit signal when there is at least one CAM match signal that is a high logic level "one" within the block or a block miss signal when all the CAM match signals are a low logic level "zero" within the block and a block binary address signal corresponding to the CAM match signals of highest priority within the block;

a priority encoding logic of block hit or miss signals of each column, each column configured to generate a column hit signal when there is at least one block hit signal within the column or a column miss signal when there are only block miss signals within the column and a column binary address corresponding to the CAM match signals of highest priority within the column; and

a priority encoding logic of column hit or miss signals of a group column, the group column configured to generate a hit signal when there is at least one column hit signal within the group column or a miss signal when there are only column miss signals within the group column and a group column binary address corresponding to the CAM match signals of highest priority within the group column.

8. The content addressable memory (CAM) and hit ahead priority encoding (HAPE) logic of claim 7, further comprising:

a block multiplexer configured to select a binary address from the block having the highest priority hit within the column as a less significant portion of the column binary address,

the priority encoding logic of block hit signals being configured to generate a block multiplexer control signal for selecting the block having the highest priority hit within the column; and

a binary address encoding logic of block hit signals configured to generate a more significant portion of the column binary address.

9. The content addressable memory (CAM) and hit ahead priority encoding (HAPE) logic of claim 7, wherein each block comprises:

a group of sub-blocks, each sub-block having an equal number of input signals, each sub-block having priority encoding and binary address encoding logic configured to generate a sub-block highest priority binary address as well as hit or miss generating logic configured to generate a sub-block hit or miss signal, the sub-block hit or miss signal being generated independently before the sub-block binary address;

a block hit or miss generating logic configured to generate a block hit or miss signal, the block hit or miss signal being generated independently before the block binary address is generated;

a sub-block multiplexer configured to select a binary address from a highest priority sub-block within the block as a less significant portion of the block binary address; and

a priority encoding logic of each sub-block hit signals configured to generate a control signal of the sub-block multiplexer; and

a binary address encoding logic of the sub-block hit signals configured to generate a more significant portion of the block binary address.

10. The content addressable memory (CAM) and hit ahead priority encoding (HAPE) logic of claim 9, wherein the priority encoding logic, the address encoding logic, and the multiplexer have logic circuitry of the same structure.

11. The content addressable memory (CAM) and hit ahead priority encoding (HAPE) logic of claim 10, wherein the hit generating logic, the priority encoding logic, the address encoding logic, and the multiplexer have dynamic NOR logic.

12. The content addressable memory (CAM) and hit ahead priority encoding (HAPE) logic of claim 8, wherein a signal for controlling the multiplexer is generated before or at the

same time that the less significant portion of the highest priority local address is generated.

13. A content addressable memory (CAM) system, comprising:

one or more columns comprising a plurality of circuit segments, at least one of the circuit segments configured to generate a first circuit segment output based on whether at least one of a plurality of circuit segment inputs received by the at least one of the circuit segments corresponds to a first logic level,

at least one of the one or more columns configured to generate first address information based on a selected one of the first circuit segment outputs that corresponds to a second logic level, to set a node to a third logic level in response to a first input signal, and to subsequently change the node to a fourth logic level in response to one or more of a plurality of second input signals.

14. The CAM system of claim 13, wherein the first circuit segment output represents circuit segment hit information.

15. The CAM system of claim 13, wherein the at least one of the plurality of circuit segment inputs represents match information.

16. The CAM system of claim 13, wherein the selected one of the first circuit segment outputs is a highest priority one of the first circuit segment outputs that corresponds to the second logic level.

17. The CAM system of claim 13, wherein:

the one or more columns are a plurality of columns, and

the plurality of circuit segments are arranged in the plurality of columns and a plurality of rows.

18. The CAM system of claim 13, wherein:

the one or more columns are a group of columns;

each column in the group configured to generate a column output based on the first circuit segment output of the at least one of the circuit segments; and

the group configured to generate second address information based on a selected one of the column outputs that corresponds to a fifth logic level.

19. The CAM system of claim 13, wherein:

the at least one of the one or more columns is configured to pre-charge the node in response to the first input signal; and

the at least one of the one or more columns is configured to subsequently discharge the node in response to the one or more of the plurality of second input signals.

20. The CAM system of claim 13, wherein the first input signal is configurable independently of the one or more of the plurality of second input signals.

21. The CAM system of claim 13, wherein the first logic level and the second logic level are the same logic level.

22. The CAM system of claim 13, wherein the one or more columns comprise:

a first logic circuit configured to generate a first logic circuit output based on the selected one of the first circuit segment outputs that corresponds to the second logic level;

a second logic circuit configured to generate a second logic circuit output based on whether the first circuit segment output corresponds to the second logic level; and

a third logic circuit configured to generate the first address information based on the selected one of the first circuit segment outputs that corresponds to the second logic level.

23. The CAM system of claim 22, wherein at least one of the first logic circuit, the second logic circuit, and the third logic circuit is configured to set the node to the third logic level in response to the first input signal, and to subsequently change the node to the fourth logic level in response to the one or more of the plurality of second input signals.

24. The CAM system of claim 22, wherein:

the at least one of the circuit segments is configured to generate a second circuit segment output representing second address information; and

the one or more columns further comprise:

a fourth logic circuit configured to select one of the second circuit segment outputs as a less significant portion of the first address information; and

a fifth logic circuit configured to generate a more significant portion of the first address information.

25. The CAM system of claim 24, wherein at least one of the fourth logic circuit and the fifth logic circuit is configured to set the node to the third logic level in response to the first input signal, and to subsequently change the node to the fourth logic level in response to the one or more of the plurality of second input signals.

26. The content addressable memory (CAM) system of claim 24, wherein the one or more columns are each configured to generate a control input for the third logic circuit before or at the same time when the second circuit segment output is generated.

27. The content addressable memory (CAM) system of claim 22, wherein:

the plurality of circuit segment inputs is divided into a plurality of subsets of the circuit segment inputs; and

the first logic circuit comprises:

a plurality of fourth logic circuits each configured to generate a fourth logic circuit output based on whether at least one of a corresponding subset of the circuit segment inputs corresponds to the first logic level; and

a fifth logic circuit configured to generate the first circuit segment output based on whether at least one of the fourth logic circuit outputs corresponds to the first logic level.

28. The CAM system of claim 27, wherein:

at least one of the fourth logic circuit and the fifth logic circuit is configured to set the node to the third logic level in response to the first input signal, and to subsequently change the node to the fourth logic level in response to the one or more of the plurality of second input signals; and

the fourth logic circuit output is an input to the fifth logic circuit.

29. A content addressable memory (CAM) system, comprising:

a circuit segment configured to generate a circuit segment output based on whether at least one of a plurality of circuit segment inputs received by the circuit segment corresponds to a first logic level,

the circuit segment configured to set a node to a second logic level in response to an input signal, and to subsequently change the node to a third logic level in response to the plurality of circuit segment inputs,

the circuit segment output corresponding to said third logic level.

30. The CAM system of claim 29, wherein at least one of the plurality of circuit segment inputs corresponds to a match line output.

31. The CAM system of claim 29, wherein the circuit segment output represents circuit segment hit information.

32. The CAM system of claim 29, wherein at least one of the plurality of circuit segment inputs represents match information.

33. The CAM system of claim 29, wherein:

the circuit segment is configured to pre-charge the node in response to the input signal; and

the circuit segment is configured to subsequently discharge the node in response to the plurality of circuit segment inputs.

34. The CAM system of claim 29, wherein the input signal is configurable independently of the plurality of circuit segment inputs.

35. The CAM system of claim 29, wherein the first logic level and the third logic level are the same logic level.

36. The CAM system of claim 29, wherein the circuit segment is a first circuit segment, and further comprising a second circuit segment configured to generate address information based on the circuit segment output.

* * * * *

Exhibit X1

Plaintiff's infringement analysis

prior to file complaint

Based on the information in Exhibit 1 and public information TetraMem has developed the multi-level resistance state resistor storage element and designed embedded RRAM IP offering to sell, TetraMem's RRAM has read the claim 29 of US patent RE45259.

In the year of 2000 the CEO of TetraMem introduced their RRAM (ReRAM) with multi-state resistance value resistor storage element. Multi-state resistance value resistor storage is not something new, the CEO of TetraMem told me that their multi-state resistance value resistor storage element has the best quality and linearity. The CEO of TetraMem introduced to me their RRAM circuit design and the use of ADC (analog to digital circuit design) to read the multi-level resistance value of the resistor. The CEO of TetraMem told me that TetraMem has very good ADC and memory read circuit design capability, they have designed. the test chip with their multi-level resistance value resistor and ADC. The CEO of TetraMem showed me their test chip manufactured and the measurement result.

The following Figure1 is the basic of RRAM (ReRAM) which is an memory with the different states of the resistor to store the binary data. Figure 1(a) shows the structure of the resistor, it is an isolator material sandwiched with two metal electrodes, some special isolate material and some special metal electrode can make the resistor to conduct with different resistance values. the value of the resistor can be used to record and store the data and information. This is the basic principle of resistance random accessible memory (RRAM). Figure 1(b) is the resistance state transfer diagram curve when applying current and voltage to the resistor. Figure 1(c) is the two resistance value of the two state-value resistor. Figure 1(d) is the three resistance value in the multi-state-value resistor of TetraMem in Exhibit 2.

Figure 2 is the array of resistance memory schematic (RRAM).

Figure 3 shows the major part of read circuit of RRAM with multi-state resistance value of resistor storage element in TrtraRam.

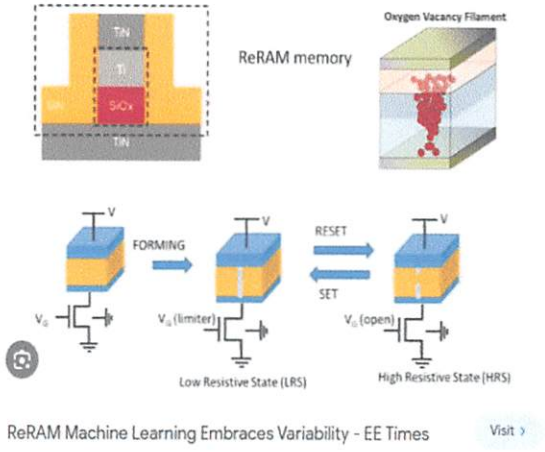


Figure 1 (a) structure of the Resistor

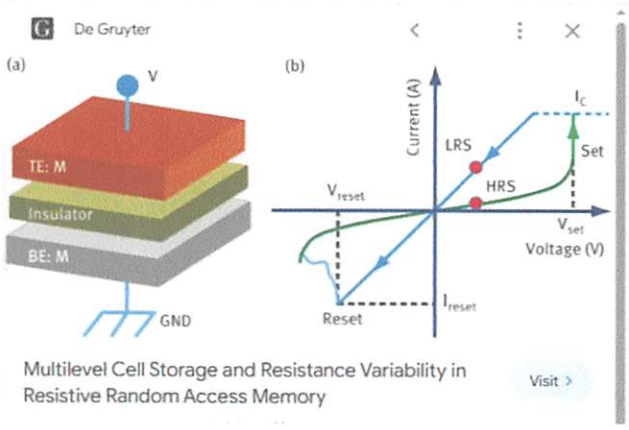


Figure 1 (b) characteristics of the Resistor

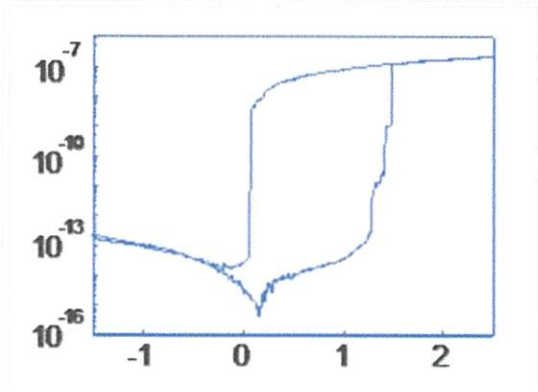


Figure 1 (c) two resistance value of the two level Resistor

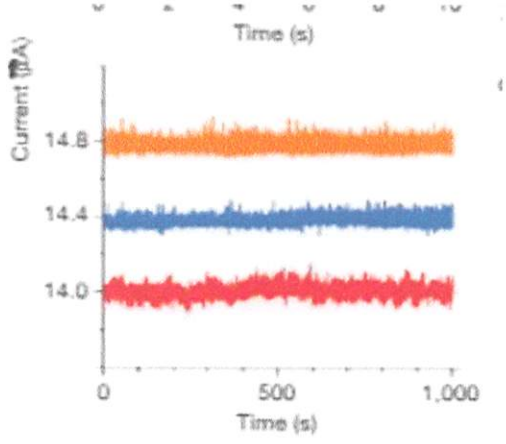


Figure 1 (d) three resistance value of the multi-level Resistor of TetraMem in Exhibit 2

Figure 1. The Structure, characteristics and values of the Resistor storage element used in the RRAM (ReRAM)

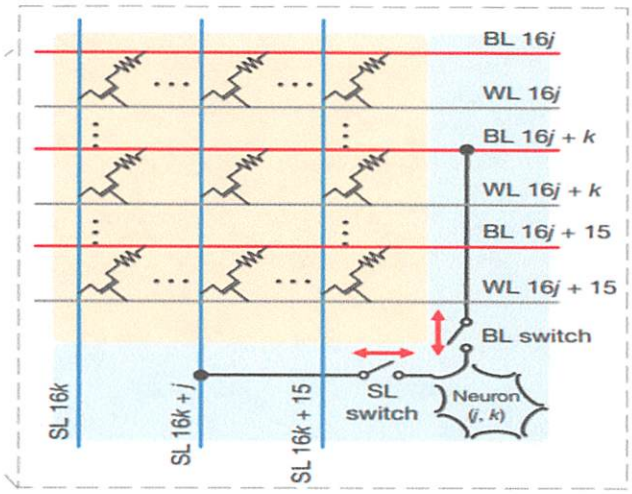


Figure 2 (a) ReRAM array

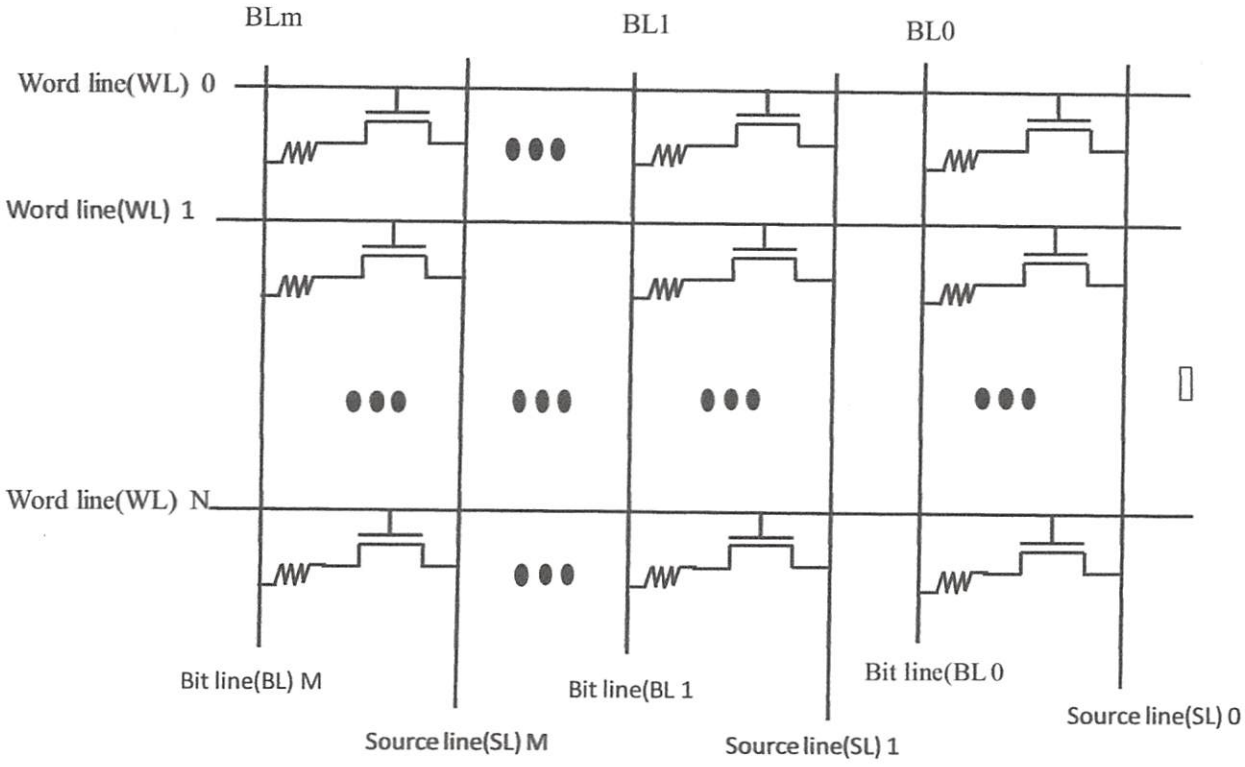


Figure 2. RRAM (ReRAM) array of M column and N row, the N column bit line and source line to read and write the data stored in the resistors, N row word line to enable reading and writing of the data stored in the resistors.

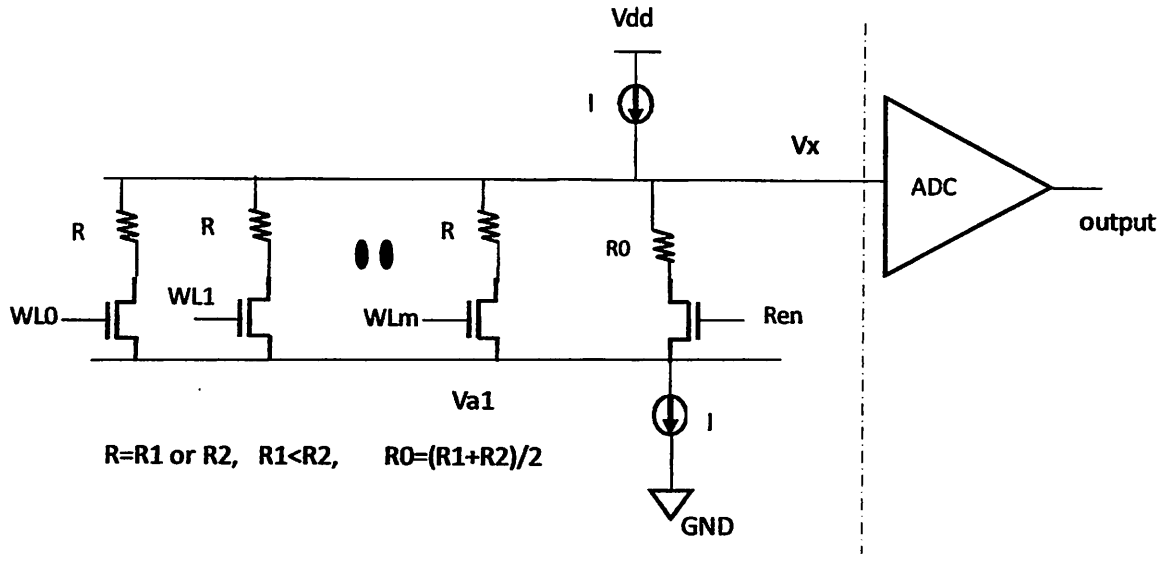


Figure 3 The read circuit of RRAM of TetraMem

Figure 3 is the circuit of reading one column in the RRAM(ReRAM) in Figure 2, it was turned with 90 degree. The left side of the dot line is the reading circuit of the RRAM(ReRAM) used in the available RRAM. For RRAM of resistors with two level resistance usually use an amplifier on the right side of the vertical dot line. But for the TetraMem's RRAM (ReRAM) using multi-level resistance of resistors an ADC (analog signal to digital signal converter) is used.

When starting reading,

Step 1: Switch read enable signal "Ren" signal to V-read, the N transistor controlled by "Ren" signal is On, the node Vx is set to voltage V_0 through the N transistor, R0, and two current source I.

Step2 If read the data controlled by WL0,simultaneously switch WL0 from low voltage to high voltage and switch Ren from high voltage to low voltage, the transistor of WL0 is ON and transistor of Ren is OFF: in the RRAM with two level resistance of the resistor storage element, the voltage of Vx will change to V1 or V2 which corresponds to the resistance state of the resistor storage element connected to the transistor controlled by WL0, in the RRAM having multi-level resistance of resistor storage element, the voltage of Vx will change to one of V1, V2, V3 ..... which corresponds to the resistance state of the resistor storage element controlled by WL0, each bit voltage of ADC's output reflects the value of Vx which corresponds to the information of the resistance value being read.

Claim chart of Claim 29 of US Patent RE45259 read " The read circuit of RRAM of TetraMem " of Figure 1.

claim	**Accused device:** RRAM IP which TetraMem has been developing and offering to sell. RRAM IP which TetraMem have the read circuit shown in Figure 3
Claim 29 *of* US patent RE45259	This claim 29 reads on schematic of "the read circuit of RRAM of TetraMem" in FIG.3
A content addressable memory (CAM) system, comprising:	This is preamble
(1) a circuit segment configured to generate a circuit segment output based on whether at least one of a plurality of circuit segment inputs received by the circuit segment corresponds to a first logic level,	The read operation of RRAM of TetraMem in Figure 3 is explained in page 5, cited as follows: Step 1: Switch read enable signal "Ren" signal to V-read, the N transistor controlled by "Ren" signal is On, the node Vx is set to voltage V_0 through the N transistor, R0, and two current source I. Step2 If read the data controlled by WL0, simultaneously switch WL0 from low voltage to high voltage and switch Ren from high voltage to low voltage, the transistor of WL0 is ON and transistor of Ren is OFF: in the RRAM having multi-level resistance of resistor storage element, the voltage of Vx will change to one of V1, V2, V3 ..... which corresponds to the resistance state of the resistor storage element controlled by WL0, each bit voltage of ADC's output reflects the value of Vx which corresponds to the information of the resistance value being read. This claim section read the schematic in Figure 3 because the generating output of the read circuit is

	based on at least one of WL0, WL1,...., WLm change its voltage level from low to high and switch on the controlled transistor and read the connected (controlled) resistance and generate the output voltage.
(2) the circuit segment configured to set a node to a second logic level in response to an input signal, and	(2) The read operation of RRAM of TetraMem in Figure 3 is explained in page 5, cited as follows: Step 1: Switch read enable signal "Ren" signal to V-read, the N transistor controlled by "Ren" signal is On, the **node Vx is set to voltage V_0** through the N transistor, R0, and two current source I. This section claim read the RRAM of TetraMem in Figure 3, voltage of node Vx is set based on the input voltage of Ren.
(3) to subsequently change the node to a third logic level in response to the plurality of circuit segment inputs, the circuit segment output corresponding to said third logic level.	(3) This section claim read step 2. after (in step 1) node Vx was set to voltage V_0 by Ren, subsequently simultaneously switch WL0 from low voltage to high voltage and switch Ren from high voltage to low voltage, the transistor of WL0 is ON and transistor of Ren is OFF: in the RRAM having multi-level resistance of resistor storage element, the voltage of Vx will change to one of V1, V2, V3 ..... which corresponds to the resistance state of the resistor storage element controlled by WL0, each bit voltage of ADC's output reflects the value of Vx which corresponds to the information of the resistance value being read.

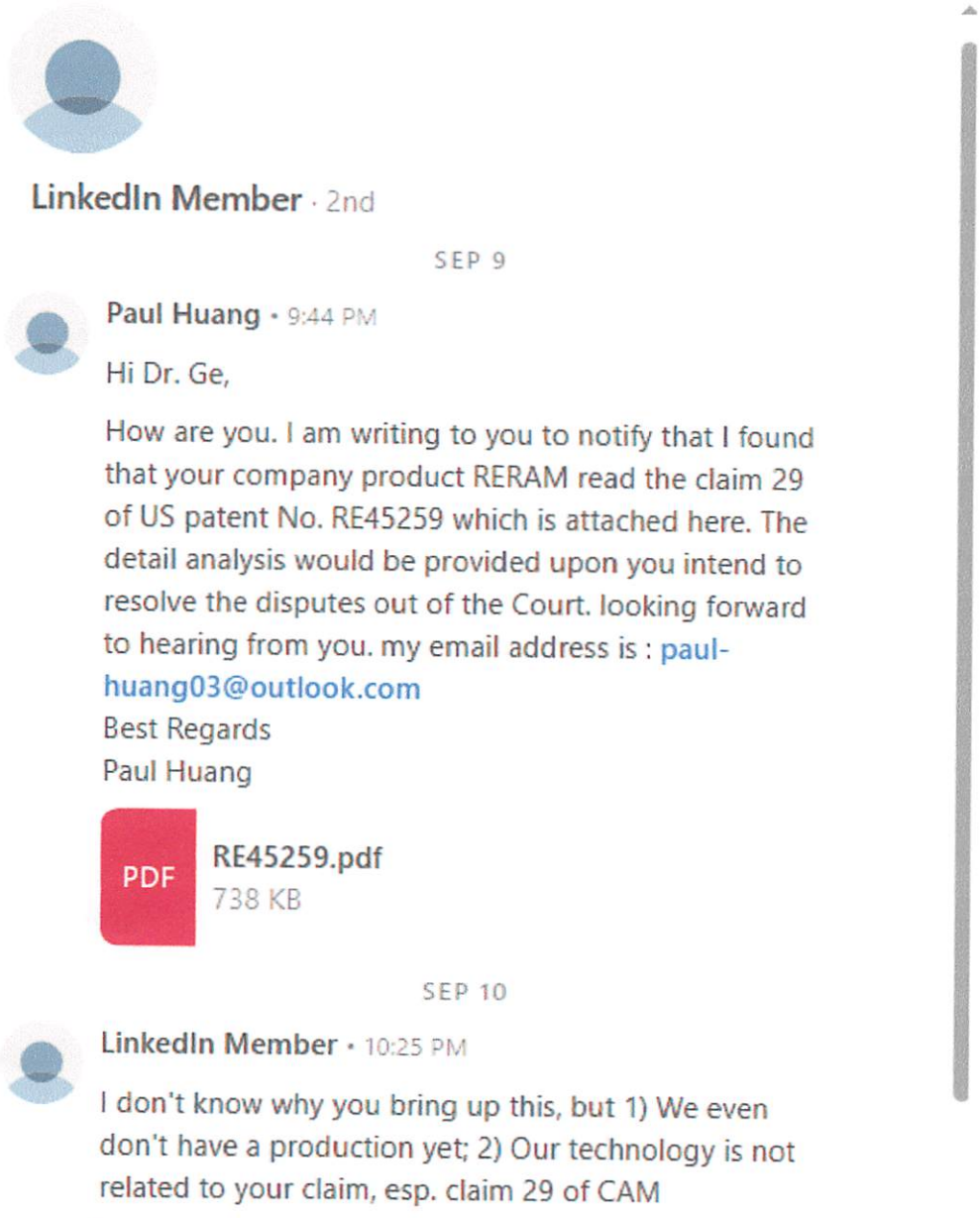
Exhibit 1

Plaintiff's knowledge and effort to discussion prior to complaint as well as the information of TetraMem

Around February 1,2020, I was referred by some one to meet Dr. Glenn (Ning) Ge, the CEO of TetraMem,Inc. at TetraMem's office. Dr. Glenn Ge explained the patented RRAM technology which TetraMem used. Comparing with the two resistance level of storage resistor , one level is at low resistance storing "state 1" and the other level is at high resistance storing "state 0", used in the RRAM in the Market , The storage resistor used in the TetraMem's RMAM can be programed to multi-resistance levels storing multi-states data rather than only "state 0" and "state 1" through applying different voltage and current on it, which is more suitable to be used in the AI or in memory computing. Multi-state resistance value resistor storage is not something new, the CEO of TetraMem told me that their multi-state resistance value resistor storage element has the best quality and linearity. The CEO of TetraMem introduced to me their RRAM circuit design and the use of ADC (analog to digital circuit design) to read the multi-level resistance value of the resistor. The CEO of TetraMem told me that TetraMem has very good ADC and memory read circuit design capability, they have designed. the test chip with their multi-level resistance value resistor and ADC. The CEO of TetraMem showed me their test chip manufactured and the measurement result.

Dr. Ge said that his company has more value because his company is not only a resistor memory company, his company is also a chip company with circuit and logic designers and design capability. I said to Dr. Glenn Ge that the unique competitive value of his company as a start-up is its patented multi-level resistance of storage resistor.

On September 9, 2023 I sent the following message through LinkedIn network to Glenn (Ning) Ge, the CEO of TetraRam



LinkedIn Member · 2nd

SEP 9

Paul Huang • 9:44 PM

Hi Dr. Ge,

How are you. I am writing to you to notify that I found that your company product RERAM read the claim 29 of US patent No. RE45259 which is attached here. The detail analysis would be provided upon you intend to resolve the disputes out of the Court. looking forward to hearing from you. my email address is : paul-huang03@outlook.com

Best Regards

Paul Huang

PDF RE45259.pdf 738 KB

SEP 10

LinkedIn Member • 10:25 PM

I don't know why you bring up this, but 1) We even don't have a production yet; 2) Our technology is not related to your claim, esp. claim 29 of CAM

On September 10 , Glenn(Ning) Ge CEO of TetraRam replied me the above message through LinkedIn network

In the same time he dragged me into black name list and I can not see his name and can only view him as LinkedLn Member

On September 21, 2023 I called Ning Ge, he did not answer my call.

On September 21, I sent Ning Ge a text message: " Hi Dr.Ge, May we meet in person and chat. Thank You. Paul". He did not reply.

On October 11,2023 I sent Mr. Ning (Glen) Ge a text message in Chinese :

Hi Dr, Ge, Three years ago We met each other , at that time I said to you:" the most value technology of your company is the technology of resistor your company developed, the other stuff such as phase lock loop and analog to digital converter design etc. can be purchased or developed through hiring regular engineers." I said this to you with very good wishes. Maybe " sesame open the door" is valuable at special place and moment.

Dr. Ge still did not reply my request.

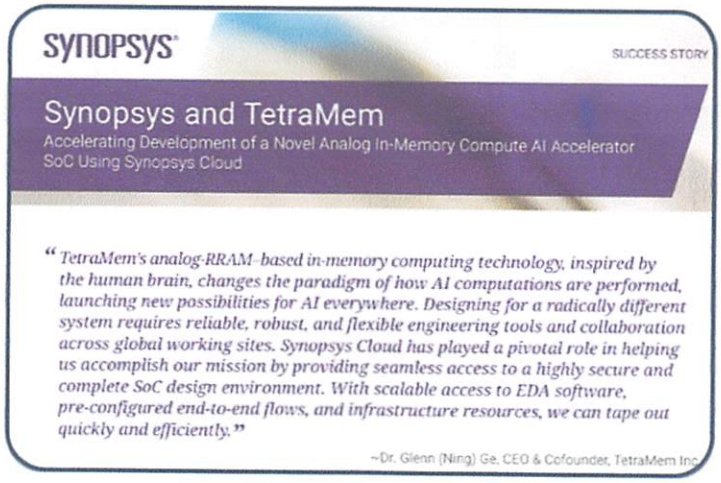
The following content is the public release on the RRAM information, offering and joint development offering.

TetraMem and Synopsys Collaborate to Accelerate Development of Groundbreaking AI Accelerator Chips with Analog In-Memory Computing

" TetraMem's analog-RRAM–based in-memory computing technology, inspired by the human brain, changes the paradigm of how AI computations are performed, launching new possibilities for AI everywhere. Designing for a radically different system requires reliable, robust, and flexible engineering tools and collaboration across global working sites. Synopsys Cloud has played a pivotal role in helping us accomplish our mission by providing seamless access to a highly secure and complete SoC design environment. With scalable access to EDA software, pre-configured end-to-end flows, and infrastructure resources, we can tape out quickly and efficiently."
~Dr. Glenn (Ning) Ge, CEO & Cofounder, TetraMem Inc.

...

Read the full Success Story at
https://www.synopsys.com/content/dam/synopsys/cloud/success-story/tetramem-success-story.pdf



"TetraMem's analog-RRAM–based in-memory computing technology, inspired by the human brain, changes the paradigm of how AI computations are performed, launching new possibilities for AI everywhere. Designing for a radically different system requires reliable, robust, and flexible engineering tools and collaboration across global working sites. Synopsys Cloud has played a pivotal role in helping us accomplish our mission by providing seamless access to a highly secure and complete SoC design environment. With scalable access to EDA software, pre-configured end-to-end flows, and infrastructure resources, we can tape out quickly and efficiently."

~Dr. Glenn (Ning) Ge, CEO & Cofounder, TetraMem Inc.

TetraMem

Business

TetraMem Inc. is a US-based company founded in Silicon Valley in 2018. It is poised to deliver the industry's most disruptive in-memory computing (IMC) technology for efficient AI applications. TetraMem is the world's only company to produce a high bit density multi-level RRAM-based computing-in-memory accelerator in commercial foundries. The technology was featured in the March 30, 2023 edition of Nature. Our team brings together complementary skill sets and technology know-how ranging from proprietary computing-in-memory devices to novel computing architectures. Learn more at tetramem.com and follow us on LinkedIn.

Challenges

- Having a global team spread across continents requires a seamless, unified mixed-signal SoC design environment for efficient collaboration.
- CAD setup flows and on-demand scalable and elastic IT resources are needed.
- The AI accelerator's complex design requires smooth integration and verification of analog in-memory computing and the digital RISC-V processor with critical time-to-market targets.

Exhibit 8

The public information claimed by TetraMem about TetraMem and its products

Dr. Glenn NingGe, CEO

Recent breakthroughs in data and algorithms have ignited transformation in the AI computing hardware realm, bringing revolutions to various industries. Healthcare, autonomous vehicles, smart cities, and cybersecurity industries are widely embracing these innovations, but not without their share of challenges.

As Dr. Glenn Ning Ge, CEO of TetraMem says, "Relying on traditional computing methodologies involving separate processing and memory unit leads to von Neumann bottlenecks and memory walls, where the system spends significant time and energy moving data for the matrix-based deep-learning applications. This issue is further aggravated by the surge in data-centric applications, which can result in reduced computation capabilities."

AI models, particularly deep neural networks, require massive computational power for vector-matrix multiplication

TetraMem ingrains all these features into its suite of solutions, comprising an analog in-memory computing chip, IP and customizable design solutions, and software toolkit and support. This analog in-memory computing chip is designed to accelerate AI algorithms, particularly for edge devices. Given its architecture, the chipset consumes significantly less power and provides faster processing times. Currently delivered as the TetraMem® MX100 SoC (system on-chip), an evaluation kit designed for customers to test its capabilities, the chipset is a gateway for numerous groundbreaking innovations that transform how AI models are used. Academic and industry partners and researchers are exploring the earlystage development and research of analog IMC using TetraMem's hardware and SDK software, giving them the opportunity to delve into this emerging field.

"Our platform isn't just about the hardware—it's a six-dimension co-design and co-optimization," says Dr. Ge. "We consider the entire AI computing ecosystem when we have a chance to re-design the new AI computing system, from materials and devices to algorithms and software, ensuring optimized performance across the board where the emerging computing paradigm meets up with emerging devices"

The blend of deep expertise, comprehensive solutions, and a client-centric approach sets TetraMem apart. An unwavering commitment to innovation and excellence ensures that it does not just meet, but exceeds client and partner expectations. For instance, Dr. Ge himself holds a PhD and three master's degrees including a MBA from The University of Michigan Ross School of Business.

"The tech landscape is fast and everevolving, and resting on one's laurels is not an option. We encourage a culture of continuous learning and innovation, guaranteeing we're always ahead of the curve," says Dr. Ge.

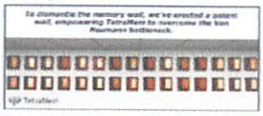
❝Relying on traditional computing methodologies involving separate processing and memory unit leads to von Neumann bottlenecks and memory walls, where the system spends significant time and energy moving data for the matrix-based deep-learning applications. Now, the

Company
TetraMem

Management
Dr. Glenn NingGe, CEO

Description
TetraMem is a pioneer in analog memristor technology and in-memory computing that addresses the challenges of AI processing in power-constrained environments with its in-memory computing technology for edge applications. This groundbreaking technology eliminates weight-data movement, substantially improving the energy efficiency and performance of AI and ML workloads.

accumulation operations. The exponential rise in computing needs calls for huge power consumption at data centers, putting a significant financial toll on most businesses. Simultaneously, intensive computation leads to excessive heat generation, and efficiently dissipating heat is challenging in compact devices without active cooling systems. Scalability and technology node limits also add to existing woes where physical limitations and quantum effects pose a serious challenge to further miniaturization.



Part of the issue lies in hardware-software co-optimization, which is often hindered by the hardships associated with developing compilers, runtime systems, and hardware architectures that work harmoniously.

"Addressing these challenges and developing an environment for AI applications to make a real difference hinges on the combination of innovation in fundamental new material science, emerging devices, semiconductor manufacturing, system architecture, and algorithm design," says Dr. Ge.

That's where TetraMem emerges as a leader.

TetraMem's hardware solutions are complemented by its comprehensive software development kit. TetraMem Instinct™ software, which facilitates the seamless development and deployment of applications using IMC technology. It also offers standard IP solutions tailored to specific needs, recognizing that different applications have varying requirements. This supports seamless integration of IMC technology into a diverse range of applications, including high-throughput, low-power-consumption devices for eyeball tracking, face detection, and motion detection. The firm's offerings in AIMC can also be harnessed to accelerate AI algorithms crucial for real-time data processing in medical applications. Rapid analysis of patient data, medical images, or even real-time monitoring can benefit from the computational efficiency and energy savings offered by TetraMem's solutions.

> We believe that technology must serve humanity. Our goal is not just to enhance machines but to uplift humanity.

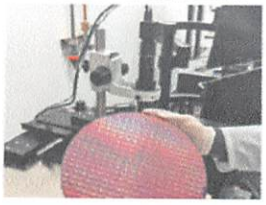
"This is 'The Transistor Moment' for AI computing after decades of von Neumann and transistor resonances," says Dr. Ge.

Strategic Partnership to Build AI Accelerator Chip

> new memory is the computer.

A testament to TetraMem's clientcentricity and continuous innovation is its ability to be attuned to market demands and emerging trends. Through regular interactions with potential clients, tech analysts, and industry leaders, it gauges real-time market needs and actively seeks feedback to understand the challenges businesses face. Coupled with a crossfunctional collaboration approach, the team ensures that its solutions are technologically advanced and a direct response to real-world challenges. Rigorous testing to assess real-world efficiency and applicability is also a priority for TetraMem.

The team focuses on agile principles, continuous development, and regular model updates, all with a commitment to sustainability. Its IMC solutions are energy-efficient, addressing environmental concerns and operational costs associated with data processing.



TetraMem

Management
Dr. Glenn NingGe, CEO

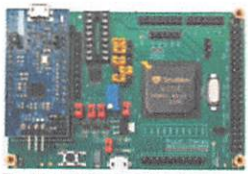
Description
TetraMem is a pioneer in analog memristor technology and in-memory computing that addresses the challenges of AI processing in power-constrained environments with its in-memory computing technology for edge applications. This groundbreaking technology eliminates weight-data movement, substantially improving the energy efficiency and performance of AI and ML workloads.

A pioneer in analog in-memory computing, TetraMem is at the forefront of this transformative shift in the AI computing paradigm. Born out of the collective vision of a team with deep expertise and experience in semiconductor design, AI applications, and emerging memory technologies, it delivers an efficient crossbar array-based neural processing unit (NPU) that uses analog in-memory computing technology. Under the leadership of Ge, who holds a PhD and three master's degrees and brings 20 years of industry experience and approximately 800 patent filings globally to his role, the team has developed the first-of-its-kind, purpose-built solution. It is inspired by the human brain's efficiency and uses the power of emerging nonvolatile memory devices to perform computation directly within the physical memory unit, accelerating AI algorithms. This architectural shift promises unparalleled, several order-of-magnitude improvement in energy efficiency and performance, especially in edge applications where real-time processing is paramount.

The 'TetraMem Way' of Spearheading Innovation

TetraMem's groundbreaking technology offers scalability and energy efficiency beyond any market-leading analog technologies. The TetraMem's NPU delivers 20–100 TOPS/W at INT8 using mature technology nodes, like 65 nm and 22 nm, and the team expects to gain 300 TOPS/W at more advanced nodes with crossbar sizes scaling up and some of other technical innovations like ADC renovation. The solution's memory-based computation ability eliminates weight-data movement and elevates the performance of AI or ML workloads. Coupled with massive parallel processing, it guarantees higher throughput than traditional architectures. The technology also enables non-volatile memory devices, where the neural network models are pre-loaded onto the chip to allow data fetching directly from sensors, without weighing down the memory.

Backed by industry leaders and solid alliances with major semiconductor companies, TetraMem is positioned to validate the technology's potential for large-scale commercial deployments. For instance, its strategic partnership with Andes Technology, a supplier of high-efficiency, low-power RISC-V processor cores, allows for delivery of a fast, highly efficient AI inference chip. The fusion of Andes' high-performance RISC-V Vector CPU with TetraMem's IMC architecture through Andes Custom Extension (ACE) enables the best performance. This amplifies the strengths of both companies, resulting in fast, energy-efficient AI inference that surpasses all limitations of traditional computing approaches and transcends the 'memory wall' and 'end of Moore's Law' constraints.



The joint collaboration aims to create a powerful chip that improves energy efficiency by at least an order of magnitude. Optimizing computations and eliminating the transfer of weight data will significantly extend the battery life of edge devices and impose a near-zero impact on thermal budgets.

TetraMem is also venturing to enhance its technology's potential for flexibility and scalability. It stays true to this mission by designing AI accelerator chips with versatility and scalability from 22 nm to 7 nm and beyond, making it easy to integrate into numerous AI-powered products and applications across multiple industries. The team has demonstrated that its compute memristor can be scaled down to 2 nm and below, carving a roadmap to many future-proof solutions.

Envisioning a Human- Centered World of AI

A team of engineers and domain specialists is at the heart of TetraMem's excellence, giving it a competitive spirit to drive advancements and stay ahead of the curve in today's tech landscape teeming with innovators. Their extensive experience and rigorous R&D process enable the firm to constantly push the envelope to ensure that its technology remains at the forefront.

"We believe that technology must serve humanity. Our goal is not just to enhance machines by adding artificial intelligence but to uplift humanity through emotional intelligence. Empathy and compassion are core values we hold dear, and this is important to machines and robots. Therefore, we strive to infuse humanity into machines and robots," says Ning Ge. "Our technology enables a world of AI that acts and feels more human."

Stepping into the Future of Efficient AI Models

The horizon is luminous for TetraMem as the upcoming era of AI computing, spanning the edge and cloud, comes into focus. Its IMC technology is poised to address the pressing challenges of energy efficiency and power constraints in a landscape that demands highly power-conscious computational strategies. Continuing its streak to steer ahead of technological innovations, the firm is building more efficient processors for CPUs and GPUs, capable of running AR, VR, and data centers with less power consumption. The goal is to deliver cuttingedge technologies that operate 10 times more efficiently than a GPU running at 500 watts with less than 50-watt power consumption.

TetraMem envisions unveiling the AI accelerator chip and making engineering samples and development kits for the new 22-nm TetraMem MX series chip available to the public by the end of 2024. The firm is also joining hands with numerous tech leaders and research entities to co-develop environmentally friendly AI solutions, shaping the narrative around green AI.

It's safe to say that TetraMem is more than just a silicon chip vendor. It is a partner for companies looking to harness the next wave of computing advancements. All of its solutions are crafted with the future in mind, ensuring that its partners are always ahead in the rapidly evolving technological landscape.

Company
TetraMem

Management
Dr. Glenn NingGe, CEO

Description
TetraMem is a pioneer in analog memristor technology and in-memory computing that addresses the challenges of AI processing in power-constrained environments with its in-memory computing technology for edge applications. This groundbreaking technology eliminates weight-data movement, substantially improving the energy efficiency and performance of AI and ML workloads.

in-memory computing technology for edge applications. This groundbreaking technology eliminates weight-data movement, substantially improving the energy efficiency and performance of AI and ML workloads.

TetraMem Delivers RISC-V AI Accelerator Tape-Out in Record Time on Synopsys Cloud

Anuj Pant

Mar 27, 2024 / 3 min read

Table of Contents

Founded in Silicon Valley in 2018, TetraMem is ready to change the landscape for AI computing by addressing one of its most vexing challenges: the need to deliver massive compute processing while keeping its power consumption in check. With its in-memory computing (IMC) technology for efficient AI applications, TetraMem is the only company producing a high-bit-density, multi-level RRAM (otherwise known as a computing memristor) accelerator in commercial foundries.

Comprised of a global team spread across different continents, TetraMem needed an efficient way to collaborate to develop its hardware solutions, as well as scalable and flexible design and verification resources. The company found its answer in cloud-based EDA solutions, namely through Synopsys Cloud. Providing cloud-native EDA tools and pre-optimized hardware platforms, Synopsys Cloud supports chip design from end to end, with advantages including end-to-end license management automation, robust security, and spot instances for high-memory EDA workloads.

"TetraMem's analog RRAM-based in-memory computing technology, inspired by the human brain, changes the paradigm of how AI computations are performed, launching new possibilities for AI everywhere," said Dr. Glenn (Ning) Ge, CEO and co-founder of TetraMem. "Designing for a radically different system requires reliable, robust, and flexible engineering tools and collaboration across global working sites."

Table of Contents

"Synopsys Cloud has played a pivotal role in helping us accomplish our mission by providing seamless access to a highly secure and complete SoC design environment. With scalable access to EDA software, pre-configured end-to-end flows, and infrastructure resources, we can tape out quickly and efficiently."

Dr. Glenn (Ning) Ge, CEO and Co-Founder, TetraMem |

Faster Time to Results and Elastic Compute Resources

The TetraMem team knew they needed smooth integration and verification of its complex AI accelerator's analog in-memory computing technology and its digital RISC-V processor. Since project demands on EDA resources can fluctuate depending on the design stage, the team recognized that having scalable, elastic, on-demand access to these resources would be advantageous. Aggressive time-to-market targets also weighed on the team.

Synopsys Cloud proved to be the answer to all these challenges, enabling TetraMem to experience:

- A jump start on hardware development through a pre-configured, end-to-end analog design flow
- An intuitive, user-friendly browser-based platform for chip design
- Faster time to results through on-demand, pay-per-use access to unlimited EDA resources without licensing constraints

option is a usage-based licensing approach offering by-the-minute pricing for EDA tools, while a cloud subscription license (CSL) option is term-based and requires upfront payment. Both PPU and CSL options use cloud credits. Synopsys Cloud can be deployed through Bring-Your-Own-Cloud (BYOC) or Software-as-a-Service (SaaS) models.

Read other Synopsys Cloud customer success stories here.

- Instant availability of scalable and advanced compute and storage resources

"We were able to achieve a very fast infrastructure setup on the Synopsys Cloud EDA environment within days," said Wenbo Yin, VP of IC Design at TetraMem. "The vast selection of EDA tools and IP available on the cloud enabled us to start the design, verification, and backend flow very quickly. The flexibility of using as many licenses as we needed enabled us to obtain fast turnarounds on simulation, verification, and backend flow, which significantly reduced the engineering hours spent on those compute-intensive tasks. The ease of global access provided a single environment for our global R&D team with seamless access."

Notes David George, TetraMem's head of global operations: "We are creating the world's most efficient AI accelerator chip, built around a RISC-V architecture and powered by our novel analog in-memory computing technology. Developing and integrating these elements requires a well-engineered platform to support an iterative and flexible design process in order to meet performance objectives with quality. With Synopsys Cloud, we are able to access a complete, pre-configured design flow that enables our global R&D teams to collaborate efficiently. This improved productivity for our entire team, and helped us architect a robust design for a first-of-its-kind technology."

In addition to the technologies available through Synopsys Cloud, the TetraMem team also tapped into Synopsys' worldwide support team. "One of our challenges from the start was bringing global team members up to speed to meet our product specifications and time-to-market goals," said Charles Wei, the company's director of SoC design. "The Synopsys sales, technical support, and R&D teams worked together to help us become familiar with the platform. Any EDA tool upgrades or new tool purchases are delivered immediately without worrying about CAD installation, global environment setup, version compatibility, etc. We trust that using Synopsys Cloud and collaborating with Synopsys for the development of our current and future SoC designs is the right choice."

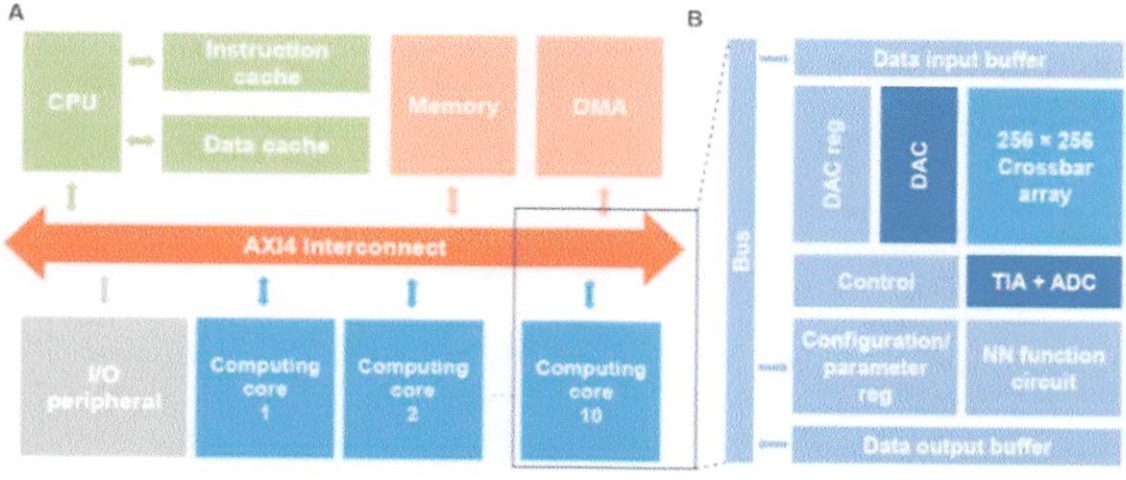
Flexible Licensing and Deployment Options

Synopsys Cloud features a FlexEDA business model with two licensing options. The pay-per-use (PPU)

TetraMem Inc. (Fremont, Calif.), a startup that is developing analog-in-ReRAM compute processor architecture, has published a paper that shows its form of analog computing is capable of executing calculations with arbitrary precision.

The ability to perform high-precision multiplication within single electronic devices that can be readily formed in arrays offers scope to reduce the power consumption of machine learning when based on artificial neural networks.

The publishing of *Programming memristor arrays with arbitrarily high precision for analog computing* in *Science* follows on from the news that TetraMem has worked with Synopsys to develop an analog-in-memory compute AI accelerator SoC. That SoC is apparently called Cullinan.



Layout of analog in-memory computing accelerator SoC with A) whole SoC and B) diagram of one of ten identical computing cores. Source: Supplementary materials to TetraMem's Science paper.

The article provides a technical underpinning to the premise that analog computing can support calculations with arbitrary precision. This plays to TetraMem's fundamental device a metal-oxide based cross-point switch, which was reported in Nature in April 2023 (see Single memristor provides 11-bit memory device).

The device is made of a mixture of Al3O2, above a layer of HfO2 sandwiched between a tantalum/titanium top electrode and a platinum bottom electrode. Each of the bilayers is less than 1nm thick so that after being laid down they appear to form a mixed layer rather than two separate continuous layers. The device was fabricated in a 240-nm diameter via above the CMOS peripheral circuitry.

Conductive channels are formed in the material by applying a set voltage and that can be at varying levels of strength. The authors propose that the conductive and insulating phases of the metal-oxide material relate to the orthorhombic phase with a high number of oxygen vacancies and the monoclinic phase without oxygen vacancies, respectively.

"We are creating the world's most efficient AI accelerator chip, built around a RISC-V architecture and powered by our novel analog in-memory computing technology," said David George, head of global operations for TetraMem, at the time of the Synopsys announcement.

TetraMem was co-founded in 2018 by Joshua Yang, Professor of Electrical and Computer Engineering at the University of Southern California, Professor Qiangfei Xia of the University of Massachusetts, Miao Hu (CTO), and Glenn Ge (CEO). The company had already produced a number of test chips in silicon by mid-2022 and at that time expected to see mass production of chips including its technology by about the end of 2023.

Related links and articles:

The article in *Science* can be accessd here:
https://www.science.org/doi/10.1126/science.adi9405

Dr. Glenn Ge

Cofounder & CEO

LinkedIn >

Dr. Miao Hu

Cofounder & CTO

LinkedIn >

Dr. Joshua Yang

Cofounder & Scientific Board Chief Advisor

Dr. Qiangfei Xia

Cofounder & Chief Process Advisor

Exhibit 2

The RRAM of TetraMem was published on the nature magazine one picture was cited by Exhibit X1

Article

Thousands of conductance levels in memristors integrated on CMOS

https://doi.org/10.1038/s41586-023-05759-5

Received: 7 August 2022

Accepted: 25 January 2023

Published online: 29 March 2023

Check for updates

Mingyi Rao[1,2,5], Hao Tang[3,5], Jiangbin Wu[4,5], Wenhao Song[4,5], Max Zhang[1], Wenbo Yin[1], Ye Zhuo[4], Fatemeh Kiani[2], Benjamin Chen[2], Xiangqi Jiang[3], Hefei Liu[4], Hung-Yu Chen[4], Rivu Midya[2], Fan Ye[2], Hao Jiang[2], Zhongrui Wang[2], Mingche Wu[1], Miao Hu[1], Han Wang[4], Qiangfei Xia[1,2], Ning Ge[1], Ju Li[3] & J. Joshua Yang[1,2,4]✉

Neural networks based on memristive devices[1–3] have the ability to improve throughput and energy efficiency for machine learning[4,5] and artificial intelligence[6], especially in edge applications[7–21]. Because training a neural network model from scratch is costly in terms of hardware resources, time and energy, it is impractical to do it individually on billions of memristive neural networks distributed at the edge. A practical approach would be to download the synaptic weights obtained from the cloud training and program them directly into memristors for the commercialization of edge applications. Some post-tuning in memristor conductance could be done afterwards or during applications to adapt to specific situations. Therefore, in neural network applications, memristors require high-precision programmability to guarantee uniform and accurate performance across a large number of memristive networks[22–28]. This requires many distinguishable conductance levels on each memristive device, not only laboratory-made devices but also devices fabricated in factories. Analog memristors with many conductance states also benefit other applications, such as neural network training, scientific computing and even 'mortal computing'[25,29,30]. Here we report 2,048 conductance levels achieved with memristors in fully integrated chips with 256 × 256 memristor arrays monolithically integrated on complementary metal–oxide–semiconductor (CMOS) circuits in a commercial foundry. We have identified the underlying physics that previously limited the number of conductance levels that could be achieved in memristors and developed electrical operation protocols to avoid such limitations. These results provide insights into the fundamental understanding of the microscopic picture of memristive switching as well as approaches to enable high-precision memristors for various applications.

Memristive-switching devices are known for their relatively large dynamical range of conductance, which can lead to a large number of discrete conductance levels. Different approaches have been developed to accurately program the devices[31]. However, only devices with fewer than 200 conductance levels have been reported so far[22,32]. There are no forbidden conductance states in the dynamical range of the device because a memristor is analog and can, in principle, achieve an infinite number of conductance levels. However, the fluctuation commonly observed at each conductance level (Fig. 1e) limits the number of distinguishable levels that can be achieved in a specific conductance range. We found that such fluctuation can be substantially suppressed, as shown in Fig. 1e,f, by applying appropriate electrical stimuli (called 'denoising' processes). Notably, this denoising process does not require any extra circuitry beyond the usual read-and-program circuits. We incorporated the denoising process into device-tuning algorithms and successfully programmed a memristor made in a standard commercial foundry (Fig. 1a–d) into 2,048 conductance levels (Fig. 1g), corresponding to a resolution of 11 bits. Conductive atomic force microscopy (C-AFM) was used to visualize the evolution of conduction channels during programming and denoising processes. We discovered that a normal switching operation (set or reset) always ends up with some incomplete conduction channels, which appear as islands or blurry edges along the main conduction channel and are less stable than the main conduction channel. First-principles calculations indicate that these incomplete channels are unstable phase boundaries with dopant levels in a range that is sensitive to nearby trapped charges, contributing to the large fluctuations of each conductance level. We showed, experimentally and theoretically, that an appropriate voltage in the denoising process either annihilates (weakens) or completes (enhances) these incomplete channels, resulting in a strong reduction in fluctuation and a

[1]TetraMem, Fremont, CA, USA. [2]Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA, USA. [3]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, USA. [5]These authors contributed equally: Mingyi Rao, Hao Tang, Jiangbin Wu, Wenhao Song ✉e-mail: jjoshuay@usc.edu

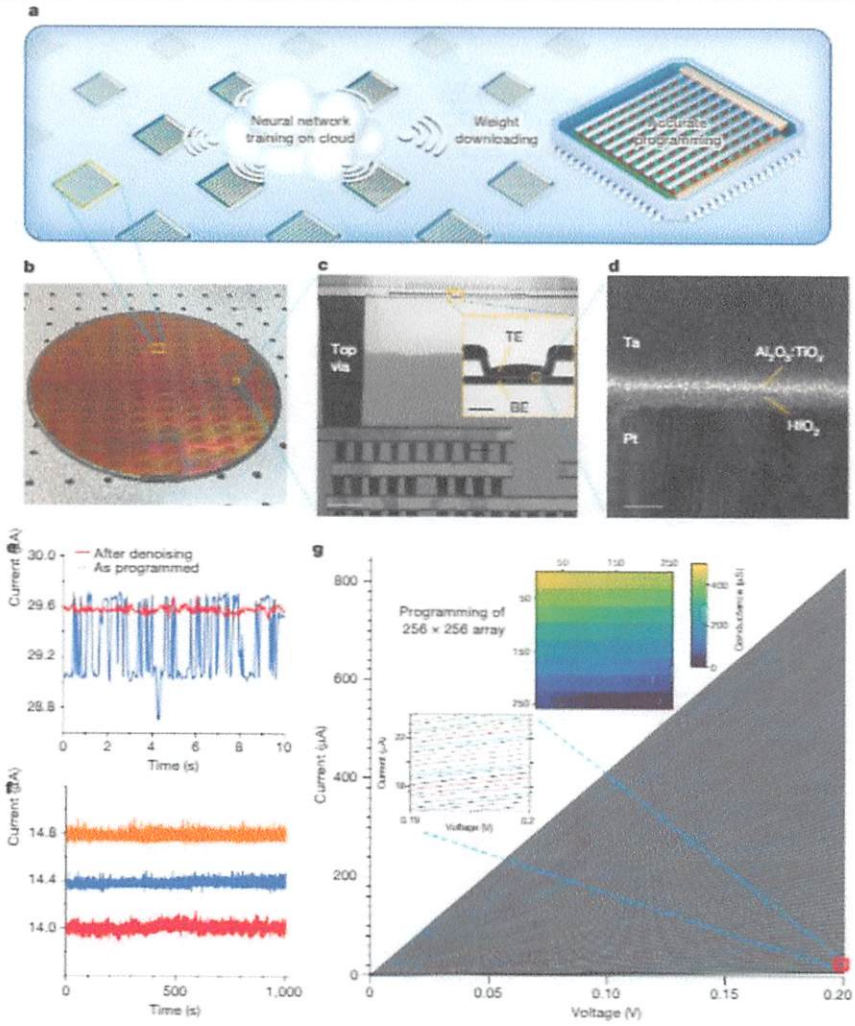


Fig. 1 | High-precision memristor for neuromorphic computing. a, Proposed scheme of the large-scale application of memristive neural networks for edge computing. Neural network training is performed in the cloud. The obtained weights are downloaded and accurately programmed into a massive number of memristor arrays distributed at the edge, which imposes high-precision requirements on memristive devices. **b**, An eight-inch wafer with memristors fabricated by a commercial semiconductor manufacturer. **c**, High-resolution transmission electron microscopy image of the cross-section view of a memristor. Pt and Ta serve as the bottom electrode (BE) and top electrode (TE), respectively. Scale bars, 1 μm and 100 nm (inset). **d**, Magnification of the memristor material stack. Scale bar, 5 nm. **e**, As-programmed (blue) and after-denoising (red) currents of a memristor are read by a constant voltage (0.2 V). The denoising process eliminated the large-amplitude RTN observed in the as-programmed state (see Methods). **f**, Magnification of three nearest-neighbour states after denoising. The current of each state was read by a constant voltage (0.2 V). No large-amplitude RTN was observed, and all of the states can be clearly distinguished. **g**, An individual memristor on the chip was tuned into 2,048 resistance levels by high-resolution off-chip driving circuitry, and each resistance level was read by a d.c. voltage sweeping from 0 to 0.2 V. The target resistance was set from 50 μS to 4,144 μS with a 2-μS interval between neighbouring levels. All readings at 0.2 V are less than 1 μS from the target conductance. Bottom inset, magnification of the resistance levels. Top inset, experimental results of an entire 256 × 256 array programmed by its 6-bit on-chip circuitry into 64 32 × 32 blocks, and each block is programmed into one of the 64 conductance levels. Each of the 256 × 256 memristors has been previously switched over one million cycles, demonstrating the high endurance and robustness of the devices.

substantial increase in memristor precision. The observed phenomena generally exist in a memristive-switching process with localized conduction channels, and the insights can be applied to most memristive systems for scientific understanding and technological applications.

Conductance levels and arrays on integrated chips

Memristors used in this study were fabricated on an eight-inch wafer by a commercial semiconductor manufacturer (Fig. 1b). Details about the

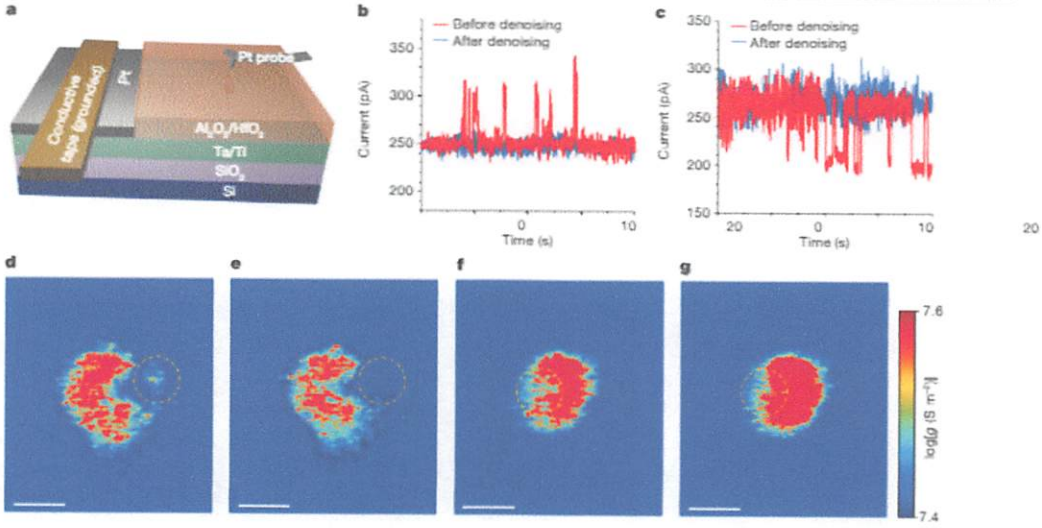


Fig. 2 | Direct observation of the evolution of conduction channels in the denoising process using C-AFM. a, Schematic of the customized memristor structure and C-AFM testing set-up. A C-AFM probe was used as the top electrode in the customized device. Because Ta easily oxidizes in air and is not a practical probe material, a Pt probe was used. This Pt probe had the same purpose as that of the bottom Pt electrode of the standard memristor that we used. To maintain the material stack of a standard memristor, the customized memristor has a reversed structure. **b**, Current readings at 0.1 V before (red) and after (blue) a denoising process using a subthreshold reset voltage. **c**, Current readings at 0.1 V before (red) and after (blue) a denoising process using a subthreshold set voltage. **d**, Conductance map measured by C-AFM scanning corresponding to the before-denoising state (red) in **b**. **e**, Conductance map corresponding to the after-denoising state (blue) in **b**. **f**, Conductance map measured by C-AFM scanning corresponding to the before-denoising state (red) in **c**. **g**, Conductance map corresponding to the after-denoising state (blue) in **c**. The dashed yellow circles in **d**–**g** highlight the changes observed before and after the denoising process. Scale bars, 10 nm.

fabrication process are provided in the Methods. Cross-section views of a memristor are shown in Fig. 1c, and the crucial resistive switching layers are magnified in Fig. 1d. The elemental image produced by electron energy-loss spectroscopy is shown in Supplementary Fig. 1. The device, which consists of a Pt bottom electrode, a Ti/Ta top electrode and a HfO_2/Al_2O_3 bilayer, was fabricated in a 240-nm via above the CMOS peripheral circuitry. The Al_2O_3 and Ti layers are designed to be thin (<1 nm) so that they seem as a mixed layer rather than two separate continuous layers. When the bottom electrode is grounded, the device can be switched by applying either a sufficiently positive voltage (for set) or a negative voltage (for reset) to the top electrode. The fluctuation level (characterized by the standard deviation of a measured current under a constant voltage) after a set or a reset operation is distributed in a wide range (Supplementary Fig. 2). The result indicates that an as-programmed state typically has large fluctuations. This considerably limits the applications of memristors, but is a characteristic of memristive materials more generally[33–36]. The data also show that a set operation tends to induce a larger fluctuation in an as-programmed state than does a reset operation. Such reading fluctuations mainly consist of random telegraph noise (RTN), which typically has step-like transitions between two or more current levels at random time points under a constant reading voltage. Such RTN generally exists in memristors. Even fluctuations that do not seem step-like may in fact be made of a RTN[37], which can be shown only when the measurement sampling rate is higher than the RTN frequency, as shown in Supplementary Fig. 3. It has been demonstrated previously by simulations that memristor RTN may be caused by charges occasionally trapping into certain defects and blocking conduction channels because of Coulomb screening[34,35]. However, experiments that directly link trapped charges, conduction channel(s) and RTN, and how to remove it, are missing. Although this is a critical issue for memristors in general, it has been unclear how to reduce the RTN in memristors. These experiments are important not only for understanding the physical origin of memristor RTN but also for revealing the entire microscopy picture of memristive switching and providing possible solutions to high-precision memristors.

We discovered that the fluctuation level could be greatly reduced by applying small voltage pulses with optimized amplitude and width. An example is given in Fig. 1e, in which an as-programmed state with a considerable fluctuation (blue) was stabilized into a low-fluctuation state (red) by denoising pulses. Using a three-level feedback algorithm devised to denoise, as shown in Supplementary Fig. 4, a single memristor was tuned into 2,048 conductance states between 50 and 4,144 μS, with a 2-μS interval between every two neighbouring states. All states were read by a voltage sweeping from 0 to 0.2 V, as shown in Fig. 1g. The bottom inset to Fig. 1g shows magnification of the current–voltage curves, which show the well-distinguishable states and the marked linearity of each state. Three nearest-neighbour states after denoising are shown in Fig. 1f, in which a constant voltage of 0.2 V reads each state for 1,000 s. The current fluctuation of every state is within 0.4 μA, corresponding to 2 μS in conductance. No significant overlap was observed in the neighbouring states. A magnification of the measurement at high-conductance states is shown in Supplementary Fig. 5. Memristors from multiple chips of an 8-inch wafer were measured, demonstrating considerable programming uniformity across the entire wafer, as shown in Supplementary Fig. 6. We further used the denoising process in the array-level programming of an entire 256 × 256 array using the on-chip circuitry. The experimentally programmed patterns are shown in Fig. 1g (top inset) and Supplementary Fig. 7. For demonstrations using the on-chip circuitry, the programming precision was limited by the precision of the on-chip analog-to-digital conversion peripheral circuitry,

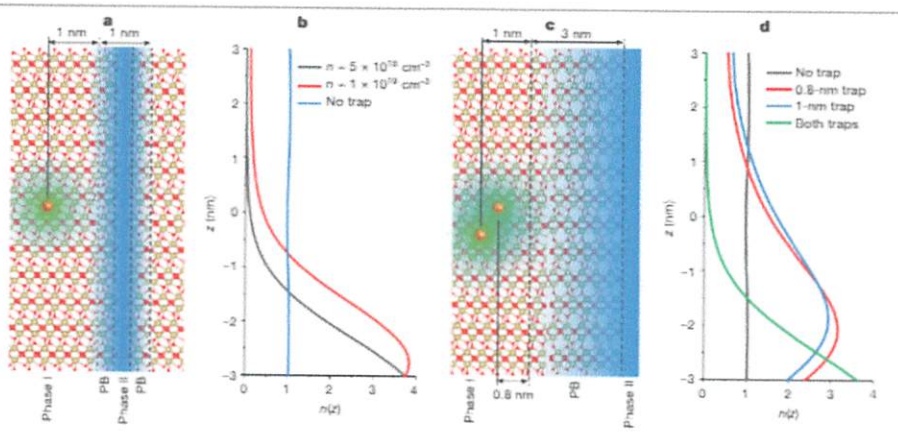


Fig. 3 | Trapped-charge-induced conductance change in incomplete conduction channels. a, The RTN-responsible defect (orange) is 1 nm away from an island-like conduction channel (blue). The channel is formed by a conductive phase region (phase II) and the phase boundary (PB) region. **b**, The transport electron wavefunction corresponding to **a**, where z denotes the position of the channel along the electron transport direction (from −3 nm to 3 nm), and $n(z)$ shows the normalized integration of the transport electron wavefunction on the plane perpendicular to the z direction, which indicates the electrical conduction at each z position. The black and red curves are $n(z)$ when the carrier density in the channel is 5×10^{19} cm^{-3} or 1×10^{19} cm^{-3} with one electron trapped at the defect, respectively, and the blue line is $n(z)$ with no electron trapped. **c**, Two defects (orange) are positioned away from a channel that is attached to the main conduction channel. The PB region is 3 nm in width. **d**, The transport electron wavefunction corresponding to **c**. The red and blue lines correspond to $n(z)$ when one electron is trapped in the defect 0.8 nm and 1 nm away from the channel, respectively, and the green and black lines correspond to $n(z)$ when both or none of the defects have trapped electrons. The carrier density in the channel for the simulation is 5×10^{19} cm^{-3}.

which was 6-bit (64 levels) in this design. The testing set-up and the schematic of the driving circuits are shown in Supplementary Fig. 8. The extra system cost caused by the denoising process is estimated in Supplementary Information Section 9. Because a relatively smaller voltage is needed for denoising than is required for typical set or reset programming, the extra energy consumption is only a small fraction of the energy needed for programming. Further studies show that the denoising operation can also reduce RTN in other material stacks, for example, a TaO_x-based memristor, as shown in Supplementary Fig. 10. Because reading noise has been observed in various resistive switching materials, the results indicate that the denoising step is an important, potentially essential, process for the training of memristive neural networks because unstable readings lead to incorrect outputs from the neural networks, and these cannot be compensated by adaptive in situ training.

Conduction channel evolution in denoising processes

Deciphering the underlying reason for the above results is essential for finding a reliable solution to the problem of unstable conductance states and understanding the dynamic process of memristive switching. Visualizing the evolution of conduction channels during electrical operations is informative for this purpose[39–42]. We used C-AFM to precisely locate the active conduction channel(s) and scan all of the surrounding regions. Details of the measurement are provided in the Methods and Supplementary Fig. 11. A customized device was fabricated for the C-AFM measurements. A schematic of its structure is shown in Fig. 2a. To use the Pt-coated C-AFM tip as the top electrode, the device was designed to have a reversed structure compared with that of the standard device shown in Fig. 1d. By grounding the bottom electrode and applying a voltage to the top electrode, the device can be operated as our standard device with opposite voltage polarities—that is, a positive voltage tends to reset the device, and a negative voltage tends to set the device. Denoising operations were also successfully performed by C-AFM, as shown in Fig. 2b,c. The conductance scanning results corresponding to the reading results of Fig. 2b are shown before (Fig. 2d) and after (Fig. 2e) denoising, and those for the reading results of Fig. 2c are shown in Fig. 2f,g. A comparison of the conductance maps in Fig. 2d,e reveals that the main part of the conduction channel (the 'complete' channel) remains nearly the same whereas the positive denoising voltage annihilates an island-like channel (the 'incomplete' channel). By contrast, the negative denoising voltage (Fig. 2f,g) reduces the noise by removing the current dips in Fig. 2c. These results indicate that the conductance of an RTN-rich state can be divided into two parts: the base conductance provided by complete channels and the RTN provided by incomplete channels. These incomplete channels had formed together with complete channels but were smaller in size. Such incomplete channels were also observed in $SrTiO_x$-based resistive switching devices[43]. A memristor can be denoised by eliminating incomplete channels (by either removing or completing them). Incomplete channels are more sensitive to voltage stimuli compared with complete channels, which makes it possible to tune the former without affecting the latter by using appropriate electrical stimuli. Further studies suggest that this is a general mechanism and can also be performed in other material stacks (Supplementary Fig. 12). It should be noted that the seemingly isolated island(s) may or may not be electrically connected with the main conduction channel beneath the surface. However, this does not change the denoising mechanisms or operation protocols.

Switching and denoising mechanisms

To understand the mechanism of denoising, we studied the microscopic origin of RTN in memristors. An important question is whether RTN is induced by an atomic effect or electronic effect. As shown in Supplementary Fig. 13, incomplete channels are consistently observed in a C-AFM scan whenever RTN is observed. Once incomplete channels are eliminated, the RTN disappears. This indicates that RTN is associated

Exhibit 3

Kramer Levin

James Hannah
Partner
T 650.752.1712
jhannah@KRAMERLEVIN.com

333 Twin Dolphin Drive
Redwood Shores, CA 94065
T 650.752.1700
F 650.752.1800

April 29, 2024

By EMAIL AND OVERNIGHT DELIVERY

Xiaohua/Paul Huang
P.O. Box 1639
Los Gatos, CA, 95301
Tel: (669) 273-5633
Email: paul-huang03@outlook.com

Re: ***Xiaohua Huang v. TetraMem Inc. – Case No. 23-cv-04936-SVK***

Dear Mr. Xiaohua/Paul Huang,

As you know, we represented TetraMem Inc. ("TetraMem") in the recent action entitled, *Xiaohua Huang v. TetraMem Inc.*, Case No. 23-cv-04936-SVK (N.D. Cal. 2023) ("Action"), which was closed on February 9, 2024. We found your February 12, 2024 and April 27, and 28, 2024 emails regarding potentially filing another lawsuit very disconcerting given the information that we have provided to you already that demonstrates that there is no infringement of your patents.

For example, both your Complaint, and First Amended Complaint (the "Complaints") contained the same baseless allegations against TetraMem. Dkt. Nos. 1, 1-1, and 15, 15-1 to 15-3. A sufficient investigation would have revealed that the claims raised therein were without any factual foundation, legally unreasonable, and frivolous. *Id.* Also, the fact that you did not remedy any of the deficiencies highlighted in TetraMem's *first* Motion to Dismiss and Motion for an Award of Attorneys' Fees and Costs shows and supports that your pleadings were filed in bad faith for an improper purpose[1]. Dkt. Nos. 8, 10, and 15, 15-1 to 15-3.

The claims in your Complaint did not meet the basic requirements or *Twombly/Iqbal*'s plausibility requirements. *Id.* Among other things, TetraMem's alleged accused devices were never specifically identified in any of your Complaints, and none of TetraMem's alleged accused devices include any of the main features of claim 29 of U.S. Patent RE45259. *See e.g.*, Dkt. Nos. 1, 8, 10, 15. As an example, your purported "claim chart" and allegations in the FAC failed to include any connection between any specific

[1] See fn. 2.

feature of the accused product(s) to the claim language. Dkt. No. 15; Dkt. No. 15-2, Ex. X1 at 6-7.

A reasonable pre-filing inquiry would demonstrate that U.S. Patent RE45259 and TetraMem's accused devices are two totally different technologies. The U.S. Patent RE45259 primarily applies to content-addressable memory (CAM), whereas TetraMem has been developing technology for in-memory computing (IMC). Moreover, TetraMem's ReRAM IP and chips are still in the development phase, and has no plan of including any CAM designs or any main features of U.S. Patent RE45259 in its future products, as outlined in the table below.

	U.S. Patent RE45259	TetraMem
Circuit Type	CAM (Content Addressable Memory)	IMC (In-Memory Computing)
Circuit Input	Voltage, Logic Level Voltage input applied to Transistor Gate	Voltage, Analog Voltage input applied to memristor in 1T1R (one transistor and one resistance switch)
Circuit Output	Voltage, Logic Level	Current, Analog Each 1T1R performs multiplication. Each wire performs current integration as a line output

Given the information that you have, any attempt to file another lawsuit against TetraMem will be in violation of the Federal Rules of Civil Procedure, as you cannot have a good faith basis for a patent infringement claim against TetraMem after conducting a reasonable pre-suit investigation into the claim(s), as required by Rules 8(a) and 11 of Federal Rules of Civil Procedure.

Based on your *pro se* litigant's experience[2], you certainly already know that:

[2] *Xiaohua Huang v. Genesis Glob. Hardware, Inc.*, No. 2:20-cv-1713-JAM-KJN, Dkt. No. 33 at 7 n.5 (E.D. Cal. Dec. 15, 2020) (noting in footnote 5 "other patent cases plaintiff ha[d] brought as a pro se litigant, which include citations to plaintiff's history of vexatious conduct and inability to follow the federal and local rules of court—despite being given multiple opportunities to do so"); *Huang v. MediaTek USA, Inc.*, 815 F. App'x 521, 525 (Fed. Cir. 2020) (affirming district court's dismissal of plaintiff's pro se infringement claims, despite "four opportunities to serve proper contentions . . . , multiple warnings and ample guidance from the district court."); *Huang v. Open-Silicon, Inc.*, No. 18-cv-00707-JSW, 2018 U.S. Dist. LEXIS 221487 (N.D. Cal Aug. 27, 2018) (denying plaintiff's motion to amend for failure to follow the local rules, failure to offer more than just boilerplate in the proposed amendments, and failure to demonstrate good faith attempts at supporting factual inferences); *Huang v. Huawei Techs. Co.*, No. 15-cv-01413-JRG, 2017 WL 1133201, at *3 (E.D. Tex. Mar. 27, 2017) (sanctioning plaintiff for his "bad faith and an abuse of the judicial process").

> Rule 11 expressly requires that an unrepresented party presenting a pleading, motion, or other paper before the court certify that he has performed "an inquiry reasonable under the circumstances" such that he can verify that (1) "it is not being presented for any improper purpose, such as to harass, cause unnecessary delay, or needlessly increase the cost of litigation," (2) "the claims ... are warranted by existing law or by a nonfrivolous argument for extending, modifying, or reversing existing law;" (3) "the factual contentions have evidentiary support or, ... will likely have evidentiary support after a reasonable opportunity for further investigation or discovery." Rule 11 of Federal Rules of Civil Procedure ("Rule 11"); *Raylon, LLC v. Complus Data Innovations, Inc.*, 700 F.3d 1361, 1366–67 (Fed. Cir. 2012) (citing Fed. R. Civ. P. 11(b)(1)–(3)).

TetraMem would be forced to pursue remedies available to it if there is another violation of Rule 11. *See Hunter Douglas Inc. v. Great Lake Woods, Inc.*, No. 15-cv-00106-REB-KLM, 2019 WL 1375675, at *16 (D. Colo. Mar. 27, 2019) (granting attorney fees under Section 285 where the defendants' conduct was "not just tactical blunders," but rather, they "advanced baseless claims and positions, and pursued claims and defenses they knew or should have known to be meritless.")[3].

Given the dismissal of the previous Action, there should not be any further litigation between the parties. Filing yet another complaint against TetraMem for patent infringement, particularly one with the same deficiencies highlighted in TetraMem's Motions[4], would be "exceptional" under Section 35. U.S.C. § 285. If it comes to another litigation, TetraMem will have no choice but to pursue all available remedies against you, including attorneys' fees and costs, pursuant to Rule 11, and 35 U.S.C. § 285.

Sincerely,

James Hannah

[3] *See also Harris Research, Inc. v. Perrine,* No. 1:05-cv-136 CW, 2009 WL 1457674, at *9 (D. Utah May 22, 2009), *report and recommendation adopted,* 2010 WL 936071 (D. Utah Mar. 15, 2010) (granting attorney fees under Section 285 against *pro se* litigants); *Finch v. Hughes Aircraft Co.*, 926 F.2d 1574, 1582 (Fed. Cir. 1991) ("While courts are particularly cautious about imposing sanctions on a pro se litigant, whose improper conduct may be attributed to ignorance of the law and proper procedures, this and other circuits have imposed sanctions in cases where **even a nonlawyer should have been aware that his conduct was frivolous**.") (emphasis added).

[4] Dkt. Nos. 8, 10, 20, 28, and 32.

Exhibit 4

Exhibit 4

Plaintiff's response to Defendant's letter in Exhibit 3

Dear Mr. Hannah,

The following is my response to the letter in April 29, 2024.

The argument in the letter on April 29, 2024 is nothing but cheating to fool the *pro se* litigant and the forged "so called evidence" to further cheat the Court.

(1) The claim is not limited to one embodiment, see Exhibit 5 of USPTO claim rule, the claim 29 of RE45259 read the ReRAM as you stated 1R, 1T (1 resistor , 1 transistor) ReMAM as in my complaint. The claim 29 of RE45259 is not limited to any one embodiment in RE45259, TCAM is the example and preamble, the claim cover the other circuits such as ReRAM of 1Rand 1T structure, because the reading operation of ReRAM of 1Rand 1T structure are read on by the claim 29 of RE45259. The letter further confirm that the ReRAM of TeTraMem (1T and 1R structure) is consistent with my complaint and the ReRAM of TeTraMem (1T and 1R structure) read the claim 29 of RE45259.

(2) The letter accused me with my past cases. I do not want to argue it here, the personals accused me with those cases are not only cheater, but violated their oath to the GOD, GOD will not like their conducts, their family and descendants will suffer for their conducts..

I do not believe the letter was written by you, the letter is nothing but evilly cheating.

I will refile the complaint against TeTraMem. The letter of April 29, 2024 still used the previous cases to cheating, defaming to harm me and to cheat

the Judge. Please stop doing so, merely for some extra incomes, to avoid your further cheating conducts will have your descendants suffer your conducts, because the ultimate justice is guaranteed by GOD.

Regards

/s/ Xiaohua Huang (Paul)

Exhibit 5

Claim interpretation

CLAIMS MUST BE GIVEN THEIR BROADEST REASONABLE INTERPRETATION IN LIGHT OF THE SPECIFICATION

During patent examination, the pending claims must be "given their broadest reasonable interpretation consistent with the specification." The Federal Circuit's en banc decision in Phillips v. AWH Corp., 415 F.3d 1303, 1316, 75 USPQ2d 1321, 1329 (Fed. Cir. 2005) expressly recognized that the USPTO employs the "broadest reasonable interpretation" standard:

The Patent and Trademark Office ("PTO") determines the scope of claims in patent applications not solely on the basis of the claim language, but upon giving claims their broadest reasonable construction "in light of the specification as it would be interpreted by one of ordinary skill in the art." In re Am. Acad. of Sci. Tech. Ctr., 367 F.3d 1359, 1364[, 70 USPQ2d 1827, 1830] (Fed. Cir. 2004). Indeed, the rules of the PTO require that application claims must "conform to the invention as set forth in the remainder of the specification and the terms and phrases used in the claims must find clear support or antecedent basis in the description so that the meaning of the terms in the claims may be ascertainable by reference to the description." 37 CFR 1.75(d)(1).

See also In re Suitco Surface, Inc., 603 F.3d 1255, 1259, 94 USPQ2d 1640, 1643 (Fed. Cir. 2010); In re Hyatt, 211 F.3d 1367, 1372, 54 USPQ2d 1664, 1667 (Fed. Cir. 2000).

Patented claims are not given the broadest reasonable interpretation during court proceedings involving infringement and validity, and can be interpreted based on a fully developed prosecution record. In contrast, an examiner must construe claim terms in the broadest reasonable manner during prosecution as is reasonably allowed in an effort to establish a clear record of what applicant intends to claim. Thus, the Office does not interpret claims when examining patent applications in the same manner as the courts. In re Morris, 127 F.3d 1048, 1054, 44 USPQ2d 1023, 1028 (Fed. Cir. 1997); In re Zletz, 893 F.2d 319, 321-22, 13 USPQ2d 1320, 1321-22 (Fed. Cir. 1989).

Because applicant has the opportunity to amend the claims during prosecution, giving a claim its broadest reasonable interpretation will reduce the possibility that the claim, once issued, will be interpreted more broadly than is justified. In re Yamamoto, 740 F.2d 1569, 1571 (Fed. Cir. 1984); In re Zletz, 893 F.2d 319, 321, 13 USPQ2d 1320, 1322 (Fed. Cir. 1989) ("During patent examination the pending claims must be interpreted as broadly as their terms reasonably allow."); In re Prater, 415 F.2d 1393, 1404-05, 162 USPQ 541, 550-51 (CCPA 1969) (Claim 9 was directed to a process of analyzing data generated by mass spectrographic analysis of a gas. The process comprised selecting the data to be analyzed by subjecting the data to a mathematical manipulation. The examiner made rejections under 35 U.S.C. 101 and 35 U.S.C. 102. In the 35 U.S.C. 102 rejection, the examiner explained that the claim was anticipated by a mental process augmented by pencil and paper markings. The court agreed that the claim was not limited to using a machine to carry out the process since the claim did not explicitly set forth the machine. The court explained that "reading a claim in light of the specification, to thereby interpret limitations explicitly recited in the claim, is a quite different thing from 'reading limitations of the specification into a claim,' to thereby narrow the scope of the claim by implicitly adding disclosed limitations which have no express basis in the claim." The court found that applicant was advocating the latter, i.e., the impermissible importation of subject matter from the specification into the claim.). See also In re Morris, 127 F.3d 1048, 1054-55, 44 USPQ2d 1023, 1027-28 (Fed. Cir. 1997) (The court held that the USPTO is not required, in the course of prosecution, to interpret claims in applications in the same manner as a court would interpret claims in an infringement suit. Rather, the "PTO applies to verbiage of the proposed claims the broadest reasonable meaning of the words in their ordinary usage as they would be understood by one of ordinary skill in the art, taking into account whatever enlightenment by way of definitions or otherwise that may be afforded by the written description contained in applicant's specification.").

The broadest reasonable interpretation does not mean the broadest possible interpretation. Rather, the meaning given to a claim term must be consistent with the ordinary and customary meaning of the term (unless the term has been given a special definition in the specification), and must be consistent with the use of the claim term in the specification and drawings. Further, the broadest reasonable interpretation of the claims must be consistent with the interpretation that those skilled in the art would reach. In re Cortright, 165 F.3d 1353, 1359, 49 USPQ2d 1464, 1468 (Fed. Cir. 1999) (The Board's construction of the claim limitation

"restore hair growth" as requiring the hair to be returned to its original state was held to be an incorrect interpretation of the limitation. The court held that, consistent with applicant's disclosure and the disclosure of three patents from analogous arts using the same phrase to require only some increase in hair growth, one of ordinary skill would construe "restore hair growth" to mean that the claimed method increases the amount of hair grown on the scalp, but does not necessarily produce a full head of hair.). Thus the focus of the inquiry regarding the meaning of a claim should be what would be reasonable from the perspective of one of ordinary skill in the art. In re Suitco Surface, Inc., 603 F.3d 1255, 1260, 94 USPQ2d 1640, 1644 (Fed. Cir. 2010); In re Buszard, 504 F.3d 1364, 84 USPQ2d 1749 (Fed. Cir. 2007). In Buszard, the claim was directed to a flame retardant composition comprising a flexible polyurethane foam reaction mixture. 504 F.3d at 1365, 84 USPQ2d at 1750. The Federal Circuit found that the Board's interpretation that equated a "flexible" foam with a crushed "rigid" foam was not reasonable. Id. at 1367, 84 USPQ2d at 1751. Persuasive argument was presented that persons experienced in the field of polyurethane foams know that a flexible mixture is different than a rigid foam mixture. Id. at 1366, 84 USPQ2d at 1751.