Dissertation

submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the
Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by
Dipl.-Phys. Gilbert Maurice Güttler
born in Bad Friedrichshall, Germany

Date of oral examination: November 14, 2017

Achieving a Higher Integration Level of Neuromorphic Hardware using Wafer Embedding

Referees: Prof. Dr. Karlheinz Meier
Prof. Dr.-Ing. Ulrich Brüning

Abstract

This thesis investigates the application of a chip packaging technology on a silicon wafer for a neuromorphic hardware system. The procedure embeds a complete silicon wafer into a printed circuit board (PCB). The embedding procedure was developed together with the Fraunhofer Institute for Reliability and Microintegration in Berlin. At first bare silicon disks were embedded into PCBs to find the correct material selection and the suitable stack up. Finally, the HICANN wafers of the Electronic Visions Group, which are employed in the BrainScaleS system, were embedded into PCBs. For that reason 20 cm wafers are thinned down to 250 µm and the redistribution layer (RDL) pads get an additional copper layer. The Deep EvolutioN in System Embedding (DENSE) board has two copper layers and uses microvias for the connections to the wafer. Existing software of the BrainScaleS system is used to test and characterize the DENSE board. The tests cover the communication to the wafer and the functionality of the analog parameter storage and the neuron circuits. It is shown that the embedding process has no influence on the circuits and on the RDLs of the wafer. Additionally, the long-term reliability of the board was tested in a climate cabinet by conducting an accelerated environmental stress test. No failures were observable on the DENSE board.

Zusammenfassung

Diese Arbeit untersucht die Anwendung einer Chip-Verpackungstechnik auf einen Silizumwafer für ein neuromorphes Hardware-System. Das Verfahren bettet einen kompletten Siliziumwafer in eine Leiterplatte ein. Das Einbettverfahren wurde zusammen mit dem Fraunhofer Insitut für Zuverlässigkeit und Mikrointegration in Berlin entwickelt. Zuerst wurden blanke Silizumscheiben in Leiterplatten eingebettet, um die korrekte Materialauswahl und den passenden Lagenaufbau zu finden. Schließlich wurden HICANN Wafer der Electronic Visions Gruppe in eine Leiterplatte eingebettet, diese werden auch im BrainScales System benutzt. Dafür werden die 20 cm Wafer auf 250 µm gedünnt und die Pads der Umverdrahtungslage erhalten eine zusätzliche Kupferschicht. Die DENSE Leiterplatte hat zwei Kuperlagen und verwendet Microvias für die Verbindungen zum Wafer. Vorhandene Software vom BrainScaleS System wird zum Testen und Charakterisieren der DENSE Leiterplatte eingesetzt. Die Tests decken die Kommunikation zum Wafer und die Funktionalität des analogen Parameterspeichers und der Neuronenschaltungen ab. Es wird gezeigt, dass das Einbettverfahren keinen Einfluss auf die Schaltkreise und die Umverdrahtungslagen des Wafers hat. Außerdem wurde die langfristige Zuverlässigkeit der Leiterplatte mit einem beschleunigten Umgebungsbelastungstest in einem Klimaschrank geprüft. Es waren keine Ausfälle auf der DENSE Leiterplatte feststellbar.

Contents

1 Introduction **1**

2 BrainScaleS System **5**

2.1 Components 5

2.1.1 Neuromorphic Silicon Wafer 5

2.1.2 The FPGA Communication PCB 10

2.1.3 MainPCB 11

2.2 Wafer Connection Technique 17

2.2.1 Wafer Post Processing 18

2.2.2 Elastomeric Connectors 20

2.2.3 FR4 Positioning Mask 23

2.3 Construction of a BrainScaleS System 23

2.4 Scalability of the BrainScaleS System 27

2.5 Conclusion 28

3 Embedding a Silicon Wafer **29**

3.1 Embedding in a PCB 29

3.2 Challenges in the Material Selection 30

3.2.1 Material Property Coefficient of Thermal Expansion 31

3.2.2 Heat Dissipation 33

3.3 Wafer Preparation 37

3.3.1 Wafer Thinning 37

3.3.2 Copper Pads Finish 38

4 Bare Silicon Lamination Tests **39**

4.1 First Lamination Test with Standard FR4 Material 39

4.2 Second Lamination Test with CIC Material 41

4.3 Third Lamination Test with low-CTE Material 43

4.4 Conclusion 47

5 DENSE Prototype **49**

5.1 Stack up 49

5.2 Layout 54

5.3 Test-Hardware 59

5.4 Connect DENSE Board with Cube Setup 60

5.5 Test Methods 64

5.5.1 LVDS Link Resistance 64

5.5.2 Switchram Test 65

5.5.3 Floating-Gate Cell Test 65
5.5.4 Neuron Calibration 66
5.5.5 Thermal Stress 69
5.5.6 Test of the Redistribution Layers 70
5.6 Experiments on HICANNv2 70
5.7 Conclusion 72

6 Results of the DENSE Prototypes **73**
6.1 Problem during the Embedding Process 73
6.2 Warpage 76
6.3 Switchram Test 76
6.4 Floating-Gate Cell Test 81
6.5 Neuron Calibration 85
6.6 Reliability Tests 91
6.7 Conclusion 94

7 Roadmap for the Embedded Wafer System **95**
7.1 Backside Connections to the Wafer 95
7.2 Higher Layer Count 97
7.3 Silicon Wafers with 30 cm Diameter 99
7.4 Application of Through-Silicon Vias 99
7.5 Cooling of the Embedded Wafer 101
7.6 Embedded Capacitors 103
7.7 Interconnection of several Modules 106
7.8 Final Embedded Wafer System 107

8 Conclusion **111**

9 Outlook **113**

Bibliography **121**

A BrainScaleS system **129**

B Bare Silicon Tests **133**

C DENSE Preparation **137**

D DENSE Measurements **139**

E Cooling Calculations **153**

F Used Hardware and Software **155**

1 Introduction

New generations of computing devices like processors either have more computing power, consume less electrical power, shrink in size or a combination of all.
Gordon Moore described this development with a simple rule saying that when referring to integrated circuits the complexity of integrated circuits doubles every 12 to 24 months [*Moore*, 1965]. Neuromorphic computing systems like the SpiNNaker Project, TrueNorth or BrainScaleS system are no exception to this development [*Furber et al.*, 2014; *Merolla et al.*, 2014; *Schemmel et al.*, 2010a]. These neuromorphic systems implement different models of neurons and synapses with optimized electronic circuits on a silicon substrate. Following different architectural approaches, these systems allow the simulation or emulation of neuronal networks. For all systems the integration density of neurons and synapses plays an important role as it mainly determines the maximum size of the implemented neuronal networks. The main premises of the Electronic Visions Group at the Heidelberg University is the physical neuron model. Every neuron in a neural network is emulated by one neuron circuit on silicon. Therefore, how can the possible neural network size be increased?
There are two possibilities to increase the network size. First of all increase the number of neurons and synapses per silicon chip by using an advanced process technology. The other option is to build more systems and interconnect them. The Electronic Visions Group follows both approaches.

In 2000 the neuromorphic engineering at the Heidelberg University (Heidelberg University) started with the EvoOpt chip which was produced in a 350 nm process It had 64 neurons and 4 096 synapses [*Schemmel et al.*, 2001].

The next chip in 2001 the Heidelberg AnaloG Evolvable neural Network (HAGEN) was still produced in a 350 nm process but with 256 neurons and 33 000 synapses [*Schemmel et al.*, 2002]. Both chips implemented the perceptron neuron model [*Rosenblatt*, 1958]. Perceptrons differ strongly from biological plausible neurons, as the information transfer is not done with spatio-temporal events like in biology with action-potentials.

In 2005 the successor chip Spikey came with a completely new-designed neuron model, based on the Leaky Integrate-and-Fire (LIF) model with conductance-based synapses [*Destexhe*, 1997]. It is a more realistic continuous-time model than the perceptron model as the neurons are not in binary states anymore. This is achieved by using analog electronic components solving the LIF neuron model differential equations. Simplified, a capacitor represents the neuron membrane and additional components emulate other cell mechanisms like leakage. The analog implementation of the neuron works with an acceleration of 10 000 compared to biological time. An experiment emulating 1 year takes around 8 h on Spikey.

The synapses were also improved with a plasticity feature, called Spike-Timing Dependent Plasticity (STDP). STDP increases or decreases the synaptic strength depending on the spike time correlations [*Schemmel et al.*, 2006].

The Spikey chip is produced in the enhanced 180 nm process. The latest version 5 has 384 neurons and 98 304 synapses in total [*Pfeil*, 2015]. First approaches to interconnect several neuromorphic chips were undertaken. The goal was to emulate neuronal networks with a higher neuron count than are available on a single chip [*Friedmann*, 2009]. One drawback was that the bandwidth between the chips was magnitudes lower than on-chip.

The next big leap came with the High-Input Count Analog Neural Network Chip (HICANN) chip and the Wafer-Scale Integration (WSI) system concept within the Fast Analog Computing with Emergent Transient States (FACETS) project [*Husmann and Zoglauer*, 2010]. The HICANN chip contains 512 Adaptive exponential integrate-and-fire Model (Adex) neurons and 112 640 synapses in a 180 nm process [*Schemmel et al.*, 2010a]. Rather than cutting the single chips out of the wafer, the WSI concept keeps the wafer as a whole. To overcome the internal chip borders, which arise from the production process, a new wafer level redistribution layer technology was developed together with the Fraunhofer Institute for Reliability and Microintegration in Berlin. The process called Post-Processing Procedure achieves a very high interconnection density and, therefore, a higher bandwidth between the separated chips.

The 384 HICANN chips sum up to 200 000 neurons and 43 million synapses. The system with all its components is called BrainScaleS. Even tough big neural networks can be emulated there is a big gap between the hardware and biology. In comparison, the brain of a honeybee contains 960 000 neurons and 1×10^9 synapses [*Menzel and Giurfa*, 2001]. To further compensate this gap the BrainScaleS systems can be interconnected to achieve bigger neuronal networks. The current BrainScaleS cluster is shown in fig. 1.1.

There are so far 20 BrainScaleS systems (BrainScaleS systems) installed and operating. The central compute cluster exchanges spikes between the systems. It is planned to send neuronal events directly between each other without the use of the compute cluster.

For the latest neuromorphic hardware chip the High-Input Count Analog Neural Network Chip Digital Learning System (HICANN-DLS) the Electronic Visions Group switched to a 65 nm process. Unfortunately, the neuron size does not scale with the feature size as they are implemented with analog components [*Aamir*, 2017]. Therefore, there is no increase in the number of neurons compared to the HICANN chip. Only the digital parts profit from the new process. A new feature of the HICANN-DLS chip is the Plasiticity Processing Unit (PPU) an embedded processor which allows the implementation of modulated STDP [*Friedmann*, 2013].

Thus, the remaining possibility for increasing the number of neurons and synapses is to build more systems. In chapter 2 it is shown that the current system architecture is not suited for a very large system cluster in the range of hundreds of systems.

This thesis presents the concept of embedding the complete silicon wafer in a

Figure 1.1: The BrainScaleS cluster in the European Neuromorphic Institute (ENI) in Heidelberg. In the first three and last two server racks are the BrainScaleS systems. Each rack offers space for four wafer modules.

PCB as a solution for a very large neuromorphic hardware cluster. It was developed together with the Fraunhofer Institute for Reliability and Microintegration (IZM) in Berlin.

Momentarily, the embedding process is used for small, single silicon chips. Commercially available are embedded switching power regulators from Texas Instruments (TI) which have a very small package size [*Instruments*, 2017]. Another demonstration is a 50 kW power module where several power Metal-Oxide Semiconductor Field-Effect Transistors (MOSFETs) are embedded in a PCB to achieve a better electrical performance [*Neeb et al.*, 2014].

At the beginning there was no experience with the embedding of silicon wafers in PCBs. First tests had to be done at the IZM to see if the silicon wafers withstands the embedding procedure. The positive outcome is shown in chapter 4. These results enabled the development of the prototype DENSE with a fully embedded HICANN wafer, which is described in chapter 5. The prototypes were tested and characterized at the Heidelberg University. In chapter 6 the positive results are presented. It is shown that the wafers are still completely functioning and that the neurons can be calibrated. Furthermore, an accelerated environmental stress test is done with one of the boards and no degradation or failures occurred.

This prototype is not ready to replace the BrainScaleS system system yet, multiple prototypes with additional features and technologies are required. The necessary steps and improvements for the Embedded Wafer System (EWS) are listed in chapter 7. At the end further applications of the embedding process are discussed.

2 BrainScaleS System

This chapter describes the BrainScaleS system which has been developed by the Electronic Visions Group. A fully assembled and operational system, also called wafer module, is shown in fig. 2.1. The section 2.1 covers the core components of a wafer module and continues with the two used interconnecting technologies, wafer-wide redistribution layers and elastomeric connectors. Followed by a description of the assembly process of a system.
The chapter finishes with the experiences obtained during the assembly and operation phase of the first systems and the resulting effects on the scalabilty of the system.

2.1 Components

A wafer module is build from 68 PCBs and 14 mechanical parts, see tables A.1 and A.2 for a detailed list of the components. All parts, except the FPGA Communication PCB (FCP board) and the WIO board, have been developed in the Electronic Visions Group. The FCP board and WIO board have been designed by the Dresden University of Technology. Moreover, the mechanical workshop of the Kirchhoff-Institute for Physics produced all mechanical parts for the modules. Only the heat sinks at the back for the cooling of the HICANN wafer and the Raspberry Pi are commercially available.
The Raspberry Pi controls and monitors the states of all module components, e.g. chip power state, wafer temperatures, FCP board power etc. The modules to digitize the 96 analog signals from the HICANN wafer are called Analogue Readout Modules (AnaRMs) and are not shown in fig. 2.1. The AnaRM is an Flyspi board with an expansion board. On the Flyspi (Flyspi) board is an Field Programmable Gate Array (FPGA) which is used for the host communication and measurement control. The expansion board is equipped with an Analog-to-Digital Converter and connected over a ribbon cable to the Analogue Breakout Board (AnaB Board). In total twelve AnaRMs can be attached to a single wafer module.

Due to the number of components only the silicon wafer, the FCP board and the MainPCB (MainPCB) are described in detail. For more information about the remaining components see [*HBP SP9 partners*, 2014].

2.1.1 Neuromorphic Silicon Wafer

The core of the wafer module is the neuromorphic silicon wafer it contains the circuits emulating the neurons and synapses. The Wafer is produced in a standard 180 nm CMOS process technology and has a diameter of 20 cm. The circuit on the wafer has

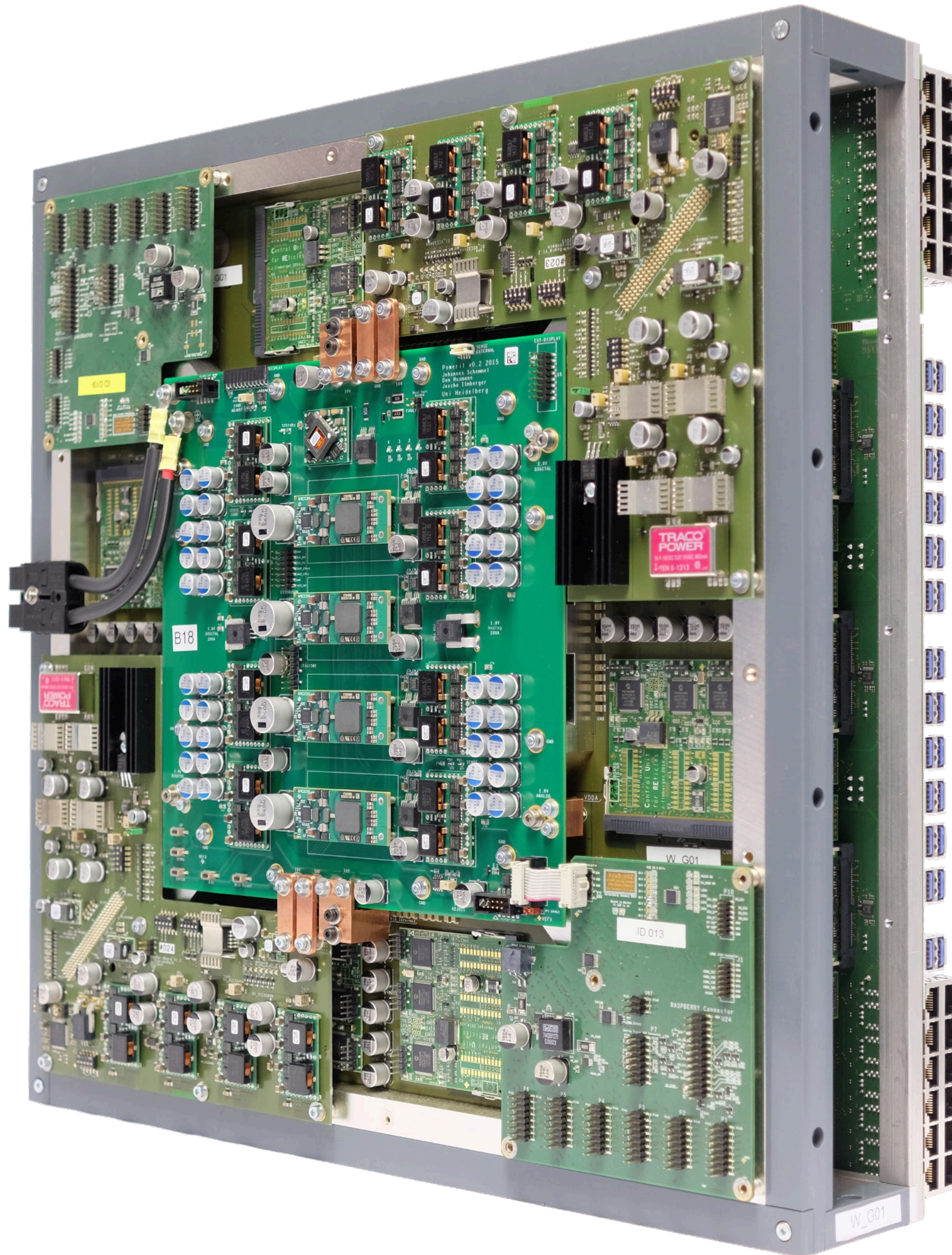


Figure 2.1: A fully assembled BrainScaleS system. At the front are the power supply and Analog Breakout boards. On the right edge are the Gigabit Ethernet and Universal Serial Bus version 3.0 (USB 3.0) sockets of the Wafer IO Boards (WIO boards) visible. The module has a height and length of 50 cm and a width of 15 cm.

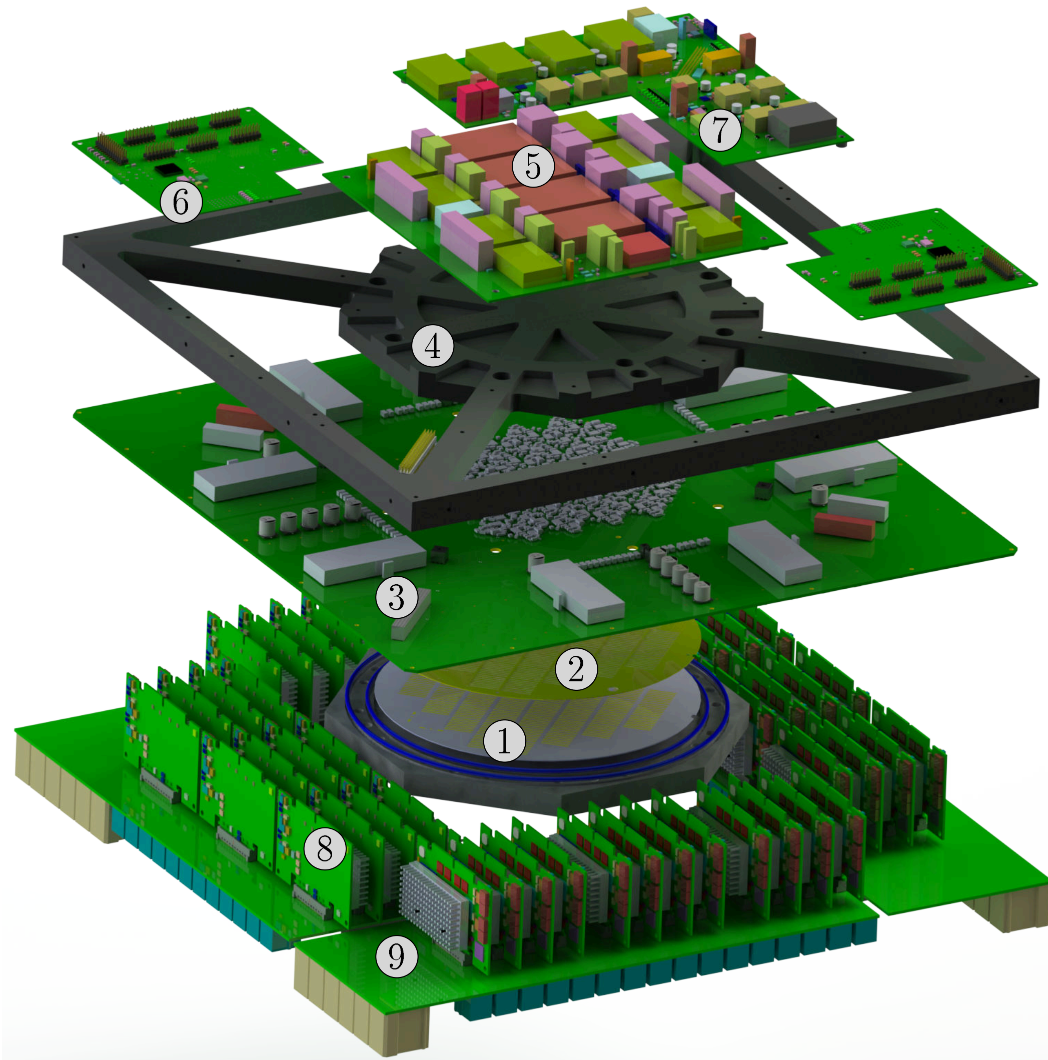


Figure 2.2: Exploded view of the BrainScaleS system. (1) Silicon wafer with the HICANN chips, (2) Positioning mask with the elastomeric strip connectors, (3) MainPCB, (4) Aluminium frame (Top Cover) for mechanical attachment and GND connection, (5) Main Power Supply PCB (PowerIt), (6) AnaB Board, (7) Auxillary Power Supply PCB (AuxPwr PCB), (8) FCP boards, (9) WIO boards. Courtesy of Dan Husmann.

been developed during the FACETS project and has been tested with single chips, which are called HICANN chips. Eight HICANN chips fit into the biggest area which can be produced on a silicon wafer at once. This group of eight is a so-called reticle. The area of a reticle is around $2 \times 2\,\mathrm{cm}^2$. In fig. 2.3 a HICANN wafer with Post-Processing, a reticle and a single HICANN chip are shown. Eventually, a silicon wafer consists of 48 reticles - or 384 HICANNs.

High-Input Count Analog Neural Network Chip Each HICANN chip has 512 neurons with 220 synapses each. In sum there are 2×10^5 neurons and 4.4×10^7 synapses on a single wafer. Furthermore, several adjacent neurons can be combined to form a bigger neuron with more synaptic inputs. The maximal combination is of 64 neurons and, therefore, achieves 14 336 synaptic inputs into one neuron. The implemented neuron circuit is designed to emulate the Adaptive exponential integrate-and-fire Model [*Brette and Gerstner*, 2005]. Its neuron design is purely analog without any clock source. Hence, the neuron runs in a continuous-time mode and consumes only power when a state-change happens, e.g. an external event arrives at the neuron. Due to the analog implementation of the neuron it runs with a speed-up factor of 10^4 compared to biological time [*HBP SP9*, 2015]. Therefore, one second in real-time takes only 0.1 ms on the HICANN chip. This allows to run more experiments in the same time which is from special interest for parameter sweeps. Furthermore, long-term evolutionary processes can be executed in a shorter time, e.g. one year in biology only takes eight hours on the HICANN chip. A detailed description of the neuron implementation is available in [*Millner*, 2012].
In contrast to the analog neuron behavior is the communication between neurons and external hardware components completely digital. An action potential released by the neuron, called spike, is represented by a digital packet. Every event contains the pre-synaptic neuron number. Additionally, a time-stamp is attached to the packet when the destination is not on the silicon wafer [*Schemmel et al.*, 2010b].

By using a 180 nm Complementary Metal-Oxide Semiconductor (CMOS) process the available space for digital circuits is limited. Therefore, only a small control and interface logic is directly implemented in the chip. Most tasks concerning the experiments are sourced out onto the FCP board, see section 2.1.2. For the communication between HICANN and FCP board two channels are available; one is a Joint Test Action Group (JTAG) interface and the other is a Low-Voltage Differential Signaling (LVDS) link. The JTAG-interface connects all eight HICANNs to one chain. It is a simple and robust interface for test and maintenance purposes, however, it is not designed for a high data throughput [*IEEE*, 2001]. The first tests of the HICANN chip use the JTAG interface, later it is only used to initiate the high-speed communication lines. The LVDS interface is designed to achieve a high and fast data transfer [*ANSI/TIA/EIA-644*, 1996]. It is used for the configuration of the HICANN chip and the transfer of spike data from and to the chip.

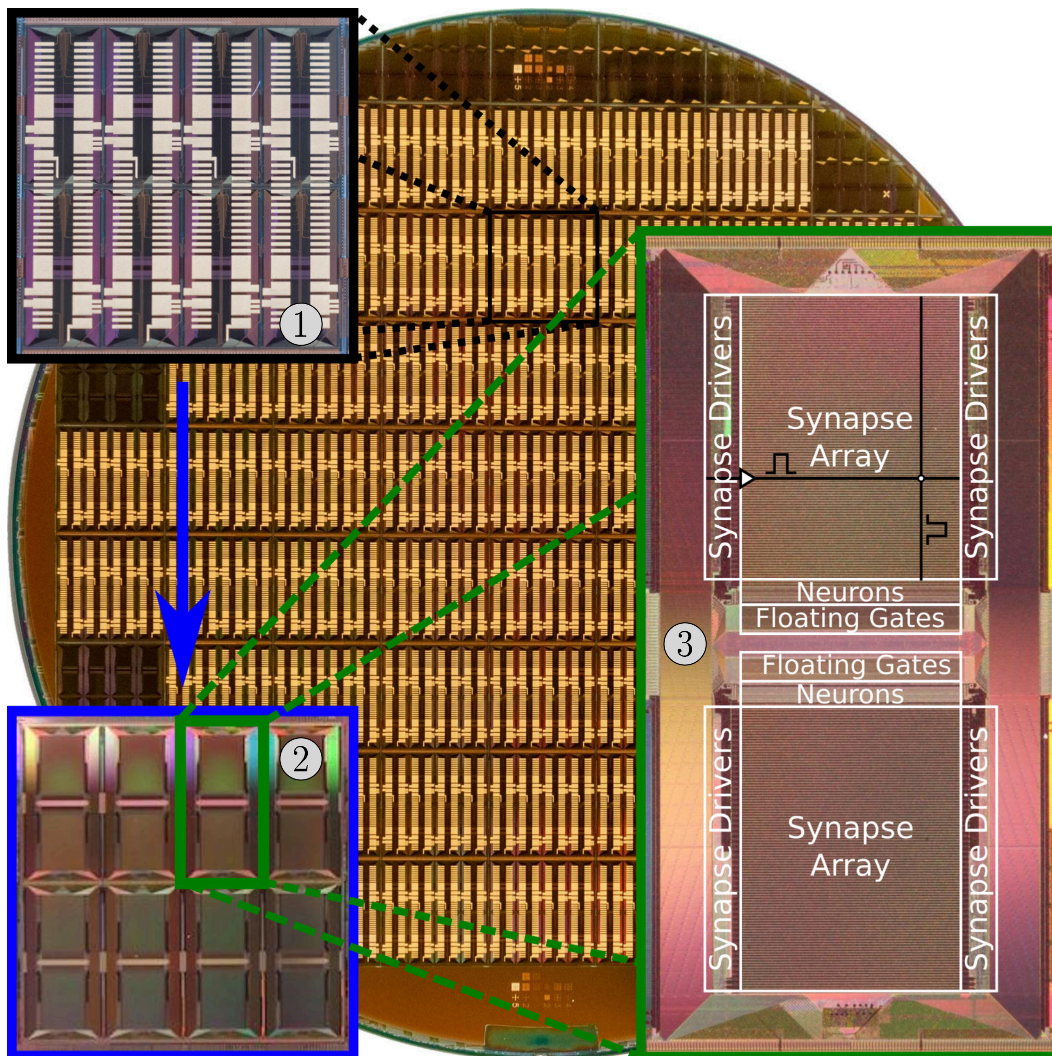


Figure 2.3: A photograph of a HICANN wafer with the Post-Processing structures on top. In the top left enlargement (1) is a reticle with its Post-Processing structures. A reticle has a length of 20.145 mm and height of 20.0482 mm. The bottom left image (2) shows the reticle without Post-Processing layers. The eight distinct HICANN chips are visible. The enlargement (3) on the right side depicts a single HICANN chip with its main components. The size of a HICANN is $5 \times 10\,\mathrm{mm}^2$.

Post-Processing Procedure Unlike the standard wafer procedure, cutting dies out of the silicon wafer, here the wafer is kept as a whole. This requires a new approach for connecting the wafer to external hardware, e.g. to power supplies and communication modules. Therefore, additional wafer-wide RDLs are introduced. The technique is also called Post-Processing. It enables connections between the separated chips on the wafer, so that data can be routed to other chips on the wafer without the interaction of external devices. This on-wafer communication is called Level-1 Communication.

As it is an essential aspect of the system architecture the Post-Processing Procedure is described in detail in section 2.2.1.

2.1.2 The FPGA Communication PCB

The FCP board has four major tasks. First of all it is responsible for the configuration of the HICANN chip, which means setting up the parameters of the neurons and synapses and the network topology. The second task is processing spike data that is stimulating the neural network and, of course, recording of neuron spike activity. Therefore, the FCP board is equipped with 1 GiB of Dynamic random-access memory (DRAM) for storing experiment data.

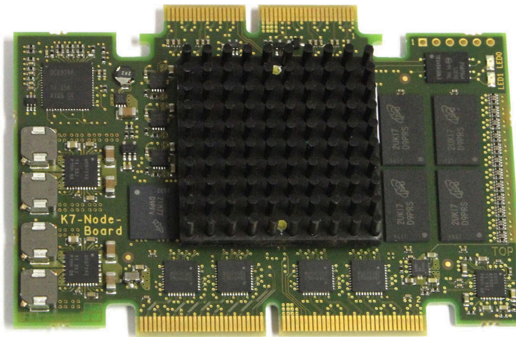


Figure 2.4: Top view of the FCP board. The Xilinix Kintex7 FPGA (Kintex7 FPGA) is located under the black heat sink. Over the lower edge connector the Kintex7 FPGA communicates with the HICANNs chips on the wafer. The upper connector carries the signals for the ethernet connection and the direct communication with adjacent FCP boards.

Additionally, the FCP board is capable of transmitting spike data to other FCP boards or wafer modules (wafer modules).

Eventually, it is responsible for the experiment execution, e.g. starting and ending of experiments.

To process these tasks the FCP board is equipped with an FPGA. The necessity for a FPGA is conditional upon the high data rate of 2 Gbit/s between one HICANN chip and the FCP board. The link's frequency is 1 GHz and capable of full-duplex communication. Furthermore, one FCP board is responsible for eight HICANN chips

at the same time [*Schemmel et al.*, 2010a]. In the worst-case the FCP board has to handle:

$$2\,\text{Gbit/s} * 8 * 2 = 32\,\text{Gbit/s} \tag{2.1}$$

For the data transfer with a conventional computer cluster is a Gigabit Ethernet interface implemented. Additionally, the FCP board has seven interface links to other FCP boards. Three of the links have a fixed destination to adjacent FCP boards, whereas the four remaining links are routed to USB 3.0 connectors on the WIO boards [*HBP SP9 partners*, 2014]. The connectors and cables of the USB 3.0 standard are only used for the physical connection, the USB 3.0 protocol is not implemented. The USB 3.0 connectors allow connections to arbitrary FCP boards or wafer modules.

2.1.3 MainPCB

The MainPCB is the central connection point of the wafer module. It connects the wafer with the FCP boards and is responsible for the power distribution to the HICANN wafer. This requires a lot of space. Therefore, the MainPCB has a length and width of 43 cm.

On the top of the board is the power supply control logic located, see fig. 2.5. Every reticle has twelve independent supply voltages. Two of them, VDDD and VDDA, have to deliver each 4 A in the worst-case, which accumulates for each voltage to 192 A for the whole wafer. These two voltages are the main supply sources for the digital and the analog neuron circuits. Especially, for the analog voltages a voltage drop in the supply chain has to be minimized. For example the neuron behavior directly depends on VDDA. Hence, the maximal allowed voltage difference over all chips is 50 mV. Therefore, ideally thick copper lanes or planes are used to reduce the voltage drop. The same points apply to the return path of the currents. To reduce the length of the GND path the Top Cover is used as GND connection. Therefore, the MainPCB has GND pads in the corners of every reticle on top. The Top Cover has the proper stamps to connect to the pads. In fig. 2.7 the pads and the stamps are shown. This way the voltage drop on the GND path is negligible. The other reticle voltages are not so demanding, they are used for the floating-gate cells, the inter-chip and off-chip communication. The table A.3 shows a complete list of all reticle voltages.

The pads to the silicon wafer (silicon wafer) requires most of the area on the bottom. The remaining space is used by the 48 connectors for the FCP boards. In fig. 2.6 the bottom of the MainPCB is shown. The most power is dissipated on the bottom side. Therefore, the FCP board connectors are aligned into one direction. This way the cool air entering the system from the lower edge can cool the FCP boards and HICANN wafer efficiently.

LVDS Communication Channels For the high speed communication between the wafer and the FCP boards LVDS [*ANSI/TIA/EIA-644*, 1996] links are used. Each

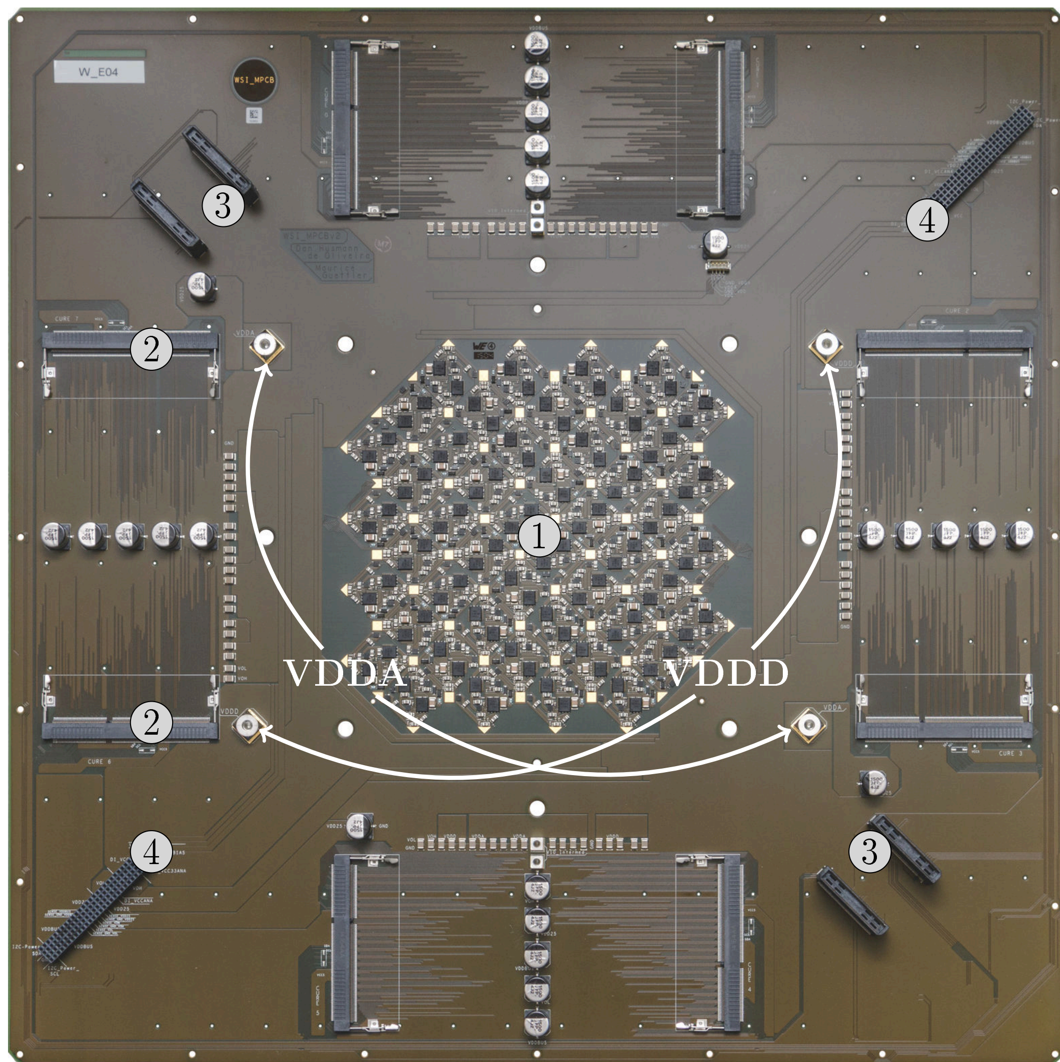


Figure 2.5: A photograph of the top of the MainPCB. The board has a length and width of 43 cm. In the center of the board are the Power Metal-Oxide Semiconductor Field-Effect Transistors (PowerFETs) (1) which switch the power supply of the reticles on the wafer. The PowerFETs are controlled by microcontrollers on the Control Unit of Reticles Boards (Cure Boards). They also measure the supply voltages after the PowerFETs. In case of an overcurrent or an overvoltage the corresponding reticle is turned off. In the top left and bottom right corner are the connectors to the AnaB Board boards (3). The AuxPwr PCBs (4) are located in the bottom left and top right corner. The main supply voltages VDDA and VDDD are generated on the PowerIt and inserted into the MainPCB via press-fit connectors.

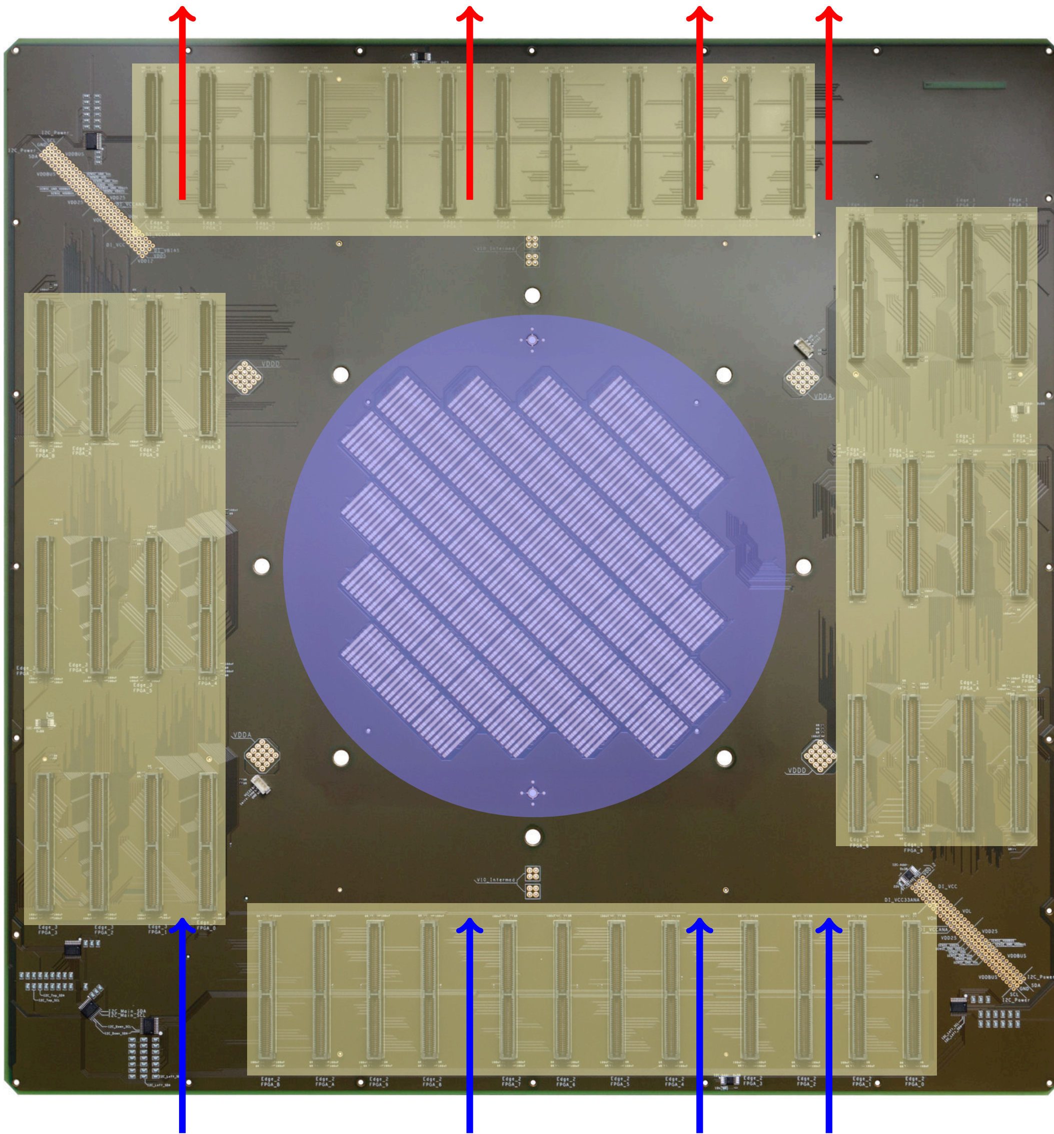


Figure 2.6: Photograph of the bottom of the MainPCB. In the center of the board are the pads for the elastomeric connectors which create the vertical connections to the wafer. The blue area represents the HICANN wafer. At the board edges are the connectors for the FCP boards (yellow areas). The forced air flow is depicted by the arrows. The cool air is inserted from the bottom (blue arrows) and the heated air (red arrows) leaves the system at the top. Therefore, all connectors are oriented in the same direction.

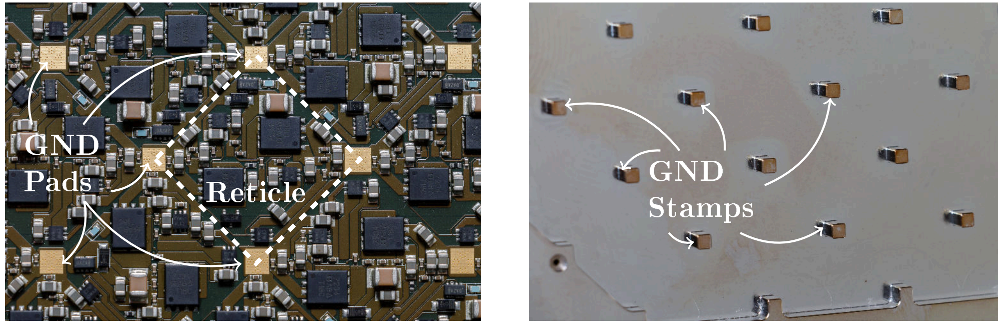


Figure 2.7: The left image is an enlargement of the PowerFET area on top of the MainPCB. On an area of $2 \times 2\,\mathrm{cm}^2$ are six PowerFETs, 24 blocking capacitors and one termination resistor placed. In the corners of the reticles are the GND pads for the stamps of the Top Cover. The stamps of the Top Cover are shown in the right image.

HICANN has one receive and one transmission channel. Every LVDS link needs its own clock, which is another LVDS pair. Additionally, the HICANN needs for the general operation a 50 MHz clock, which is routed only once and distributed at the end to all eight chips. In sum a reticle has 33 LVDS line pairs. Hence, the MainPCB has to route 3168 differential lines from the silicon wafer to the FCP boards.

Analog Measurement Lines There are 96 single-ended lines with a controlled impedance of 50 Ω. Every reticle has two of them, which are shared among all eight HICANN chips. These analog measurement lines allow the user to look at different neuron parameters like membrane voltage or floating-gate values.

Flatness of Contact Area Another requirement to the board is its flatness over the wafer contact area. The connection between the wafer and the MainPCB is described in detail in section 2.2. Here the importance is that the flatness is below 100 µm. Otherwise it is not possible to achieve reliable connections over the complete area.

Stack up The standard board stack ups do not fulfil the requirements for a high signal count and a good power distribution, therefore, the stack up was developed in close cooperation with Würth Elektronik GmbH - Circuit Board Technology. The amount of lines demands the use of microvias, which interconnect two adjacent layers without blocking all other layers. To save routing space the microvias need to be stacked on top of each other. Therefore, the microvias have to be completely filled with copper. Unfortunately, the height of the dielectric layer between the copper layers is limited to 100 µm by the microvias. Therefore, smaller lines are needed to achieve a differential impedance of 100 Ω [*Würth Elektronik*, 2015]. Furthermore, the maximum number of layers with microvia interconnections is limited. Every additional layer with microvias means a complete run through the production line.

During that the board gets heated up and a new copper layer is added under high pressure. This process changes the properties of the former processed layers. They get more and more rigid and the probability for breaks during the compression rises. Thus, the maximum number of heating and pressure cycles is set to five. This allows a reliable production process [*Keller*, 2011]. The figure 2.8 depicts the final stack

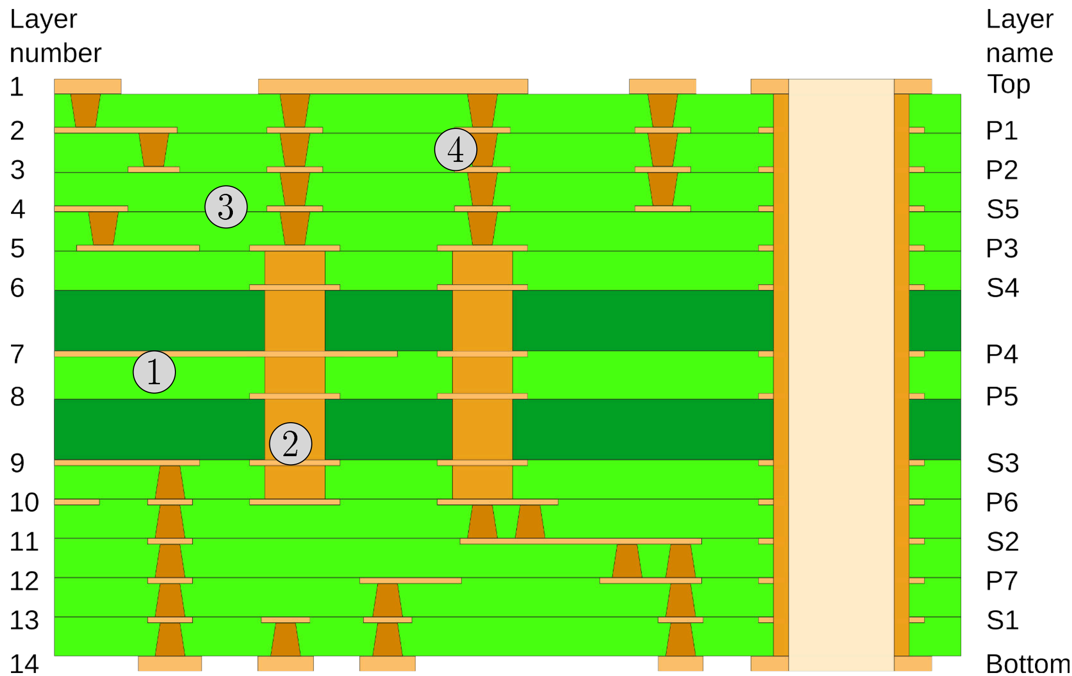


Figure 2.8: The Stack up of the MainPCB. At the beginning the two FR4 (FR4) cores (1) and the layers P3 and P6 are laminated. Then the Buried Vias (2) are produced. In the next steps the preimpregnated bondings (prepregs) layers (3) are added. This is done symmetrically around the inner layers in sequential steps. After every new prepreg layer the microvias (4) to the layer below are added.

up of the MainPCB. The microvia layers are symmetrically build around the two FR4 cores in sequential cycles. At the end there are 14 copper layers. Most of the layers (P1 to P7) are used for the power distribution and as reference layers for the high-speed signal lines. This leaves only five available layers for signal routing.
The line parameters for the differential and single-ended lines have been calculated with the Allegro PCB (Allegro PCB) calculator [*Cadence*, 2014]. A crosscheck calculation was done by Würth Elektronik GmbH - Circuit Board Technology (Würth Elektronik) too. The calculated values are in tables A.4 and A.5.

Due to the complex layout the development of the board took two persons six months. The design was partitioned in smaller modules. The reticles, for example, were single modules and each module had its own layout. In the end the modules were combined in the big MainPCB layout. This allowed the parallel development of the

MainPCB which would otherwise not be possible and prolong the development time. Additionally, the design was automated by the use of SKILL scripts [*Cadence*, 2017b]. The placement of the modules was done by a script. It had to be repeated several times to optimize the layout. This way faster iterations cycles were possible. Another useful script calculated the intersection points of lines which were on different layers and placed vias according to the design constraints. This was used to connected the lines from the reticles with the FCP board connectors which were in total 3 220 lines. The script finished in minutes what would have taken a person at least a week.

Unfortunately, not all chips of the HICANN wafer are connected over their LVDS links to their corresponding FCP board boards. The two reticles in the middle of the wafer are only accessible via the JTAG interface. These chips cannot be used in neuronal networks for experiments but at least the switches for the on-wafer communication are configurable. Furthermore, no routing capabilities are lost, also the mapping process has not to find redirections around the chips.

Wafer Power Supply The supply voltages for the silicon wafer are generated on the PowerIt and the AuxPwr PCBs. The PowerIt was developed by Dr. Johannes Schemmel and the AuxPwr PCB by Lars Sterzenbach [*Sterzenbach*, 2015]. It is possible to switch every reticle on the wafer individually by PowerFETs in the supply lines. The PowerFETs are placed on top of the MainPCB. In fig. 2.9 the supply scheme of a single reticle is depicted. If there is a defect on the silicon wafer, the PowerFETs cut off this chip area from the supply voltages. This prevents other chips from getting influenced or even destroyed by defects in other areas. The PowerFETs are controlled by microcontrollers, which are placed on the Cure Boards [*Ilmberg*, 2014]. To detect a fault in the power supply, the microcontrollers have measurement lines going to every input and output pin of the PowerFETs on the MainPCB. In total there are $48 * (2 * 12) = 1152$ voltage measurement lines. At the moment the microcontroller digitizes only the voltages after the PowerFET, due to routing issues on the Cure Board board. Nevertheless, a reliable monitoring is possible.

Production The size and the stack up made it difficult to find a manufacturer who could produce the PCB. Eventually, only the PCB manufacturer Würth Elektronik succeeded in the production of faultless boards. Necessary to that end was a special treatment of the project at Würth Elektronik production plant. Therefore, the board run through the production line with a special marking and was checked more often than a standard PCB to see problems as early as possible in the production line. The special measures extend the production time to six weeks for a single board. This was not a problem because the goal was to receive faultless boards and not the achieving of a high PCB throughput.

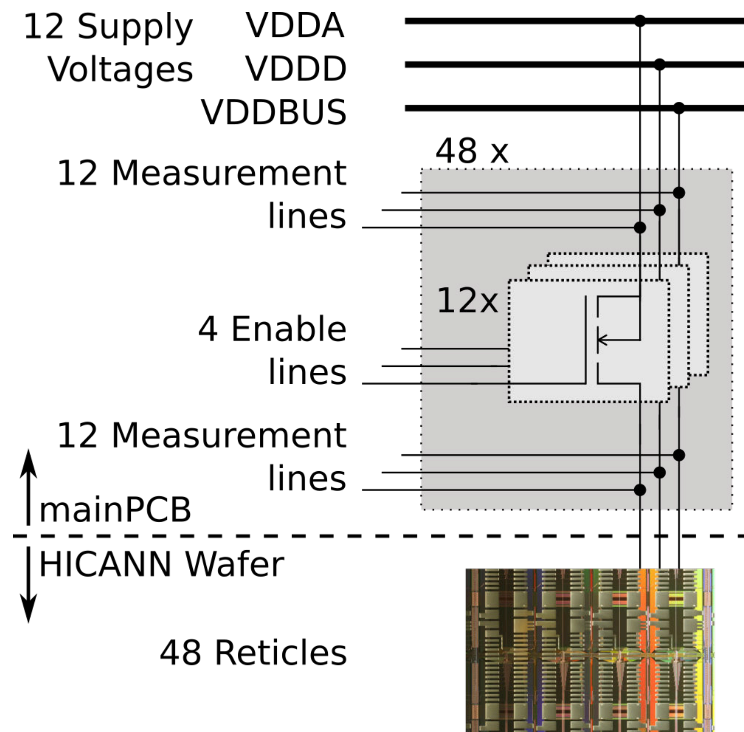


Figure 2.9: The image shows the power supply of a single reticle on the wafer. There are PowerFETs in the each of the supply lines to one reticle. This way every reticle is individually switchable. The voltages are divided into four groups and controlled by the enable lines. Additionally, there are measurement lines before and after every PowerFETs going to the Cure Boards.

2.2 Wafer Connection Technique

There are several connection techniques for standard-sized silicon chips like Wedge Bonding and Flip-Chip-Bonding. Unfortunately, these techniques are not feasible for connecting a whole silicon wafer. For example Flip-Chip-Bonding is soldering process where solder bumps are placed on the silicon surface and then resoldered on a substrate. After the soldering all joints are fixed and it is not possible to repair broken ones. Only by detaching the wafer and reballing it, the soldering process can be redone. This induces thermal stress to the silicon wafer each time and the possibility for errors increases.
The Wedge Bonding is the standard process for creating connections to silicon chips. Wires are drawn from the chip pads to other pads on a substrate which are mostly PCBs. The diameter of the wires is in the range of 25 µm. Therefore, the connections between chip and PCB are very fragile and have to be molded into a compound. Like for the Flip-Chip Bonding the connection between the wafer and the PCB is fix. A

new silicon wafer always needs new PCBs.
At the time of the assessment of the connection techniques the requirements to the methods were; an easy construction and a stress-free replaceability of the silicon wafer. Both methods do not satisfy these requirements.
In the FACETS project another connection technique using elastomeric connectors has been developed and evaluated. The elastomeric connectors are a reliable and well-known method to connect liquid-crystal displays (LCDs) with PCBs. To use this technique with a silicon wafer, the wafer needs additional processing steps. In the subsequent BrainScaleS (BrainScaleS) project the final system with these connectors started its operation.

2.2.1 Wafer Post Processing

The development of Post-Processing procedure started within the FACETS project [*Husmann and Zoglauer*, 2010]. It emerged from the collaboration of the IZM and the Electronic Visions Group. The objectives were to create horizontal links between the separated chips on a silicon wafer and provide the mating structures to connect to the MainPCB through the elastomeric strip connectors.
The advantage of the redistribution layers is that it uses lithography techniques from chip production, therefore, the horizontal links can have a much higher density. Currently, the width and spacing are 4 µm. Enlargement views of these lines are shown in fig. 2.10. On a PCB the state-of-the-art technology achieves line widths and spaces of 40 µm [*AT&S*, 2017]. Staying on the wafer also has advantages for the signal transmission. The distance between transmitter and receiver is shorter, therefore, the lines do not need to match a certain line impedance [*Johnson and Graham*, 2011]. Furthermore, a shorter trace has a lower line capacitance and, thus, the transceiver needs less energy to switch the logic state of the line.

The redistribution consists of two copper layers. For the maximum reliability it is recommended not to mix different structure scales. Hence, the layer direct above the silicon wafer is used for the fine-pitch connections across chip borders. It allows to route 1280 lines over the vertical and 2048 lines over the horizontal edges of the reticle to the next neighbor. The coarser pad structures are placed on the second RDL. The total height of the redistribution layer is 20 µm [*Schemmel et al.*, 2010b].

Post-Processing Layer Stack Up The Post-Processing stack up is depicted in fig. 2.11.
At first a layer of Benzocyclobutene (BCB) is applied onto the silicon wafer, which is etched away over the passivation openings. The openings have a diameter of 5 µm and a depth of 10 µm [*HBP SP9 partners*, 2014]. On the BCB layer the fine-pitch connections for the connections across chip borders are created, to connect to these lines the holes to the silicon wafer get completely filled with copper. On chip routing is possible, in contrast to the fine-pitch lines the copper height is increased. On top of the first routing layer comes again a BCB layer as separation material to the next layer. The second layer is used for the coarser pad structures for the connection to

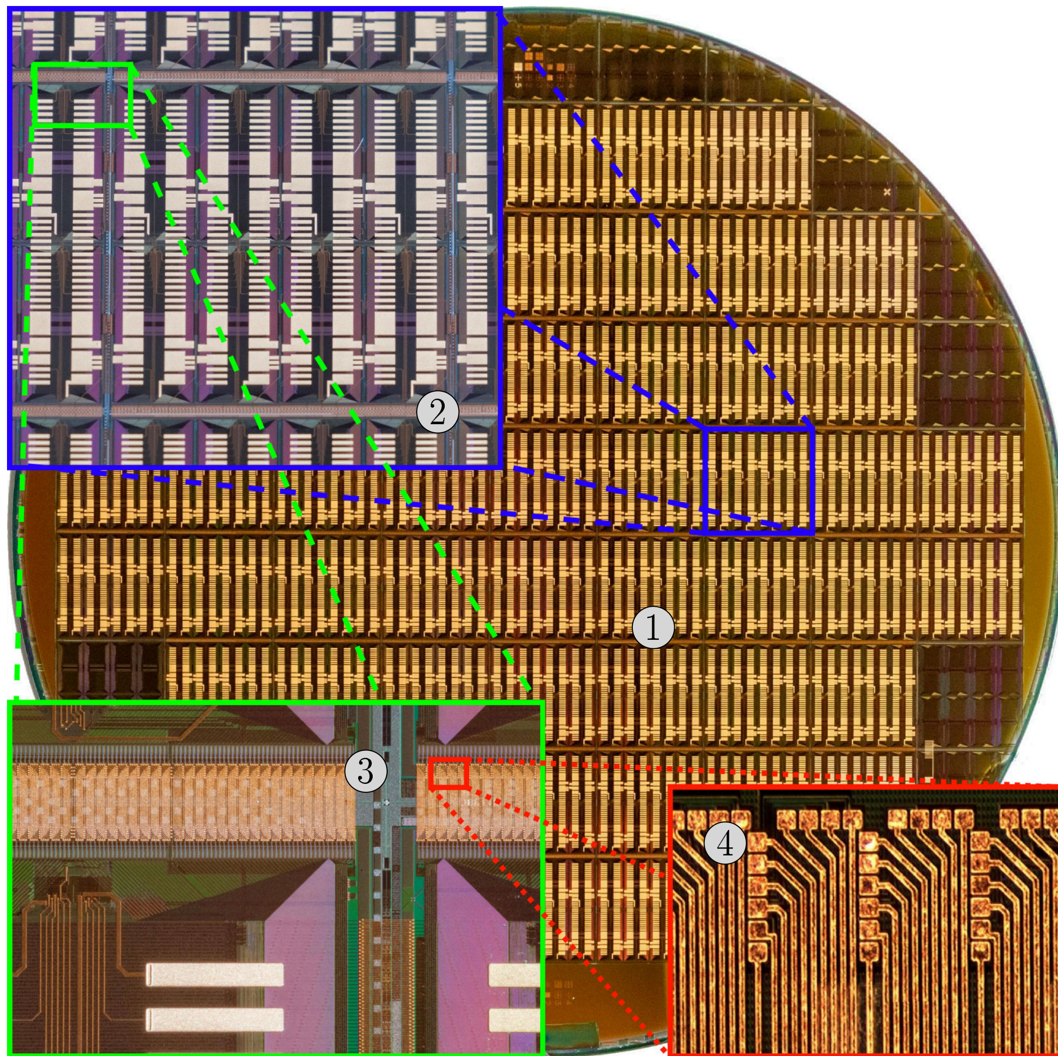


Figure 2.10: Enlargement views of the Post-Processing structures on the wafer. In the background (1) is the HICANN wafer. The top left enlargement view (2) shows the RDL pads of a reticle. In the bottom left photograph (3) the corner of a reticle is shown The line density on the vertical edge is higher than on the horizontal edge. In the vertical direction 2048 lines cross the edge, whereas, the horizontal edge 1280 lines cross. The vertical fine-pitch connections between two chips are visible in (4).

the MainPCB. The copper of the pads is covered with a small layer of gold. This prevents the oxidation of the pads and creates a better electrical connection with the elastomeric connectors [*Zoschke et al.*, 2017].

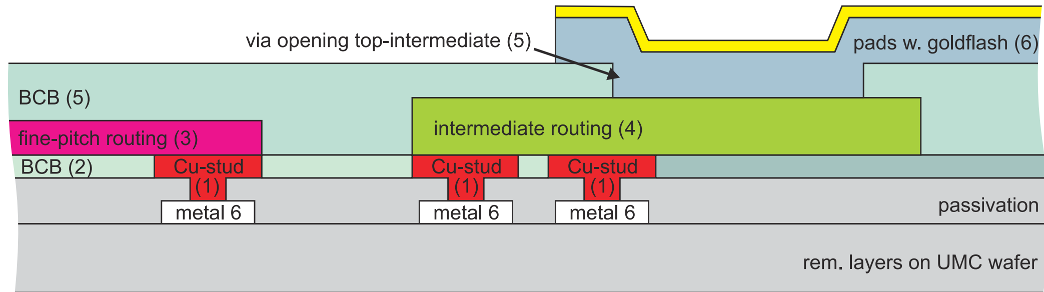


Figure 2.11: Stack up of the redistribution layer. At first the copper studs (1) are created which connect the metall-6 pads to the first routing layer. On the routing layer are two types of copper lines. There are the fine-pitch lines (3) with a line width of 4 µm used as interconnections between adjacent reticles. The second type are intermediate routes with a larger copper thickness. On this thicker copper lines vias (5) connect through the second BCB layer (5) from the top. The pads on top are covered with a goldflash to reduce the contact resistance and stop oxidation of the copper. Courtesy of Kai Zoschke (IZM).

Reticle Pad Design For the different signal requirements customized pad sizes are required. The pads for the power supply with a high current capacity need to be larger than the pads for the high-speed signals with only a small current. Therefore, the pads for data signals and the supply voltages with a small current capacity have the size of $1200 \times 200\,\mu m^2$. For voltages with a maximum current of 500 mA the pads are $1200 \times 1400\,\mu m^2$.

Two vertical HICANNs use a combined Post-Processing layout. In fig. 2.12 the pinout of two HICANNs is shown. They share certain signals for simplification like the clock TCK of the JTAG interface. Additionally, a few connections from one chip to the other are routed in the first RDL. For example the JTAG output signal TDO from the upper HICANN goes to the JTAG input TDI of the lower HICANN. The pin layout was developed by Dr. Andreas Grübl.

2.2.2 Elastomeric Connectors

The vertical connection between the silicon wafer and MainPCB is established via the ZEBRA connectors. These elastomeric strip connectors consist of conductive and non-conductive layers. The layers have a pitch of 100 µm and, therefore, allow a high signal routing density with a very low resistance [*Fujipoly*]. The figure 2.13 depicts the buildup of a connector. For the purpose in the WSI system the connectors are

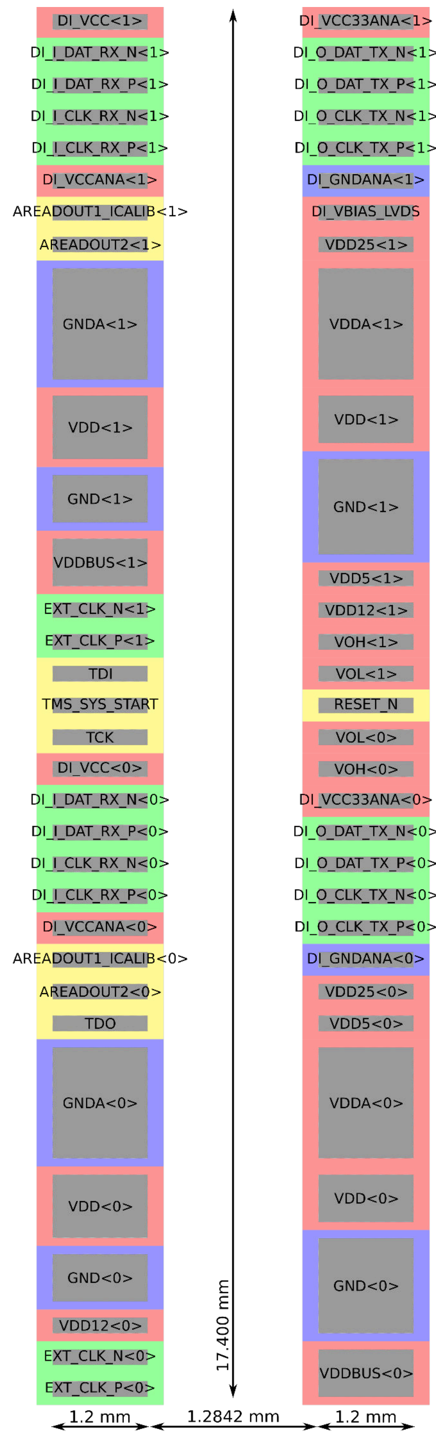


Figure 2.12: Pinout of the top redistribution layer for two HICANNs. Signal names with a '1' belong to the top HICANN, '0' corresponds to the bottom one. The red pads are supply voltages, blue the GND pads, yellow the single-ended lines and the green the high-speed links.

18 mm long, 1.04 mm wide and 1 mm high. Thus, the width of connectors is smaller

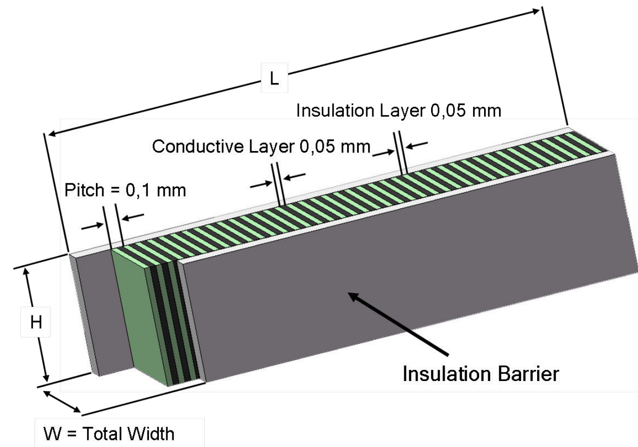


Figure 2.13: Schematic image of an elastomer connector. The conductive and insulation layers are 50 µm wide. A vertical connection is established when the elastomer connector is compressed to 80 % of its height. The dimension of an elastomer connector is L=18 mm, W=1.04 mm and H=1 mm.

than the actual pads on the silicon wafer and the MainPCB. Moreover, the connectors are longer than a column of Post-Processing pads. This way the alignment clearance is increased. The pads on the silicon wafer and MainPCB are bigger than a single conductive layer in the stripe connector. Therefore, no special alignment for the elastormeric connectors is needed. The smallest pad on the wafer and PCB is 200 µm wide. This guarantees that at least one complete conductive layer is on the pad.
There are small silver balls included in the conductive layers. These silver balls form a vertical connection when a force is applied in the z-direction. The working principle is depicted in fig. 2.14.
In the diploma-thesis of Holger Zoglauer the elastomeric connectors were tested extensively for the application in the WSI system [*Zoglauer*, 2009]. It has been shown that single strips have a high current capacity and can be used for the power supply connections.

So far the elastomeric connectors showed no aging process in the systems. Once they are installed in a system, they work flawlessly. However, if the module is reopened after a longer period, the connectors are not used anymore and replaced with new ones. The compressed connectors are not as flexible as at the beginning. They do not recover to the full height of 1 mm after the force is removed. It is unknown how a mix of new and old connectors would behave and if it would work. Therefore, it is easier to use new elastomeric connectors with a well-known behavior.

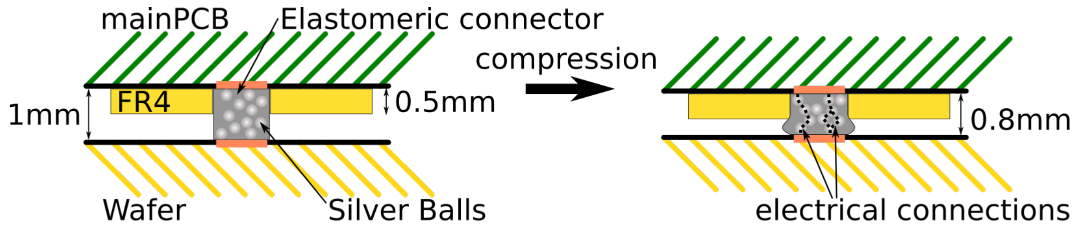


Figure 2.14: Working principle of an elastomeric connector. In the connector are small silver balls embedded. At first there is no electrical connection between the MainPCB and the silicon wafer. Only after squeezing the connector to 80 % of its original height electrical connection are created.

2.2.3 FR4 Positioning Mask

On the silicon wafer two HICANNs share two elastomeric connectors. For the complete wafer this adds up to 384 elastomer connectors. This amount of connectors needs some sort of positioning mask. Otherwise their position could change during the assembly process, especially, during the compression phase. Therefore, a positioning mask is cut out of an FR4 sheet. The disk has 384 slots to hold the elastomeric connectors in place. The manufacturing of the mask is done by the mechanics workshop of the Kirchhoff-Institute for Physics.
As the elastomer connectors get compressed to 80 % of their height, the disk thickness has to be below this. Hence, the thickness of the positioning mask is 0.5 mm and leaves enough space for the compression of the elastomeric connectors. The material of the positioning mask is FR4 because of its good working properties. The FR4 material is stable during the milling process. Especially, the strip between two elastomeric connectors is a weak point of the mask as it is only 1.5 mm wide. Here the glass-fibers play an important role. The fibers make the mask inherently stable, so that it is flexible but also rigid enough for the small strips not to break up. There are other materials with the same properties, like Teflon. These materials are more expensive than FR4 and, therefore, not used.

2.3 Construction of a BrainScaleS System

The assembly of a wafer module starts in the clean room, where the mating of the silicon wafer and the MainPCB is done.

At first a positioning mask has to be filled with 384 elastomeric connectors. Each connector needs to be inspected optically for defects. The common errors are wrong cuts or broken patterns of conductive and insulating layers. The examples of defects are shown in fig. 2.15.

The filling of one mask takes around two hours for a semiskilled person. Next the mask is screwed to the bottom of the MainPCB. In parallel the Wafer Bracket is fixated to the adjustment device and the silicon wafer is laid into it. Now both get united, therefore, the MainPCB with the positioning mask is flipped over and placed

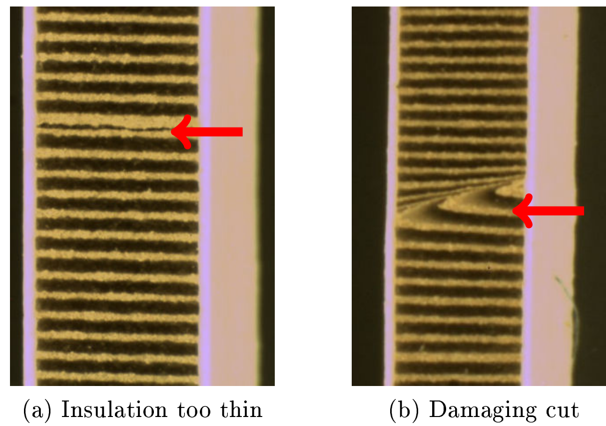


Figure 2.15: Two generic images showing defect elastomeric connectors. In (a) the connector has an insulation layer which is too thin. In (b) the connector was not cut properly and ruined the layer structure. Using a connector like (b) would create shorts between adjacent pads.

over the silicon wafer.

There are two fiducial marks on the wafer and crosses of copper wires on the MainPCB for the proper alignment, see fig. 2.16. The adjustment device can move the Top Cover with the MainPCB in x- and y-direction. If there is an angular error between them, it is possible to rotate the Wafer Bracket (Wafer Bracket) with the wafer. The objective is to place the crosses of the copper wires over the wafer marks.

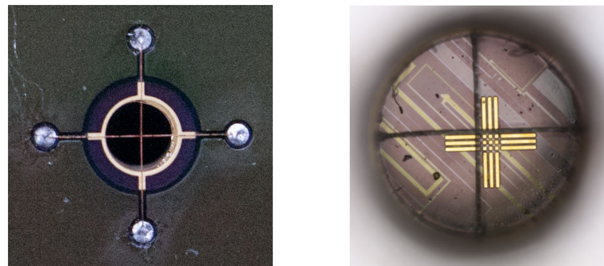


Figure 2.16: The left image shows the wires on the bottom of the MainPCB. In the right image the view through the Top Cover on the HICANN wafer is shown. The cross of the MainPCB should be above the center of the fiducial mark. The squares on the wafer are $50 \times 50\,\mu\text{m}^2$.

When it is achieved the screws are inserted into the Top Cover and tightened. In the process of tightening it is important to lower the Top Cover as even as possible. Therefore, the screws get fixated in small steps and at the same time the crosses are checked to stay in position.

Additionally, there are test boards developed by Sabanci-University and Heidelberg

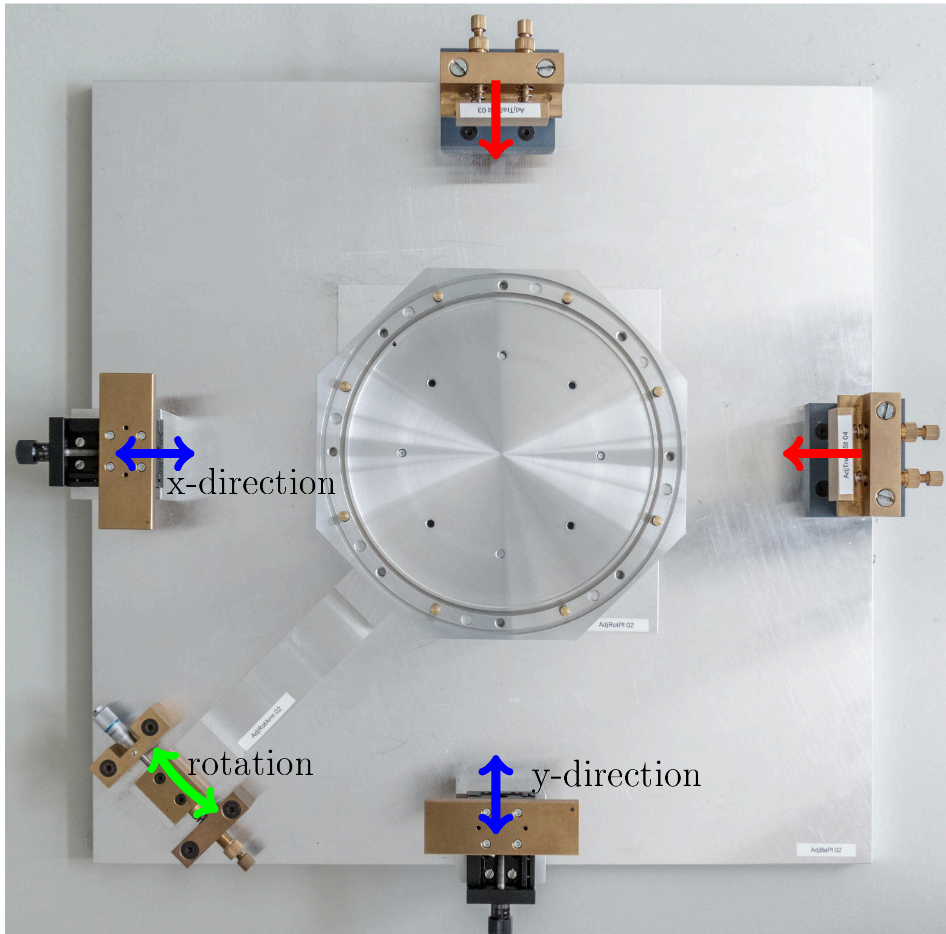


Figure 2.17: The Adjustment device in the clean room. In the center is the wafer bracket which holds the HICANN wafer. The top and right block constantly push against the Top Cover (red arrows). Thus, the position of the Top Cover together with the MainPCB is determined by the position of the other blocks at the left and lower edge. They can be shifted with micrometer screws (blue arrows). A rotation error can be compensated with the micrometer screw in the lower left corner (green arrow) which rotates the silicon wafer.

University, which check if there is an electrical connection between the silicon wafer and the MainPCB. The boards are attached into the FCP board connectors on the MainPCB. They measure every termination resistance of the LVDS lines on the wafer and the diode forward voltage of the Electrostatic Discharge (ESD) structures in the single-ended lines, e.g. JTAG lines. The LVDS lines should have a resistance of around 120 Ω which mainly comes from the termination resistor between the p- and n-line on the wafer. The elastormeric connector and the copper lines on the MainPCB have a negligible resistance compared to the 120 Ω. A forward voltage of 0.6 V at 2 mA is correct for the ESD diodes. The result of a test run is shown in

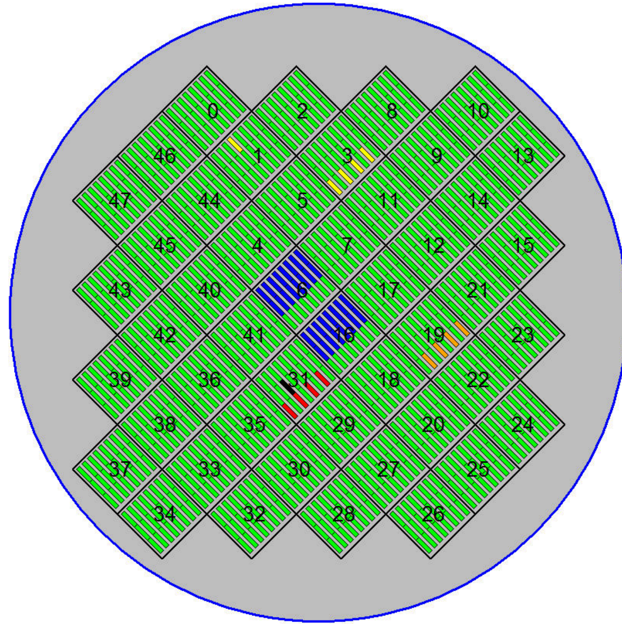


Figure 2.18: The result of a connection test. The visualization shows a representation of the wafer and depicts the location of errors by colors. Every elastomeric connector is divided into three parts. The two upper boxes represent the termination resistances of the receiver and transmission lines. For example the blue boxes in reticle 6 and 16 show that there is no high-speed connection available. The lower box is assigned to the diode tests. The reticles 19 and 31, for example, have the problems with the JTAG lines. A green box means that there were no errors. The other colors code the amount of errors, starting from one error(yellow) to not working at all(black). Courtesy of Dan Husmann.

fig. 2.18. The software controlling of the boards and the visualization was written by Dan Husmann. If the tests are sufficiently well, then the final step can be done. Otherwise, if there are too many errors, the system has to be reopened. Depending on the problem it may necessary to conduct former tasks again, e.g. the alignment procedure. It may also possible that single elastomeric connectors show problems

which were not recognized during the optical inspection and have to be exchanged. When the test returns a positive result, then the final step is to replace the air between the silicon wafer and the MainPCB with nitrogen. The nitrogen stops a possible oxidation of the elastomeric connectors with their silver balls. Furthermore, without the air there is no condensation of water drops which could destroy the electronic circuits.

2.4 Scalability of the BrainScaleS System

It is intended to increase the number of neuromorphic hardware systems within the Human Brain Project [*HBP*, 2015]. The question that arises is: Is it feasible to upscale the current system architecture?
New system aspects become more important that were not for the current system size.

Knowledge Transfer and Distribution For a fast assembly of the desired system size it is desirable to distribute the knowledge of the assembly process over several persons. This way tasks can run in parallel and if one falls ill, the whole process is not blocked.

For the assembly of the 20 BrainScaleS systems a collaboration with the Sabanci-University (Sabanci-University) in Turkey was started. The intention was to ship all components to the group in Turkey, they assemble the systems locally and send them back ready for use. Unfortunately, a correct assembly could not be achieved after six months and the collaboration ended.

Eventually, three student assistants from the Heidelberg University were hired and supervised by Dan Husmann. Their task was the assembly of the 20 BrainScaleS systems in the clean room as described in section 2.3. Only six systems were assembled by the students. The reason for that was the high number of different steps and their complexity. It took a lot more supervised training than thought of to gain the required experience to assemble a single system by themselves.

Process Control Once installed in a cabinet rack only small changes to the system are possible. Later exchanges of components result in a time consuming disassembly which should be avoided. Therefore, a continuous process control is mandatory. Furthermore, these tests need to be easy applicable and have a comprehensible output for an unskilled worker. Such a test was developed by Christian Mauch, it tests the communication between the FCP boards and the silicon wafer [*Mauch*, 2016].

Manufacturing and Assembly Time The most complex mechanical part of the system is the Top Cover. Its production takes 18 h of machine time [*Bing*, 2017]. The work time for 50 systems is 900 h or 37.5 d if one CNC machine runs 24 h a day. Then the Top Cover needs a silver electroplating which is done by an external company. This adds at least three additional work days before it is ready to be

assembled with the other components to one wafer module.
The complete assembly of a single wafer module takes around one day including all tests. Thus, for the target size of 50 systems it would take at least 50 work days for the completion. One third of the time is used for the filling of the positioning masks with the elastomeric connectors. The other two third of the time are occupied with the assembly of the wafers with the MainPCBs which is the most difficult step. It can only be done by skilled personal.

Currently, there is no way to automate the necessary tasks and reduce the assembly time. The only improvement would be a simplification of the process and hand it over to unskilled personal. In the best case an another wafer contacting technique replaces the time consuming steps by an automated process which works without any special interventions of personal.

Space Requirements Obviously, more systems need more space. Currently, 20 systems occupy one standard 20ft ISO-container. They are mounted in five cabinet racks, each holding four wafer modules. To reach the smaller target value of 50 modules 3 containers would be necessary. For a target value of 500 modules 30 containers would be needed which is too much.

At the moment there are no direct interconnections between the systems. Only via the compute cluster neuronal networks on different wafer modules can exchange spikes. This induces some latency in the communication which is not wanted for neuronal networks running in a time-continuous mode. Later it is intended that spikes are routed directly between the modules having only the physical latency of the connecting wire. So, additionally, the distance between the modules gets important and should be as short as possible.

2.5 Conclusion

The Wafer-Scale Integration approach is a feasible method to create high density interconnections between separated chips on a silicon substrate. So far no problems occurred which were related to the Post-Processing process. Furthermore, the elastomeric connectors as a full wafer connection technique establish reliable connections between HICANN wafer and MainPCB.

Nevertheless, one wafer module occupies a volume of 37 L. This is not a problem for a single and isolated prototype system. With an increased system number and a high connection density between the modules the current system architecture has several drawbacks. Therefore, a new system design is required which is smaller, easier and scalable to achieve a neuromorphic hardware cluster with hundreds of systems.

3 Embedding a Silicon Wafer

The embedding of chips is an ever growing market, especially, the smartphone and automotive industries demand smaller packages [*Vardaman*, 2017]. There are several embedding concepts available on the market. They can be divided into two main categories by the materials used in the embedding process.

The first category places the chips on a carrier and over-molds them with a compound mass. In general it is referred as Fan-Out Wafer Level Packaging (FOWLP), as the final packages are bigger than the actual embedded chip. The manufactures invented their own process names, e.g. Infineon calls its process embedded Wafer Level Ball Grid Array (eWLB) and Taiwan Semiconductor Manufacturing Company (TSMC) names it Integrated Fan-Out Wafer Level Packaging (InFO-WLP) [*Meyer et al.*, 2008; *Tseng et al.*, 2016]. In Apple's iPhone7, for example, the InFO-WLP combines the A10 processor and the DRAM to a very thin Package-on-Package (PoP) [*TechInsights*, 2017].

The other approach takes PCB laminates as substrates. In contrast to the FOWLP this process is located at the PCB manufacturer's site. Here also every company has its own process name. At Tokyo Denkikagaku Kogyo (TDK) it is called Semicondutor Embedded in Substrate (SESUB), Austria Technologie & Systemtechnik (AT&S) offers a process called Embedded Components Packaging (ECP) and also the IZM developed a embedding process with organic laminates [*Fujioka*, 2015; *Kriechbaum et al.*, 2011; *Boettcher et al.*, 2008]. The ECP technology is used for small power regulators from TI [*Instruments*, 2017]. The SESUB procedure is employed, for example, in TDK's bluetooth modules [*TDK*, 2017].

All the commercially available products only embed small chips regardless of the technology. There are no experiences or rules for the embedding of complete silicon wafers. In this thesis only the process of the IZM is used and evaluated as a successor for the Wafer-Scale Integration concept of the Electronic Visions Group.

3.1 Embedding in a PCB

The procedure of embedding a silicon wafer was developed together with the IZM in Berlin. The main concept is to laminate the silicon wafer into a PCB and use standard PCB materials and PCB manufacturing techniques in the procedure.

Before the silicon wafer can be embedded it has to undergo certain preparations like thinning and copper application. These topics are covered in more detail in the sections 3.3.1 and 3.3.2. After the preparation steps the silicon wafer is inserted at the beginning of the board manufacturing.

It is placed in a sandwich structure between two prepreg layers, so that it builds a symmetrical stack up. This is in contrast to the general embedding procedures where the chip is bonded to a substrate or copper foil. Here the wafer is hold in position only by an additional frame around it. The frame and the prepreg layers are common PCB materials. At last thin copper foils are placed on top and on bottom of the buildup. The stack up is depicted in fig. 3.1.

Figure 3.1: Generic stack up of the embedding process. The PCB layers are symmetrical added around the silicon wafer which is placed in the center of the stack up.

The next step is the actual lamination in the PCB lamination press. For protection and later easy removal press pads are attached on each side.

The profile of the temperature and pressure in the lamination press is depicted in fig. 3.2. At first the residual air is pumped out of the stack up. This way no air bubbles get stuck inside the board. If the air stays inside the board the signal integrity of the lines is influenced as the surrounding material properties change. Furthermore, during the lamination process air bubbles can destroy a board as the volume increase of the air is much higher than the expansion of the solid materials in the PCB.

Next a low pressure is applied and simultaneously the temperature is slowly increased. Before the resin of the prepreg changes its viscosity and enters a glassy state a higher pressure is set. This ensures that the resin later is distributed equally in the package. The temperature rise is stopped above the glass transition temperature T_g which is around 200 °C and kept for 2-3 h. At this point the resin has a low viscosity and fills every hole, e.g. the small gap between the silicon wafer and the frame. In parts the resin also flows out of the board. Due to the high pressure and the resin loss the board gets thinner. The high pressure is reduced back to the lower value after 30 min. When the viscosity of the resin is high again the pressure is completely removed.

From this point on the board is handled like every other PCB. The integration of vias is done by lasers or drills and the copper patterning is conducted in typical electroplating and etching processes.

3.2 Challenges in the Material Selection

In this section the two major challenges in embedding a silicon wafer are addressed.

The first subject is the material selection for the stack up. The materials play an important role in this process. The lamination of the wafer creates a strong coupling

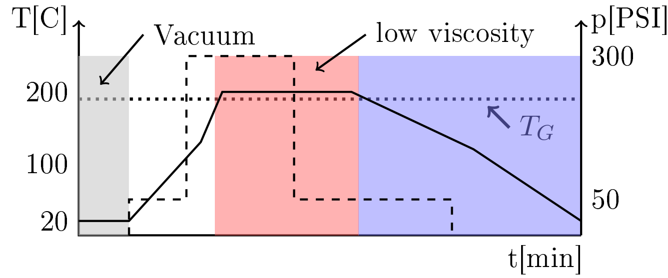


Figure 3.2: Schematic procedure of a lamination run. The solid line is the temperature profile, the dashed line represents the pressure profile. A vacuum is applied so that no air bubbles are in the stack up. In the temperature range between 180-195 °C (red area) the resin has a low viscosity. The temperature is kept above the glass transition temperature (T_g) for 2-3 h. It flows into holes and creates connections to the surrounding materials. In the following cooling phase (blue area) the resin returns to its rigid state and the layers are glued together. Procedure and values are taken from the ISOLA FR408HR proceeding datasheet [*Isola*, 2017]. For other materials the numbers can differ but in general the process is the same.

between the silicon and its surrounding materials. Therefore, the properties of the different materials ideally should be the same. Especially, the thermal expansion is a critical property. If the coefficients of thermal expansion differ to much, a temperature change induces stress to joints between materials. In PCBs this can lead to a breaking of vias [*Nowak et al.*, 2011]. Therefore, the reliability of embedded components is subject to research [*Schwerz et al.*, 2014].

The second subject is the heat transfer from the silicon wafer to the surface. It is related to the material selection as the heat source of the system; the silicon wafer; sits in the center of the board and the heat dissipation depends on material properties. Thus, a good heat transport to the outside also reduces the internal stress induced by the different coefficients of thermal expansion.

3.2.1 Material Property Coefficient of Thermal Expansion

The coefficient of thermal expansion is an important material property. A high mismatch between the coefficients of thermal expansion leads to a high warpage of the board after the lamination process [*Chen et al.*, 2016]. If the warpage is too high, further lithography steps like copper patterning could be impossible. Additionally, the mismatch induces a high internal stress with forces above the tensile strength which can break traces and vias.

The coefficient of thermal expansion (CTE) is defined as

$$\alpha = \frac{L(T_C) - L(0)}{L(0) * T_C} \tag{3.1}$$

$$= \frac{\Delta L}{L(0) * T_C} \tag{3.2}$$

Thus, the relative length expansion is

$$\Delta L = \alpha * L(0) * T_C \tag{3.3}$$

whereas $L(0)$ is the initial length and T_C the temperature difference.
In fig. 3.3 a generalized thermal lifecycle of a board is depicted. The highest temperature difference is reached during the manufacturing process, which is about 200 K. Therefore, the differences in the thermal expansion of the diverse materials is the biggest too. For a temperature rise of 200 K the wafer thickness of 250 µm increases only about 0.13 µm. In contrast to that a piece of FR408HR with the same thickness would increase about 2.75 µm.

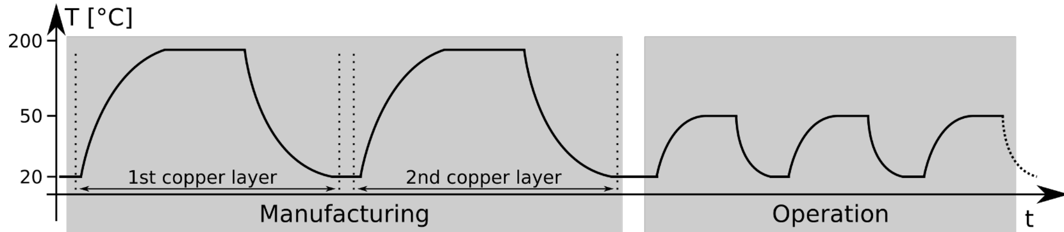


Figure 3.3: Schematic of the thermal lifecycle of an embedded wafer. The highest temperature difference is during the manufacturing process. The board has to be heat up for every additional copper layer that is added to the stack up. During the later operation time the single thermal stress is reduced but repetitively.

An overview about the various coefficients of thermal expansion is in table 3.1. During the cooling phase of the manufacturing process the resin of the prepregs cures below a certain temperature and glues everything together.

Nevertheless, the materials shrink further with an decreasing temperature which induces a static mechanical strain within the board. This cycle of heating and cooling gets repeated with every additional layer that is added to the stack up. The lamination process combined with the different coefficients of thermal expansion are responsible for the warpage of the board. This has to be avoided during production, because the structuring of the copper planes needs a flat surface. Otherwise the exposed geometries are not within the allowed tolerances, for example lines do not end where they should. In the worst case the copper layers are not aligned to each other and vias will not interconnect the layers. Later in the operation phase the temperature differences are much smaller. The circuits on the silicon wafer are

Material	Coefficient of thermal expansion [$10^{-6}\,\frac{1}{\mathrm{K}}$]
Cu(∗)	16.5
Si(∗)	2.6
FR408HR(Z-axis)(†)	55
FR408HR(X-/Y-axis)(†)	16
GEA679FG(S)(Z-axis)(§)	20 - 30
GEA679FG(S)(X-/Y-axis)(§)	12 - 14
Fe(∗)	11.8
Ni(∗)	13.4
Invar(Fe-36Ni)(∗)	0.6163
CIC(‡)	2.4 - 5.6
MCL-E-770G(LH)(Z-axis)($)	8 - 13
MCL-E-770G(LH)(X-/Y-axis)($)	1.5 - 2.0
E-Glass Fiber(#)	5 - 6
BCB(%)	42

Table 3.1: Material coefficients for thermal linear expansion.
(∗) taken from [*Haynes*, 2016]. (†) taken from [*Isola*, 2017]. (‡) taken from [*CCI Eurolam*, 2017] for a 12.5/75/12.5 ratio. ($) taken from [*Hitachi*, 2017b]. (§) taken from [*Hitachi*, 2017a]. (#) taken from [*Mangroli and Vasoya*, 2008]. (%) taken from [*Dow Chemical Company*, 2017].

designed to work at a temperature of $50 \pm 10\,°\mathrm{C}$ [*Schemmel*, 2016]. The resulting temperature difference between a turned-off and operating system is in the range of 30 K. Hence, the absolute thermal expansion is reduced. On the other hand these temperature cycles happen more often and so slowly weaken the materials and interconnections.
A well-tuned material selection is essential for a reliable and long lasting package. Within the project several material compositions were tested at the IZM in terms of board warpage, thermal reliability and lithographic feasibility. The results are presented in chapter 4.

3.2.2 Heat Dissipation

In the current WSI system the heat of the wafer is directly transferred away via the aluminum bracket and the attached heat sinks. This way it is possible to keep the wafer temperature constant at a power dissipation of 500 W. For an embedded wafer the heat has to go through the prepreg and copper layers before any kind of heat sink can cool the system. Heat transport is like an electrical current, it goes from a high temperature area to a cooler area and takes the path with the highest thermal conductivity. The thermal conductivity is a material-dependent parameter.

In table 3.2 the typical values of the used materials are listed.

Material	Thermal conductivity $[\frac{\mathrm{W}}{\mathrm{m\,K}}]$
Air(∗)	0.0262
Al(∗)	236
Cu(∗)	401
Si(∗)	124
FR408HR(†)	0.4
MCL-E-679FG(S)§)	0.6 - 0.7
Fe(∗)	80.2
Ni(∗)	90.7
Invar(Fe-36Ni)(∗)	13.835
MCL-E-770G(LH)(§)	0.6 - 0.7
E-Glass Fiber(#)	0.3 - 0.4
Solder Mask(%)	0.2

Table 3.2: Material coefficients for thermal conductivity.
(∗) taken from [*Haynes*, 2016]. (†) taken from [*Isola*, 2017]. (§) taken from [*Hitachi Chemical Company Ltd.*, 2017]. (#) taken from [*Mangroli and Vasoya*, 2008]. (%) taken from [*Cree*, 2017].

The heat dissipation through an material is defined as

$$\frac{dQ}{dt} = -\lambda * A * \frac{dT}{dx} \tag{3.4}$$

λ is the thermal conductivity, A the cross sectional area, $\frac{dT}{dx}$ the temperature gradient and $\frac{dQ}{dt}$ the dissipated power per time.
This can be solved for dT.

$$dT = -\frac{\frac{dQ}{dt}}{\lambda * A} * dx \tag{3.5}$$

$$\int_{T_0}^{T_1} dT = -\frac{\dot{Q}}{\lambda * A} * \int_0^l dx \tag{3.6}$$

l being the length of the object

$$\tag{3.7}$$

$$\Delta T = -\frac{\dot{Q}}{\lambda * A} * l \tag{3.8}$$

With eq. (3.8) the required temperature difference can be calculated, where $\dot{Q}$ is the dissipated power per time, *lambda* the thermal conductivity, A the cross section area and l the length of the material. Assuming in the final board are four copper

layers on each side. Each copper layer comes with an additional prepreg layer with a 100 µm thickness. The thermal conductivity value is taken of the Hitachi MCL-E-770G(LH) material. At this point the copper is neglected for the calculations as its thermal conductivity is 1000-fold higher in comparison to the prepreg material. It is intended to cool the board only from one side like in the WSI system. Hence, there is a 400 µm thick layer of prepreg material. At the moment the HICANN wafer can consume up to 500 W of energy constantly. The radius of the wafer is 10 cm. From eq. (3.8) the resulting temperature difference over the prepreg layer is 9.1 K.

The value can be improved by introducing thermal vias to the backside of the wafer. First the thermal resistance of all thermal vias is calculated. The thermal resistance is defined by eq. (3.9) with d being the depth of the via, λ is the thermal conductivity and A is the cross section area. The area of all vias is calculated by the area of a single via times the number of total vias n.

$$R_{th,vias} = \frac{d}{\lambda * A} \tag{3.9}$$

$$= \frac{d}{\lambda * n * (\pi * r_{via}^2)} \tag{3.10}$$

Due to the thermal vias the area for the heat transfer through the prepreg is reduced and has to be taken into account.

$$R_{th,prepreg} = \frac{d}{\lambda * (\pi * r_{wafer}^2 - n * (\pi * r_{via}^2))} \tag{3.11}$$

The total resistance is calculated like for electrical resistances. The resistances of the prepreg and the thermal vias are parallel to each other which leads to

$$\frac{1}{R_{total}} = \frac{1}{R_{th,vias}} + \frac{1}{R_{th,prepreg}} \tag{3.12}$$

The distance between equally distributed vias is an interesting parameter for the board design. Therefore, the wafer area is divided by the number of vias. This return the area occupied by one via. For simplicity the area is assumed to be a circle. This way the distance d between two vias can be estimated by eq. (3.14).

$$\pi * l^2 = A_{via} = \frac{\pi * r_{wafer}^2}{x} \tag{3.13}$$

$$\pi * l^2 = \frac{\pi * r_{wafer}^2}{x}$$

$$l^2 = \frac{r^2}{n}$$

So the distance between two vias is

$$d = 2 * l$$
$$= 2 * \sqrt{\frac{r^2}{n}} \tag{3.14}$$

In table 3.3 the thermal resistances and the temperature difference are calculated for different numbers of thermal vias. From 10 000 vias upwards the thermal vias have

# vias	R_{vias} [K/W]	$R_{prepreg}$ [K/W]	R_{total} [K/W]	ΔT [K]	Via density [vias/cm^2]	Distance [mm]
0	0	0.0182	0.0182	9.1	0	-
1000	0.3528	0.0182	0.0173	8.65	3.2	6.325
10000	0.0353	0.0182	0.0120	6.00	31.8	2.000
40000	0.0088	0.0183	0.0059	2.97	127.3	1.000
50000	0.0071	0.0183	0.0051	2.55	159.2	0.894
100000	0.0035	0.0184	0.0030	1.48	318.3	0.632

Table 3.3: Influence of thermal vias to the heat transport off the silicon wafer to the backside.

a significant influence on the temperature difference. Taken the distance between the vias into account a good compromise is around 40 000 to 50 000 vias.

The last layer between the surrounding cooling medium and the board is the solder mask. A typical thickness of the solder mask is 35 µm [*Böttcher*, 2016]. The material has a low thermal conductivity value around 0.2 W K/m, see table 3.2. Due to the thin layer and the big area the temperature difference to the cooling medium is 2.8 K.

This results in a temperature gradient from the wafer surface to the bottom of the board of around 12 K. To keep the wafer on a working temperature of 50 °C the temperature at the solder mask has to be below 38 °C. The use of thermal vias decreases the required temperature difference to 6 K resulting in a surface temperature of 44 °C.

Hence, it is no problem to get the heat out of the board through the backside. Next step is to remove the heat from the board surface. So far it is not decided if it is done by air or water cooling. Both methods are capable of cooling it down as the power density is just 2.6 W/cm^2. There are computer processors commercially available with a higher heat density which are still air cooled, e.g. the Intel i7-975 processor with a heat density of 49 W/cm^2 [*Intel*, 2017a].

If not only the silicon wafer area is used, the value decreases further. This can be done by heat spreading with the copper layers so the cooling area can be extended.

3.3 Wafer Preparation

Before the wafers can be embedded in the PCB they have to be prepared. Two procedures are necessary, one is wafer thinning and the other is the pad preparation for the electroplating process.

3.3.1 Wafer Thinning

The thickness of the wafer for the BrainScaleS system is 750 µm, which is not optimal for the embedding process. Thick wafers are more rigid and break earlier under mechanical strain than thin wafers. A thin wafer withstands bending without problems. Such a bending could happen during the manufacturing of the board, due to the different material properties. Furthermore, the stack up surrounding the wafer is easier and more reliable. There are less layers of material required and the behaviour of the stack up is more predictable during the lamination process.

Unfortunately, the wafer thinning itself has an effect on the later wafer bow. The

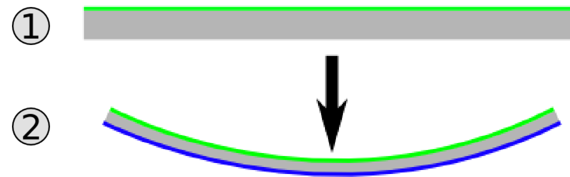


Figure 3.4: The effect of back-side grinding to the silicon wafer. The gray material is the polycrystalline silicon substrate, the green layer is the active device layer with the metal wiring on top and the blue layer is a silicon oxide layer which was created through the grinding process. At the beginning (1) the silicon wafer is rigid and flat. After thinning the wafer is concave (2). Different methods can reduce the bow.

ideal grinding removes the silicon from the back-side without inducing any damages to the surface. Unfortunately, the grinding process creates tiny cracks and defects in the back-side which widen the back-side surface and cause the silicon wafer to bend. The result is depicted in fig. 3.4. Therefore, the grinding is done in several steps to minimize the wafer bow. The wafer thinning is done at the IZM in Berlin. Before the actual thinning the wafer edge is trimmed which increases the stability and mitigates a chipping of the silicon edges [*Garrou et al.*, 2014]. The grinding process is done with a fine grained wheel. The silicon is removed with different feed rates. In the first step 472 µm of silicon at a rate of 0.5 µm/s are taken away. Due to the high removal rate the surface has a high roughness. Therefore, the roughness is reduced in two runs with lower removal rates. The next 10 µm of silicon are removed with a rate of 0.25 µm/s. The last removal of 5 µm is done with 0.1 µm/s [*Zoschke*, 2017]. The lower feed rates at the end of the grinding process help to lower the warpage of the silicon wafer. After the thinning to 250 µm the wafer is flat enough for the

embedding process.

If the wafer is thinned further, especially, below 100 µm additional stress relief methods are needed. A finer grain wheel can remove topological damages to the surface but it cannot reduce the stress due to the change in the chemical composition. Before the grinding procedure there was polycrystalline silicon on the back. After the grinding there is a thin layer of silicon oxide. This increases the spacing between the atoms and, thus, widens the lattice. The effect gets more dominant for wafers thinner than 100 µm. The damages can be reduced by a reactive etching or a polishing process, like plasma dry etching of chemical mechanical planarization [*Garrou et al.*, 2014] These methods remove material with a very low rate of 1-10 µm/min. Nowadays, there are silicon wafers with a thickness of 5 µm [*Chasin et al.*, 2016]. The drawback of these ultra thin wafers is that they need a carrier wafer for handling. Without a carrier wafer no further processing would be possible.

Hence, a wafer thickness of 250 µm was chosen. The wafers are thin enough for the embedding process, but are still easy to handle.

3.3.2 Copper Pads Finish

The pads of the Post-Processing Procedure (Post-Processing Procedure) have a thin gold finish, which is good for the WSI system where the pads could react with the air and oxide. Whereas for the embedding - especially for the later electroplating of the vias - thick copper pads are desired. A thin pad could easily be damaged during the laser drilling. A thicker pad has a higher tolerance against failures. The electroplating works at its best, if the substrate is also copper. Therefore, all Post-Processing Procedure pads get an extra layer of copper on top, which is about 10-12 µm thick [*Zoschke*, 2016].

The copper is applied in a semi-additive deposit procedure. The first step is a sputter deposition of the complete wafer with an adherent layer of copper. Next the photoresist is patterned in a lithography step. In a electroplating process the copper is added in the openings of the photoresist. The last step is the removal of the remaining photoresist and the underlying adherent layer.

Both of the tasks, the wafer tinning and the copper deposit, are done by the IZM in Berlin.

4 Bare Silicon Lamination Tests

Before the active wafers were embedded the IZM did test runs with bare silicon disks. They had the same physical properties, i.e. the diameter and thickness, but had no active components on it. These tests should find a suitable material composition in regards to warpage, via and process parameters. The results are presented in the following sections.

4.1 First Lamination Test with Standard FR4 Material

The first test run should show how the different coefficients of thermal expansion influence the package and what the optimal wafer thickness is. Therefore, two lamination runs were done with a 200 µm thick silicon disk and another with a 250 µm thick disk. The diameter of the silicon disks is like the current HICANN wafer 20 cm. The prepreg was the standard PCB material MCL-E-679FG [*Hitachi*, 2017a]. The wafer was surrounded by an rigid FR4 frame of the same material. The coefficients of thermal expansion of the prepreg are 5- to 12-times bigger than the value of silicon, see table 3.1. The stack up is depicted in fig. 4.1.

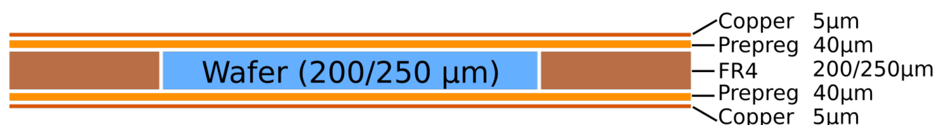


Figure 4.1: Stack up of the first embedding test.

As expected the board is not flat after the lamination process, see fig. 4.2. The board shows a convex bow. The reason is the difference in the coefficients of thermal expansion. The FR4 material expands more than the silicon which creates internal stress in the board. The stress is still present after the board is cooled down, because the resin from the prepreg cured at a higher temperature of 180 °C. In the cooling phase the materials shrink further and the internal stress increases as there is a strong coupling between the materials. Therefore, the complete board gets bowed. The warpage was so strong that further lithography steps were not possible.

After removing the FR4 frame the package returned to a flat surface. Hence, the warpage was mainly induced by the rigid FR4 frame during the lamination process. The flat surface allowed it to run the disk through an etching process, so that copper structures were created on top and on bottom, see fig. 4.3.

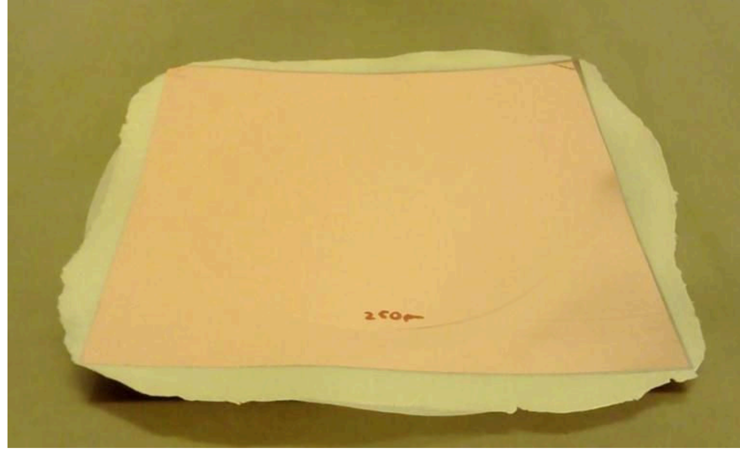
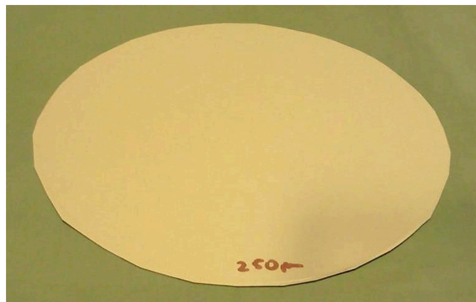


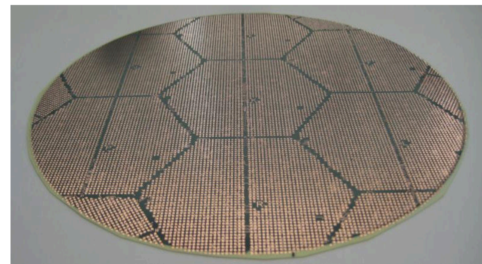
Figure 4.2: The package with the 250 µm thick wafer after the lamination process. The surface is not flat enough, it cannot be properly processed further. The result for the 200 µm thick wafer is the same. Courtesy of Lars Böttcher (IZM).

At the end the disk was x-rayed to see if the silicon disk was still intact. No cracks were found at all, see fig. B.1. Hence, the compression during the lamination process did not break the silicon disk.

Conclusion This test run showed that the process of embedding is no risk to the silicon disk. Another material composition is required because of the warpage. Without the FR4 and prepreg frame the board returns to a flat surface capable of furhter processing.



(a) After cutting out



(b) Processed package

Figure 4.3: The package became flat after the removal of the FR4 frame (a) and could pass an etching process (b). Courtesy of Lars Böttcher (IZM).

4.2 Second Lamination Test with CIC Material

In this run the stack up stayed the same as before, but the FR4 core material was exchanged with a sheet of Copper-Invar-Copper (CIC). The prepreg material is again the Hitachi MCL-E-679FG.

The CIC material consists of three layers. The first and third layer is plain copper, the second layer is Invar. Invar is an alloy of 65 % iron and 35 % nickel. This alloy has a very low CTE of $0.6 \times 10^{-6}\,\frac{1}{\mathrm{K}}$, in combination with the copper layers its final CTE is around $2 \times 10^{-6}\,\frac{1}{\mathrm{K}}$, see table 3.1. The CIC layer is also a good thermal heat spreader due to the copper. In contrast to that the Invar is negligible as it has only around 3 % of the thermal conductivity of copper, see table 3.2. As the CIC has no direct connection to the silicon wafer the heat spreading effect is not used at the moment. Furthermore, the CIC sheet can later be used as a GND plane due to its good electrical properties. It is no special treatment of the CIC material needed as it already has thick copper layers to connect to. Nevertheless, the focus of this test runs lie on the embedding process and mainly on the reduction of warpage.

The change in the material selection reduced the amount of warpage of the board after the lamination procedure significantly.

Another difference to the previous run is the silicon disk. The complete RDL stack up from the HICANN wafer is applied, meaning it has the fine-pitch lines and the contact pads. Furthermore, the contact pads got a 12 µm thick copper coating which is needed for a successful electroplating process, see section 3.3.2.

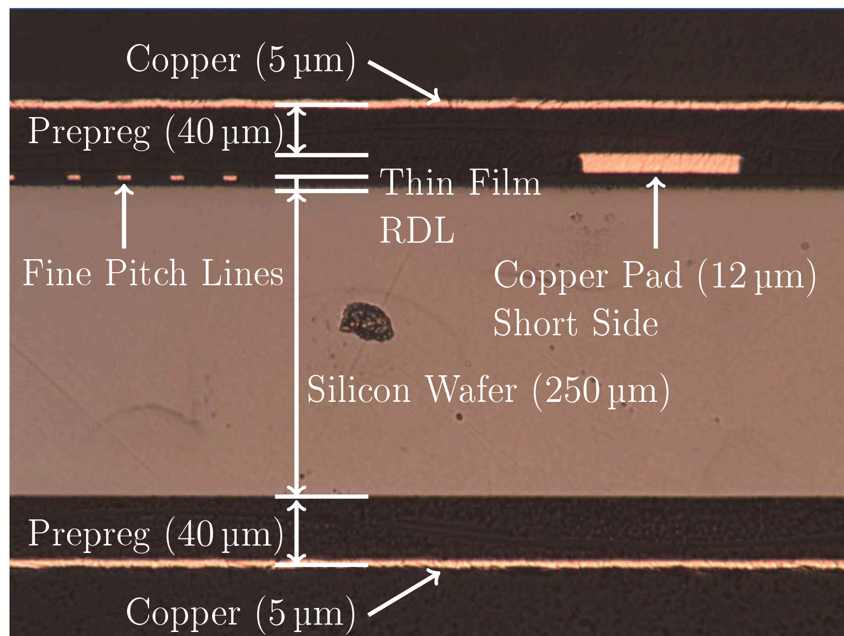


Figure 4.4: Cross section through the board. From the RDL the short side of a contact pads is visible. Courtesy of Kai Zoschke (IZM).

To see if the RDLs were damaged during the lamination process, two cross section images were taken. In fig. 4.4 the cross section shows the short side of a big copper contact pad and several fine pitch lines. A cross section image orthogonal to the previous image is shown in fig. 4.5. The long side of a contact pad is visible. There are no signs that the lamination process altered the RDLs in any way. The fine pitch lines are all on the same level, the distance between copper pad and silicon stays constant and the pad itself is intact.

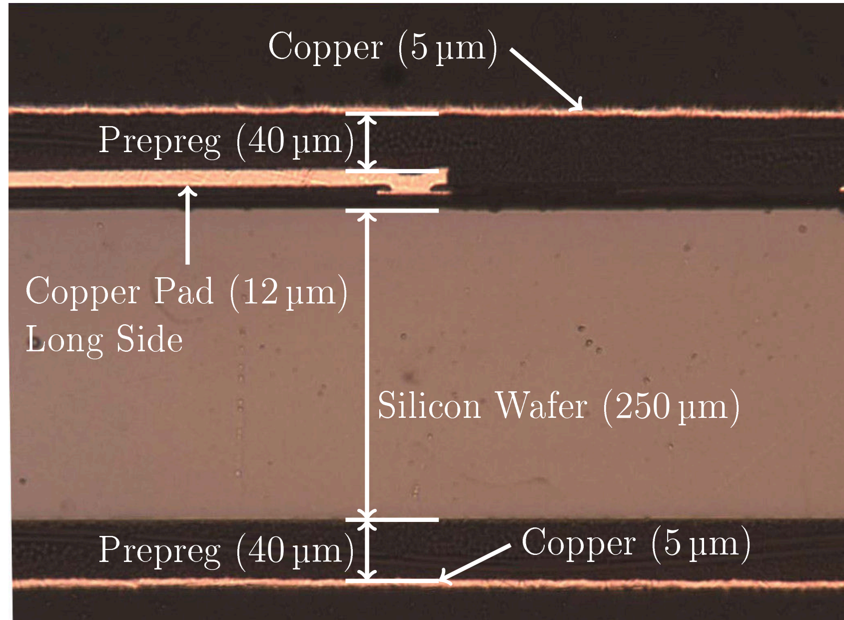


Figure 4.5: Cross section through the board. The long side of a contact pad is visible. Courtesy of Kai Zoschke (IZM).

The next step was to test the laser drilling for the vias connecting to the wafer. The laser works at a wavelength of 355 nm which is in the ultra-violet region. Its repetition rate is 200 kHz and has a beam size of 10 µm [*Schmoll*, 2017].

A hole is drilled in three phases [*Zoschke*, 2015]. First the copper on top is opened, to remove the copper a beam energy of 2 W is required. In the next phase the 40 µm prepreg is ablated. For that the laser power is reduced, because the prepreg has a decomposition temperature of around 350 °C whereas the copper's melting point is at 1084 °C [*Hitachi*, 2017a; *Haynes*, 2016]. Thus, for the ablation of the prepreg a beam energy of 1 W is chosen. In the last phase the laser power is set to 0.5 W to have more control over the material removal. The amount of removed material per time is low, on the other hand the accuracy of the depth is increased. Furthermore, the damage to the underlying copper pad is minimized which is of advantage for the later electroplating process. The final diameter of the hole is 60 µm. The result is shown in fig. 4.6.

At the end the board was x-rayed. The images showed an intact silicon wafer.

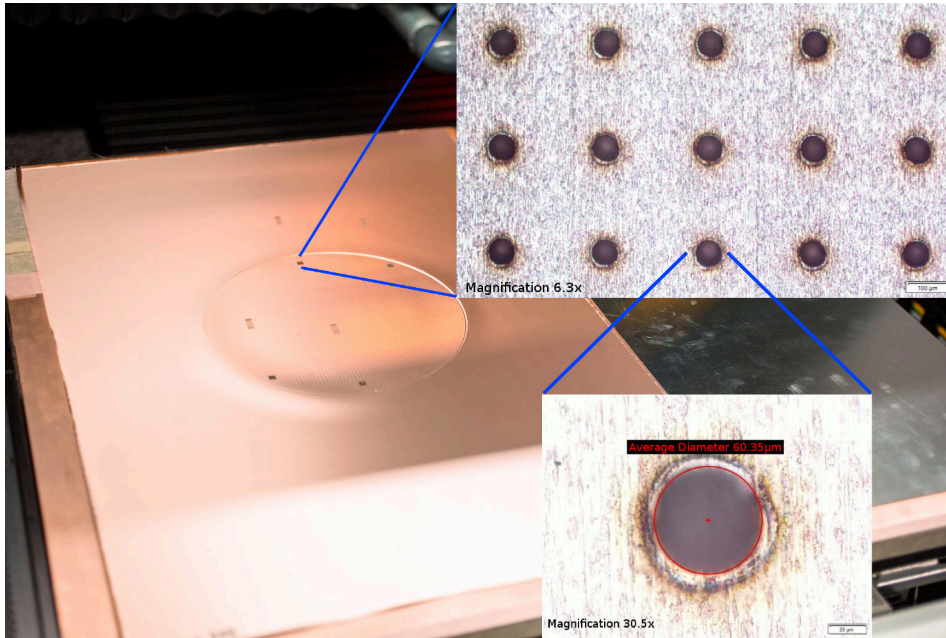


Figure 4.6: Result of the second lamination run. The warpage of the board is low and it was possible to drill clean holes with a diameter of 60 µm. Courtesy of Lars Böttcher (IZM). Images are modified.

Also the RDLs with the thick copper application looked fine. In fig. B.2 an x-ray image shows a segment of the silicon wafer.

Conclusion The application of the CIC material improved the mechanical behavior of the board during and after the lamination process. The warpage was in a range to allow lithography steps on the complete board and not only on a cut out piece like in the previous test run. It is possible to drill the holes for the wafer contacting. Again no silicon disk was destroyed by the lamination process.

4.3 Third Lamination Test with low-CTE Material

In the third test run two major things changed.

First the prepreg material is replaced. It is now the MCL-E-770G(LH) which has an ultra low CTE of $2 \times 10^{-6} \, \frac{1}{\mathrm{K}}$ [*Hitachi*, 2017b]. The material shows a significant reduction in the warpage after processing than other materials. It is achieved by improving the composition of soft and hard segments within the resin [*Kotake et al.*, 2014]. The core material is again a sheet of CIC. Hence, CIC, prepreg and silicon wafer have coefficients of thermal expansion close to each other.

The second thing that changed is the stack up. The core material is not 250 µm thick anymore, it is reduced to 150 µm. Therefore, additional prepreg layers are

added symmetrical around the core. The prepreg layers are each 50 µm thick and compensate the thinner CIC sheet. They have a cut out in the shape of the silicon wafer like the CIC sheet. This makes the stack up more flexible during the lamination process. A minor change is the thickness of the other prepreg layers which cover the whole board. They are also 50 µm thick now. This way only one version of the material is needed.
The new stack up is depicted in fig. 4.7.

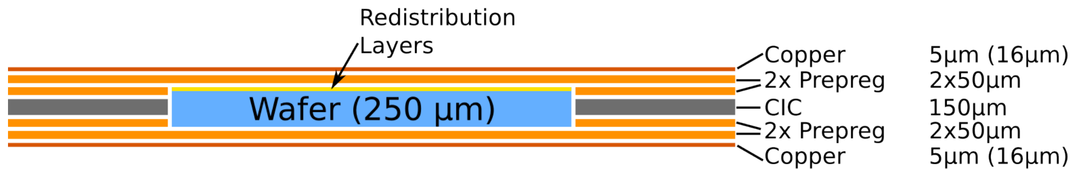


Figure 4.7: New stack up of the third lamination test which uses the low CTE prepreg material.

The silicon wafer is the same as before. It is a bare silicon disk which gets the complete RDL stack up of the HICANN wafer.

After the lamination process the bow of the board is very low. It is in the range of the second test run. Thus, a full board lithography process is possible. The next step is to check the influence of the electroplating procedure on to the board.

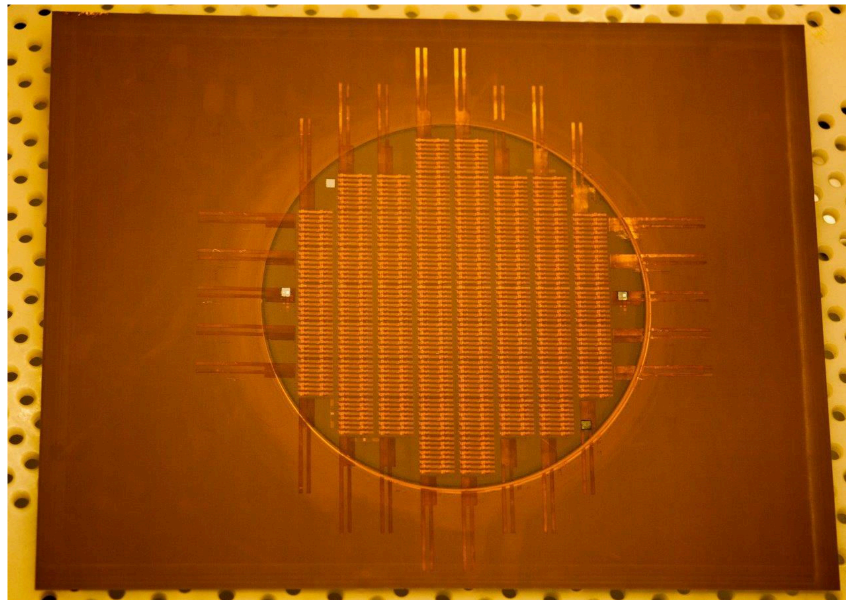


Figure 4.8: Third prototype after the electroplating and etching procedures. The board shows only a low convex bow. Courtesy of Lars Böttcher (IZM).

At first the holes for the contact vias to the wafer are drilled. The parameters are

the same as before, see section 4.2. The holes are completely filled with copper in a galvanic bath. Next a photoresist is coated on top and the final copper pattern is applied by exposing the photoresist. The unexposed film areas are washed away and the copper in the open areas is removed by an etching process.
The resulting board is shown in fig. 4.8.
The reliability of the vias is tested with daisy chains running across the wafer. The chain length depends on the number of reticles involved. From each reticle 16 RDL contact pads are used meaning 32 vias are in the chain. The longest chain combines 64 pads with 128 vias.

The quality of the etching process is checked with an optical inspection. There the line widths and diameter of the via landing pads are measured. In fig. 4.9 the etching result is shown. The lines have a width of around 40 µm which is in the range of tolerance. The same is true for the via hole diameter and the via landing pad.

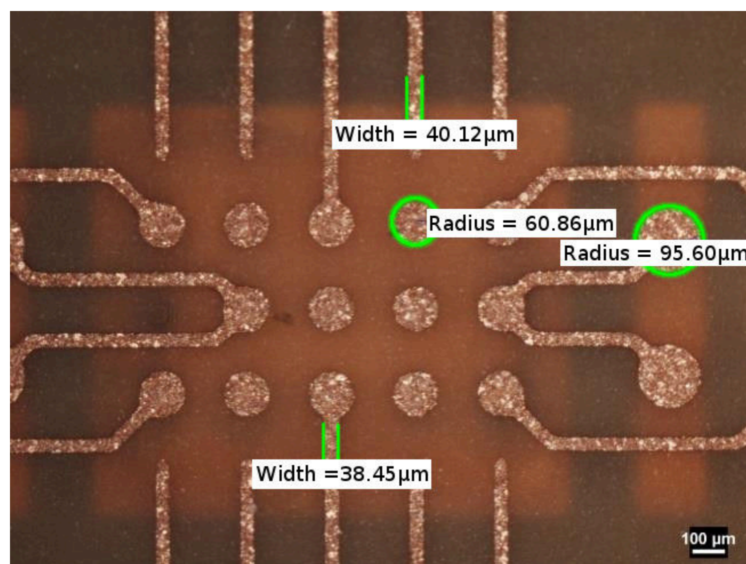
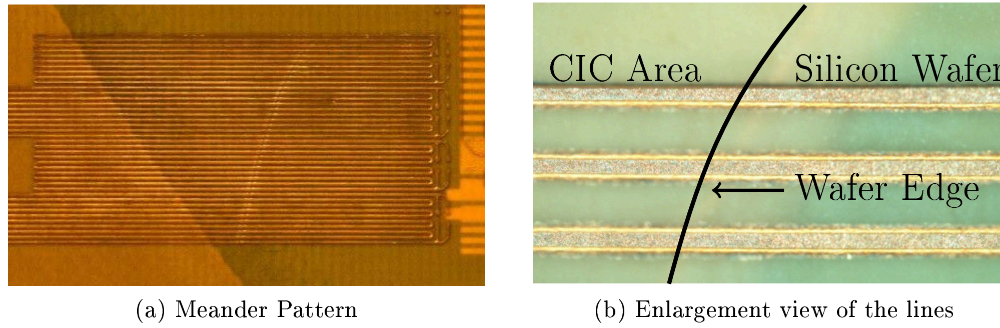


Figure 4.9: Optical inspection of the etched lines and via landing pads. All measured values are within the range of tolerance. Courtesy of Lars Böttcher (IZM). Image is modified.

Additionally, cooper lines run in a meander pattern across the junction between the silicon wafer and the CIC/prepreg stack. The optical inspection of the copper lines shows that there is no problem with the transition between the two areas, see fig. 4.10. Furthermore, the meander structures have test points to measure the line's resistance which are used in the thermal stress tests.

Thermal Stress Test The boards of the third lamination test were put in a climate cabinet and thermally stressed. Therefore, the temperature periodically changed between 0 °C and 100 °C with a dwell time of 15 min. The test was conducted at the IZM in Berlin.



(a) Meander Pattern (b) Enlargement view of the lines

Figure 4.10: Photographs of copper lines crossing the wafer edge. There is no influence of the different materials on the lines visible.

The goal was to see the influence of the temperature cycles on the copper lines and wafer contact vias. Therefore, the resistance of the meander lines and daisy chains is measured after a certain amount of cycles. In figs. B.3 and B.4 the layout and the measurement points of both tests are depicted. The results of the tests are shown in fig. 4.11.

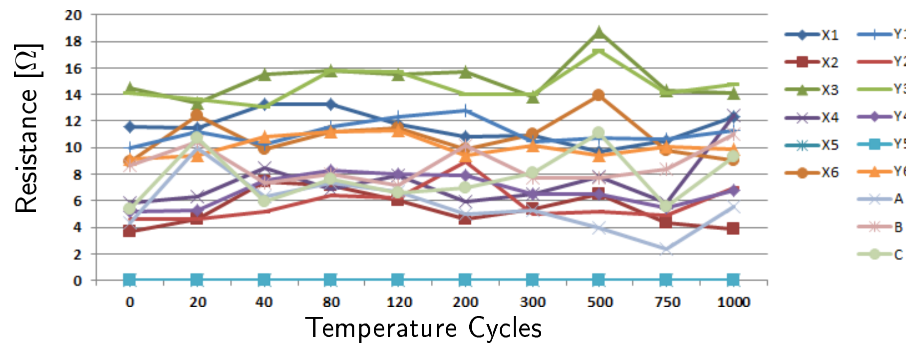


Figure 4.11: Resistance of the meander lines and daisy chains during the accelerated environmental stress test in the climate cabinet. Courtesy of Kai Zoschke (IZM).

Only one line (Y5) was completely broken from the beginning. It was one of the meander lines which went over the wafer edge. The resistance was constantly 0 Ω which suggests that there was a short in the meander line. This could have happen during the etching process, perhaps the etching time was to short. There is no long-term degradation of the lines observable. The maximal difference in the resistance between two measurements of a line is 6 Ω which is in an acceptable range. An interesting behavior would be a sudden rise or drop of the resistance value. A value of MΩ would mean that a line is open or that a via to the wafer is broken. Fortunately, this case never occurred on the boards during the test runs [*Zoschke*, 2016].

Conclusion In this run two prototypes had been produced. Both showed no problems with the embedding process and the later processing. It was possible to create vias to the RDL pads and pattern the copper on top. Furthermore, the boards withstood a thermal stress test.

4.4 Conclusion

It was shown that the embedding of a bare silicon wafer in a PCB is possible. After three iterations a suitable stack up and material selection for a board with two copper layers and a 20 cm silicon wafer was found. After the embedding procedure the warpage is low enough for further lithography steps. On the copper layers lines with a width of 40 µm were produced and the connection to the wafer through vias was established.

5 DENSE Prototype

At the beginning of 2016 the Electronic Visions Group received new silicon wafers with the High-Input Count Analog Neural Network Chip version 4.1 (HICANNv41). These wafers replaced the High-Input Count Analog Neural Network Chip version 2 (HICANNv2) wafers installed in the BrainScaleS systems. Therefore, the HICANNv2 wafers could be used for new tasks. The HICANNv2 wafers are good candidates for embedding tests with active wafers. The circuits are well understood and software for controlling and testing is available.

Hence, the first embedding prototype, the DENSE board, is equipped with a HICANNv2 wafer. In comparison to the BrainScaleS system the DENSE prototype is reduced in complexity and functionality.

The layout of the DENSE board and prototype system was developed in the Electronic Visions Group in Heidelberg. The embedding and production of the DENSE board is done at the IZM in Berlin.

The chapter starts with an introduction to the physical stack up and the layout considerations for the DENSE board. In section 5.3 the complete test system is depicted, followed by a description of the connection principle between the DENSE board and the test hardware. At the end the different test methods to characterize the prototype system are described.

Objectives The major task of this prototype is the analysis of the influence of the lamination process to the RDL and the electronic circuits on the wafer, especially, on the neuron circuits. This allows an estimation for the yield of the process. Furthermore, the reliability of the board will be checked with a climate cabinet test. For the first time this board will have connections from the top copper layer to the bottom, which have to be checked. Another new feature in this prototype setup are the solder-less connectors which create the connection between the DENSE board and the test hardware. There is no experience with this type of connector in the Electronic Visions Group, so this will show if they can be used for future projects.

This prototype is not intended be used for heat dissipation tests. The power consumption of a single HICANN chip is too low and the layout is not optimized for the powering of multiple chips at the same time.

5.1 Stack up

The stack up is the same as for the last silicon-only prototype. It is depicted in fig. 5.1. The difference is the wafer inside the board. This time it is the HICANNv2 wafer with RDLs. Moreover, the copper layers are covered with a solder resist.

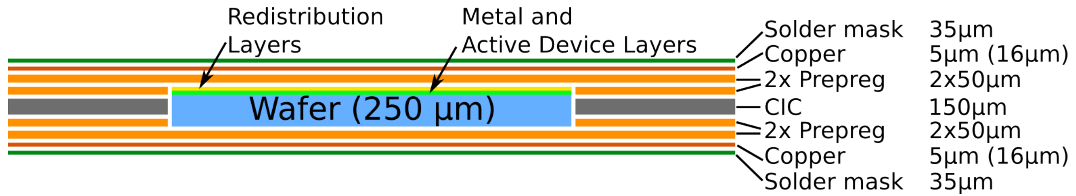


Figure 5.1: Stack up of DENSE prototype. All layers except the solder mask are combined in the lamination step. In the lamination process a 5 µm copper foil is used, which grows in the electroplating step to around 16 µm.

The HICANNv2 wafer has a thickness of 750 µm. For the embedding it has to be thinned down to 250 µm.

Around the wafer is one sheet of CIC of height 150 µm. In the center of the CIC material is a cut out in the shape of the wafer. However, it is not only a circular cut-out like for the bare silicon disks it mimics the wafer's notch. The cut out and the notch keep the wafer at its position during the lamination process.

On top and bottom of the CIC sheet comes a 50 µm sheet of prepreg with the same wafer cut out. The surrounding material heights add up to 250 µm which is the same height as the wafer.

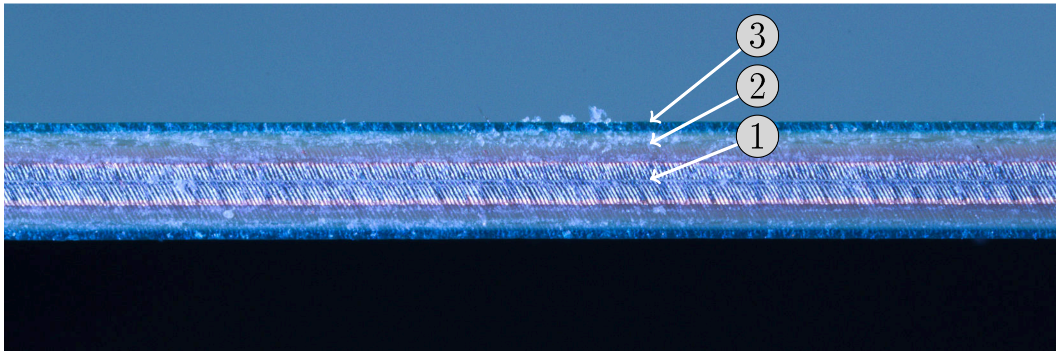


Figure 5.2: Image of the cross-section through the DENSE board. The metallic area in the center is the CIC material (1). The inner material is Invar and the outer material is copper. Next comes the colorless prepreg material (2). On top of the prepreg is the copper layer, but it is not routed to the board edges. Therefore, it is not visible in this image. At last come the solder mask layers (3).

At last one sheet of prepreg and one 5 µm copper foil are placed on top and on bottom of the stack up. The prepreg layer is again 50 µm thick. The thickness of the copper layers is 16-18 µm after the electroplating process [*Böttcher*, 2016].

Due to the thin prepreg layers the cooling of the prototype board is not a problem. Eight HICANN chips distributed over the wafer can be used at the same time so that the maximum heat dissipation is low. An aluminium block attached to the back of

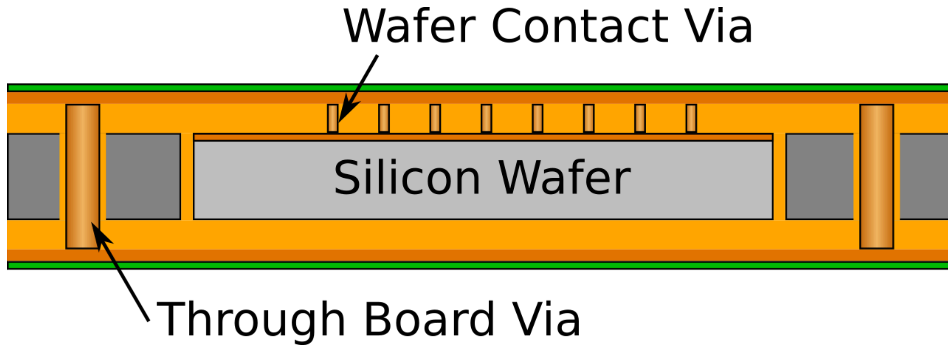


Figure 5.3: There are two types of vias used in the DENSE board. Microvias create the vertical connections between the wafer topside and the copper layer. The other via type connects the topside with the bottom side of the board.

the board is sufficient.

Wafer Contact Vias The connections from the top copper layer to the RDL pads on the wafer are created by microvias. The holes are drilled with an UV laser beam. The diameter of the holes is 60 µm and the landing pads are 100 µm. In the later electroplating process the holes get completely filled with copper. The complete filling is important for the power supply of the chip as it increases the current carrying capacity of the via. Furthermore, the filing enables the stacking of microvias. The resistance of a filled microvia is calculated with:

$$\begin{aligned} R_{MV} &= \rho * \frac{l}{A} \\ &= \rho * \frac{l}{\pi * r^2} \end{aligned} \tag{5.1}$$

The parameters from the DENSE board are via depth l =50 µm, hole radius r =30 µm and the specific conductivity for copper $\rho = 1.721 \times 10^{-2} \frac{\Omega\,\mathrm{mm}^2}{\mathrm{m}}$ [*Haynes*, 2016]
This results in a resistance of

$$R_{MV} = 0.304\mathrm{m}\Omega \tag{5.2}$$

There is no standard definition specifying the maximum current through a via. Different criteria are possible, e.g. voltage drop or power loss. In this case the voltage drop is a reasonable criterion, because the analog circuit's behavior depends on the analog voltages. The table 5.1 lists the voltage drop for various currents. Even though for 1 A the voltage drop is only 0.304 mV a maximum value of 0.25 A is chosen. This provides a safety margin for cases that a via has a bad contact or degrades over time.

Current [A]	0.1	0.2	0.5	1.0	5.0
Voltage Drop [mV]	0.030	0.061	0.152	0.304	1.522

Table 5.1: Voltage drop over a single microvia for different currents.

Plated Through Holes Due to the conducting CIC layer outside the wafer area the implementation of plated through holes (PTHs) is different from the standard procedure for PCB plated through holes. Typical PTHs are drilled after the stack up is finished. If this procedure would be used for the DENSE board, the board had many shorts as all via walls were connected to each other over the CIC material. Therefore, the group at the IZM thought of a different procedure, which creates short-free plated through holes. In contrast to a standard PCB the CIC sheet is drilled before the lamination process. The diameter of these pre-drilled holes is 400 µm. The holes get filled during the lamination process with the liquid resin from the surrounding material. When the lamination process is finished the actual vias can be produced. These vias are smaller than the holes in the CIC sheet, they have a final diameter of 100 µm. The resin in the holes serves as an insulator between the CIC and the copper. Thus there are no shorts between the vias and the CIC material [*Böttcher*, 2015]. In fig. 5.4 the procedure creating PTHs is depicted.

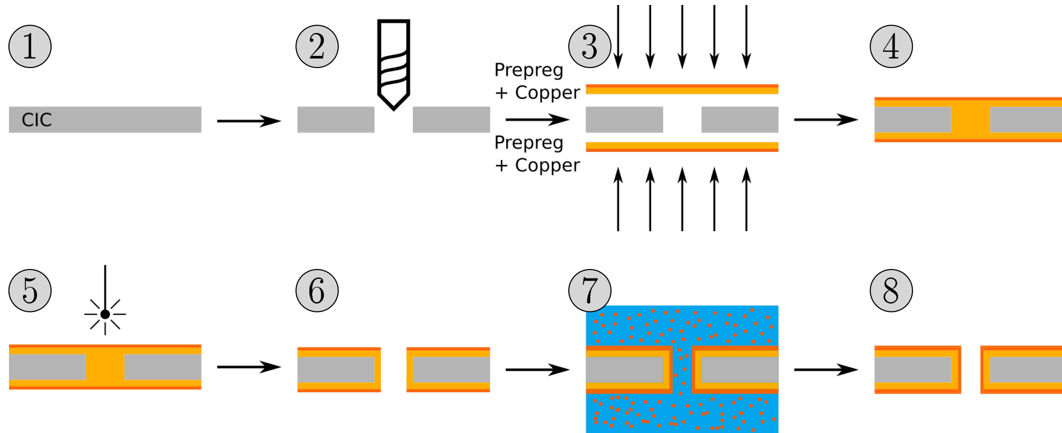


Figure 5.4: Schematic process of adding vias through the board. First there is the unprocessed sheet of CIC (1), which is drilled in the second step at the positions where later the PTHs will be (2). After the preparation the standard lamination procedure continues. When the stack up is completed (4) the holes for the vias are drilled (5). In the next step the copper on the prepreg gets processed. Then an electroplating bath (7) creates the copper wall inside the hole. The via is completed and a short-free connection from top to bottom is created (8).

These vias have a final copper wall thickness of 10-15 µm. The via resistance can

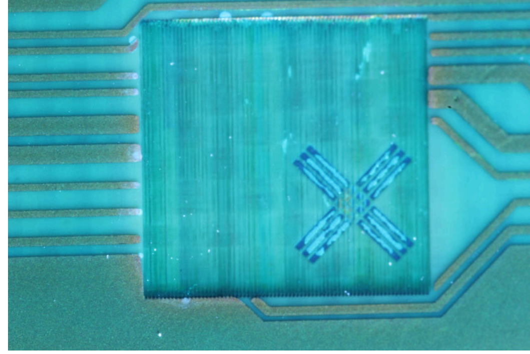


Figure 5.5: Fiducial mark on the silicon wafer covered by the solder resist. The vertical scratches in the opening are from the removal of the copper and the prepreg on top of the fiducial marks.

be calculated like for the wafer contact vias

$$R_{PTH} = \rho * \frac{l}{\pi * ((r+t)^2 - r^2)} \tag{5.3}$$

$$= \rho * \frac{l}{\pi * (2 * r * t + t^2)} \tag{5.4}$$

with the parameters: via depth $l = 450\,\mu m$, hole radius $r = 50\,\mu m$ and wall thickness $t = 15\,\mu m$

$$R_{PTH} = 1.4\,m\Omega \tag{5.5}$$

The resistance of a single plated through hole is $1.4\,m\Omega$ which is more than 4-times the resistance of the wafer contact via. Nevertheless, the voltage drop is negligible and so a maximum current of 0.25 A is chosen. This eases the layout process as the number of required vias for a supply voltage is the same for both types.

Fiducial Marks The HICANN wafer has two fiducial marks on the top RDL. In the WSI system both are used to align the wafer to the MainPCB.

Here the fiducial marks are used as reference points for the layout. For that reason the copper and prepreg on top of the marks has to be removed after the lamination step. The Laser Direct Imaging (LDI) machine adjusts the layout depending on the position of these fiducial marks. From the position of the fiducial marks the LDI machine automatically calculates X- and Y-offset values, scaling and rotation factors. These correction values are applied to the layout data [*Orbotech*].

5.2 Layout

The layout of the DENSE board is simple due to the minimalist stack up. To increase the yield of testable chips all can be used independently and without interference of each other. Every chip has its own power supply and no signals are shared between them. Additionally, for the operation of the board no connectors are used which need soldering.

The board is 360 mm long and 240 mm wide. Its thickness is 450 µm. In fig. 5.6 the layout of the DENSE board is shown. The HICANN wafer is placed in the center of the board.

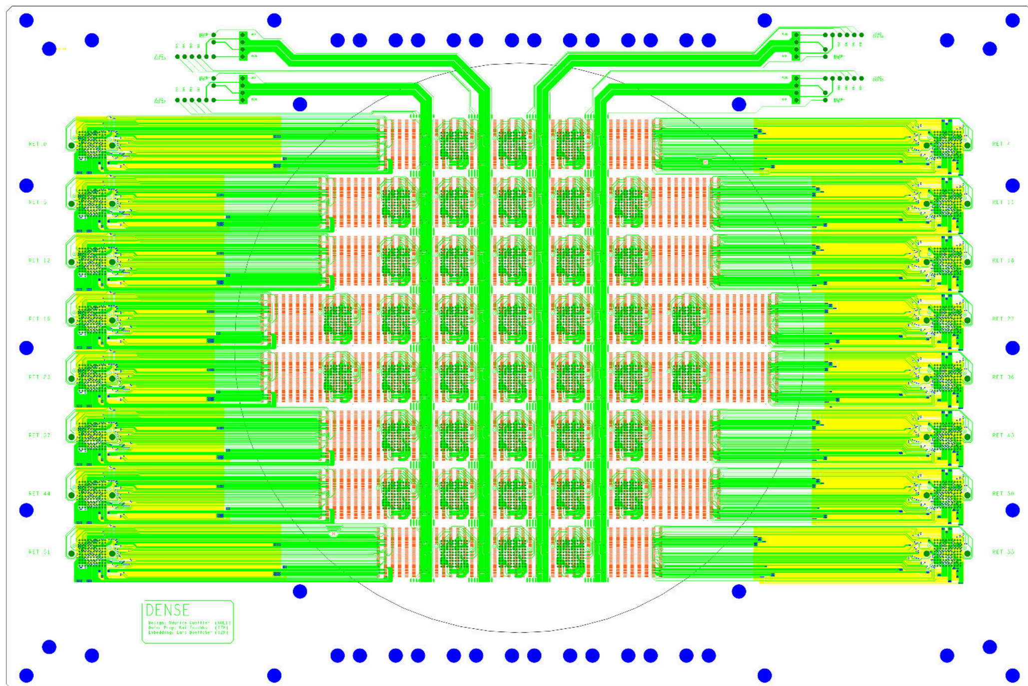


Figure 5.6: Layout of DENSE board. The copper structures on top are green, the structures on bottom are yellow and the RDL pads on the wafer are orange. The black circle depicts the embedded wafer. The HICANN chips at the wafer edge are routed off the wafer area and use PTHs.

Cadence Allegro The layout was created with Cadence Allegro [*Cadence*, 2017a]. The reason was to reuse scripts and techniques which were developed for the Main-PCB [*Güttler*, 2010]. These techniques split a big and flat project into a hierarchical project with several smaller sub-modules. The reticles on the wafer, for example, are an useful unit size for a module. Each of the sub-modules can developed independently. The modules can be added to the layout multiple times and, hence,

save time. Changes applied to a module, apply to every instance in the layout automatically. Additionally, with the use of SKILL (SKILL) scripts [*Cadence*, 2017b] the construction of the layout can be automated. The placement of the modules is managed by a SKILL script, which gets the information about the position of the modules from a simple text-file.

Routing scheme The entire layout is shown in fig. 5.6. The wafer sits in the middle of the board and its wafer notch faces to the south. In the WSI system the wafer is rotated by 45°. Hence, the routing area for a single reticle is increased by the factor $\sqrt{2}$ as the diagonal of the reticle is usable. Unfortunately, the development of the layout is more complicated than in the not rotated case. Therefore, to keep the layout of the DENSE board simple the wafer is not rotated.

On each reticle two HICANNs are chosen which are yet connected over the JTAG chain in the Post-Processing layer. These two HICANNs are mapped on one ZA1 connector. The signals of two HICANNs use almost all of the pads on one ZA1 connector. Hence, additional HICANNs cannot be mapped on the same connector. Additionally, for test purposes 16 HICANNs are interconnected over JTAG and go to dedicated connectors.

The reticle layout can be divided into two groups. There are the reticles where the connector is directly above the HICANN chip and, in contrast, there are reticles where the connector is at the board edge. The layout for the first group is depicted in fig. 5.7. The connector sits on top of the HICANN chip and the signals and voltages from below are routed to the next possible pad on the ZA1 connector. One ZA1 connector occupies almost the complete reticle area. For the inner reticles two modules were created: one with the additional JTAG connections and one without. These two modules were used for all 80 HICANN chips which save a lot of time.

The layout of a reticle with its ZA1 connector outside the wafer area is shown in fig. 5.8. These chips use the through-board vias for routing purposes. Otherwise, it would not be feasible to connect the signals to the pads as the space between the pads is too small. With the through-board vias it is possible to access the ZA1 pads from the back of the DENSE board.
One problem of routing the signals and supply voltages out of the wafer area is the limited space. All connections to and from the wafer go through the top copper layer. For the 12 supply voltages, two GND nets and 25 signal lines are only 2 cm of space available which is the length of one reticle edge. Outside the wafer area the back of the board can be used too. For the ZA1 connectors at the board edge again modules were created. One module for each side and then copied seven times.

Impedance On the DENSE board are two types of impedance-controlled lines. A $50\,\Omega$ single-ended line and a $100\,\Omega$ edge-coupled differential stripline. In tables 5.2 and 5.3 the impedance values for the differential and single-ended lines are listed. The calculations were done for top due to the symmetrical stack up the results apply for the bottom too. The values were calculated with the Sonnet field solver [*Sonnet*,

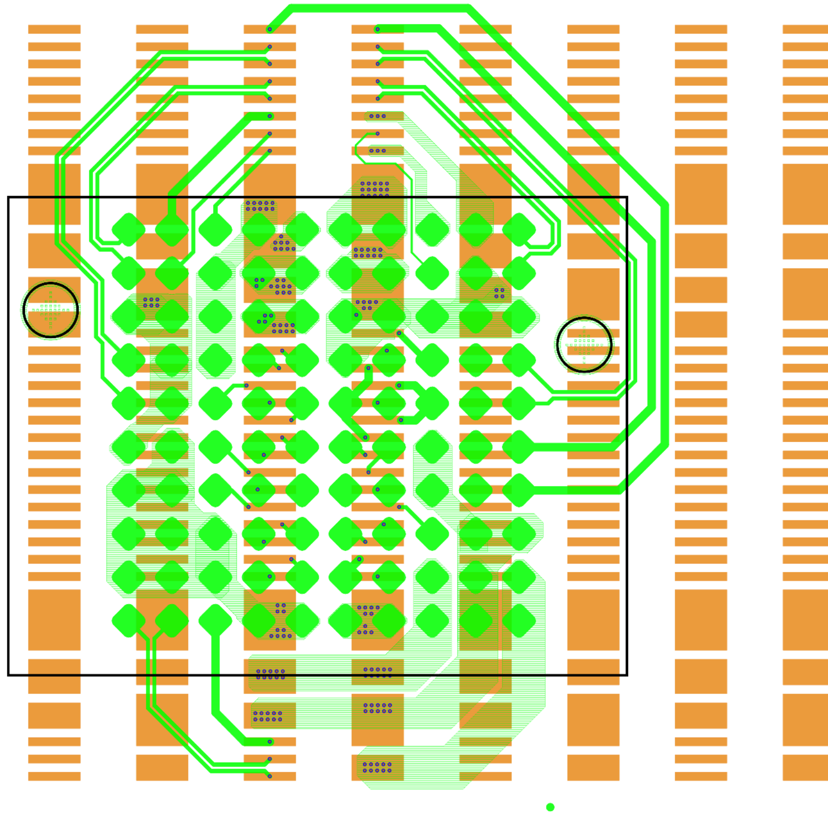


Figure 5.7: Layout of an inner wafer reticle. The wafer RDL pads are the orange rectangles, the top copper structures are green and the wafer contact vias are the blue dots.. The green squares are the pads of the ZA1 connector. The size of the ZA1 connector is depicted by the black line. The black circles are the openings in the ZA1 connector to the fiducial marks on the DENSE board.

2017]. The line geometries are simulated for two areas on the board. One area is over the silicon wafer which has one prepreg layer between the BCB and top. Additionally, a metal layer is on top that is not connected to any voltage and, hence, a floating plane.
The second area is outside the silicon wafer. The CIC sheet is used as a reference plane. There are two layers of prepreg between the CIC sheet and the top copper layer. The detailed stack up properties are listed in tables C.1 and C.2.

Clearly, not all calculated values are in range of the target impedance. Especially, the differential lines above the silicon wafer and the single-ended lines over the CIC material have a big deviation, e.g. 43 % for the single-ended line. The other lines are just over the 10 % tolerance value. Whereas 10 % is the limit for the impedance value that is acceptable for the links. Unfortunately, this mistake was discovered after the production of the board. A second calculation with the 2D field solver from Allegro

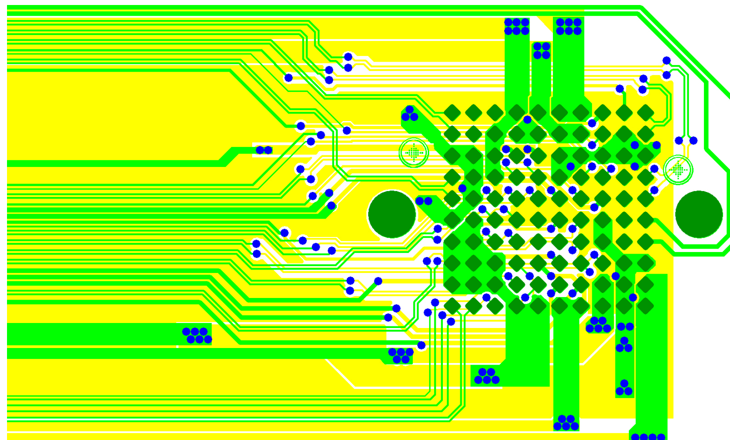


Figure 5.8: Layout of ZA1 connector outside the wafer area. The dark green squares are the ZA1 pins on top (green). The blue points are the PTHs to the back of the board (yellow).

[*Cadence*, 2017a] returned better impedance values. Nevertheless, they still were out of the 10 % range. In the Allegro Impedance Calculator the impedance values for the inner reticles could not be calculated because it is not possible to add a floating metal layer to the stack up. This floating plane is the ZA1 holder lying on top of the DENSE board. Therefore, only values for the CIC area were calculated with the Allegro tool.

Layer	Line Width [µm]	Line Separation [µm]	Sonnet Impedance [Ω]	Allegro Impedance [Ω]
Top (over silicon)	100	50	84.0	-*
Top (over CIC)	80	90	110.7	104.1

Table 5.2: Line geometries of the differential lines on the DENSE board. The target differential impedance is 100 Ω.
(*) : Stack up with floating metal plane on top is not possible to simulate in the Allegro Impedance Calculator.

One reason for the mismatch is that the calculations where done without solder mask in the stack up. The solder mask is only 35 µm thick, but in this stack up with the thin prepregs it has an high impact. Another mistake was the assumption that there is air above the inner reticles. The mechanic setup was designed after the layout of the DENSE board was finished. The ZA1 holder above the DENSE board serves as a shielding layer for the signal lines which reduces the impedance of the

Layer	Line Width [µm]	Sonnet Impedance [Ω]	Allegro Impedance [Ω]
Top (over silicon)	100	57.1	-*
Top (over CIC)	80	71.5	68.9

Table 5.3: Line geometries of the single-ended lines on the DENSE board. The target impedance is 50 Ω.
(*) : see table 5.2.

differential lines too.

The corrected line geometries for the stack up are in table 5.4 and table 5.5. Again the Allegro Impedance Calculator was used to cross-check the values from the Sonnet software. Both tools return sufficient well values for the area above the CIC material. A difference of 5 Ω is acceptable for impedance calculations between the tools. They use different approaches for the impedance calculation. In the Sonnet tool the complete stack up is divided into small cells which leads to discretization errors [*Merrill*, 2015]. On the other hand the Allegro Impedance Calculator is a static solver and, therefore, some assumptions are put into the calculations which can lead to deviations [*Cadence*, 2009].

Layer	Line Width [µm]	Line Separation [µm]	Sonnet Impedance [Ω]	Allegro Impedance [Ω]
Top (over silicon)	100	125	100.8	-*
Top (over CIC)	130	120	99.1	93.8

Table 5.4: Correct line geometries for differential lines with a target impedance of 100 Ω.
(*) : see table 5.2.

It is possible to measure the impedance of signal lines on a board with test structures. These structures would be additional to the existing layout and only for this purpose. On the DENSE board it is not provided as there is no interest in the exact impedance value.

Layer	Line Width [µm]	Sonnet Impedance [Ω]	Allegro Impedance [Ω]
Top (over silicon)	120	52.5	-*
Top (over CIC)	150	54.9	53.0

Table 5.5: Correct line geometries for the 50 Ω single-ended lines.
(*) : see table 5.2.

5.3 Test-Hardware

The HICANN chip cannot be used alone. It needs an FCP board for the communication with a host computer. The Electronic Visions Group together with the Dresden University of Technology (Dresden University of Technology) group developed a test device for single HICANN chips, called the Cube Setup. The signals and supplies are the same as for a dual HICANN chip on the DENSE board. Therefore, the easiest way to communicate with the chips on the DENSE board is by using the existing Cube Setup. All it needs are some additional boards to adapt the connector of theCube Setup to the connector on the DENSE board. The boards are described in section 5.4.

In fig. 5.9 all developed components except the ZA1 connectors for the prototype system are shown. For the design of the components the 3D computer-aided design (CAD) program SolidWorks was used [*Dassualt Systèmes*, 2016]. All the mechanical components like the DENSE mechanic frame, the ZA1 connector stencil, the vertical holder and the Double Dense Connector board (DDC) holder were produced at the mechanics workshop of the Kirchhoff-Institute for Physics (Kirchhoff-Institute for Physics).

Cube Setup The Cube Setup is a replication of the BrainScaleS system on a smaller scale. Instead of a complete wafer it is equipped with single HICANN test chips. The HICANN chips communicate with FCP boards like in the BrainScaleS system. There are four FCP boards, but only two of them can be connected to eight HICANNs each. In contrast to the BrainScaleS system the only missing thing are the interconnections between the HICANNs.

The user of the Cube Setup perceives no difference in the usage of the Cube Setup than in comparison to the BrainScaleS system. The software stack is completely the same.

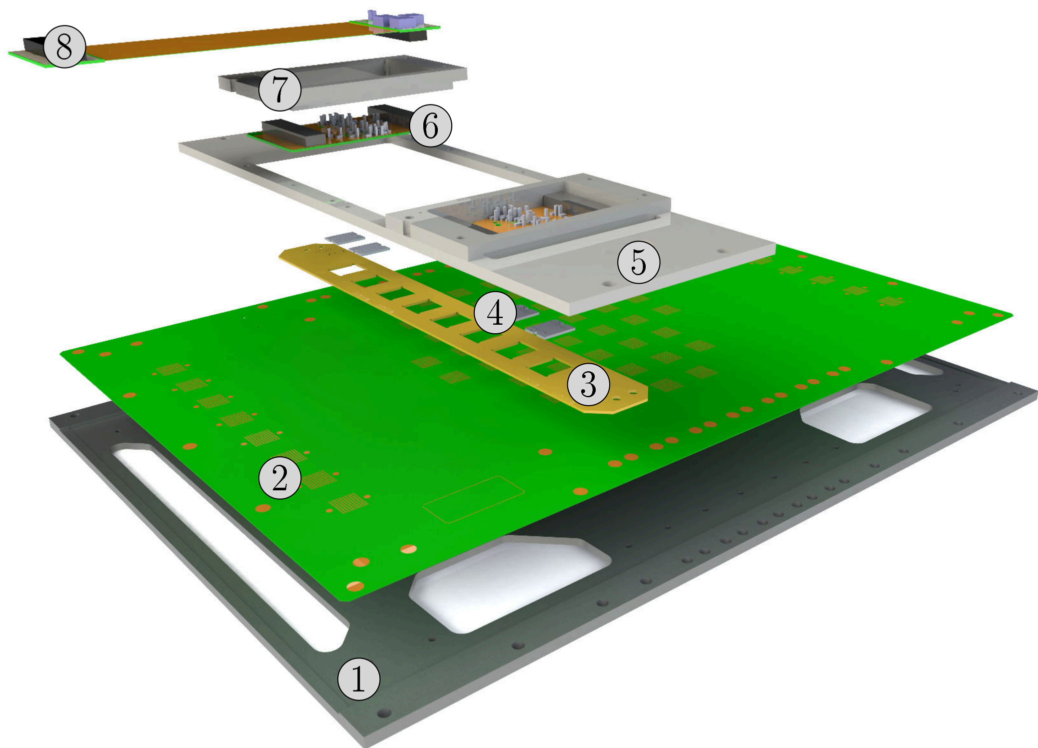


Figure 5.9: Explosion view of the DENSE prototype system. (1) DENSE mechanic frame, (2) DENSE board, (3) ZA1 stencil, (4) Samtec ZA1 connector, (5) DDC vertical holder, (6) Double Dense Connector board, (7) DDC holder, (8) IBoard-to-DDC flex-rigid board

5.4 Connect DENSE Board with Cube Setup

The new key component in the prototype system is the ZA1 connector. It is used as a vertical connector between the DENSE and the DDC board like the elastomeric connectors in the WSI system. The goal is to have solder-less and reliable connections, whereas the connector can be reused. The DDC board connects to the iBoard via the IBoard-to-DDC flex-rigid board (Flexadapter).

In fig. 5.10 an assembled DENSE prototype connected to the Cube Setup is shown.

IBoard-to-DDC flex-rigid board In principle, the Flexadapter is an extension cable. It is a flex-rigid board with two rigid areas at the ends of the board. In between is a 20 cm long flexible area with two copper layers. On one side is the connector to the iBoard and on the other is the connector to the DDC board. The pin-out of both connectors is identical.

On the board there are ten differential pairs and eight single-ended lines which

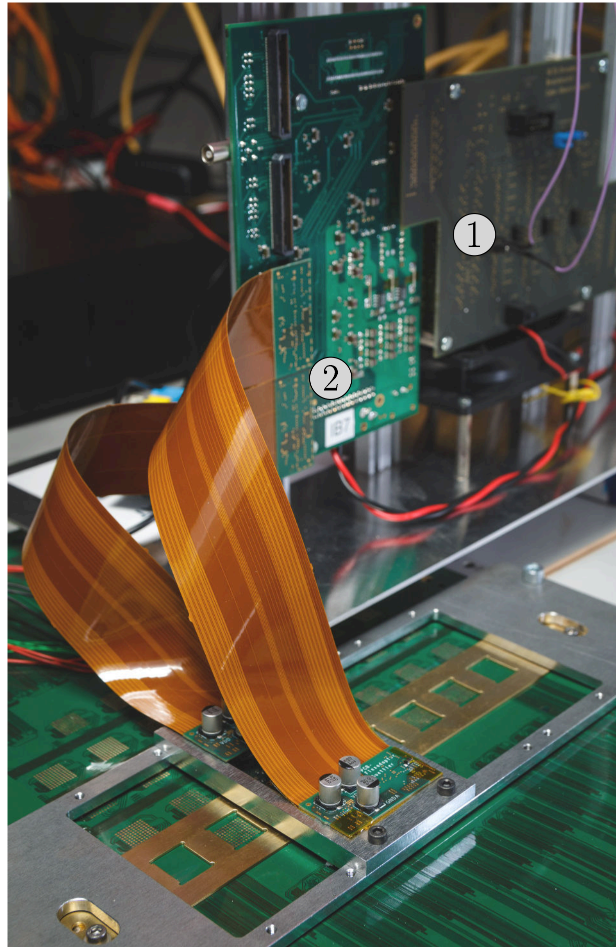


Figure 5.10: Prototype system in the laboratory. The two Flexadapters connect the DDC board with the iBoard. In this configuration the FCP board board can communicate with four HICANNs. An additional DDC board and two Flexadapter boards would allow the maximum connectivity with eight HICANNs.
In the background is the Cube Setup (1) with one iBoard (2).

have to be impedance-controlled. The signals run on the top flex layer and GND shapes on the bottom flex layer are reference planes.

Double Dense Connector board The DDC PCB has six copper layers and has a length of 7.5 cm and a width of 4.0 cm. It is responsible for the adaptation of the pad array from the DENSE board to the Flexadapter connector. Additionally, it is equipped with blocking capacitors for the power supplies of the chips. There are two capacitors for every voltage: a 0402 capacitor with 100 nF and a 0603 capacitor with 10 µF.

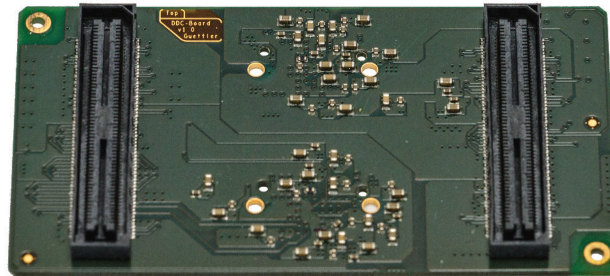


Figure 5.11: Photograph of the Double Dense Connector board. The left and right Samtec connectors go to the Flexadapter. In the middle are the blocking capacitors for the four HICANN chips. On bottom is the same pad structure as on top of the DENSE board.

Due to the size of the Flexadapter connector it is reasonable to design the board so that it covers two reticles on the DENSE board. Hence, it is possible to communicate with four HICANNs over one DDC board.

It is possible to place a second board next to the first one and, thus, to be able to talk with eight HICANNs at the same time.

Samtec ZA1 The ZA1 connector creates vertical connections between two boards. The connector consists of an interposer board with spring contacts on both sides. The pitch between the pins is 1 mm. The pads have a width and height of 0.62 mm and are rotated by 45° [*Samtec*, 2017b]. The maximum tolerance in one direction is about 0.4 mm to have a reliable contact. Here the smallest configuration with 10 x 10 springs is chosen, see fig. 5.12. Other configurations are larger than the available reticle area.

The spring contacts of the ZA1 connector need to be compressed to establish a reliable vertical connection. The force for the compression is applied from the fixated DDC board. The final vertical position of the DDC board is calculated to be at the point where the ZA1 connector is compressed to 1 mm.

The standard way to align a ZA1 connector is by using register pins, which go through the holes in the connector and all other components. This is feasible for the chips of the DENSE board, where the ZA1 connector pads are outside the wafer area.

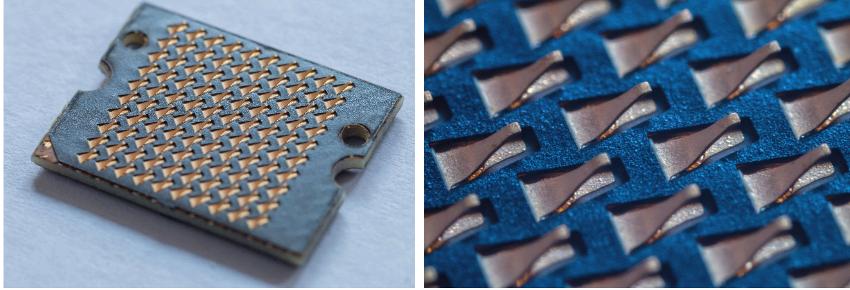


Figure 5.12: The left image shows the top of the ZA1 connector. This connector configuration is the smallest of the series, it has ten rows with ten spring contacts each. On the bottom is the same arrangement of spring contacts as on top. Through the holes are the fiducial marks on the DENSE board visible. The right image is an enlargement on the spring contacts.

There holes for the register pins are possible. Unfortunately, for the chips with the pads directly on top of the wafer it is not possible to add the holes. These holes had to be drilled though the silicon wafer which would destroy the circuits on it. Therefore, another method for the alignment for the ZA1 connectors was developed. A stencil with cut-outs for the connectors is produced. It is fixated at the upper and lower edge of the DENSE board and has slots for eight ZA1 connectors. The advantage is that once the position is correct, the position of all eight slots is correct. The height of the stencil is exactly 1 mm like the height of a compressed ZA1 connector. Hence, the stencil prevents an over compression of the spring contacts on the interposer.

Construction In contrast to the BrainScaleS system can the construction of the DENSE prototype setup can be done in any laboratory and does not need to be in a clean room. The silicon wafer is air-tight encapsulated and mechanically protected by the PCB.

First, the DENSE board is screwed on tightly with the mechanic frame. Next the ZA1 stencil gets positioned. One ZA1 connector is inserted into one of the slots in the stencil. Now the stencil is moved until the fiducial marks on the DENSE are visible through the holes in the ZA1 connector and then fixated. Additional ZA1 connectors can be put into free slots. The vertical holder is placed over the stencil and screwed down with the mechanic frame. The upper and lower edge of the DENSE frame is higher than the stack of DENSE board and ZA1 holder, so that the vertical holder lies on top of them. Then the package of DDC and DDC holder is positioned within the vertical holder.

By tightening the screws of the DDC holder the DDC board is lowered down. During this process the ZA1 connector gets compressed to 1 mm. At the lowest position the distance between the DDC board and the DENSE frame is the height of the DENSE board and the ZA1 connector, see fig. 5.13. At last the Flexadapter is plugged into the DDC and the iBoard. The prototype is ready and the HICANNs

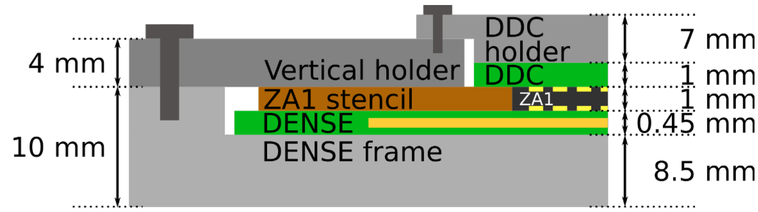


Figure 5.13: Cross section of the DENSE prototype system. The edge of the DENSE frame is as high as the stack of DENSE board and ZA1 connector.

can be turned on.

5.5 Test Methods

The accessible chips on the DENSE board are comparable to the single HICANN test chips. Therefore, the tools testing and characterizing a single HICANN can be used for the DENSE prototype. It saves time and also helps improving the software stack.

The tests used are described in the following sections. At first simple tests should prove that it is possible to communicate with the chips. Then the analog parameter storage is checked for proper working. At last a neuron calibration should show that the neuron behaves like expected from simulation.

5.5.1 LVDS Link Resistance

Every high-speed LVDS link has a termination resistor between its n- and p-line, see fig. 5.14. It is intended to match the output and input resistance of the driver and receiver circuit with the differential line impedance Z_{diff}. The resistance is measured easily with a multimeter. The correct value is around 120 Ω.

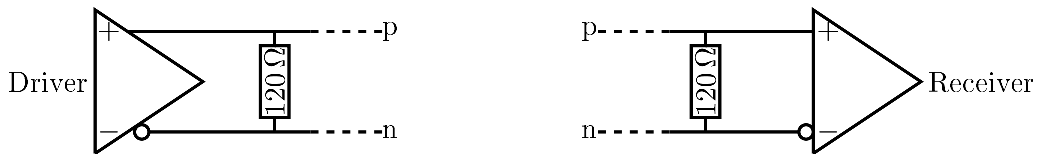


Figure 5.14: Termination resistors in the LVDS links. The dashed lines go to pads on top of the DENSE board where they can easily be accessed.

This simple test shows if the vias between the silicon wafer and the copper layer exist and work properly.

Every pin of the chip has ESD diodes at his input. It would be possible to operate these diodes in forward-mode and measure the voltage drop over the diodes like on the test boards for the BrainScaleS system. This way additional vias to the wafer could be checked, e.g. of the power supply pads. On the other hand the effort is higher

than for the termination resistor as an additional test board has to be developed, produced and tested. Therefore, this measurements were not conducted.

5.5.2 Switchram Test

The first digital test performed is the Switchram test written by Andreas Grübl. The test writes random data in the custom Static Random Access Memory (SRAM) cells of the crossbar switches. The data is read back and compared with the sent data. The test itself is simple, but it covers several topics. At first it checks the communication with the wafer. It reads the JTAG-IDs of the HICANN chips and the FCP board. This ensures that the communication via the JTAG channel is working and stable. Without it the initialization of the high-speed LVDS links would not be possible. The exchange of the link properties between the FCP board and HICANN chip is done via JTAG. The Automatic Repeat reQuest (ARQ) protocol can be used as an additional communication layer on top of the physical link layer [*Philipp*, 2008; *Karasenko*, 2014].

If the communication channel is set up the data is sent down to the chip. Each packet contains an address and a data word of 16-bit width. In total 112 addresses are tested. Next the data is read back and checked if it is correct.

Errors can occur on different levels. A HICANN ID of 0x00000000 or 0xffffffff, for example, is most-likely an open connection or a short in the TDI/TDO line.

Another advantage of the test is that the communication can fallback to a lower level. If there is a problem with the high-speed link, the software returns to the JTAG interface. The test takes longer as the throughput of the JTAG interface is much lower than of the high-speed links.

5.5.3 Floating-Gate Cell Test

The Floating-gate MOSFET (FGMOS) is a modified field-effect transistor which has a source and drain port like a standard MOSFET. In contrast to the MOSFET is has two gates. One is completely isolated, called the floating gate, and the other lies on top of it, the control gate. The state of the floating gate defines the behavior of the transistor. If there is no charge on the floating gate, the cell behaves like a typical MOSFET. By adding electrons to the floating gate the threshold voltage of the transistor gets shifted. The amount of charge on the floating gate controls the resistance of the underlying channel. The charging or discharging is done by Fowler-Nordheim-Tunneling or hot-carrier injection. A detailed introduction can be found in [*Pavan et al.*, 2004]. In the BrainScaleS system the floating gate cells are used as an analog parameter storage. There are voltage cells which have a programmable output voltage between 0 V and 1.8 V Additionally, there are cells with a current mirror in the output stage, which converts the stored voltage into a current.

The HICANN chip has four floating gate blocks. Each block consists of 24 rows where every row has 129 floating gate cells. The even rows are voltage cells and the

uneven are current cells. In total there are $4 * 24 * 129 = 12384$ storage cells on one HICANN chip [*HBP SP9 partners*, 2014].

The configuration of the neuron's behavior is mainly done via the floating gate cells. They store the resting and reset potential, the refractory period etc. Due to the large number of cells involved problems induced by the lamination process would be visible. The conducted test was written by Sebastian Schmitt. It draws for each cell a random value, then the cells are programmed to these values. Afterwards the stored value is read back with the AnaRM. The software compares the digitized value with the initially chosen value. Hence, a divergent behavior of the cells would be visible. The error pattern would be the same as for the Switchram test. Broken cells return a wrong value.

5.5.4 Neuron Calibration

The most complex task is the calibration of a HICANN chip. The goal of the calibration is to have a stable and reproducible set of parameters for each neuron, so that it behaves like the underlying neuron model and that the variations between neurons are minimized. At the time of this work the calibration was capable to match the neuron's behavior to the Leaky Integrate-and-Fire model. The extension to the full capabilities of the Adaptive exponential integrate-and-fire Model is work by Mitja Kleider [*Kleider*, 2017]. The calibration software is developed by numerous persons of the Electronic Visions Group. A detailed description of the neuron calibration is in [*Schwartz*, 2013; *Schmidt*, 2014]. Both are based on the HICANNv2 chip and are not calibrating the synaptic input of the neuron. The synaptic input calibration was developed by Christoph Koke for the newer HICANNv41 chip [*Koke*, 2017]. In his work he also improved the calibration of the neuron, which was implemented in the HICANNv2 calibration too.

In the following paragraph an introduction into the calibration process is given.

The general behavior of the circuits is proven with software simulations before the chip production. Unfortunately, the components on the silicon substrate, like transistors or capacitors, are not perfect. They have small time-independent variations coming from the manufacturing process, e.g. variation in size or doping. In general, these fixed pattern variations do not disturb the behavior of the digital circuits but in analog circuits they can have an influence. There are also time-dependent variations like changes in temperature and supply voltages. However, the system is specified to work in a defined temperature range around 50 °C and the power supplies deliver voltages with a minimum of fluctuations. Hence, these parameters can be neglected for the calibration process.

Each neuron has 23 individual parameters to compensate the effects of the process variations. The neuron parameters are listed in table A.6. As mentioned in the previous section the neuron parameters are stored in floating gate cells.

The procedure for the calibration is split into two parts. In the first part the behavior depending on a parameter is recorded. Therefor, the parameter is set to a value and the reaction is recorded. This can be repeated multiple times and for

different working points.
The observation possibilities on the HICANN chip are limited, not every part of the neuron circuit can be individually recorded. It is possible to record the floating gate cell values and the membrane voltage of the neurons. Thus, certain parameters are only indirectly measurable. These parameters are observed by their influence on the membrane voltage. From this data correction values are calculated, which should compensate the inhomogeneities.

The second part of the calibration is the evaluation run. It is verifies the obtained correction values. The same method as before is executed only with the corrected parameters. It is checked if the result lies within an excepted range. If this is not the case, the neuron is put on a blacklist. At the end the software returns a complete list of neurons which could not be calibrated. This is a good indicator for the state of the HICANN chip. Furthermore, the calibration software creates a database with the offset values, fit and adaptation parameters for the circuits.

In the following the calibration of the three simplest parameters is described.

Resting Potential E_L The first parameter that is calibrated is the neuron's resting potential E_L. It is a very simple procedure, but depicts the working principle well.

The resting potential is controlled by one voltage floating gate cell. At first the floating gate cell is programmed to a specific target value. The neuron is left in its resting state, so all inputs are disabled. Then the mean membrane voltage is measured. This is repeated for different target values. The fit result provides an offset value which is used as correction value for future settings.

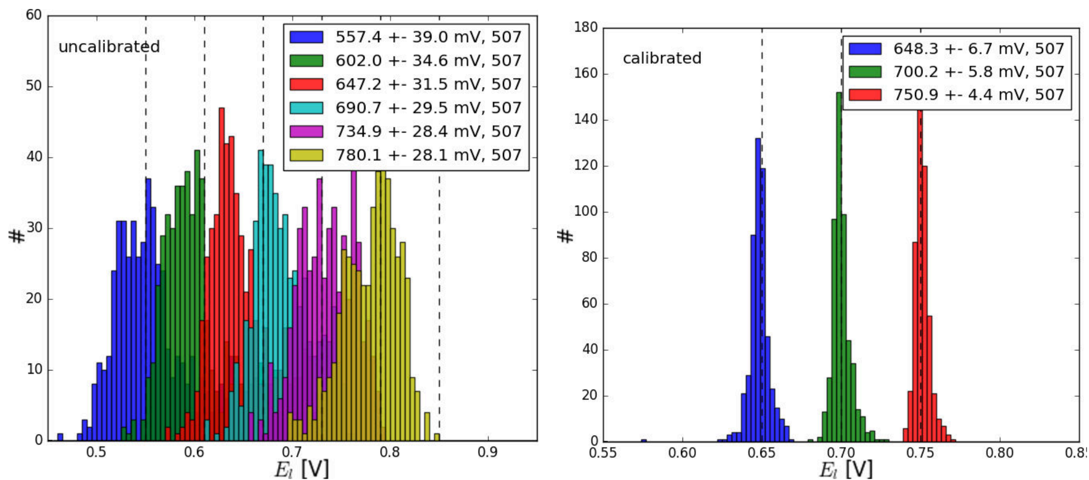


Figure 5.15: Result of the calibration on the resting potential distribution. The distributions are narrower and their mean value is closer to the target value. In the right plot not all target values are shown.

The result of one calibration run is shown in fig. 5.15. Through the calibration the mean value over all neurons and the spread of the distribution is improved.

Reset Potential V_{Reset} The reset potential V_{Reset} is one of the voltages which cannot be directly measured with the AnaRM. It is only visible in the membrane trace for a short time period after the neuron spiked. Therefore, the resting potential E_l is set above the threshold V_t, so that the neuron is constantly spiking. The time for the reset potential to be visible is defined by the refractory period τ_{ref}. The longest period is chosen to increase the visibility to the maximum and give the membrane voltage enough time to settle to V_{Reset}. In fig. 5.16 the membrane trace of a single neuron is shown. For the calibration only the trace between the drop of the mem-

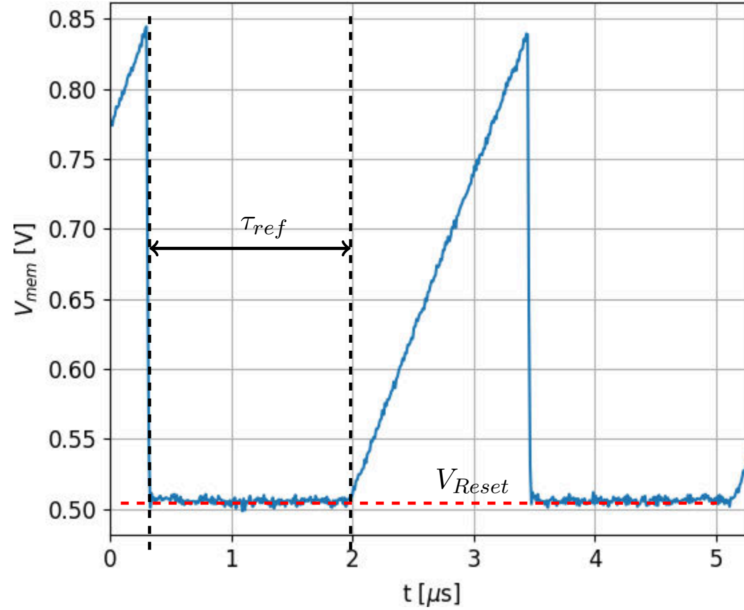


Figure 5.16: Single trace from a firing neuron. The neuron spikes at $t_0 = 0.2\,\mu s$ and stays at the reset potential V_{Reset} for the refractory time period τ_{ref}. From this part of the trace the reset potential is measured. This trace is from HICANN 133 on DENSE002.

brane voltage and the next rise is taken, from this part of the trace the mean value is calculated. A detailed description of the extraction method is in [*Schmidt*, 2014]. This is repeated for all neurons in one floating gate block. Then the mean over all of the values is calculated. The reason is that there is only one reset potential V_{Reset} for all neurons in one floating gate block.

Threshold Potential V_t The method for the threshold potential V_t is the same as for the reset potential only this time the algorithm looks for the local maximum in the membrane trace. The spike trigger circuit is a simple comparator which evaluates the membrane voltage against the set threshold voltage V_t. Hence, the information of the threshold voltage is decoded in the height of the membrane voltage before a spike is triggered. In fig. 5.17 the principle is shown for three different settings of

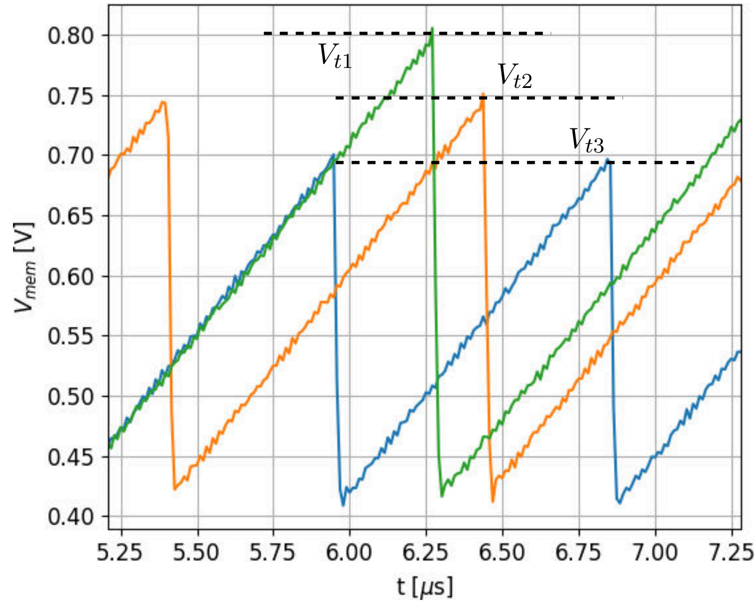


Figure 5.17: Threshold measurement for three different settings. V_{t1} is set to 0.8 V, V_{t2} to 0.75 V and V_{t3} to 0.7 V. These traces are from HICANN 245 on DENSE002.

the threshold potential V_t. The refractory period is changed to the smallest value to record as many spikes as possible to increase the quality of the statistic.

There are more sophisticated calibration algorithms for other parts of the neuron circuit. The software calibrates the neuron parameters V_{Reset}, V_t, E_l, the excitatory synaptic reverse potential E_{synx} and the inhibitory synaptic reversal potential E_{syni}.

The calibration software is a holistic tool for the characterization of a single chip. It shows deviations from the expected behavior, therefore, is a good test method for the digital and analog circuits on the HICANN chip. If there was an influence of the embedding process on the neuron circuits, it would be visible in the calibration data.

5.5.5 Thermal Stress

One of the critical points in the usage of wafer embedding is the long-term reliability of the connections to the wafer. The is no experience of how the board will change during the lifetime.

The main reason for failures is the thermal stress to the connections because of the different coefficients of thermal expansion. As depicted in fig. 3.3 with each heating and cooling cycle the probability of a connection problem increases. The steady thermal states are not a problem, the temperature changes induce the most stress to the board as the materials expand differently.

To simulate the lifetime of a board it is placed in the climate cabinet MK53 of the ASIC laboratory at the Kirchhoff-Institute for Physics [*Binder*, 2017]. In the climate

cabinet the board is repeatedly heated up and cooled down. In the later use-case the temperature of the board is between room temperature 20 °C and 60 °C [*Schemmel*, 2016]. For the climate cabinet test the temperature range is extended to 15 °C and 90 °C. This is almost the same temperature range of commercially available products which is from 0 °C to 85 °C [*Intel*, 2017b].

Another question is, can the board withstand several cycles in a soldering furnace without problems? For the test the soldering furnace of the electronics workshop of the Kirchhoff-Institute for Physics is used. It is a Quicky 450 from ASSCON Systemtechnik with a self-programmed temperature profile. The furnace is programmed to hold a maximum temperature of 230 °C for 2.5 min [*Schmitt*, 2017]. Possible errors could be a delamination of the layers or a tearing off the contact vias from the silicon wafer.

5.5.6 Test of the Redistribution Layers

The main purpose for the application of RDLs on the wafer was the interconnecting of separated HICANN chips. Therefore, fine-pitch lines run across the chip borders. Unfortunately, these fine-pitch lines cannot be tested with the current prototype setup. For a test two adjacent HICANNs from different reticles have to be connected to the Cube Setup. Moreover, they have to be connected over the horizontal fine-pitch lines. Only this way the existing software can be reused. Other methods are possible but required additional hardware and software adaptations. However, the requirement to connect two horizontal adjacent HICANNs with the Cube Setup is not feasible with only one copper layer. It is not possible to route all signals to the ZA1 connector. Therefore, a new board with at least two copper layers is needed to conduct tests with the fine-pitch lines of the RDLs.

There are lines routed on the RDL inside the HICANN area, e.g. the JTAG data output TDO of the upper HICANN goes to the JTAG data input TDI of the lower HICANN. Additionally, the JTAG clock TCK and mode select line TMS run in the RDL. If one of the lines breaks during the embedding process, the JTAG communication of the two chips would show failures. This would indicate that there is a problem with the RDLs. Furthermore, the power supply is routed through the RDLs stack up. If the embedding destroys connections of the supply voltages, it is observable by malfunctions in the chip circuits. For example, without the VDD12 voltage the floating gate cells value cannot be changed. This would be visible in the data of the floating gate test. Hence, it is possible to evaluate the influence of embedding procedure on the RDLs without the fine-pitch lines.

5.6 Experiments on HICANNv2

In principle it is possible to run neuronal network experiments on the HICANN chip. Unfortunately, the implementation of the synaptic input in the version 2 of the HICANN chip has some restrictions. First the dynamic range of the synaptic weights is small. Although, the smallest weight in the synapse is chosen the synaptic

input circuit saturates after a few synaptic events. For the biggest weight only one event is needed to saturate the input. The reason is an undersized capacitor in the circuit which integrates the events. Furthermore, the synaptic conductance is weak compared to the leakage conductance. This means a lot more events are needed to excite a neuron to spike. Detailed descriptions of the various effects can be found in [*Kiene*, 2014; *Koke*, 2017]. The restrictions have an effect on the neuron usage too. In fig. 5.18 a rasterplot of a simple feed-forward network from the work of Dominik Schmidt [*Schmidt*, 2014] is shown. The network consists of several populations. Every population has 12 neurons which are connected in a all-to-all scheme with the next population. Hence, every neuron sees the complete neuronal activity of the previous population. From the rasterplot it is visible that the neurons quickly return in an inactive state, if they do not receive any input. Therefore, the application of

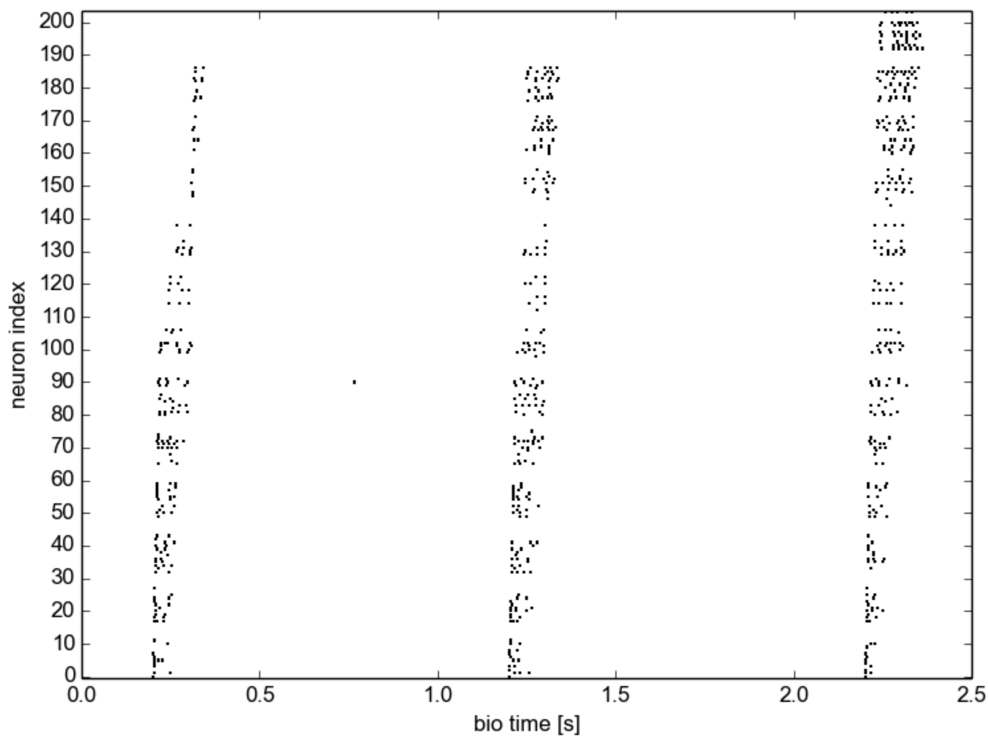


Figure 5.18: Rasterplot of a feed-forward neural network emulation with 200 neurons across two calibrated HICANN chips. Spikes are sent to population 0 (located on the lower end) in one second intervals and propagate upwards through the chain. Figure and caption taken from [*Schmidt*, 2014].

the neurons with their synapses in network experiments is limited. Together with the floating gate variations the neurons do not work reliable. The effort to set up

and tune a neuronal network can be seen in the theses of Dominik Schmidt, Tobias Nonnenmacher and Denis Alevi [*Schmidt*, 2014; *Nonnenmacher*, 2015; *Alevi*, 2015].

In 2015 a new HICANN chip was produced which has improvements in the synaptic circuit. This chip's neurons have a larger working range and a more reliable behavior. Thus, allowing more reasonable experiments like a classification of handwritten numbers of the MNIST dataset with a deep neuronal network [*LeCun and Cortes*, 1998; *Schmitt et al.*, 2017].

The focus of this thesis is the embedding of HICANN wafers in PCBs and not the tuning of neuronal networks by hand. This tuning has to be redone for every chip and consumes a lot of time. While the gain of information about the embedding procedure is low. Therefore, no neuronal experiments are run on the embedded HICANN wafers. The chosen tests allow a characterization of the HICANN chips without the usage of the synapses. The results gathered from the floating gate and neuron circuits are applicable on the synapse circuits. It is not intended to use the DENSE prototype system in the daily productive environment. Hence, until it is required to run neuronal networks with synaptic events on the DENSE prototypes the HICANNv2 chip is sufficient for the embedding experiments.

5.7 Conclusion

The DENSE prototype is the first PCB with an embedded HICANN wafer. The main goal of this prototype is to see the influence of the embedding process on the circuits and RDLs. With 122 HICANNs a fourth of the wafer is accessible. The chips are equally distributed over the wafer. If there are local problems due to the embedding process, they are detectable with the chosen tests. Especially, the high number of floating gate cells and neurons circuits allows an assessment of the embedding procedure.

6 Results of the DENSE Prototypes

This chapter presents the test results of the DENSE boards. In total two HICANN wafers were embedded in PCBs. The DENSE002 board is shown in fig. 6.1.

At first a problem which occurred in the lamination process and its solution is described. Then the results of the software tests for both boards are presented. At the end the results of the reliability tests are shown.
All tests were conducted and evaluated at the Electronic Visions Group in Heidelberg. In the appendix F the used hardware, software modules and all settings for the experiments are listed.

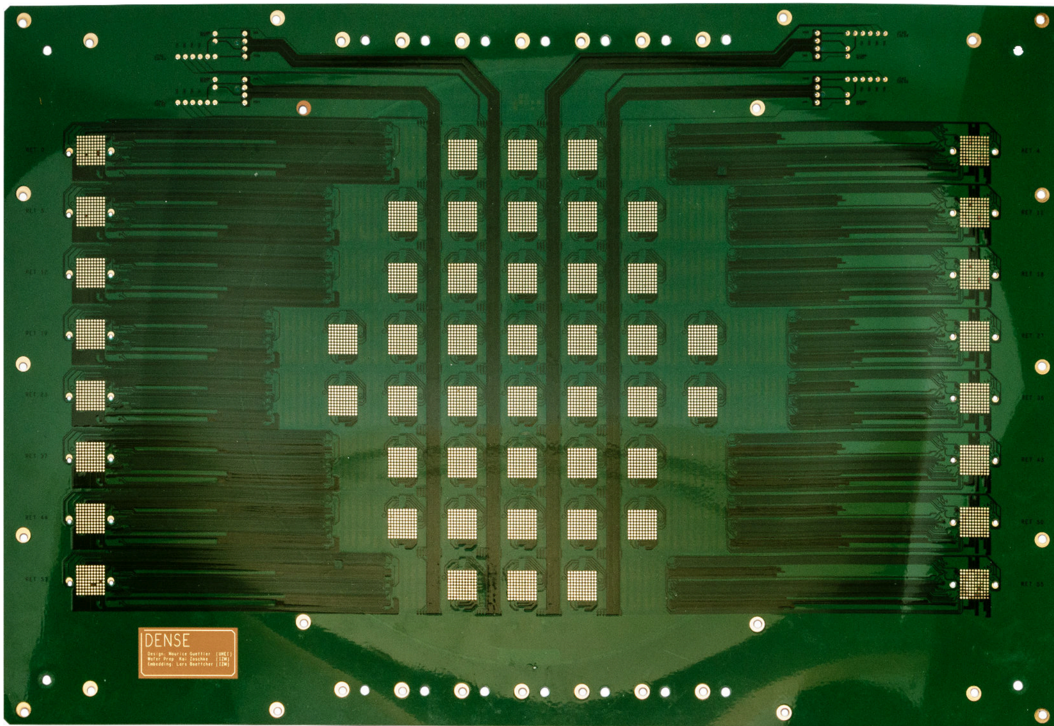


Figure 6.1: Top view of the DENSE board.

6.1 Problem during the Embedding Process

For the first time there was a problem with the embedding process. In the DENSE001 board it led to a breaking of the silicon wafer. The wafer has a single crack. It goes

from top to the wafer notch and divides it into two parts. The crack is shown fig. 6.2. Fortunately, some of the chips are still working and not effected. The reason for the crack was a rotation of the silicon wafer in the PCB. In the other DENSE002 board the wafer also rotated. Here the wafer did not break during the lamination procedure.

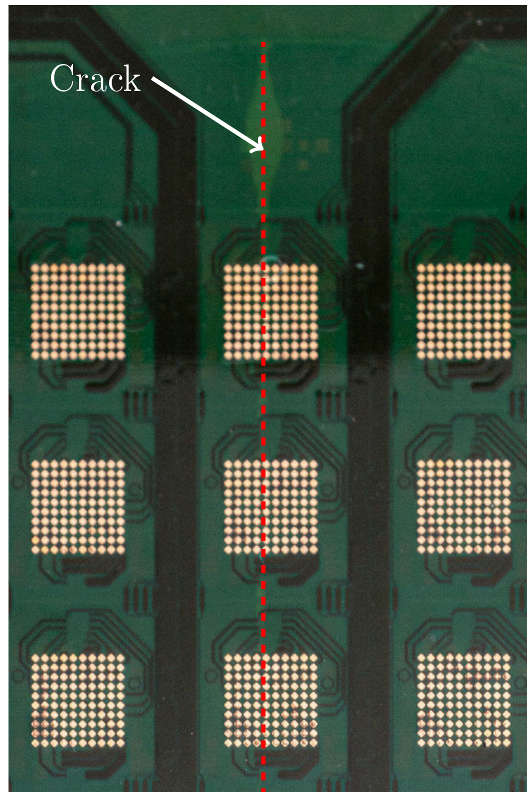


Figure 6.2: The DENSE board with the broken HICANN wafer. The crack goes from top to the wafer notch and divides the wafer into two parts.

Wafer Rotation The rotation and breaking of the silicon wafer happened during the lamination step of the board. It begins with the preparation of the CIC sheet. The first step is to cut out the shape of the wafer and drill the holes for the later PTHs in the CIC sheet. The cut-out for the wafer also contains the counterpart to the wafer notch. The notch together with the CIC sheet should keep the wafer in its position during the lamination process. The silicon wafer is placed into the cut-out and on each side a layer of prepreg and copper is applied. Then the stack up is put in the PCB lamination press. In the course of this the wafer slipped out of its position. The angle and the direction of the rotation are random. The wafer in the DENSE001 board is counter-clockwise and in the DENSE002 board clockwise rotated.

The manufacturing of the copper layout is not a problem. It is automatically

adjusted to the positions of the fiducial marks on the wafer. This is also true for the microvias contacting the silicon wafer. The problem are the vias going through the CIC sheet from top to bottom. These connections are created by two drilling steps. The first drilling happens during the CIC preparation and later after the lamination procedure for the actual vias again. Ideally, the holes of both runs are at the same positions. The positions of the first holes cannot be adjusted to the new wafer position. Therefore, they do not fit with the new layout which is rotated accordingly to the wafer. A drilling of new holes is not reasonable because the new holes would have contact to the CIC material and, hence, create shorts between the signals. The complete process is depicted in fig. 6.3.

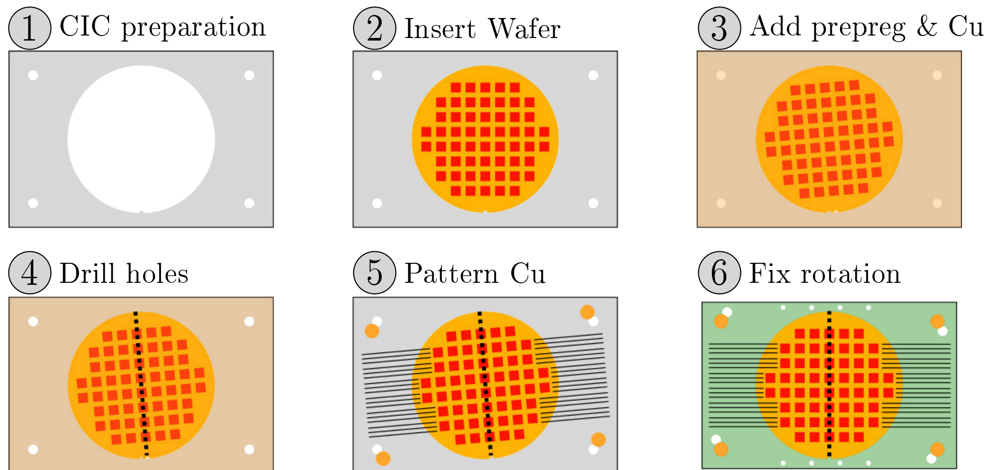


Figure 6.3: Schematic procedure of embedding a wafer with the focus on the rotation and the resulting problems. The dashed line represents the crack of the wafer in DENSE001.

Consequences The wafer rotation only effects the connectors outside the wafer area. The layout and the wafer contact vias of the inner reticles are correct. For the usage in the prototype setup the chips need to be in a vertical line as the mechanic system is not capable of compensating rotated connectors. Therefore, in a final step the board has to be cut out so that the wafer is rotated back into its intended position. This transition is done between the steps five and six in fig. 6.3.

All chips connected to the connector pads outside the wafer area are not usable. This reduces the total number of usable chips by 32 and 80 chips remain for tests. This applies for both boards.

Solution Due to the rotation the counterpart of the wafer notch on the CIC sheet gets under the silicon wafer. This creates an uneven distribution of the applied pressure during the lamination process. For the DENSE001 board this resulted in a breaking of the silicon wafer.

A solution is to remove the single notch with a flat, see fig. 6.4. A notch requires less substrate area but the area around the notch is not used on the HICANN wafer. Therefore, there is no drawback in introducing a flat. In the WSI system the notch was at first also used for the positioning of the wafer in the bracket. In the end the pin was removed due to variations of the notch's size between different silicon wafers. The idea of the flat is that even if the wafer is not in the correct position it slides

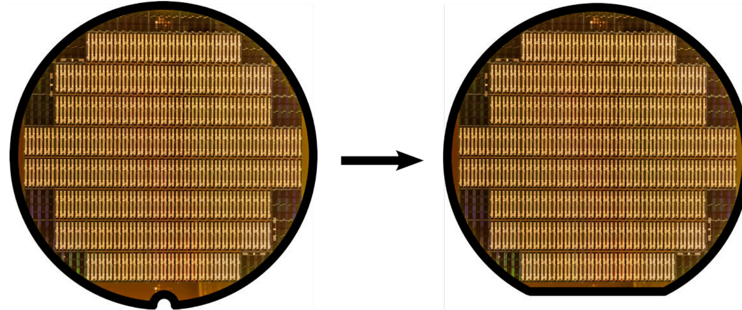


Figure 6.4: Solution to prevent the silicon wafer from sliding out of its position and rotating. The wafer notch is replaced with a flat. For a better presentation the notch and the flat are bigger drawn than they are in reality.

back into the correct position during the compression of the layers. The test with the flat will be done with the next prototype.

6.2 Warpage

There is a low warpage of the boards visible. The board edges are a few millimeters higher than the center of the board. The warpage is for both boards in the same direction towards the wafer top side. Nevertheless, the boards can be used in the prototype setup. The corners of the board are fixated with the mechanic frame. Additionally, the mechanic components for the connection to the DENSE board also push the board back into a flat state. Due to the thin board and the silicon wafer the applied pressure neither breaks the board nor the wafer.
So far no changes in the warpage during and after the tests was observable.

6.3 Switchram Test

At the beginning not all mechanical components were ready. In order to have some early results, it was decided to solder wires directly on the connector pads. As a consequence, simple JTAG tests could be run, e.g. read back of the HICANN-ID. In summary the tests were positive and showed no problems. Unfortunately, the pads are not usable anymore with the ZA1 connectors after the soldering procedure. For that reason on four of the 80 accessible HICANN chips on the DENSE001 no further tests are possible.

Results of DENSE001 The results of the Switchram test of DENSE001 with activated ARQ are depicted in fig. 6.5. A detailed list with the error messages of the single HICANN chips is in table D.1. The wafer crack is in the center of the silicon wafer and goes from top straight to the wafer notch. It is in the area of HICANN chips in the middle column, i.e. from 9 to 437. All these chips do not work properly. In all cases a wrong HICANN-id is returned. It is either 0x00000000 or 0xffffffff.

To check if the contacts to the wafer pads are broken all termination resistors of the LVDS links on the chips were measured. Two termination resistors could not be measured, they had an open connection. This was on HICANN number nine. The open connection would lead to problems with the high-speed communication but it does not describe the JTAG error. If there was an open connection in the TDO line, the received HICANN-id should be 0x00000000 and not 0xffffffff. On all other chips the termination resistance was correct, see table D.2. Thus, the wafer contact vias seem to have no problems. There is also no short between the data output pin TDO and the supply voltages detectable, see table D.3.

The supply voltage of the prototype setup is measured. It is 13.8 V and four working HICANN chips have a current consumption of 0.6 A. The broken chips consume 0.3 A more than the functioning chips. Hence, somewhere has to be a short which is responsible for the JTAG erros.

Without destroying the board and the wafer it is not possible to further specify the failure location. The board has to be sawed and looked at the cross sections. The failure could be on the substrate level or in the metal layers on top of the silicon. It is most-likely related to the breaking of the wafer due to the location of the chips and the crack.

The HICANNs 177, 213, 373 and 401 have problems with the JTAG communication. The chips returned a wrong HICANN-id. The same measurements of the termination resistors and the resistance between TDO and the supply voltages were conducted than before, see tables D.2 and D.3. There were no problems of failures detectable. Therefore, the error has to be in the chip. It could be possible that there are distortions on the chip from the breaking of the wafer.

The HICANN chip number 433 shows always link transmission errors. During the initialization process test data is sent down to the chip and immediately returned to the FPGA without any processing of the data. It checks if the found link parameters allow a stable and faultless communication. Here the data received by the FPGA is corrupt. Unfortunately, there is no explanation for this behavior so far. The termination resistances were in the correct range and an optical inspection showed nothing of significance. Therefore, the reason has to be within the HICANN chip because everything else is identical to the other chips on the board. It is possible that the data is corrupted at the input of the HICANN chip or during the transportation from the input receiver to the output transceiver. The remaining HICANNs 101, 117, 161, 197 and 389 have problems with the initialization of the high-speed links. They cannot find the appropriate link parameters which is indicated by a returned link status of 0x0a. This error occurs also on the WSI systems. At the moment no reason could be found for this error [*Mauch*, 2017].

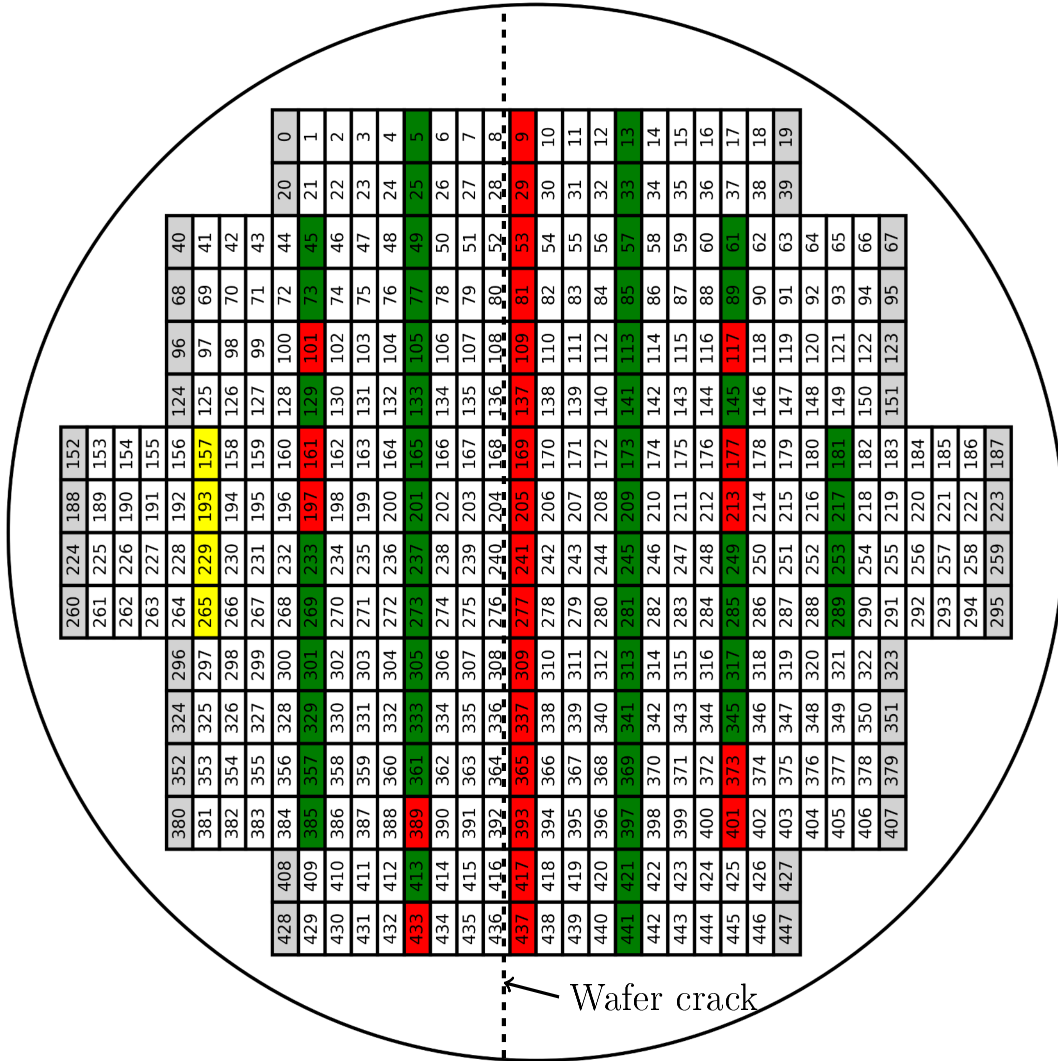


Figure 6.5: Wafermap of the Switchram test results of the DENSE001 board. The green HICANNs passed the test and the red ones failed it. There are 50 functioning HICANNs and 26 with problems. The inital JTAG tests made the yellow HICANNs not usable. The gray HICANNs are not accessible due to the wafer rotation.

In total around 65 % of the HICANN chips are working properly. The error pattern does not indicate that there is a systematic problem induced by the embedding process. Most of the errors seem to be related with the wafer crack. For cross validation of the results there is the DENSE002 board which has an intact silicon wafer.

Result of DENSE002 In comparison to the DENSE001 board the second prototype DENSE002 has a lot less errors. A wafer map with the test results of the Switchram test is shown in fig. 6.6. Only three HICANN chips, which corresponds to 4 % of all chips, are not working properly. The location of the errors do not overlap with errors on the first board. In table D.4 the detailed error messages are listed. The chips 5 and 29 return a false HICANN-id over JTAG. It is always 0xffffffff which seems to be a short to one of the supply voltages. This could not be validated by resistance measurements between TDO and the supply voltages, see table D.6. The termination resistors in the LVDS links have the correct values. Therefore, the connections to the wafer look fine. From the data a definite answer for the reason of problem is not possible.

The Switchram test fails on the HICANN 253 due to transmission errors on the high-speed links. On this HICANN not all termination resistances could be measured. In table D.5 the values are listed. For the clock line in sending direction (TX_Clk) it was not possible to measure the resistance so that it is an open connection. A closer look on the corresponding ZA1 pads revealed the reason for the open line. The RDL pads are not directly under the ZA1 pads. Therefore, copper lines run from the microvias at the RDL pad locations to the ZA1 pads. Here, an etching problem is visible. One of the lines misses a piece of its trace close to the ZA1 pad. In fig. 6.7 an enlargement of this area is shown. There are two possible reasons for the missing piece. First, the copper line had at this position a loose connection to the prepreg layer and was torn off during the etch bath. Second, the board was too long in the etch bath and removed in this area more copper than intended. The latter happens if the copper distribution is unbalanced across the board. This could be the case for the DENSE board. In the outer area of the board a lot more copper has to be etched away and, hence, needs more time. On the other hand in the area above the wafer is less etching required to create the copper structures. The location of the failure is at the edge between the inner wafer area with a low etch rate and the outer area with a high etch rate.
It can be solved by copper thieving. This method fills the empty areas with a checkerboard pattern of copper which balances the etch rate across the board.

Nevertheless, the failures on the DENSE002 board seem not to be related to the embedding procedure.

Conclusion There are no general problems with the communication over the JTAG and the high-speed links visible. Furthermore, the impedance mismatch of the signal lines is not a problem. The reason for that could be the length of the signal lines on

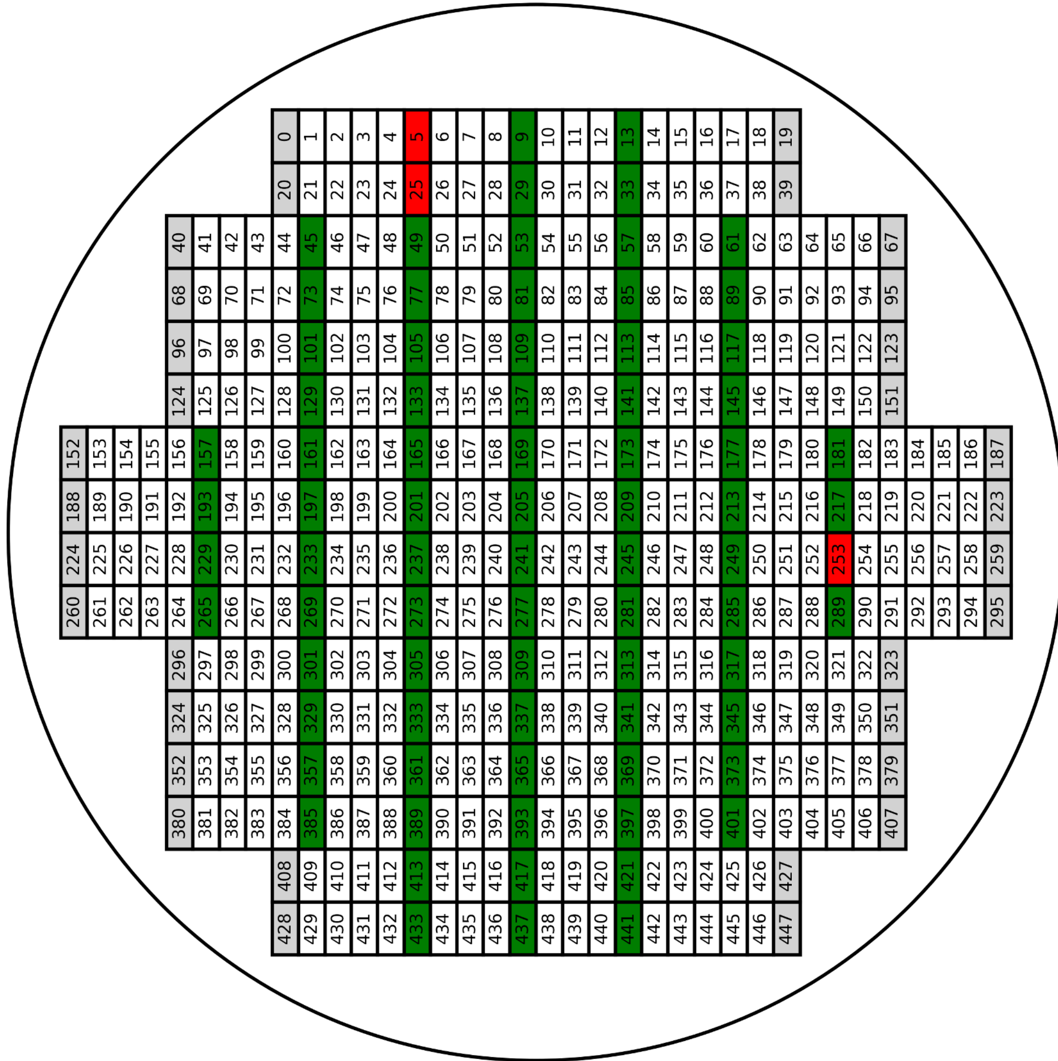


Figure 6.6: Result of the Switchram tests with activated ARQ of the DENSE002 board. The green HICANNs passed the test and the red ones failed it. There are 77 working HICANN chips and three broken chips. The gray HICANNs are not usable due to the wafer rotation and the white HICANN are no accessible.

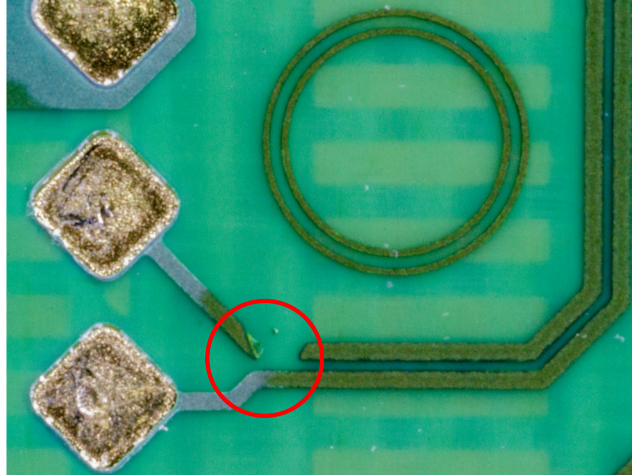


Figure 6.7: Photograph showing a missing piece of a trace. It is the sending clock line of HICANN 253 on the DENSE002 board. This is the reason for the transmission errors in the Switchram test.

the inner reticles. Most of the lines have their wafer contact vias directly in the ZA1 pads and the other lines are short. The longest trace has a length of 13.7 mm.
The high error count on the DENSE001 board is most likely related to the breaking of the silicon wafer as the second prototype shows almost no problems.
The missing trace on the DENSE002 board is an optimization problem of the board layout and not specifically related to the embedding procedure.

6.4 Floating-Gate Cell Test

The Floating-gate measurement is more complex than the Switchram test. It involves more circuits like the analog readout and the floating gate controller. On the other hand the test can easily be evaluated by measuring the voltage of the floating gate cells. In fig. 6.8 a typical result of the test is shown. The name of the rows are listed in D.7.

All four floating gate blocks are plotted separately and each rectangle in a block represents a single floating gate cell. The colors indicate the difference between the target value and the real value of the floating gate cell. The ideal result would be 0 V as a perfect match between set point and real value. In real world applications this is difficult to achieve as the programming time and the steps are finite. Christoph Koke achieved deviations in the range 1-8 mV for the voltage cells with optimized controller settings [*Koke*, 2017]. These settings were not yet available for these test runs so that the deviations are higher. It is not subject to this thesis to analyse the floating gate cell performance, therefore, the larger spread is accepted. The interest lies on a general perspective and a wafer-wide behavior analysis.

The set points are once randomly chosen for each cell and then reused on all chips.

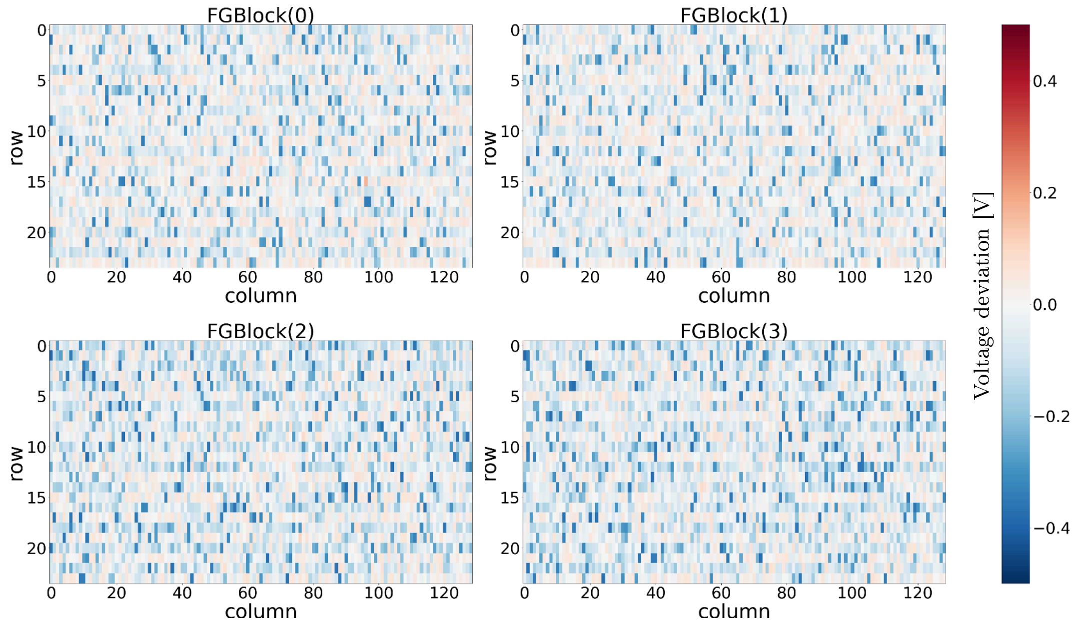


Figure 6.8: Floating gate test of HICANN 301 on the DENSE001. The color of a cell depicts the voltage deviation of the real value from the target value.

Random values have the advantage that the complete parameter range is tested at once. If only one value for all cells is chosen, the test has to be run with several values to dismiss the case that the set point is by chance the value of a broken cell. Furthermore, random values are less prone to spatial problems in the floating gate blocks.

The differences between chips should only be the device variations and the trial-to-trial differences.

Limiting Analog Readout Buffer In fig. 6.8 a high number of large negative deviations is visible. A comparison between the set value and the real value shows that the large negative deviations are always ocurring with set points above 1.4 V. This is not a problem of the floating gate cells they can reach 1.8 V. The problem is in the analog readout. The floating gate cell is not measured directly. There is an additional buffer circuit in the readout line. This buffer is intended to shield the floating gate cell output from disturbances due to the analog measurements. In the current implementation the buffer can only reach an output voltage of 1.4 V. Therefore, all input voltages above 1.4 V are clipped to the maximum output voltage. In fig. D.1 the distribution of the measured values is presented, the peak and cut-off at 1.4 V is clearly visible. For a better assessment of the results another visualization method was chosen. The data is plotted in a histogram showing the frequency of the deviations. In fig. 6.9 the gray histogram shows the data of fig. 6.8. From the histogram the minimal and maximal deviation is directly visible. Furthermore, the bin

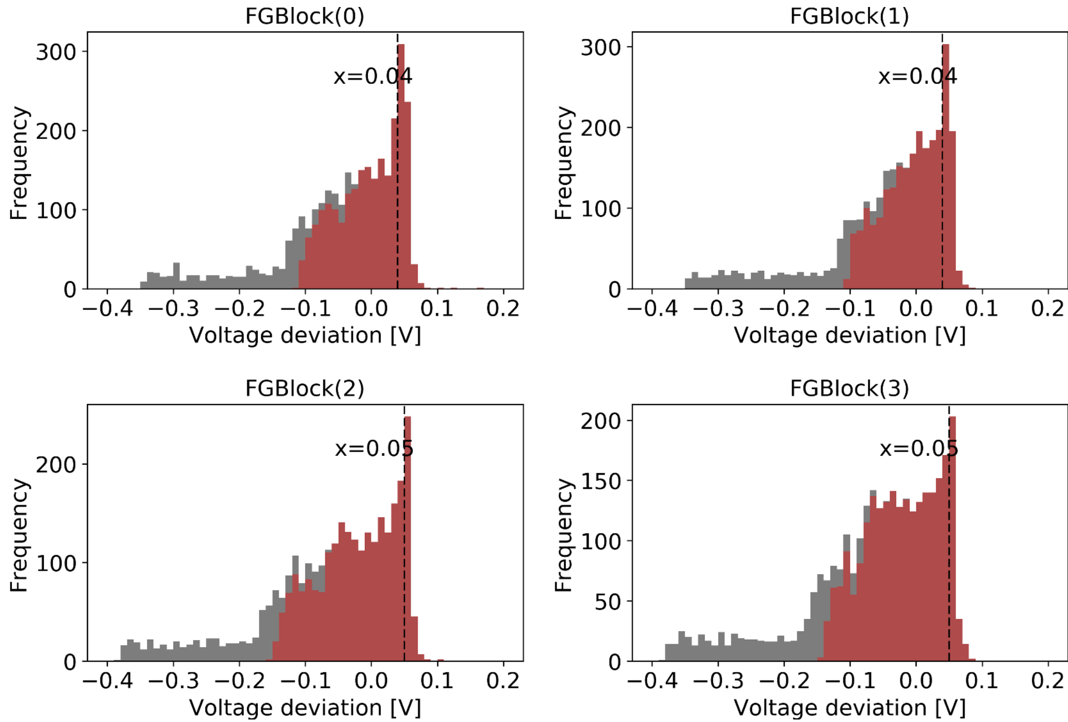


Figure 6.9: Histogram of floating gate test for HICANN 301 on DENSE001. The gray histogram shows all values of the test and the red histogram only uses the values with a set value below 1.4 V.

with the most frequent value is marked. Typical distributions have a peak around 40-50 mV. Above this peak is a clear cut-off and the frequency goes to zero. On the other side there is not such a clear limit. Here the frequency slowly decreases until a deviation of −0.13 V is reached. Then it drops down to a lower level where the frequency stays constant until −0.4 V. This plateau comes from the equally distributed random numbers between 1.4 V and 1.8 V which lead to equally distributed errors from 0 V to −0.4 V. In fig. 6.9 the red histogram is based on the same data, but without the deviations of cells with a target value above 1.4 V. The spread of the deviations is reduced. The new range is from −0.2 V to 0.1 V. As expected the area from −0.4 V to 0 V has changed. The plateau from −0.4 V to −0.2 V has vanished.

For the evaluations the cells with target values above 1.4 V are sorted out and the remaining cells are used.

Different floating gate Block Behavior Some of the floating gate tests showed an unusual behavior between the different blocks like in fig. 6.10. The floating gate blocks zero and one have a normal distribution of deviations. The blocks two and three have more and stronger deviations into the negative regime, see fig. D.3. The result was stable for several test runs and did not change. The only difference between the two groups is the analog readout. Block zero and one belong to the

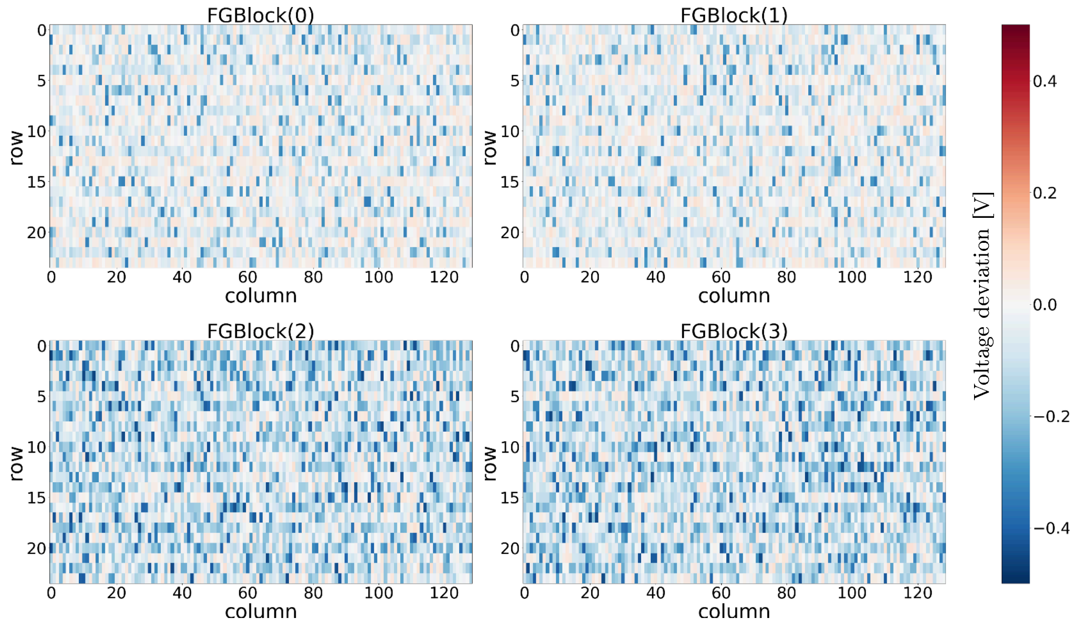


Figure 6.10: Floating gate test of HICANN 341 on DENSE001. The floating gate blocks two and three show a lot more negative deviations than the blocks zero and one.

upper floating gate group and are read via one of the two analog readout lines. However, the floating gate blocks two and three are measured via the other analog readout line. After checking the cable configuration of the prototype setup the offset was gone. It was most likely a bad contact resistance of the cable. In later tests it never showed up again.

This was the only problem which occurred on the DENSE001 board. On the other HICANN chips no problems were visible. The floating gate cells worked properly.

Broken floating gate blocks On the DENSE002 board the tests revealed a new problem of the floating gate cells. There the floating gate blocks of HICANN 173 were completely broken. The result is shown in fig. 6.11. The measured data of the rows are either at nearly 0 V or around 1.4 V. Whereas, the 1.4 V probably is 1.8 V as this is the maximum output voltage of the floating gate cell. The row-wise pattern indicates that the programming of the cells is not working. All four floating gate blocks of the HICANN chip show the same behavior.

The histogram of the data is shown in fig. 6.12. The distributions differ from chips with working floating gate blocks. It is broader and spreads from −1.4 V to 1.4 V. Even though the other digital parts of the chip may work, the neurons and synapses cannot be set up in the correct way as the analog parameter storage is not configurable.

The problem with broken floating gate blocks occurred only once on the DENSE002

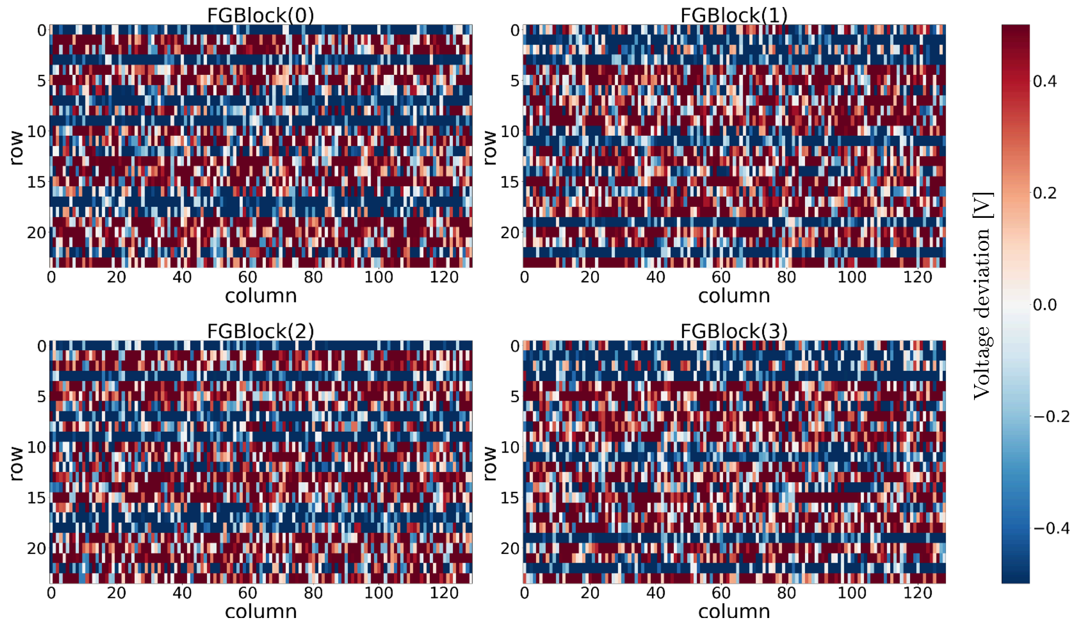


Figure 6.11: Result of the floating gate test of HICANN 173 on DENSE002. There is a row-wise pattern, the deviations are either at negative or positive maximal value.

board. Due to the low occurrence of the error, it seems not to be related to the embedding process. If it is related to the embedding process, it had to effect more chips on the wafer.

Conclusion Overall the tests of the floating gate cells on the HICANN chips were a success. It was no influence of the embedding process visible. Almost all floating gate blocks worked as expected only one chip had completely broken floating gate cell blocks.

6.5 Neuron Calibration

The neuron calibration was the test with the highest time consumption. One HICANN needs 40 min for a complete run. Half an hour is spend on the calibration and the remaining 10 min on the evaluation and plotting of the results. The current prototype system uses two Flexadapter boards. Therefore, four HICANNs can be accessed with this configuration at the same time. After around 160 min all four HICANN chips are calibrated. A complete wafer with 80 chips takes around 53 h. Unfortunately, the system has to be modified by hand to switch to four other HICANNs. The ZA1 connectors and the DDC board have to be moved to the next slots. This modification takes around 10 min.

The calibration software creates a list with black-listed neurons. These neurons

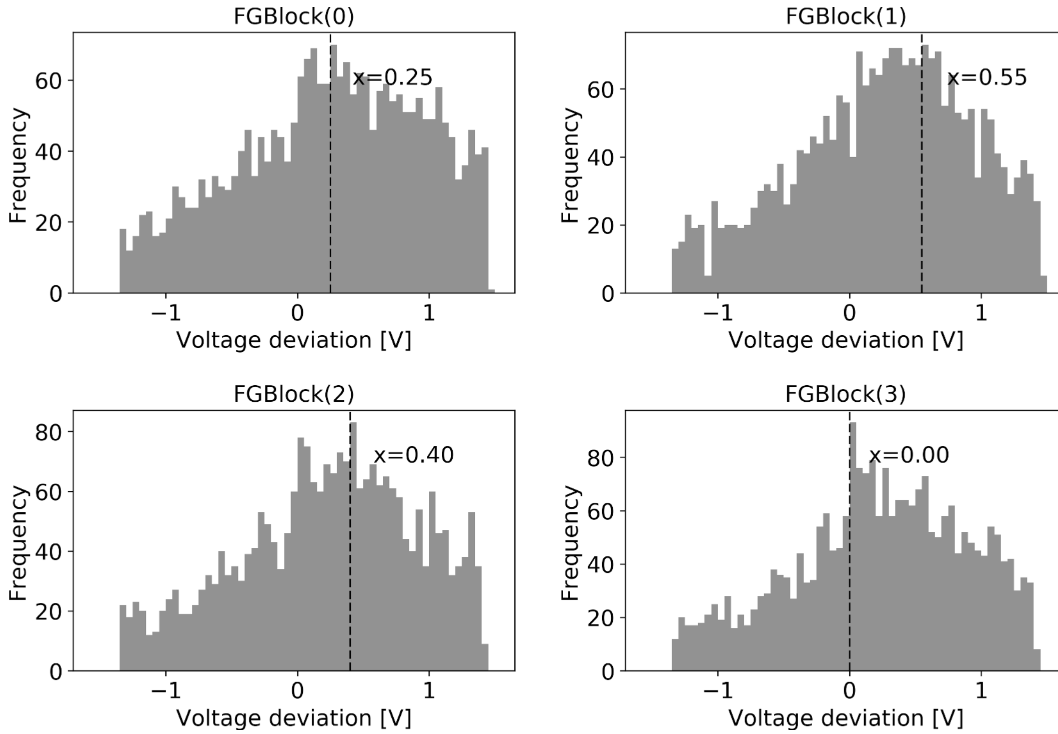


Figure 6.12: Corresponding histogram to the floating gate test of HICANN 173 on DENSE002.

did not behave in the intended way so that they are not usable for experiments. The number of neurons marked as defective is a good indicator for the state of the HICANN chip. Therefore, it is used in a wafer visualization. In fig. 6.13 a heat map of DENSE001 is shown. The number of defect neurons is color-coded. A blue HICANN has a low number of sorted out neurons and a red HICANN has a high number. The neuron calibration can only work on chips with a functioning high-speed communication. Therefore, all chips which had problems with the Switchram test are excluded. Moreover, the control software for the chips can only work if all HICANNs in a JTAG-chain have a functioning high-speed communication. If only one chip has a problem, the software throws an exception and ends the experiment or calibration run. The smallest chain consists of two HICANN chips, in the prototype setup two of these are linked together. The HICANN 101 on DENSE001, for example, has a problem with its high-speed communication. His direct neighbor HICANN 129 has no problem but due to the software restriction it cannot be used. Depending on the position of the DDC board the HICANNs above or below the broken HICANN cannot be used too. In this case either the HICANNs 45 and 73 or 161 and 197 are involved.

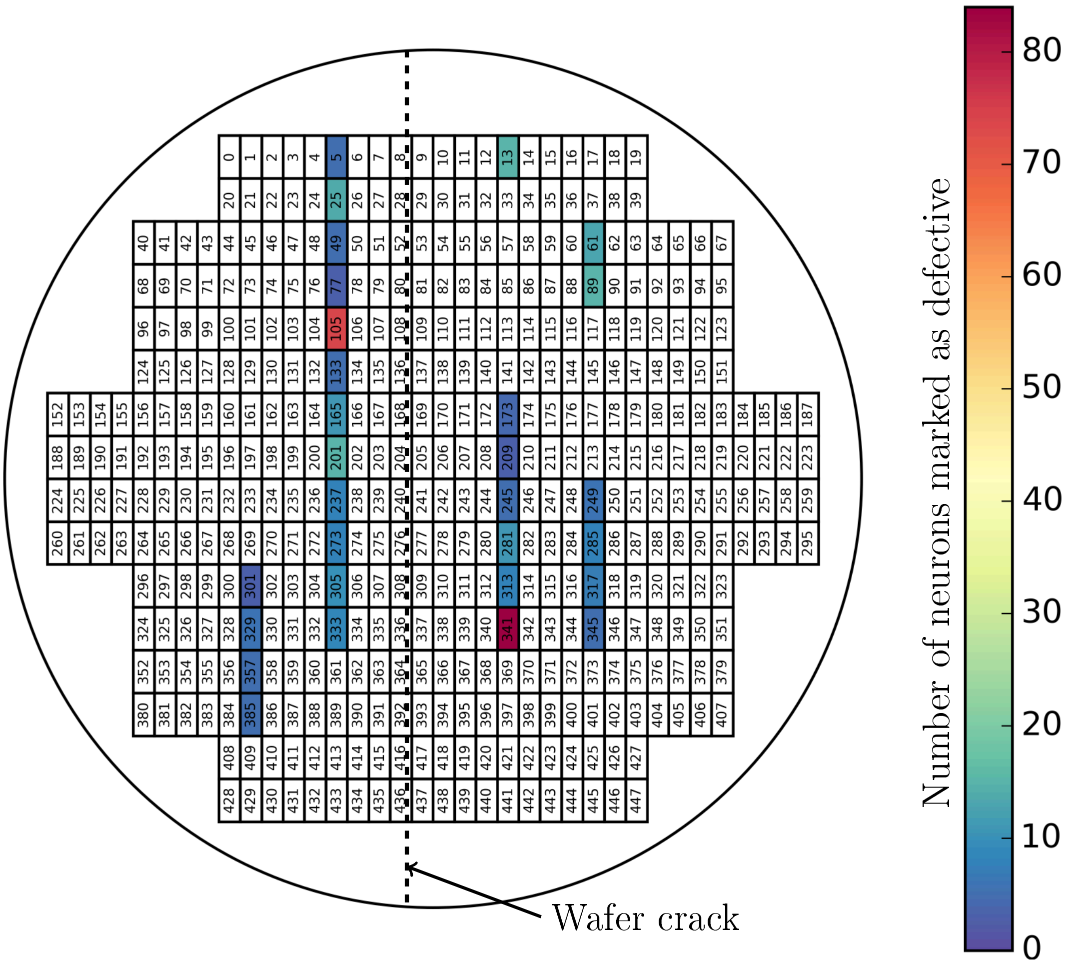


Figure 6.13: DENSE001 heat map of defect neurons according to the calibration software. The HICANNs 105 and 341 have a higher number of neurons which were sorted out than the other chips.

Neuron Calibration on DENSE001 On DENSE001 not all chips were tested because the Switchram test already revealed that not all chips are working. Therefore, the board was chosen to undergo the thermal stress procedure. While the DENSE001 board is in the climate cabinet the other DENSE board can be characterized. Thus, only a subset of 30 HICANN chips was calibrated as a basis for later comparisons.

On most of the HICANNs the calibration software excludes only a small number of neurons in the range of 5 to 15 neurons. On the BrainScaleS system Alexander Kugele did wafer wide neuron calibrations with comparable results. There calibrations with up to 10 % of black-listed neurons are accepted [*Kugele*, 2017]. Thus, the limit of accepted loss is set to 50 neurons. There are two HICANNs 105 and 341 which have a higher number of defect neurons. Both have around 80 neurons marked as defective. Most to the plots from the calibration look fine and show no specifics. Only in the figure of E_{syni} there are neurons that do not behave like expected. In figs. D.2 and D.3 the histograms of the floating gate block tests for both HICANNs are shown. The lower floating gate blocks have the typical error pattern of a broader distribution which are related to the analog readout problem, see section 6.4. A lot of the values are lower than their actual value. If the value are too low and outside the accepted range, the calibration software sorts the neurons out Therefore, the reason for the high number of black-listed neurons is the analog readout and not the embedding process.

Neuron Calibration on DENSE002 On the second DENSE board a neuron calibration of all possible HICANN chips was conducted. Only three HICANNs 5, 25 and 253 were excluded due to problems with the communication, see fig. 6.6. Unfortunately, the problem with HICANN 253 effected the three surrounding HICANNs 181, 217 and 289. The HICANNs 5 and 25 did not effect other HICANNs.

In total 74 HICANNs were calibrated. The results are shown in fig. 6.14. In table D.10 the number of neurons marked as defective are listed for each HICANN. Most of the chips are fine and have a low defect neuron count below 25 neurons. These chips were not further investigated as their error numbers are in the accepted range. In contrast to the DENSE001 board there are six HICANN chips with an elevated error number between 70 and 100. These are the HICANNs 161, 173, 229, 245, 361 and 413.

Two of the six HICANNs have a very high defect neuron count that is HICANN 173 and 229. For the HICANN 173 not a single neuron could be calibrated and, therefore, is completely black-listed. The HICANN 229 has 503 sorted-out neurons so is almost not usable at all.

HICANN 173 The reason for the HICANN 173 are broken floating gate blocks which was already detected in the floating gate tests, see fig. 6.11. None of the cells can be programmed to a reasonable value, so the neuron cannot function at all. Therefore, all neurons marked as defect.

Figure 6.14: DENSE002 heat map of defect neurons according to the calibration software. The HICANNs 5, 25, 181, 217, 253 and 289 are excluded due to communication problems.

HICANN 229 For HICANN 229 the floating gate test revealed no errors, see fig. D.5. The plots from the calibration are not very distinct. Most of the parameters look fine, e.g. the threshold voltage V_t or the resting potential E_l. On the other hand there are parameters like V_{Reset} which show an abnormal behavior, compare fig. 6.15. At first in the uncalibrated case all values are located around the same voltage of 0.78 V. In the evaluation run the values are around 0.1 V and the target

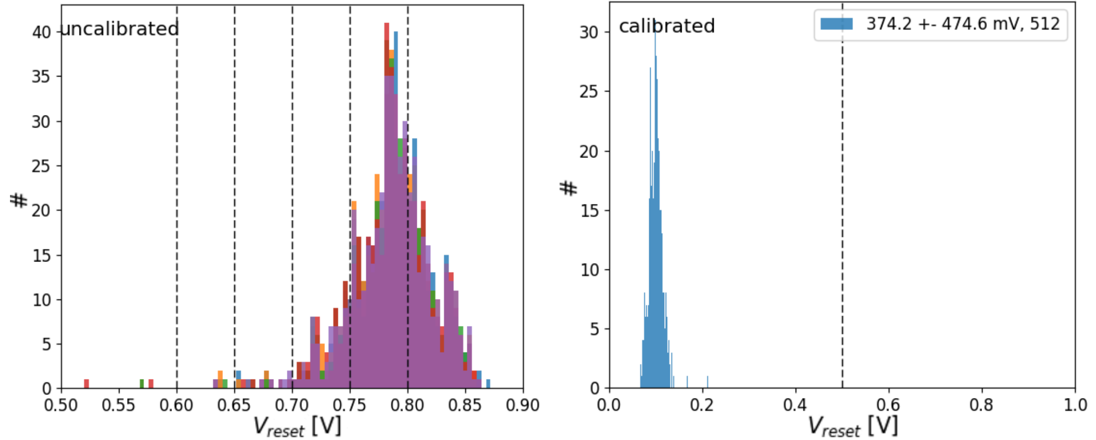
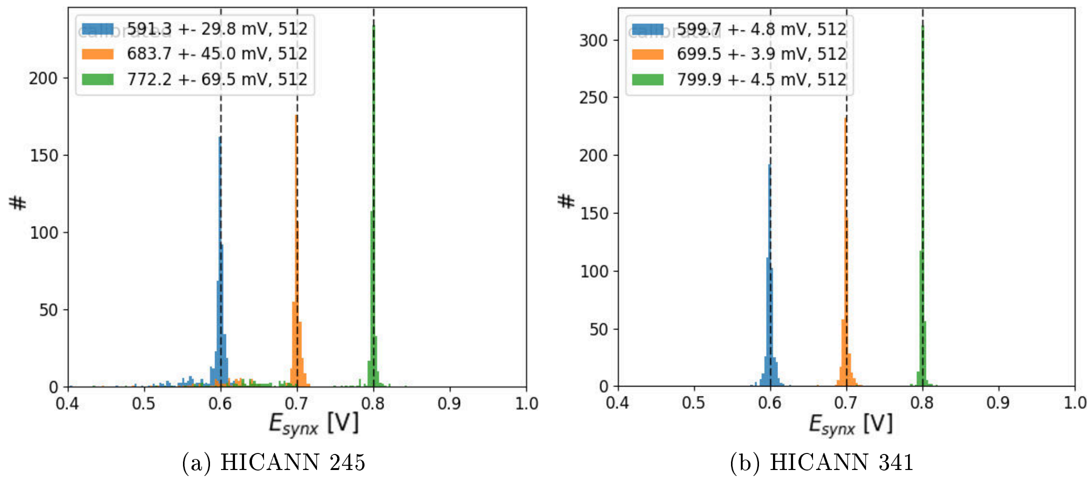


Figure 6.15: Different behavior of V_{reset} on HICANN 229 between the uncalibrated and the calibrated run. In the left plot all values are around 0.78 V and in the evaluation run they are around 0.1 V and, hence, the calibration software marks the neurons as defect.

value of 0.5 V is not reached. Therefore, the calibration software marks all neurons as defect. The V_{Reset} parameter is shared between all neuron circuits, hence, it is possible that only these single floating gate cells are not working. There are single transistor failures on silicon wafers possible as the production yield is not 100 %. On the other hand the probability that always the same cell in all four blocks is broken is low. Therefore, the reason has to be something else in the chip which is not visible from the data. The error is not occurring on multiple HICANNs so it seems not related to the embedding process.

HICANN 161, 245, 361, 413 The HICANN 245 has around 100 neurons marked as defective. In fig. 6.16 the parameter E_{synx} is compared with HICANN 341. The HICANN 245 has more neurons that do not reach the programmed voltage and, hence, lead to a wider spread of up to 70 mV. The deviation seems to be voltage dependent as the spread grows with higher E_{synx}. The HICANN 341 does not show this behavior and has a smaller spread of 5 mV for the E_{synx} parameter. The calibration software marks all neurons which are far away from the target value as defective. The floating gate test histograms look like the ones for the HICANNs 105 and 341 on the DENSE001 board. In fig. D.6 the histograms of HICANN 245 are



(a) HICANN 245

(b) HICANN 341

Figure 6.16: Results of the calibration on E_{synx} for HICANN 245 and 341. HICANN has more neurons that do not reach the programmed voltage and, hence, the distribution has a bigger spread. Whereas for HICANN 341 less neurons have a deviation to lower values.

shown. There the floating gate blocks two and three have more deviations at the lower end of the distribution which lead to more sorted out neurons. Therefore, the problem is in the analog readout.

For the HICANNs 161, 361 and 413 it is the same behavior is visible. The distributions in the histograms for the floating gate blocks two and three are broader, see figs. D.4, D.7 and D.8.

Conclusion The neuron calibration showed that the HICANN chips are in a good condition. Only on the DENSE002 board were two chips which could not be used at all. The reasons for the failures are not related to the embedding process. The number of sorted-out neurons on the other HICANN chips is low and in the expected range from the experiences gathered with the WSI system.

6.6 Reliability Tests

For the reliability tests two kinds of thermal stresses were applied to the DENSE001. First an accelerated environmental stress test is simulated in a climate cabinet. The board had to undergo 393 thermal cycles. After the thermal stress the same test procedure as before with a Switchram, a floating gate test and a neuron calibration was done. To reduce the test time only twelve HICANN chips were tested. The chips 105 and 341 were chosen as they had the most defect neurons at the beginning. Additionally, the chips 301, 329, 357 and 385 were tested because they are close to the wafer edge. In fig. 6.13 the initial state is shown. After the first 50 cycles the

board was taken out of the climate cabinet and tested. All chips worked properly and no failures were observable. The changes in the numbers of black-listed neurons are very small, the biggest difference was for HICANN 105 with five neurons. These changes are negligible, the variations are within the trial-to-trial variations. A list with detailed defect numbers can be found in tables D.8 and D.9.

In the second run 343 thermal cycles in the climate cabinet were conducted. After the run the board was tested again. The Switchram test was successful and the floating gate-test showed no abnormalities. For the neuron calibration the numbers are almost the same. The biggest change happened again on HICANN 105. This time the calibration software swapped eight neurons from the sorted-out list to the functioning neurons list.

At last the board was put in a soldering furnace and had to undergo two soldering cycles. The Switchram and floating gate test showed no failures. Also the changes in the numbers of neurons marked as defective are still in the range of trial-to-trial variations. The heatmap of the DENSE001 board after the soldering furnace is shown in fig. 6.17. Four additional HICANNs were tested and calibrated. They showed no failures and had a low number of sorted out neurons.

In fig. 6.18 the changes in the number of black-listed neurons over all tests is shown. There is no general trend visible. The thermal stress tests had no influence on the circuits. The board itself looks fine, there are no signs for a delamination of the different layers.

Reliability of the Wafer Contact Vias A critical point in the stack up is the connection to the wafer. Every communication line is attached to the wafer by only one microvia. If this microvia is teared off the RDL pad, the communication is broken. It is not possible to repair via defects. Therefore, it has to be a reliable connection which holds over the complete lifetime of the board.

So far no failures or problems with the wafer connections occurred. Even after the thermal stress tests in the climate cabinet and the soldering furnace not a single via was broken.

ZA1 connector The ZA1 connector is a possible candidate for the inter-module connection. Therefore, it is also interesting how reliable the ZA1 connectors work. For the prototype system two connectors were ordered.
All tests where conducted with only those two connectors. On the DENSE boards one connector was used for half of the chips which are 40 compression cycles. For the thermal stress tests ten compression cycles were done. Additionally, some tests were done twice and a reposition of the connectors was required which is assumed to add 25 %.

$$
\begin{aligned}
T =& (2 * 40 + 10) * 1.25 \\
=& 112.5
\end{aligned}
$$

Figure 6.17: DENSE001 heat map of the defect neurons according to the calibration software after the climate cabinet tests and two cycles in the soldering furnace. In comparison to numbers of the initial calibration run the changes are in the accepted range.

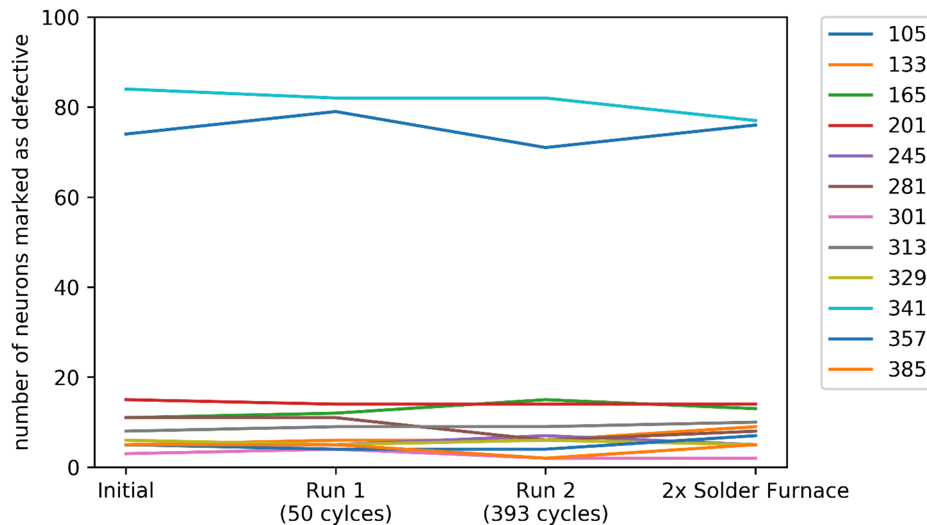


Figure 6.18: Temporal changes of the numbers of the neurons marked as defective by the calibration software. The run 1 and 2 are conducted in the climate cabinet. The last test was done after two cycles in the soldering furnace.

In total each of the connector has been compressed more than 100 times. No failure of a vertical connection through the connector was observable. The spring contacts behave as for the first time. All springs return in their initial position after removing the compression force. Hence, the connectors are reusable and show no degradation.

6.7 Conclusion

The results show no influence of the embedding process on the circuits or the RDLs of the HICANN wafer. There is a difference visible between the DENSE001 and DENSE002 results. A lot of the chips on DENSE001 are not working but it is most likely related to the breaking of the wafer. The rotation of the wafer was the only problem occurring during the production of the boards. For that a possible solution is already available.

The yield of functioning chips is over 96 % on the DENSE002 board. The etch problem on the DENSE002 board can be fixed by changes in the board layout. All other problems are occurring on the WSI system too and, therefore, are not specifically related to the embedding process. Furthermore, the board and circuits withstood the thermal stress of an accelerated lifetime simulation in a climate cabinet and two runs in a soldering furnace. The use of single vias in signal lines showed no reliability problems.

Thus, the embedding of the HICANN wafers overall was a complete success.

7 Roadmap for the Embedded Wafer System

At the moment the DENSE board has only one copper layer. This limits the usage of the wafer. For example, the chips cannot use the on-wafer communication lines. Therefore, further development is required for an Embedded Wafer System as a replacement of the current BrainScaleS system. In this chapter the required steps for a useful Embedded Wafer System (EWS) are presented.

7.1 Backside Connections to the Wafer

The standard wafers only have contact pads on top, therefore, the layers below the embedded wafer are not used efficiently in a board layout. On the DENSE board are no electrical lines or shapes below the wafer, only in the area outside the wafer the bottom layer is used for routing. By accessing the silicon wafer from the back the utilization of the available area can be improved. Furthermore, the vias can be used for the heat dissipation from the wafer to the back of the board as the copper in the vias has a higher heat conductivity than the prepreg material.

In July 2017 the production of the next prototype DENSE version 2 started. Again a HICANN wafer is embedded. The stack up and materials are the same as for the DENSE prototype.

In comparison to the former board two major points have changed. First, the new fixation of the silicon wafer is tested, see section 6.1. Second, backside via connections to the wafer are implemented. In fig. 7.1 the layout of the back is shown. The vias have the same geometry as the microvias on top. The diameter of the hole is 60 µm and the surrounding pad has a diameter of 100 µm. The holes are drilled with a UV-laser and filled by an electroplating process. The complete back of the wafer is covered with a copper layer. This enhances the deposit of copper in the holes during the electroplating process. The copper layer is not structured so all vias are on the same potential.

For a heat dissipation test the supply voltages VDDA and VDDD of six HICANNs on some reticles are connected to pads on top. These chips have no communication channels and are in an unknown and uncontrollable state after the power-up. The power consumption varies, but it will be higher than for only one powered HICANN. The wafer area is divided into four areas, each with a different via density. In fig. 7.1 an enlargement view shows the different via densities. The via densities of the areas are listed in table 7.1. The densities from the highest to the lowest area differ by one order. In the area with the highest density are 18 vias/cm^2.

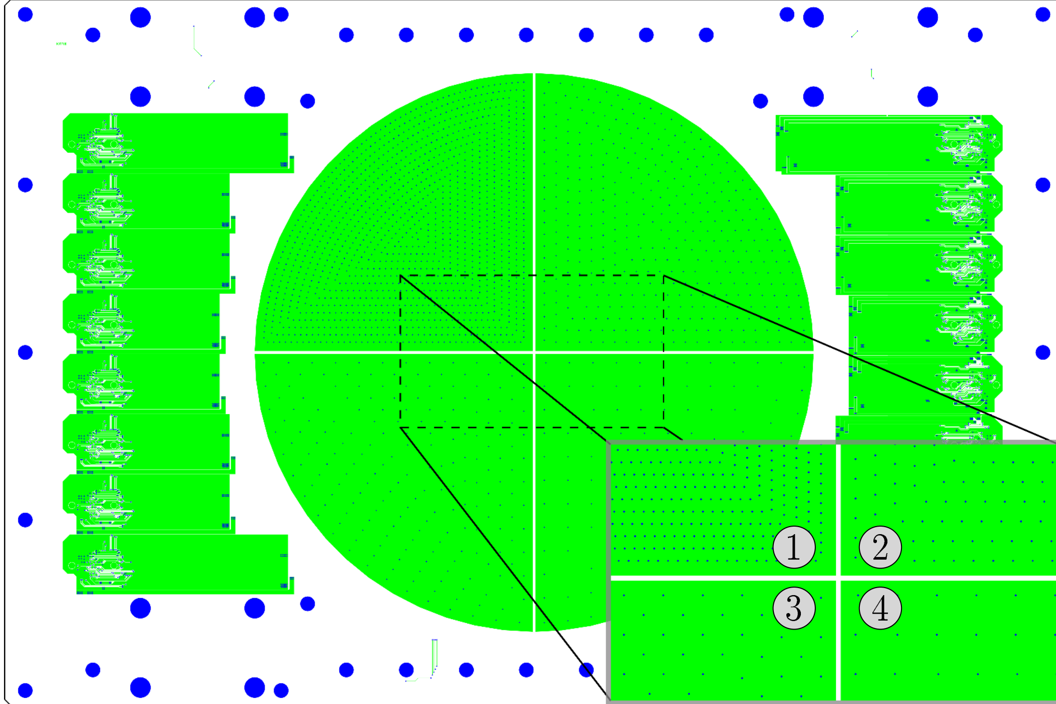


Figure 7.1: Layout of the Deep EvolutioN in System Embedding version 2 (DENSEv2) back. At the bottom right is an enlargement view of the center where the four via arrays are visible. The via density decreases from area one to area four. The green structures are the copper lines and shapes. The blue circles represent the vias.

	# vias	Via density [vias/cm^2]
top left	1266	18.1
top right	405	5.8
bottom left	197	2.8
bottom right	109	1.6

Table 7.1: Via densities in the four quadrants. The area of each quadrant is 69.93 cm^2.

For comparison on top of the DENSE board a fourth of all HICANN chips is connected by 12 500 vias. This corresponds to a density of 39.8 vias/cm^2.

When the prototype arrives at the Heidelberg University a heat dissipation test will be conducted. The special HICANNs for the heating are connected to a laboratory power supply and heat up each quadrant on the silicon wafer. The current consumption is read back from the laboratory power supply and with that the dissipated power is calculated. This prototype will also allow the evaluation of different cooling systems for the board, e.g. air or water cooling.

7.2 Higher Layer Count

The next improvement are more copper layers for additional routing area. These layers are necessary to route all signals from the wafer to the board edges. The build up will be a sequential adding of single layers around the wafer. This allows an implementation of microvias between adjacent layers. Hence, this limits the maximum number of layers because with every additional layer the existing inner layers get more rigid and the probability of breaks and delamination increases. So far all prototypes were only stressed with one compression and heating cycle.

In the WSI system the MainPCB went through five lamination cycles. Assuming a DENSE prototype could withstand the same number of cycles a reasonable stack up and layer assignment is depicted in fig. 7.2. It has in total four copper layers, whereas two layers are assigned to the signal routing.

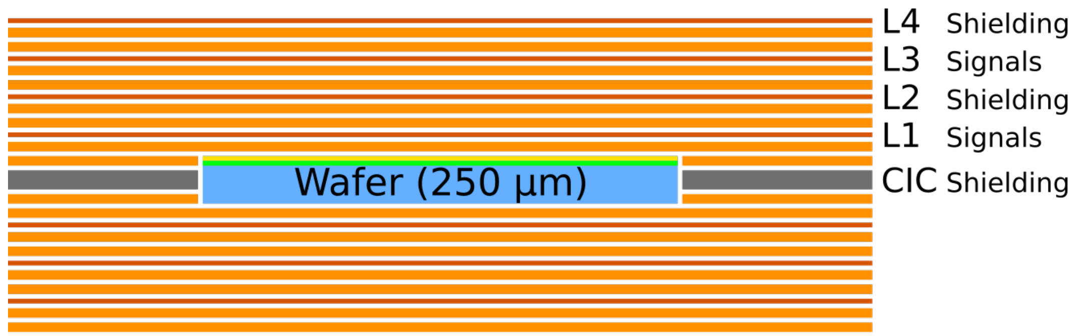


Figure 7.2: Reasonable stack up for a Dense High Layer Count (DenseHLC) board. The layers L1 and L3 are used for the signal routing. Whereas the CIC, L2 and L4 are the reference layers of the signal layers. The single-ended lines decrease with thinner prepregs by 15 µm. In contrast to that grow the differential lines by 35 µm.

For a proper shielding and as reference plane of the signal lines outside the silicon wafer the CIC material is used. Therefore, the CIC has to be connected by vias to the GND net. This way one shielding layer is saved and the first applied copper layer (L1) is utilized for signal routing.

The prepreg between the copper layers has a height of 100 µm. The height has a

direct influence on the line geometries. By using thinner prepregs the line widths also have to decrease in order to achieve the same impedance value. In table 7.2 the geometries for single-ended and differential signals are presented. The calculations were done with the Allegro Impedance Calculator [*Cadence*, 2017a].

Line type	Prepreg thickness [µm]	Line Spacing [µm]	Line Separation [µm]	Impedance [Ω]
single-ended	100	75	-	52.0
single-ended	50	60	-	51.2
differential	100	75	150	98.5
differential	50	60	200	98.1

Table 7.2: Line geometries for the signal layers L1 and L3 in the DenseHLC board. The geometries are calculated for different prepreg thicknesses.

The single-ended lines save routing area with the thinner prepregs. Nevertheless, in total the differential lines need more spacing.

Routing area of a Reticle At the moment every HICANN has four differential links, a shared differential clock and a shared JTAG interface. In total one reticle has 33 differential links and four single-ended lines. There is a rule of thumb saying to leave 2-3-times of space between the adjacent lines [*Montrose*, 1996]. Therefore, a spacing of 150 µm for the single-ended lines and 300 µm for the differential lines is recommended. The total space for all signals lines of a reticle is:

$$
\begin{aligned}
l_{SE} &= 4 * 75\,\mu\text{m} + 4 * 150\,\mu\text{m} \\
&= 900\,\mu\text{m} \\
l_{Diff} &= 2 * 33 * 75\,\mu\text{m} + 33 * 150\,\mu\text{m} + 34 * 300\,\mu\text{m} \\
&= 20\,100\,\mu\text{m} \\
l_{Total} &= l_{SE} + l_{Diff} \\
&= 21\,000\,\mu\text{m}
\end{aligned}
$$

The signal lines of one reticle occupy an area of 21 mm which is the edge length of a reticle. Therefore, one reticle needs one layer for its own signals. On the current HICANN wafer are four HICANNs one after the other. This implies that it is not possible to route the current wafer map with four layers. At least four signal layers and 4 additional shielding layers are required which means eight lamination cycles. These are too many heating and compression cycles. For reliability concerns the number should be lower. The solution is a reduction of signal lines. Two reticles behind one another have to be routed on one layer and do not occupy more than one reticle edge together. The JTAG interface cannot be removed as it is essential for initial tests and configuration. A reduction to one global clock and half of the

differential links results in an area consumption of 22.8 mm for two reticles. With some tuning of the spacings this would allow to route all chips from the current HICANN wafer to the FCP boards. This means that the chip designers have to take this system limitation into account for their future circuit development.

The DENSE board has one copper layer on each side. For the final system at least four more layers are needed according to the above estimation. The increase of layers has to be tested in several steps as there is no experience with such stack ups. A reasonable approach would be to test with each new prototype one additional copper layer. Every new layer increases the number of accessible HICANN chips and also the test coverage over the wafer.

7.3 Silicon Wafers with 30 cm Diameter

The Electronic Visions Group has started the development of new HICANN-DLS chips in a 65 nm process. In contrast to the 180 nm process where the chips are produced on wafers with a diameter of 20 cm the new chips are on 30 cm wafers. For a BrainScaleS system upgrade the wafers are cut out to fit in the existing WSI mechanic framework.

Later these 30 cm wafers have to be embedded. Therefore, tests with such wafers have to be done as the current board dimensions are not sufficient for a 30 cm wafer. In the DENSE board the wafer has a margin of 2 cm to the nearest board edge and 8 cm to the other edges. For the 30 cm wafer a higher margin will be required. The actual value depends on the specific wafer map, which is not available at the moment, and the amount of signal lines that have to be routed. The maximum board dimension that can be produced in the PCB lamination press at the IZM is $610 \times 700\,\mathrm{mm}^2$ [*IZM*, 2017].

A possible procedure could be to test the embedding process for 30 cm wafers like for the DENSE prototype with bare silicon disks and only one copper layer. The materials and the stack up do not change. Only the wafer diameter increases. If the test succeeds the next prototype would still be a bare silicon disk but with the full stack up and backside connections to the wafer. A stepwise layer increase would not be needed as it is already tested with the 20 cm wafers. After a successful production the final test is the embedding of a 30 cm HICANN-DLS wafer.

7.4 Application of Through-Silicon Vias

through-silicon vias create electrical connections through the silicon substrate to the back of the chip. Typically through-silicon vias (TSVs) are used in stacked chips like memory modules or image sensors [*Kim and Kim*, 2014; *Sukegawa et al.*, 2013]. The advantages of TSVs are shorter connections and a higher connection density between stacked chips. In comparison to traditional wire-bonds TSVs have a lower inductance and a lower impedance [*Pak et al.*, 2008]. A disadvantage of TSVs is their influence on the surrounding devices. Only a thin dielectric liner separates the

conductive material, mainly copper, from the semiconductor silicon. Thus, signals or distortions couple through the dielectric material into the silicon and perturb nearby devices. There are several techniques to reduce these influences [*Khan et al.*, 2013]. Nevertheless, the implementation of TSVs for high-speed signals is more complex than for simple power connections. Therefore, in the DENSE board the TSVs could be used for the power supply of the electronic circuits. The copper layers below the silicon wafer would be utilized for more than heat dissipation and heat spreading.

TSVs can be implemented at different stages of the silicon wafer production [*Garrou et al.*, 2008]. They can be inserted before the front end of line (FEOL) production, before the back end of line (BEOL) or post-BEOL. The process steps before the first metal wiring are called FEOL process. On the contrary the BEOL begins with the interconnecting of the transistors the metal layer. To keep the costs for prototype chips low it is reasonable to apply the TSVs only on the wafer-scale level. Thus, they are created post-BEOL but before the RDLs. Therefore, the TSVs are placeable in the scribe lines on the silicon wafer. The area is not used for active components as it is not usable after the chip separation. It is the ideal place for the TSVs. In fig. 7.3 a possible implementation of TSVs in the scribe line is shown. The advantage is that it does not require a redesign of the chip layout and can be applied after the BEOL when the silicon wafer is finished. Unfortunately, it requires a redesign of the RDLs on top of the wafer. Currently, the fine-pitch lines between the separated HICANN chips run directly above the scribe line from one chip to another. These lines have to move to the second RDL so that the TSVs in the scribe line can be accessed with vias from the first RDL.

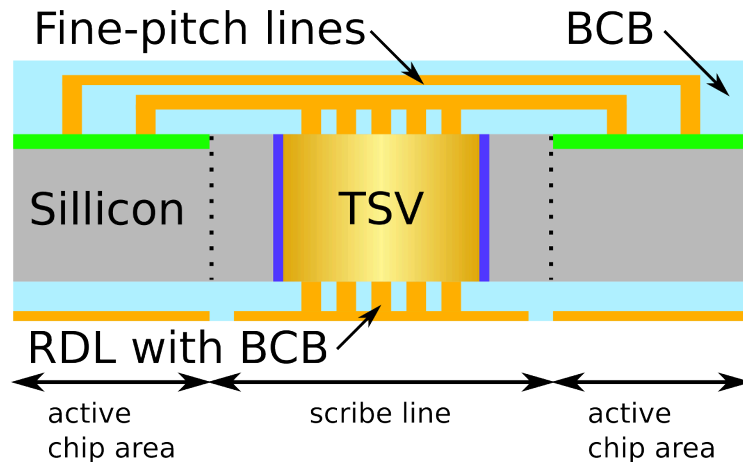


Figure 7.3: Possible implementation of TSVs in the scribe lines of a wafer.

The TSVs are created before the RDLs are applied and the wafer is thinned. The via depth is greater than the final wafer thickness of 250 µm but it does not go through the entire silicon. Then the vias get completely filled with copper. This way it is possible to place vias from the RDL direct on the TSV. After the filling the RDLs are added to the wafer. Next the wafer is thinned down in this step the

TSVs are exposed to the back. Eventually, a RDL on the back creates the copper pads and the connections to the TSVs. The microvias from the back can connect to this copper pads.

The best way is to use the TSVs only for the power supply of the wafer. This way the copper layers above the wafer are used for the signal routing and the layers below the wafer for the power supply. Thus, the complexity is reduced as no high-speed signal run through the silicon wafer and disturb the surrounding devices.

TSV parameters The design of the TSVs depends on the size of the scribe lines and the thickness of the silicon wafer. The width of the scribe line is chosen by the manufacturer in this case TSMC for the 65 nm process. This limits the maximum size of a TSV. On the 180 nm wafer the scribe line is in vertical direction 420 nm and in horizontal direction 250 µm wide [*Grübl*, 2017]. The minimum TSV diameter is determined by the aspect ratio and the silicon thickness. The aspect ratio is defined as the ratio of the TSV depth to the diameter. A typical value is around 10 : 1 [*ITRS*, 2011]. For a 200 µm thick wafer the minimum diameter is 20 µm.
The TSVs are used for the power supply and, therefore, large holes are preferred due to a lower resistance and inductance. Moreover, it is easier to completely fill large holes with copper.

The first tests could use the current HICANN wafers. There the TSVs could be implemented in the active device area which is not used. The big supply pads of the Post-Processing could be used as target pads. Here, the TSVs are added from the back after the wafer is already thinned. This test could be done with any DENSE prototype. The final design can only be tested when the wafer map of the 65 nm wafer is known.

7.5 Cooling of the Embedded Wafer

The cooling of the boards is an important aspect of the final system design. In section 3.2.2 it was shown that it is no problem to bring the heat to the surface of the board. The next step is to find a cooling solution for the board which is small, easy to install, maintainable and works reliably. In principle there are two possible cooling materials: air and water. There are more materials available that work on different mechanisms, like the two phase-cooling medium Novec 649 from 3M. It has a very low boiling point of 49 °C and a heat of vaporization of 88 J/(kg K). For comparison water has a much higher heat of vaporization at 100 °C. The drawback of other cooling materials than air and water is that they typically need a special handling. The Novec 649 medium, for example, needs an air-tight housing as the liquid changes into a gaseous state and has to stay in the cooling box [*3M*, 2017].

The cooling properties of air and water are listed in table 7.3. In comparison the water shows better cooling properties. From the specific heat capacity c definition the amount of heat can be calculated which is needed to heat a volume V by a

Cooling medium	Specific heat capacity [J/(kg K)]	Density [kg/m^3]	Heat of vaporization [kJ/kg]
Air∗	1005	1.204	-
Water∗	4182	998	2257(100 °C)
Novotec 649#	1103	1600	88(49 °C)

Table 7.3: Cooling properties of different mediums. The values are for standard condition. (∗) Data taken from [*Kuchling*, 2007]. (#) Data taken from [*3M*, 2017].

specific ΔT

$$c = \frac{\Delta Q}{m * \Delta T} \tag{7.1}$$

$$= \frac{\Delta Q}{\rho * V * \Delta T} \tag{7.2}$$

$$\frac{\Delta Q}{V * \Delta T} = c * \rho \tag{7.3}$$

The amount of heat per volume and temperature increase is proportional to the specific heat capacity c and the density ρ of the material, see eq. (7.3). From the values in table 7.3 water takes 3.5-times more energy for the same volume and temperature rise than air.

Although air is not as a good as water in terms of cooling properties it is easier to handle. Therefore, it is worth while to check if the cooling is achievable by air. The heat transfer coefficient h from a metal wall to air is comparable to a board without the solder mask and an air flow above the copper layer. For a high volumetric flow rate of air above the wall the coefficient is in the range of 58-290 W/(m^2 K) [*Kuchling*, 2007]. The equation for the transfered heat can be solved for the required temperature difference

$$Q = h * A * dt * \Delta T \tag{7.4}$$

$$\Delta T = \frac{Q}{h * A * dt} \tag{7.5}$$

$$= \frac{\dot{Q}}{h * A} \tag{7.6}$$

In table E.1 the temperature difference for a set of dissipated power and heat transfer h values is calculated with eq. (7.6). A high h value of around 200-290 W/(m^2 K) is necessary to get a small temperature difference between the air and the board. Assuming a cooling area of $30 \times 30\,\mathrm{cm}^2$, a heat dissipation of 500 W/s and an h of 250 W/(m^2 K) the required temperature difference is around 15 K.

In contrast to that water running through a pipe can take up to 2300-

4700 W/(m^2 K) [*Kuchling*, 2007]. For the estimation of the required temperature difference the cooling area is again $30 \times 30\,\text{cm}^2$, the pipe diameter is 5 mm and it is running in meanders 20-times over the cooling area like in fig. 7.4. The curves are

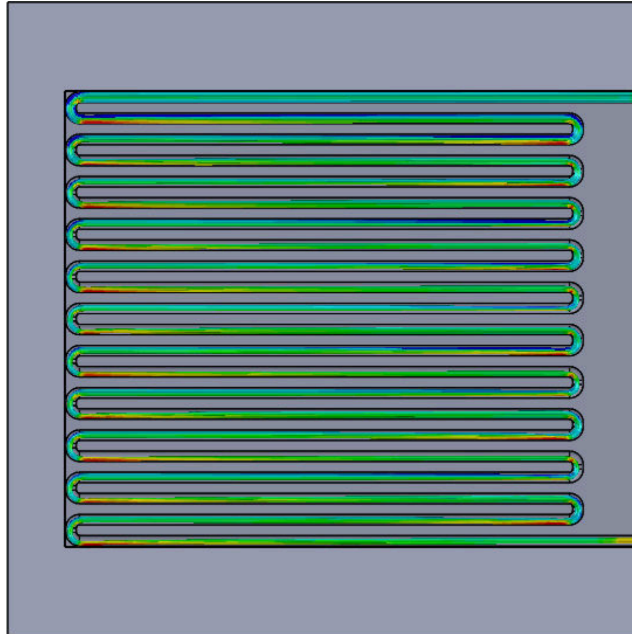


Figure 7.4: The mechanic components of the system are drawn with Solidworks. Additionally, the software can run flow and thermal simulation of the components. The image shows a flow simulation of a water cooling plate.

neglected for the calculations. A low h of 2300 W/(m^2 K) and a low volumetric flow rate of 1.7 L/min results in a required temperature difference of 10 K between the water and the aluminum block.

Hence, there is a clear advantage of water cooling in comparison with air cooling. The water cooling module can be smaller than the corresponding air cooling module. Nevertheless, the requirements for water cooling are much higher as more components and more safety measures are needed. For the final evaluation of the cooling system simulations of both types need to be performed. Furthermore, real tests with the DENSEv2 and following prototypes are possible and should be conducted.

7.6 Embedded Capacitors

Every power supply going in a chip on the wafer needs decoupling capacitors. These capacitors suppress high-frequency noise and compensate short peaks in the current consumption. Hence, the circuit sees a stabilized and clean voltage at its input. In the WSI system there are two capacitors for each voltage on a reticle. For the short-time current supply a 47 µF capacitor in a 0805-packaging is used. A 100 nF capacitor

in a 0402-packaging is chosen for the noise rejection. There are possibilities to create capacitors on silicon chips like Metall-Insulator-Metal capacitors (MIM caps) in the BEOL or MOSFET gate capacitors in the FEOL. The drawback of these kinds of capacitors is the low ratio of capacitance to area consumption. In the 180 nm process a gate capacitor achieves 800 fF [*Millner*, 2012]. Replacing one 100 nF capacitor requires 125 000 gate capacitors. Intel's 22 nm process with nine metal layers reaches a ratio of 20 fF/µm^2 for MIM caps [*Ingerly et al.*, 2012]. There a 100 nF capacitance occupies an area of

$$A = \frac{100\,\text{nF}}{20\,\text{fF}/\mu\text{m}^2} = 5\,\text{mm}^2$$

Therefore, in the EWS the only way to have decoupling capacitors with a reasonable capacitance is by using external capacitors like in the WSI system. However, the later EWSs should be stackable which requires that the surface on top and bottom is flat. Hence, the capacitors have to be embedded in the board.

There are thin capacitors specifically for the embedding process available. For example, the LPSC series is produced on silicon and has a thickness of 100 nm [*Ipdia*, 2017]. The capacitance ranges from 10 pF to 3.3 µF. Unfortunately, there is no 47 µF capacitor in this package size. Therefore, multiple capacitors with the highest values have to be taken. Standard surface-mounted capacitors could be used but their height is 10 times larger. Therefore, they are not considered.

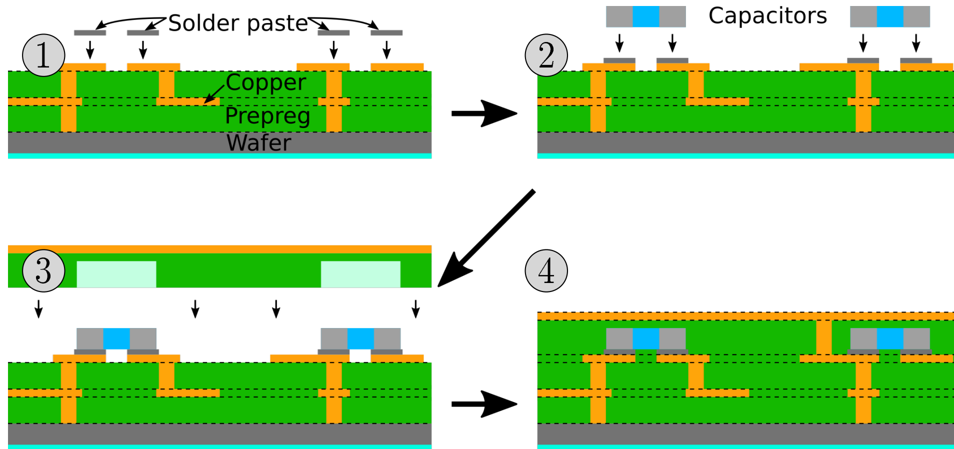


Figure 7.5: Embedding capacitors with the soldering method. In the first step solder paste is put on the pads (1). Then the capacitors are placed on the board and put in a soldering furnace (2). Like in the lamination of the silicon wafer the surrounding prepreg layer has cut outs for the capacitors. On top comes another prepreg layer and a copper foil. Then the board is placed in the PCB lamination press (3) . The last step creates the microvias to the previous copper layer (4).

There are several possibilities for the integration of capacitors in the PCB. Here the two most reasonable methods are described. First, the capacitors are soldered on the existing stack up and then they are encapsulated with a layer of prepreg and a copper foil in another lamination step. In fig. 7.5 the process is depicted. Second, the capacitors are glued on the PCB and like in the first approach a layer of prepreg and copper material is applied which is then put in the lamination press. So far the capacitors have no electrical connections. These are created by vias from the outer copper layer. Two types of vias are required; one to the pads of the capacitor and one to the previous copper layer. This way no soldering run is required which reduces the thermal stress to the board during production. On the other hand the complexity of the board increases. The procedure is depicted in fig. 7.6.

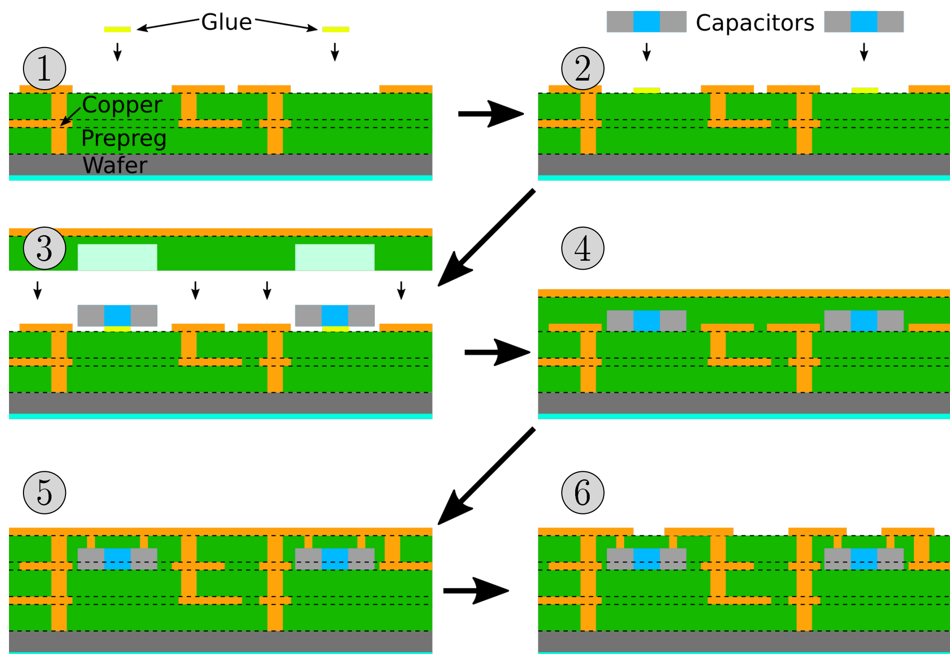


Figure 7.6: Embedding capacitors with the adhesive method. The first step is to place an adhesive on the PCB which holds the capacitors added in the second step. The capacitors are placed only on the prepreg layer and have no electrical connection so far. The next steps are like for any other additional copper layer with the exception that the prepreg has cut outs for the capacitors. In the fifth step the electrical connections to the capacitors are created by microvias. Two types of vias are created; one to the capacitor pad and one to the underlying copper layer. In the final step the copper on the outside is structured.

The embedding of capacitors can be tested without a silicon wafer in a first run. Therefore, both approach could be tested.
Later the capacitors are placed on the back of PCB where the power supply to the

wafer is routed. For the tests in a board with a silicon wafer it does not matter if they are connected to the wafer, therefore, they can be embedded in any prototype which has at least four copper layers. The wafer is cooled from the back, hence, the capacitors are in the path of the heat dissipation and will disturb the transport. This has to be taken into account for the design of the cooling module.

7.7 Interconnection of several Modules

The interconnection of the modules increases the neuronal network size that can be emulated. Between the systems there are two possible connection types.
Several of the embedded wafers are packaged together in a bigger module. In this module there are direct connections between adjacent boards. These are called short-range connections. On the other hand there are long-range connections. These are used to interconnect boards from different modules.

Short-range Connections The direct connections between boards are created by Samtec ZA1 connectors. The connector was extensively used with the DENSE prototype and showed positive properties like easy handling and reusability. The ZA1 connectors are available in different pin configurations from 10x10 up to 10x40 and in different heights. This way a lot of signals can be routed from one system to another.

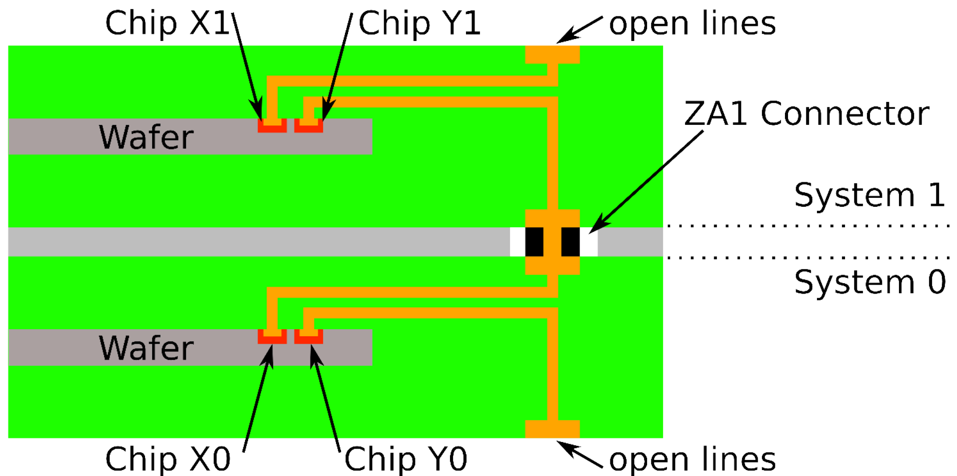


Figure 7.7: Interconnecting scheme for two systems next to each other. The routing scheme is fixed and cannot be changed after production of the boards. In the image chip X0 of system 0 is connected to chip Y1 on system 1. Without special adapters some chips of the first and last system in a module do not have external connections. In this case the chips Y0 and X1 have open lines. The ZA1 connector is used between the two boards.

The positioning of the connectors can be achieved with the cooling plates on the

back of the boards. The alignment of the boards with respect to each other is rather simple as the pitch of the connector pins is 1 mm. In fig. 7.7 the connecting principle between two boards is depicted. The short-range connections can be tested with two prototypes. Unfortunately, the current HICANN chips need an FPGA for the communication with each other. A direct HICANN to HICANN connection without the FPGA is not possible yet. Therefore, only simple daisy chains can be created and tested until the direct communication between the chips is implemented.

Long-range Connections So far it is not intended to solder any connectors on the board. Therefore, the long-range connections need solder-less connectors. There are edge connectors available, e.g. the HSEC8-DV from Samtec. They are used on the MainPCB for the FCP boards and the experiences are positive. The mating connector is 7 mm wide [*Samtec*, 2017a]. If the connector is used on different boards at the same position, then the distance between the boards has to be at least 8 mm with a safety margin of 1 mm.
If the distance is below 7 mm, a solution would be to create different boards. The edge connectors of the different boards would be at different positions so that they are not next to each other. This would mean to reduce the number of connections going from one board to another and increase the complexity of the construction of the modules.

On the other hand the DENSE board withstood the thermal stress in the soldering furnace with out damages. Moreover, the cooling of the system will require some space. For that reason a low profile connector on the board could simplify the system design. The Samtec LP-Series has a minimum stack height of 4 mm which could fit in the space between two Embedded Wafer Systems. On a flex-rigid board the mating connector could be placed and connected to another system.

At the moment there is no final solution available as there are too many open questions. It also depends on the type of links between the final chips and on the number of required lines. With the current hardware daisy chains with different connection techniques between the boards could be tested to find the solution for the final setup.

7.8 Final Embedded Wafer System

In the end all features are combined in the EWS. This leads to very compact modules which can be produced in high quantities and are easy to assemble into a server rack. A possible stack up for the silicon wafer embedded in a PCB is shown in fig. 7.8. The board is divided into two parts. On the left are the connectors for the long-range connections. Whereas, the power supply of the wafer and the short-range connections are on the right of the board.

At the back of the PCB is the cooling module. It will be thicker than the PCB. Cooling by air will require more space than a water cooling solution. For the visualization a water cooling solution is chosen. The module is a stack of five Embedded

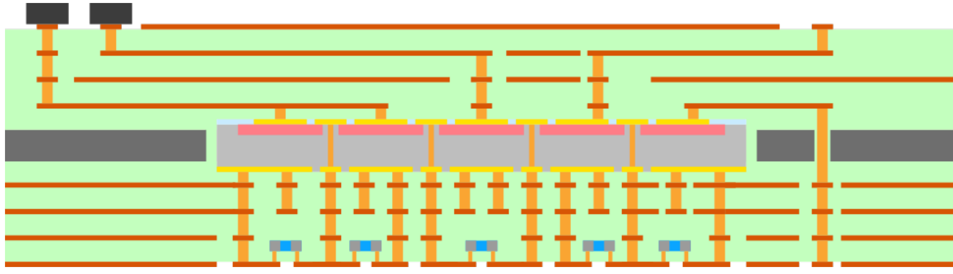


Figure 7.8: Possible stack up of the final EWS. It contains all features like backside connections, several copper layers, TSVs and embedded capacitors.

Wafer Systems and each has its own cooling module. In fig. 7.9 a complete module is depicted. The number of five is chosen due to the weight of a module and the effort it takes to replace and to disassemble one module. A cooling block of aluminum with the size $40 \times 40 \times 0.6\,\mathrm{cm}^3$ weighs 2.6 kg. In total a module weighs at least 13 kg. The modules are assembled outside the rack and, eventually, placed in the server rack and connected to the power and water supply and other modules.

Assuming the PCB is around 1 mm thick and the cooling module is 6 mm thick. The fixation of one module is partly recessed in the cooling block, therefore, it is estimated the fixation takes 3 mm of space on each side of the module. Then the

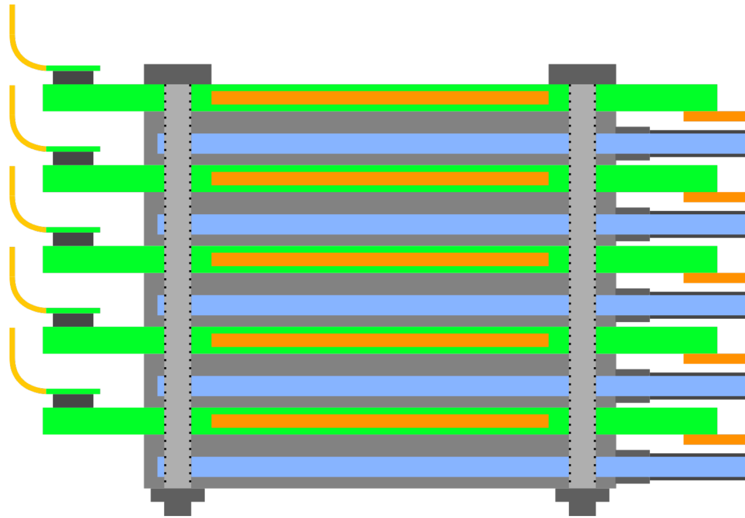


Figure 7.9: Module of five Embedded Wafer Systems with water cooling modules. The complete module weighs at least 13 kg.

resulting thickness of one module is:

$$5 * (1\,\mathrm{mm} + 6\,\mathrm{mm}) + 2 * 3\,\mathrm{mm} = 41\,\mathrm{mm}$$

For the assembly in the rack there should be a safety margin of 3 mm between the

modules. A standard 19-inch server rack has an opening in the front of 45 cm. Therefore, the number of systems side by side is

$$(45\,\text{mm} - 3\,\text{mm})/44\,\text{mm} = 10$$

which means 50 systems can be placed on one frame side by side. A server rack has a height of 42 rack units, a rack unit is 44.45 mm. The EWS module occupies nine rack units. For the positioning frame above and below the modules one rack unit is estimated. Thus, ten rack units are occupied by one EWS module, which allows four of these frames per rack. In total there are $4 * 10 * 5 = 200$ Embedded Wafer Systems in one rack. For comparison in the WSI system are only four wafer modules per rack. Therefore, the Embedded Wafer System enables a higher integration level of system.

Conclusion To reach the final Embedded Wafer System a lot of development has to be done. Partly this has already started like the DENSEv2 board with backside connections. Furthermore, some of the points can be developed in parallel. For example, with each new board the cooling can be tested and improved and with two boards the connection principles can be evaluated.
At the moment the concept looks reasonable and the distinct steps seem feasible.

8 Conclusion

In this work complete silicon wafers were embedded into a printed circuit board (PCB) the very first time. The project was developed and conducted together with the Fraunhofer Institute for Reliability and Microintegration (IZM) in Berlin. At first the correct material selection for the PCB had to be found. Therefore, bare silicon disks with no active components were embedded. After three iterations of lamination a composition of an ultra low coefficient of thermal expansion (CTE) organic preimpregnated bonding (prepreg) and a Copper-Invar-Copper (CIC) material showed the best properties. The last prototype board had a low warpage and allowed a patterning of the copper layers. In the same runs the symmetrical stack up with the silicon wafer in the center of the board was improved. It turned out that a thin sheet of CIC with layers of prepreg on top and bottom is the best surrounding frame for the silicon wafer. Furthermore, the stack up was unaffected by the thermal stress which was induced in a climate cabinet. The result was no failures emerged in the meander and daisy-chain measurements.
Before the wafer can be embedded it has to be prepared in order to withstand the lamination process. Hence, the wafer is thinned down to 250 µm to be more flexible. Additionally, the pads of the redistribution layer (RDL) get a copper finish for a better deposit of copper in the later electroplating procedure. The lamination step combines the silicon wafer and the different layers to the final board. In this step the resin of the prepreg layers is heated up and becomes a glassy state which during cool down cures and holds the parts together. The applied pressure does not harm the wafer as long as there no shear forces. None of the bare silicon disks broke in the embedding tests. One of two High-Input Count Analog Neural Network Chip (HICANN) wafers broke during the lamination step. The reason for that was the rotation of the wafer in the course of this the counterpart of the wafer notch got under the wafer. This led to an uneven distribution of the pressure. The solution is to replace the notch by a flat, this will be tested with the next prototype.
The manufacturing process of the plated through holes (PTHs) from top to bottom of the board had to be modified due to the electrically conductive CIC material. Therefore, the CIC sheet is drilled before the lamination step, the holes get filled with resin during the lamination process and the resin serves as an insulator between the CIC and the copper of the actual PTHs. Unfortunately, the PTHs could not be used because both wafers rotated in the lamination step.

Two Deep EvolutioN in System Embedding (DENSE) boards with embedded HICANN wafers were tested in Heidelberg. The Switchram test revealed that the communication with 35 % of the chips in the DENSE001 board with the broken wafer is not functioning. The reason for that is most likely the breaking of the wafer.

For comparison in the DENSE002 prototype only three out of 80 chips are not working properly which is less than 4 %. The defect chips have errors which also occur in the Wafer-Scale Integration (WSI) system and, therefore, seem not to be related with the embedding process.
Furthermore, there were no impairments observable on the digital and analog circuits. The floating gate test checks the differences between the target values and the actual values in the analog parameter storage. While a few floating gate blocks were broken, there are no signs that the failures are induced by the embedding process. The deviations of the floating gate cells were in the accepted range for HICANN chips.
At the end a neuron calibration was conducted on the working chips. Most of the neurons could be calibrated and only a small percentage was black-listed by the software, which was expected for the HICANN chips. The chips with a higher number of sorted out neurons than usual could be explained by other defects on the chips or were observed on the WSI system too. No defect is related to the embedding of the wafer.
The reliability of the boards was tested in a climate cabinet and a soldering furnace. In the climate cabinet the DENSE board was repetitively thermally stressed. Almost 400 cycles were driven between 15 °C and 90 °C. Additionally, the board had to undergo two soldering runs in the furnace with a maximum temperature of 230 °C. The thermal stress had no influence on the number of black-listed neurons from the calibration software. In particular, no failures of the wafer contact vias occurred.

In comparison to the WSI concept the embedding of wafers showed several advantages. Time consuming tasks of the WSI system like the filling of the positioning mask and the alignment of the wafer and the MainPCB are removed. In the Embedded Wafer System the contact between wafer and board is established with microvias. These are automatically created at the correct positions because the drilling machine takes the fiducial marks on the wafer as reference points and, accordingly, adjusts the positions of the holes. Hence, the process has a higher level of automation. So it can be easier scaled up to a large number of systems.

At the moment the limiting factor for the functionality of the embedded wafers is the number of copper layers. On the DENSE board only a fourth of the HICANN chips could be accessed. For this reason the on-wafer communication over the RDL fine-pitch lines could not be test. The next prototypes will have more copper layers and, thus, more HICANNs can be used. Overall, this thesis showed that it is feasible to embed silicon wafers into PCBs and that it is worth doing further development.

9 Outlook

The embedding of wafers implies some restrictions for the wafers, thus, it is not reasonable to embed every wafer. The heat dissipation, for example, is a critical factor for the embedding. The HICANN chip has a low power consumption and, therefore, a low heat density. In contrast to that a silicon wafer with high performance processors has a high power density. It would not be feasible to cool the wafer down in an embedded version. Furthermore, the connection density has to be low. Otherwise, a fan-out of the signals is not possible. A reticle on the HICANN wafer has only 70 digital signals on a $2 \times 2\,\mathrm{cm}^2$ area. The fan-out of several modern processors next to each other is not possible as their signal density is too high.
Therefore, conceivable fields of application are systems with a low power consumption, a moderate signal density and short connections between the chips.

SpiNNaker Project The SpiNNaker Project is the second neuromorphic hardware platform in the Human Brain Project. In comparison to the Electronic Visions Group the SpiNNaker Project uses a digital approach for the neuron and synapse implementation in silicon. In contrast to the BrainScaleS system single chips are placed on PCBs. A single chip integrates 18 ARM9 processor cores and a packet router. The neuron and synapse models are defined by the loaded software which is executed on the processors. Thus, different models can be implemented and easily exchanged.
The power density is with $1\,\mathrm{W/cm}^2$ of the same order as the BrainScaleS system [*Furber et al.*, 2014]. The system is cooled by forced-air cooling. In fig. 9.1 a 48-node SpiNNaker Project module is shown. A 19-inch rack cabinet can be equipped with 120 of these boards.
The embedding of a silicon wafer with SpiNNaker Project chips does not reduce the total system height. There are other components than the cooling limiting the module's height like the Ethernet and Serial AT attachment (SATA) connectors. The density of modules in a rack cabinet would not increase. On the other hand the number of SpiNNaker Project chips on a wafer is higher than on the 48-nodes board. Thus, the embedding of a SpiNNaker Project wafer increases the chip density per board.
The connections between the chips on the 48-node board are standard PCB lines. Hence, they can be created in an embedded system too. The higher packaging density of the chips leads to shorter lines between the chips, which decreases the required power to drive the lines.
The current embedding procedure has only one copper layer which is not enough to achieve the fan-out of the signals and the distribution of the power supply. Thus,

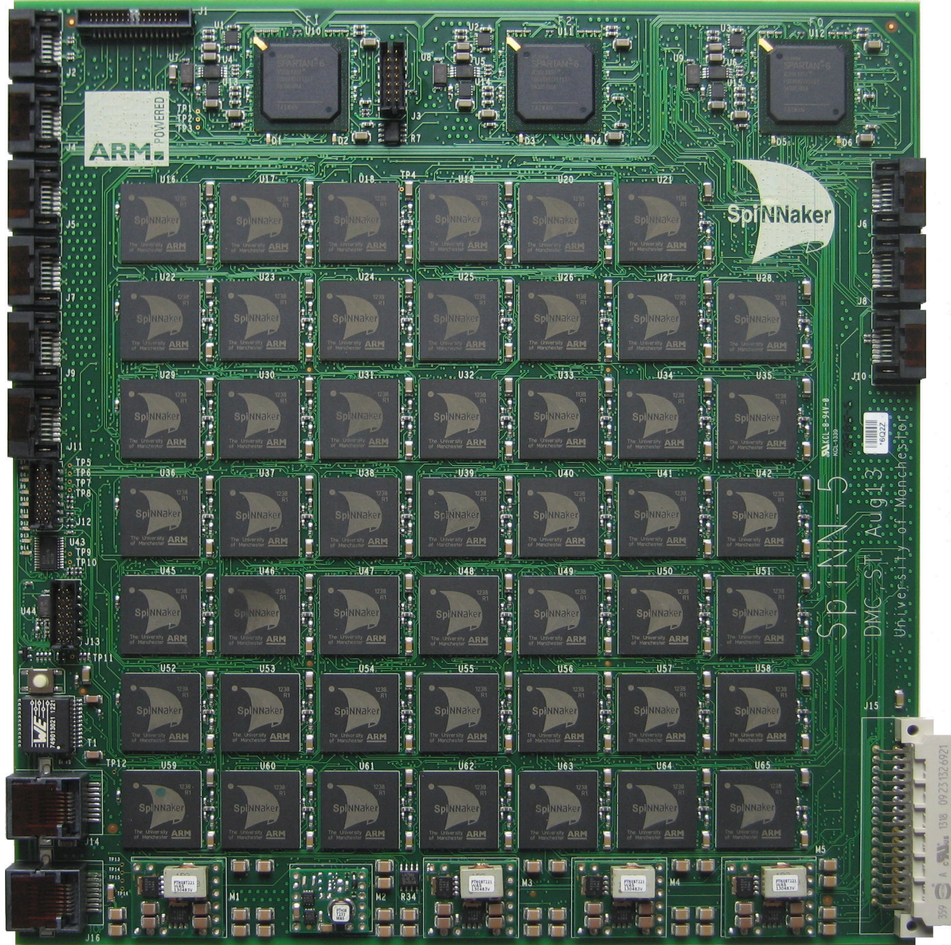


Figure 9.1: A 48-node SpiNNaker board with 864 cores. Over the SATA connectors multiple system can be interconnected and build larger neuronal networks. Courtesy of John Woods (University of Manchester).

more copper layers are needed to be able to cross lines on different layers. The connection to the wafer is established with microvias like in the DENSE board. The design of the SpiNNaker Project chip does not need to be changed as the microvias can be placed on the available chip pads. Only an additional copper finish has to be applied onto the pads for the filling of the microvias in the electroplating process.
A disadvantage of the embedding of a SpiNNaker Project wafer in comparison to the single chip node is the exchangeability of broken chips. On the 48-node board a ball grid array (BGA) chip is easily replaced with a new one. The single chips are tested prior to packaging. Hence, only known good die are packaged and assembled onto the board resulting in a fully functioning module. This procedure is not possible with chips on a silicon wafer. The chips on a wafer can also be tested before the embedding process but they cannot be replaced. In the BrainScaleS system errors

are expected and precautions are implemented, for example, turning off chip areas and routing around defective chips. Additionally, the cooling with air might not be sufficient to transport the heat away from an embedded SpiNNaker Project wafer. The chips are placed on a wafer as dense as possible, therefore, the heat density increases in comparison to the single chip setup.
The embedding of a SpiNNaker Project wafer increases the neuron and synapse density on a single board as there are more chips in the same area. On the other hand the system complexity increases. The cooling of the system has to be adapted to the higher heat dissipation. Furthermore, additional methods for error handling in the chips have to be devised.
Overall it is a worth while approach as the required effort for the implementation is low compared to the gain in the integration density.

Hybrid approach of WSI and Wafer Embedding The roadmap for the large-scale BrainScaleS system replacement is a long-term goal and will not finish before the end of the Human Brain Project. On the other hand the chip development in the Electronic Visions Group continues and new generations of HICANN wafers will be produced. For the planned wafers with the 65 nm process the current WSI concept will be reused. The available mechanical components and PCBs are kept. Especially, the MainPCBs (MainPCBs) are reused because a redesign and production of new boards takes at least one year. Therefore, the RDL on top of the 65 nm wafer needs to have the same pin-out like the 180 nm wafer.
The 65 nm wafers are produced on 30 cm silicon disks. Hence, the 20 cm wafers - the size in the current system - are cut out of the 30 cm wafers. If the final BrainScaleS system replacement is not ready, a hybrid approach of the WSI concept and the wafer embedding technique could be an intermediate step to increase the integration density. It would use the full 30 cm wafers. The wafers are embedded in a PCB with several layers and backside connections but are sill integrated in a WSI system. In fig. 9.2 a possible structure of a hybrid system is depicted. The embedding is not used as a replacement of the MainPCB but rather as another packaging step of the wafer. The elastomeric connectors are still used for the vertical connection with a new MainPCB. The restrictions for a new MainPCB will be the same as before and, therefore, the stack up will be the same. So it does not reduce the manufacturing time of the MainPCB. The major advantages are an easier development of the MainPCB and a all chips on the wafer could be connected with high-speed lines.
On the lowest copper layers of the MainPCB where the elastomeric connector pads are placed a lot of signal- and power-redistribution is done. Parts of it could be moved to the copper layers on the embedded wafer. The combination of the same power supply nets reduces the number of pads going to the MainPCB. Furthermore, each HICANN chip in a reticle has its own Joint Test Action Group (JTAG) clock TCK and mode select signal TMS. By distributing these signals on the embedded copper layers the number of signal connections going through the elastomeric connectors is reduced.

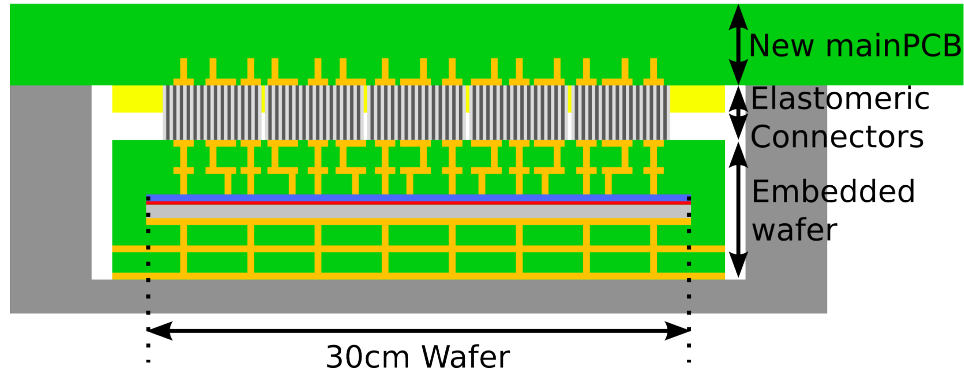


Figure 9.2: Possible setup of a hybrid approach system. The 30 cm HICANN wafer is embedded in the PCB which is connected via the elastomeric connectors to a redesigned MainPCB.

The saved space could be used to increase the height of the signal pads, this way the alignment of the wafer and the MainPCB would be easier and less error-prone. Eventually, the signal pads could be placed in a way that the routing on the MainPCB is more uniform and, hence, easier.
It would be possible to add more RDLs to achieve these advantages too. However, for the RDLs additional full-wafer masks are required, which are expensive and every change requires a new set of masks. The embedding does not need any masks for the production because Laser Direct Imaging (LDI) is used. The layout can be improved with every new board.
In this use case the wafer embedding is a fan-out packaging technology and the package can have an arbitrary shape. It could be a circular shape again or a quadratic shape.
The embedding of the wafer is not a problem for the heat dissipation. The back of the wafer can be connected with thermal vias and the copper layers below the wafer can be used for heat spreading. Furthermore, the prepreg layers are only 50 µm thick and do not disturb the heat transfer.

Depending on the target value of embedded wafers the production can be done at the IZM in Berlin. A low number of wafers can be embedded in Berlin, but for a high-volume production another PCB manufacturer has to be found.

The combination of the WSI concept and the embedding technique facilitates the transition from 20 cm to 30 cm wafers. It reduces the complexity of the signal- and power-redistribution on the MainPCB. The development of the board becomes easier and takes less time. Furthermore, it should be possible to connect all reticles to their corresponding FPGA Communication PCBs with high-speed lines.
The minimal requirements for the embedding process be an improvement in a hybrid approach system are backside vias and at least four copper layers. A prototype with backside vias is currently in production and the first test results are expected until the end of the year. Then, the next step is to add a second copper layer to the stack

up which will be tested in 2018. The tests will embed 20 cm wafers. If this tests succeed, then the embedding tests with 30 cm bare silicon disks could start.
The hybrid approach will not reach the targeted integration level of the final Embedded Wafer System (EWS) but it could be an intermediate solution to bridge the time until the final BrainScaleS system replacement is ready.

Neuromorphic Physical Model version 3 (NM-PM-3) In the Framework Partnership Agreement the Electronic Visions Group proposed two possible future developments for the neuromorphic hardware cluster until the end of the Human Brain Project [*HBP*, 2015]. One is based on the current WSI concept and has a maximum cluster size of 50 systems, whereas the limiting factors of the cluster size are the system volume and the complexity of the MainPCB and the assembly process. The hybrid approach could simplify the MainPCB. This way the development and production of the board is less sophisticated. Nevertheless, it would not remove the elastomeric connectors and the alignment procedure, therefore, still requires a lot of human interaction.

The only improvement in terms of cluster size comes with the second proposal in the Framework Partnership Agreement (FPA). It plans to increase the cluster size to 500 systems. This is only achievable with a higher level of automation and integration of the systems. Therefore, the concept is based on the embedding of silicon wafers in PCBs. It removes the elastomeric connectors and most of the tasks that have to be done by hand like the alignment of the wafer and the MainPCB. Microvias replace the elastomeric connectors to create the connections to the silicon wafer. The position of the microvias is adjusted accordingly to the fiducial marks on the wafer automatically by the drilling machine. The same holds true for the alignment of the board layout to the wafer.
The goal is to have boards that just need to be attached to a cooling module and then are ready for the assembly in a standard 19-inch server rack. In the rack the cables for the power supply and external communication are plugged in.

Aside from the further development of the embedding process the neuromorphic chips need to fulfill certain criteria. The chip needs to be more power efficient in comparison to the current HICANN chip. It was shown that a heat dissipation of 500 W is feasible to cool down. For that amount of heat a water cooling of the system is recommended, as air cooling needs more space. At the moment the power consumption for a complete wafer is targeted to be around 100 W which can be cooled by air. Another critical point is the number of signal lines leaving the wafer. The embedded wafer cannot have an arbitrary number of copper layers in the PCB stack up. For comparison the current HICANN design has to save half of its Low-Voltage Differential Signaling (LVDS) links to fit into an Embedded Wafer System with four copper layers.

The embedding procedure itself does not increase the system density. Other components of the BrainScaleS system like the FPGA Communication PCBs (FCP boards) and the power supply boards determine the overall system height. The

power regulators either have to be put on the wafer directly, which would require additional circuit development, or use low-profile power modules like the VTM48MP0-series with the drawback of soldering on the board [*Vicor*, 2017]. To remove the FCP board the communication scheme of the neuromorphic chips has to be changed. It is more efficient, if there is no difference between the chip-to-chip communication and the chip-to-host communication. Therefore, a solution would be the implementation of a router circuit which is placed on the chip and is responsible for the off-wafer communication. [*Schemmel*, 2014]. Some of the routers would be connected to host computers for the transfer of the network configuration and the neural events. The other router chips can connect to other modules in the same rack or to wafers in other racks. Ideally, the assembled racks with EWSs are placed in a toroidal structure.

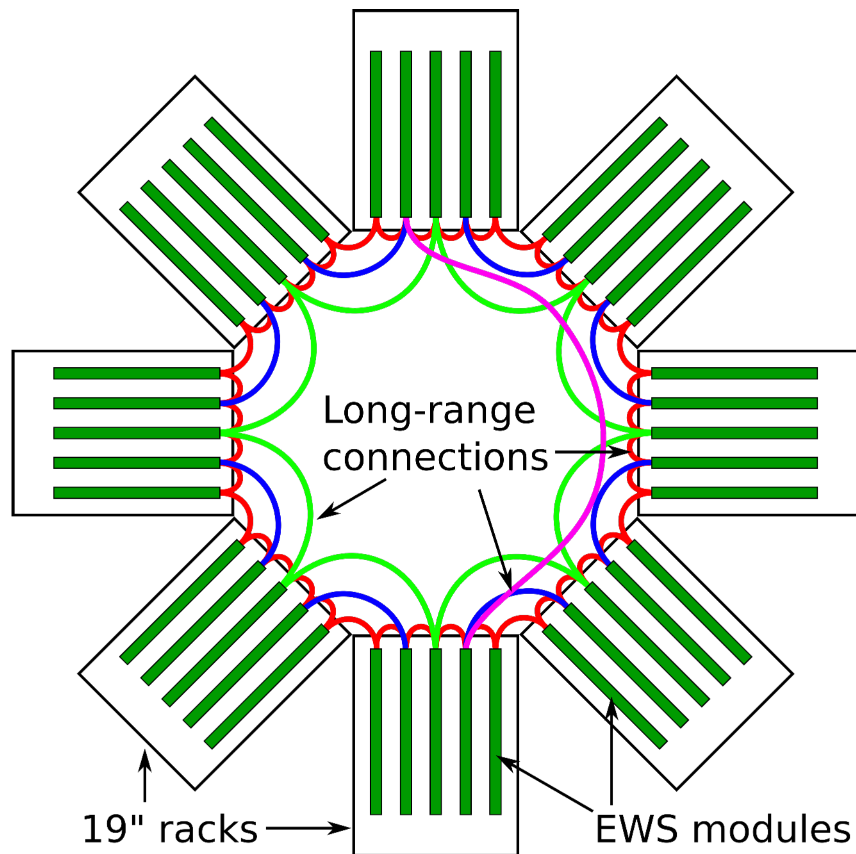


Figure 9.3: Top view on eight 19-inch server racks in a toroidal structure. For visualization each rack has five EWS modules on one level. The cables between the modules are laid in the inner area. The shortest connection are to adjacent modules (red lines), the longest is between opposite racks (magenta line).
A rack has several levels, hence, there are vertical connections between different levels too.

In fig. 9.3 a possible placement of racks is depicted. A toroidal structure has two major advantages over placing the racks next to each other. The distance between adjacent modules is constant, therefore, only one cable length is required. This reduces the costs of the cables as only one cable length has to be ordered. Furthermore, the longest distance is between opposite racks and is only half as long as in a standard rack configuration. Thus, the maximum latency induced by the interconnection network is bisected which is decisive for a time-continuous working neuron. The final wiring diagram is not yet developed as it depends, for example, on the number of links available on one Embedded Wafer System.

The integration of the FCP board functionality and the other external components on the wafer is essential for the reduction of the system height. This process also reduces the total number of components in a system. The final system consists of only a silicon wafer embedded into a PCB, a cooling module and some connectors. So the width of an Embedded Wafer System is minimized to around 1 cm. Thus, allowing a higher integration level of the systems.

The potential to achieve a highly integrated neuromorphic hardware cluster is available. A lot of progress is still needed; on the printed circuit board and on the silicon chip. Additionally, the design of the printed circuit boards and the neuromorphic chips get more and more closely interlinked and influence each other.
Anyhow, there will be exciting developments in the system design of the BrainScaleS system.

Bibliography

3M, *Novec 649 Data Sheet*, 3M, 2017, [Online]. Available: http://multimedia.3m.com/mws/media/6043950/3mtm-novectm-649-engineered-fluid.pdf?fn=Novec%20649_Datenblatt_052014.pdf.

Aamir, S. A., Personal communication, 2017.

Alevi, D., Investigating competitive dynamics in a recurrent neural network on neuromorphic hardware, Bachelor thesis, Ruprecht-Karls-Universität Heidelberg, 2015.

ANSI/TIA/EIA-644, *Electrical Characteristics of Low Voltage Differential Signalling (LVDS)*, 1996.

AT&S, *HDI Microvia PCBs*, Accessed: 2017-05-02, 2017, [Online]. Available: http://www.ats.net/products-technology/product-portfolio/hdi-microvia-pcbs/.

Binder, *MK 53 Data Sheet*, Binder, 2017.

Bing, J., Personal communication, 2017.

Boettcher, L., D. Manessis, A. Ostmann, S. Karaszkiewicz, and H. Reichl, Embedding of chips for system in package realization - technology and applications, in *2008 3rd International Microsystems, Packaging, Assembly Circuits Technology Conference*, pp. 383–386, doi:10.1109/IMPACT.2008.4783892, 2008.

Brette, R., and W. Gerstner, Adaptive exponential integrate-and-fire model as an effective description of neuronal activity, *J. Neurophysiol.*, *94*, 3637 – 3642, doi:NA, 2005.

Böttcher, L., Personal communication, 2015.

Böttcher, L., Personal communication, 2016.

Cadence, D. S. I., *Best Practices: Working with Allegro PCB SI Field Solvers*, 2009.

Cadence, D. S. I., *Effective Use off Cross Section Parameters in Allegro PCB Editor/SI*, 2014.

Cadence, D. S. I., Allegro pcb designer, www.cadence.com, 2017a.

Cadence, I., Design Systems, Allegro skill, www.cadence.com, 2017b.

CCI Eurolam, Copper-invar-copper, Accessed: 2017-05-17, 2017.

Chasin, A., et al., Impact of wafer thinning on front-end reliability for 3d integration, in *2016 IEEE International Reliability Physics Symposium (IRPS)*, pp. 6B–2–1–6B–2–6, doi:10.1109/IRPS.2016.7574562, 2016.

Chen, T. Y., M. K. Shih, M. H. Chen, W. C. Chen, S. Uegaki, C. L. Kao, P. F. Yang, and C. P. Hung, Warpage characterization analysis for embedded package technology, in *2016 International Conference on Electronics Packaging (ICEP)*, pp. 518–521, doi:10.1109/ICEP.2016.7486881, 2016.

Cree, I., *Optimizing PCB Thermal Performance for Cree XLamp LEDs*, Acessed: 2017-07-26, 2017, [Online]. Available: `http://www.cree.com/led-components/media/documents/XLamp_PCB_Thermal.pdf`.

Dassualt Systèmes, *SolidWorks*, 2016, [Online]. Available: `http://www.solidworks.com`.

Destexhe, A., Conductance-based integrate-and-fire models, *Neural Comput.*, *9*(3), 503–514, doi:http://dx.doi.org/10.1162/neco.1997.9.3.503, 1997.

Dow Chemical Company, L., *Cyclotene 4000 Series Processing Procedures*, Accessed: 2017-05-29, 2017, [Online]. Available: `http://www.dow.com/webapps/lit/litorder.asp?filepath=cyclotene/pdfs/noreg/618-00285.pdf&pdf=true`.

Friedmann, S., Extending a hardware neural network beyond chip boundaries, Diploma thesis, Ruprecht-Karls-Universität Heidelberg, hD-KIP-09-41, 2009.

Friedmann, S., A new approach to learning in neuromorphic hardware, Ph.D. thesis, Ruprecht-Karls-Universität Heidelberg, 2013.

Fujioka, H., *SESUB Technology - Embedded Package*, Accessed: 2017-07-24, 2015, [Online]. Available: `http://http://prod.semicontaiwan.org/en/sites/semicontaiwan.org/files/data15/docs/3_5._sesub-technology_introduction_for_semicon_taiwan2015.pdf`.

Fujipoly, *Zebra Elastomeric Connector*, [Online]. Available: `http://www.fujipoly.com/usa/products/zebra-elastomeric-connectors/high-performance-silver-connectors.html`.

Furber, S. B., F. Galluppi, S. Temple, and L. A. Plana, The spinnaker project, *Proceedings of the IEEE*, *102*(5), 652–665, doi:10.1109/JPROC.2014.2304638, 2014.

Garrou, P., C. Bower, and P. Ramm (Eds.), *Handbook of 3D integration*, vol. 1:, Wiley-VCH, Weinheim, 2008.

Garrou, P., M. Koyanagi, and P. Ramm (Eds.), *Handbook of 3D Integration, Volume 3: 3D Process Technology*, Wiley, 2014.

Grübl, A., Personal communication, 2017.

Güttler, M., Konzeptoptimierung und Entwicklung einer hochintegrierten Leiterplatte, Diploma thesis (German), University of Heidelberg, HD-KIP-10-68, 2010.

Haynes, W. M. (Ed.), *2016 - 2017*, CRC handbook of chemistry and physics ; 97th edition, 97th edition ed., CRC Press, Boca Raton, 2016.

HBP, *Framwork Partnership Agreement*, Human Brain Project, 2015.

HBP SP9, *Guidebook*, Human Brain Project, 2015, [Online]. Available: https://electronicvisions.github.io/hbp-sp9-guidebook/.

HBP SP9 partners, *Neuromorphic Platform Specification*, Human Brain Project, 2014.

Hitachi, L., Chemical Company, *MCL-E-679FG Data Sheet*, Accessed: 2017-05-19, 2017a, [Online]. Available: http://www.hitachi-chem.co.jp/englisch/products/bm/b02/files/bm_b02_004.pdf.

Hitachi, L., Chemical Company, *MCL-E-770G(LH) Data Sheet*, Accessed: 2017-05-17, 2017b, [Online]. Available: http://www.hitachi-chem.co.jp/englisch/products/bm/b02/files/bm_b02_006.pdf.

Hitachi Chemical Company Ltd., Personal communication, 2017.

Husmann, D., and H. Zoglauer, A wafer-scale-intagration system(WSI), FACETS project internal documentation, 2010.

IEEE, IEEE Standard Test Access Port and Boundary-Scan Architecture, *IEEE Std 1149.1-2001*, pp. i –200, doi:10.1109/IEEESTD.2001.92950, 2001.

Ilmberg, J., Development of the control unit for reticles (cure) for the wafer-scale integration system (wsi) v2.1, 2014.

Ingerly, D., et al., Low-k interconnect stack with metal-insulator-metal capacitors for 22nm high volume manufacturing, in *2012 IEEE International Interconnect Technology Conference*, pp. 1–3, doi:10.1109/IITC.2012.6251663, 2012.

Instruments, T., *TPS8267x Data Sheet*, Accessed: 2017-05-25, 2017, [Online]. Available: http://www.ti.com/lit/ds/symlink/tps826745.pdf.

Intel, C., *Intel Core i7-975 Processor Extreme Edition Specification*, Accessed: 2017-07-22, 2017a, [Online]. Available: http://ark.intel.com/products/37153/Intel-Core-i7-975-Processor-Extreme-Edition-8M-Cache-3_33-GHz-6_40-GTs-Intel-QPI.

Intel, C., *Enhanced Temperature Device Support*, Accessed: 2017-06-07, 2017b, [Online]. Available: https://www.altera.com/products/common/temperature/ind-temp.html.

Ipdia, *LPSC-Series*, Ipdia, 2017, [Online]. Available: http://www.ipdia.com/download.php?file=/3d_silicon_capacitors/Commercial%20Leaflet_LPSC_V1.1._Murata.pdf&cat=products.

Isola, G., *FR408HR Data Sheet*, Accessed: 2017-05-17, 2017, [Online]. Available: http://www.isola-group.com/wp-content/uploads/FR408HR-Laminate-and-Prepreg-Data-Sheet-0317.pdf.

Isola, G., *FR408HR Processing Guide*, Acessed: 2017-07-26, 2017, [Online]. Available: http://www.isola-group.com/wp-content/uploads/FR408HR-High-Performance-Laminate-and-Prepreg-Processing-Guide.pdf.

ITRS, *International Technology Roadmap for Semiconductors*, 2011 ed., 2011.

IZM, Pcb lamination press, Accessed: 2017-07-30, 2017.

Johnson, H. W., and M. Graham, *High speed digital design*, 29. [print.] ed., XI, 447 S. pp., Prentice Hall PTR, Upper Saddle River, NJ, 2011.

Karasenko, V., A communication infrastructure for a neuromorphic system, Master's thesis (English), University of Heidelberg, 2014.

Keller, S., Personal communication, 2011.

Khan, N. H., S. M. Alam, and S. Hassoun, GND plugs: A superior technology to mitigate TSV-induced substrate noise, *IEEE Transactions on Components, Packaging and Manufacturing Technology*, *3*(5), 849–857, doi:10.1109/tcpmt.2013.2241178, 2013.

Kiene, G., Evaluating the synaptic input of a neuromorphic circuit, Bachelor thesis, University of Heidelberg, 2014.

Kim, J., and Y. Kim, Hbm: Memory solution for bandwidth-hungry processors, in *2014 IEEE Hot Chips 26 Symposium (HCS)*, pp. 1–24, doi:10.1109/HOTCHIPS.2014.7478812, 2014.

Kleider, M., Neuron Circuit Characterization in a Neuromorphic System, Ph.D. thesis, University of Heidelberg, 2017.

Koke, C., Device variability in synapses of neuromorphic circuits, Ph.D. thesis, Universität Heidelberg, 2017.

Kotake, T., H. Murai, S. Takanezawa, M. Miyatake, M. Takekoshi, and M. Ose, Ultra low cte (1.8 ppm/c) core material for next generation thin csp, in *2014 IEEE 64th Electronic Components and Technology Conference (ECTC)*, pp. 1407–1410, doi:10.1109/ECTC.2014.6897477, 2014.

Kriechbaum, A., H. Stahr, M. Biribauer, N. Haslebner, and M. Morianz, ECP – Embedded Component Packaging Technology, in *2011 SEMICON Europa Conference*, 2011.

Kuchling, H., *Taschenbuch der Physik*, 19., aktualisierte aufl. ed., 711 S. pp., Fachbuchverl. Leipzig im Carl-Hanser-Verl., 2007.

Kugele, A., Master thesis, University of Heidelberg, In preparation, 2017.

LeCun, Y., and C. Cortes, The mnist database of handwritten digits, 1998.

Mangroli, A., and K. Vasoya, Optimizing thermal and mechanical performance in pcbs, Accessed: 2017-05-18, 2008.

Mauch, C., Personal communication, 2016.

Mauch, C., Personal communication, 2017.

Menzel, R., and M. Giurfa, Cognitive architecture of a mini-brain: the honeybee, *Trends in Cognitive Sciences*, *5*(2), 62–71, doi:10.1016/S1364-6613(00)01601-6, 2001.

Merolla, P. A., et al., A million spiking-neuron integrated circuit with a scalable communication network and interface, *Science*, *345*(6197), 668–673, doi:10.1126/science.1254642, 2014.

Merrill, J., Sonnet support personal communication, 2015.

Meyer, T., G. Ofner, S. Bradl, M. Brunnbauer, and R. Hagen, Embedded wafer level ball grid array (ewlb), in *2008 10th Electronics Packaging Technology Conference*, pp. 994–998, doi:10.1109/EPTC.2008.4763559, 2008.

Millner, S., Development of a multi-compartment neuron model emulation, Ph.D. thesis, Ruprecht-Karls University Heidelberg, 2012.

Montrose, M. I., *Printed circuit board design techniques for EMC compliance*, IEEE Press series on electronics technology, IEEE Press [u.a.], Piscataway, NJ, 1996.

Moore, G. E., Cramming more components onto integrated circuits, *Electronics*, *38*(8), 1965.

Neeb, C., J. Teichrib, R. W. D. Doncker, L. Boettcher, and A. Ostmann, A 50 kw igbt power module for automotive applications with extremely low dc-link inductance, in *2014 16th European Conference on Power Electronics and Applications*, pp. 1–10, doi:10.1109/EPE.2014.6910953, 2014.

Nonnenmacher, T., Characterization of spike-timing dependent plasticity in neuromorphic hardware, Master thesis, Ruprecht-Karls-Universität Heidelberg, 2015.

Nowak, T., R. Schacht, H. Walter, B. Wunderle, M. A. Ras, D. May, O. Wittler, and K. D. Lang, Experimental and numerical reinvestigation for lifetime-estimation of plated through holes in printed circuit boards, in *2011 17th International Workshop on Thermal Investigations of ICs and Systems (THERMINIC)*, pp. 1–6, 2011.

Orbotech, *Paragon 9000*, [Online]. Available: https://www.orbotech.com/Direct_Imaging/.

Pak, J. S., J. Kim, J. Lee, H. Lee, K. Park, and J. Kim, Sharing power distribution networks for enhanced power integrity by using through-silicon-via, in *2008 Electrical Design of Advanced Packaging and Systems Symposium*, Institute of Electrical & Electronics Engineers (IEEE), doi:10.1109/edaps.2008.4735985, 2008.

Pavan, P., L. Larcher, and A. Marmiroli, *Floating gate devices*, XIII, 129 S. pp., Kluwer Academic, Boston [u.a.], includes bibliographical references and index, 2004.

Pfeil, T., Exploring the potential of brain-inspired computing, Ph.D. thesis, Ruprecht-Karls-Universität Heidelberg, 2015.

Philipp, S., Design and implementation of a multi-class network architecture for hardware neural networks, Ph.D. thesis, Ruprecht-Karls Universität Heidelberg, 2008.

Rosenblatt, F., The perceptron: a probabilistic model for information storage and organization in the brain, *Psychological Review*, *65*, 386–408, 1958.

Samtec, *HSEC8-DV-Series Data Sheet*, Samtec, 2017a, [Online]. Available: http://suddendocs.samtec.com/catalog_english/hsec8_dv.pdf.

Samtec, *ZA1-Series Data Sheet*, Samtec, 2017b, [Online]. Available: http://suddendocs.samtec.com/catalog_english/za1.pdf.

Schemmel, J., Personal communication, 2014.

Schemmel, J., Personal communication, 2016.

Schemmel, J., K. Meier, and F. Schürmann, *A VLSI Implementation of an Analog Neural Network Suited for Genetic Algorithms*, pp. 50–61, Springer Berlin Heidelberg, Berlin, Heidelberg, doi:10.1007/3-540-45443-8_5, 2001.

Schemmel, J., F. Schürmann, S. Hohmann, and K. Meier, An integrated mixed-mode neural network architecture for megasynapse ANNs, in *Proceedings of the 2002 International Joint Conference on Neural Networks IJCNN'02*, pp. 2704–2710, IEEE Press, 2002.

Schemmel, J., A. Grübl, K. Meier, and E. Muller, Implementing synaptic plasticity in a VLSI spiking neural network model, in *Proceedings of the 2006 International Joint Conference on Neural Networks (IJCNN)*, IEEE Press, 2006.

Schemmel, J., D. Brüderle, A. Grübl, M. Hock, K. Meier, and S. Millner, A wafer-scale neuromorphic hardware system for large-scale neural modeling, in *Proceedings of the 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1947–1950, 2010a.

Schemmel, J., A. Grübl, and S. Millner, Specification of the HICANN microchip, FACETS project internal documentation, 2010b.

Schmidt, D., Automated characterization of a wafer-scale neuromorphic hardware system, Master thesis, Ruprecht-Karls-Universität Heidelberg, 2014.

Schmitt, K., Personal communication, 2017.

Schmitt, S., et al., Neuromorphic hardware in the loop: Training a deep spiking network on the brainscales wafer-scale system, *Proceedings of the 2017 IEEE International Joint Conference on Neural Networks*, doi:10.1109/IJCNN.2017.7966125, 2017.

Schmoll, M. G., *Pico Drill Datasheet*, 2017, [Online]. Available: `http://www.schmoll-laser.de/fileadmin/pdf/Schmoll-Lasertechnology.pdf`.

Schwartz, M.-O., Reproducing biologically realistic regimes on a highly-accelerated neuromorphic hardware system, Ph.D. thesis, Universität Heidelberg, 2013.

Schwerz, R., M. Roellig, S. Osmolovskyi, and K. J. Wolter, Reliability assessment of discrete passive components embedded into pcb core, in *2014 15th International Conference on Thermal, Mechanical and Mulit-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE)*, pp. 1–7, doi:10.1109/EuroSimE.2014.6813776, 2014.

Sonnet, S. I., Sonnet 3D planar EM solver, www.sonnetsoftware.com, 2017.

Sterzenbach, L., Entwicklung einer selbstüberwachenden Spannungsversorgung für ein auf Wafer-Ebene integriertes neuromorphes Hardware-System, 2015.

Sukegawa, S., et al., A 1/4-inch 8mpixel back-illuminated stacked cmos image sensor, in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, pp. 484–485, doi:10.1109/ISSCC.2013.6487825, 2013.

TDK, *SESUB-PAN-D14580*, Accessed: 2017-07-24, 2017, [Online]. Available: `https://product.tdk.com/info/en/documents/data_sheet/sesub-pan-d14580_en.pdf`.

TechInsights, I., *Apple iPhone 7 Teardown*, Accessed: 2017-07-23, 2017, [Online]. Available: `http://www.techinsights.com/about-techinsights/overview/blog/apple-iphone-7-teardown/`.

Tseng, C. F., C. S. Liu, C. H. Wu, and D. Yu, Info (wafer level integrated fan-out) technology, in *2016 IEEE 66th Electronic Components and Technology Conference (ECTC)*, pp. 1–6, doi:10.1109/ECTC.2016.65, 2016.

Vardaman, E. J., Fo-wlp market and technology trends, in *2017 International Conference on Electronics Packaging (ICEP)*, pp. 318–320, doi:10.23919/ICEP.2017.7939385, 2017.

Vicor, C., *48V Direct to CPU*, 2017, [Online]. Available: http://www.vicorpower.com/industries-computing/48v-direct-to-cpu.

Würth Elektronik, *HDI Design Guide*, 2015.

Zoglauer, H., Entwicklung und testergebnisse eines prototypensystems für die wafer-scale-integration, Diploma thesis (German), University of Heidelberg, HD-KIP-09-28, 2009.

Zoschke, K., *IZM Progress Report SP9 - Second Quaterly Meeting*, 2015.

Zoschke, K., Personal communication, 2016.

Zoschke, K., Personal communication, 2017.

Zoschke, K., M. Güttler, L. Böttcher, A. Grübl, D. Husmann, J. Schemmel, K. Meier, and O. Ehrmann, Full wafer redistribution and wafer embedding as key technologies for a multi-scale neuromorphic hardware cluster, In preparation, Accepted: 2017-07-05, 2017.

A BrainScaleS system

Name	Quantity
HICANN wafer	1
FCP board	48
WIO board	4
MainPCB	1
AnaB boards	2
AuxPwr boards	2
CURE boards	8
PowerIt	1
Sum	66

Table A.1: List of all electronic components used in one wafer module.

Name	Quantity
Top Cover	1
Wafer Bracket	1
Elastomeric Connector Positioning Mask	1
Vertical rods	4
Horizontal rods	8
Sum	15

Table A.2: List of mechanical components used in one wafer module.

Voltage Name	Voltage [V]	Maximum Current [A]	Maximum voltage difference [mV]
VDDD+VDDOUT	1.8	8	50
VDDA	1.8	8	50
DI_VCC	1.8	1	100
DI_VCCANA	1.8	0.5	100
DI_VCC33ANA	3.3	0.5	100
DI_VBIAS	1.25	< 0.01	100
VOL	0.6 - 1	0.4	50
VOH	0.8 - 1.2	0.4	50
VDDBUS	1.2	2.4	50
VDD25	2.5	0.5	50
VDD5	5	< 0.01	50
VDD12	11 - 11.5	< 0.01	50

Table A.3: List of supply voltages on a reticle. The voltage difference between all reticles on a wafer has to be smaller than the maximum voltage difference.

Layer	Line Width [µm]	Separation [µm]
Top	80	100
S5	80	200
S4	90	200
S3	90	200
S2	80	200
S1	80	200
Bottom	80	100

Table A.4: Line geometries for the differential lines on the MainPCB. The target differential impedance of the lines is $100\,\Omega$.

Layer	Line Width [µm]
Top	100
S5	100
S4	100
S3	100
S2	90
S1	90
Bottom	90

Table A.5: Line properties for the single-ended lines with an impedance of 50 Ω on the MainPCB.

Model Name	Hardware Name	HW range	Floating Gate
resting potential	V_{rest}	0-1.2 V	E_l
spike threshold	V_t	0-1.2 V	V_t
reset potential	V_{reset}	0-1.2 V	$V_{reset}(*)$
excitatory synaptic reversal potential	E_{synx}	0-1.2 V	E_{synx}
inhibitory synaptic reversal potential	E_{syni}	0-1.2 V	E_{syni}
excitatory synaptic time constant	τ_{synx}		V_{syntcx}
inhibitory synaptic time constant	τ_{syni}		V_{syntci}
refractory period	τ_{ref}	0.01-6 µs	I_{pl}
leakage conductance	g_l	0.1-6 µs	I_{gl}
membrance time constant	τ_{mem}	0.5-15 µs	I_{gl}
adaptation coupling parameter	a		$I_{gladapt}$
adaptation time constant	τ_w		I_{radapt}
spike-triggered adaptation	b		I_{fire}
effective threshold potential	V_{exp}	0-1.2 V	V_{exp}
slope factor	Δ_t		I_{bexp}

Table A.6: Neuron parameters of the Adaptive exponential integrate-and-fire Model (Adex) on the HICANN chip.
Taken from [*Schmidt*, 2014]

B Bare Silicon Tests

X-ray Images

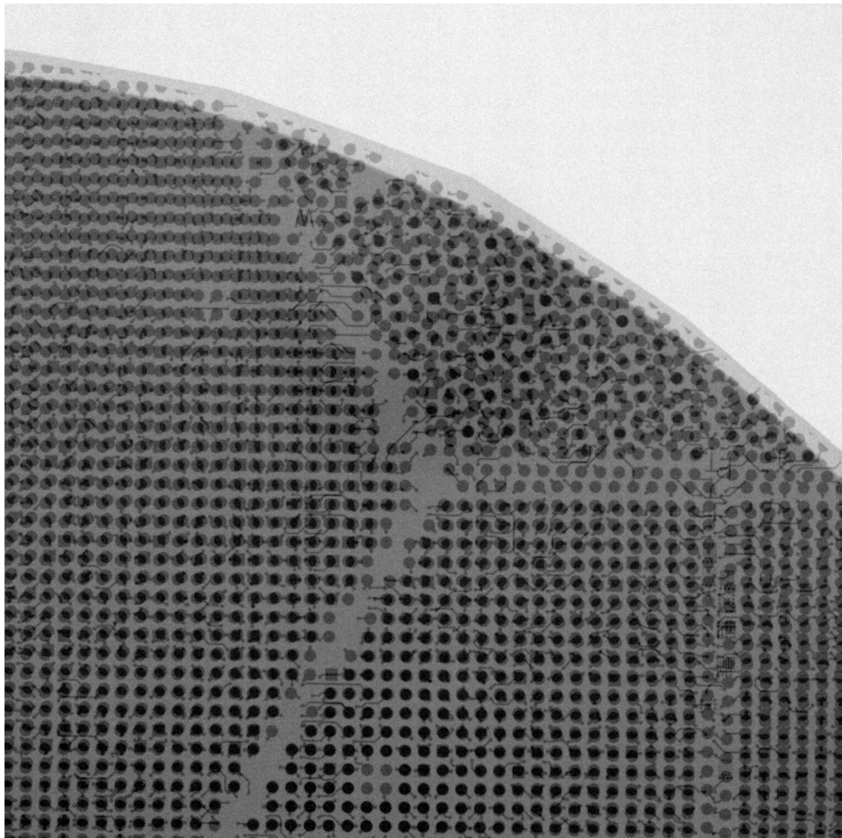


Figure B.1: X-ray image of the first lamination prototype. The wafer was cut out of the board and the copper structures were etched on top and on bottom. Courtesy of Kai Zoschke (IZM).

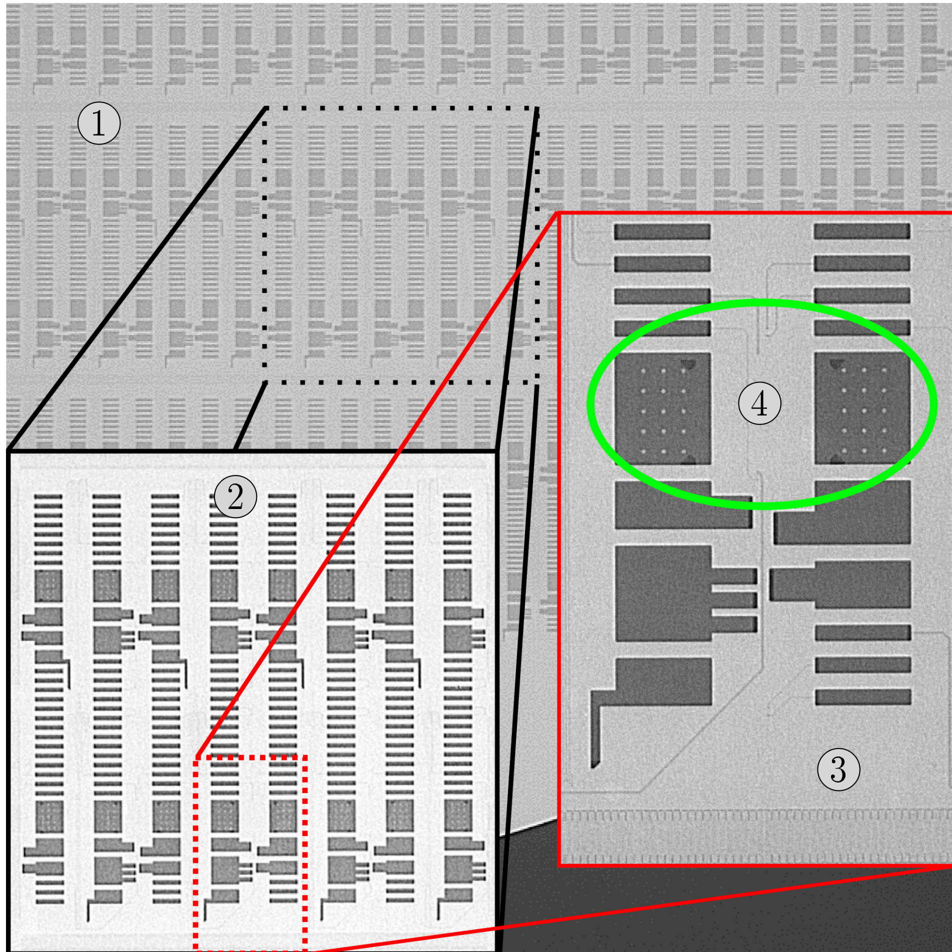


Figure B.2: An X-ray image of the second board after the laser drilling tests. In the background (1) is the silicon wafer, (2) is an enlargement view of one reticle and (3) shows the lower contact pads. At (4) are the 60 µm holes which are at the correct position. Courtesy of Kai Zoschke (IZM). The images are modified.

Test Points for the Accelerated Environmental Stress Tests

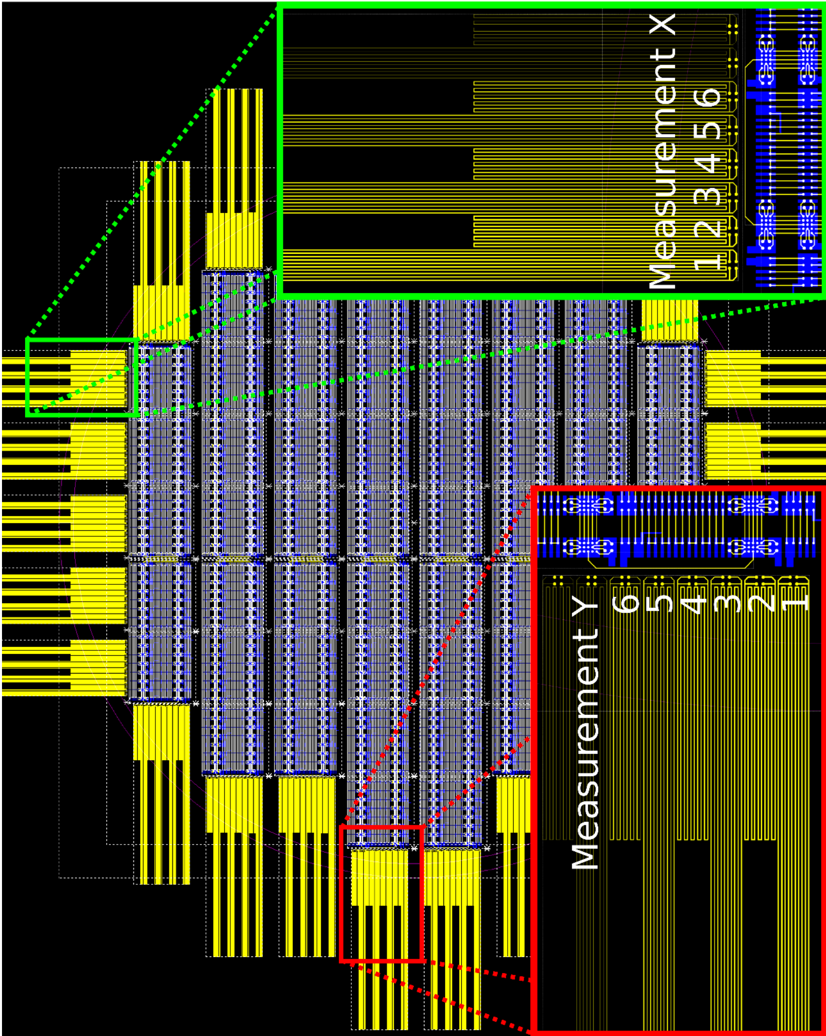


Figure B.3: Layout of the third lamination test board. Additionally, enhanced views on the test pins for the resistance measurements are shown. The measurement lines go across the wafer edge several times. Courtesy of Kai Zoschke (IZM).

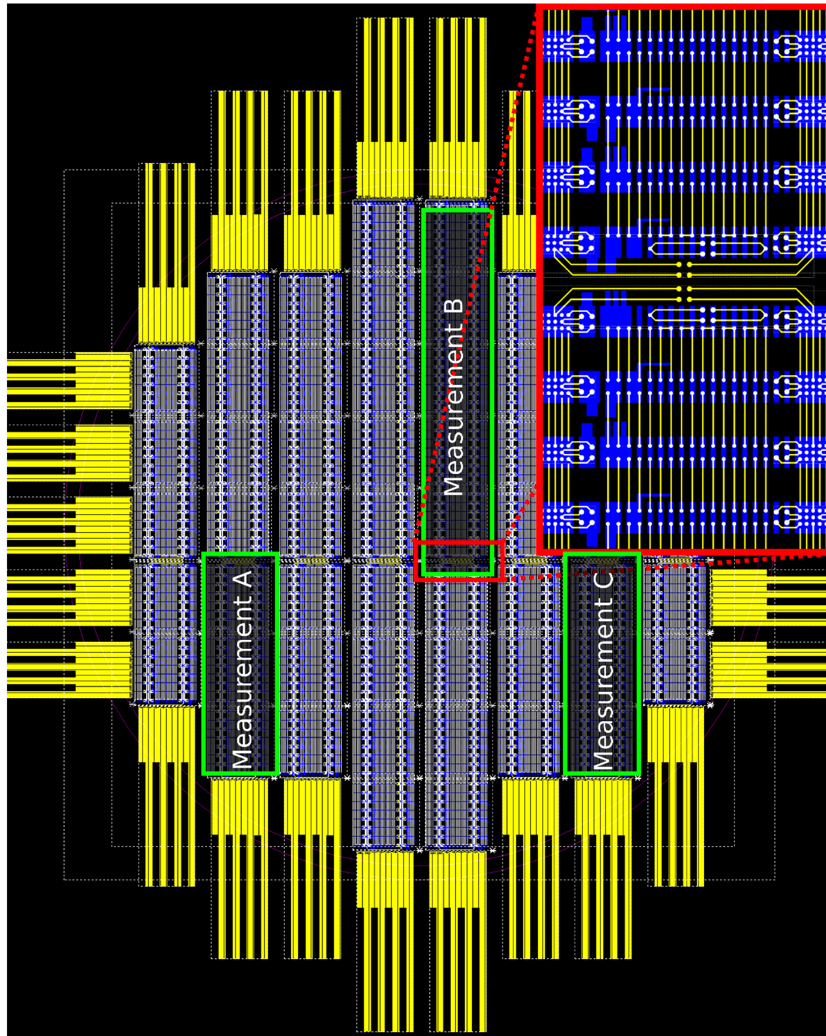


Figure B.4: Layout of the third lamination test board. The enhanced view shows the test points for the daisy chain measurements on the wafer. Courtesy of Kai Zoschke (IZM).

C DENSE Preparation

Sonnet Simulation Stack Up

Layer	Material	Thickness [µm]	Dielectric Constant
0	air	$40 * 10^3$	1.0
1	metall layer	$1 * 10^3$	
2	soldermask	35	3.5
3	copper	16	-
4	prepreg ($)	50	4.0
5	Benzocyclobutene (BCB) (†)	20	2.65

Table C.1: Stack up for the calculation of the impedance on the inner reticles.
($) Hitachi MCL-E-770G(LH) [*Hitachi*, 2017b].
(†) Data taken from Cyclotene 4000 [*Dow Chemical Company*, 2017].

Layer	Material	Thickness [µm]	Dielectric Constant
0	air	$40 * 10^3$	1.0
1	soldermask	35	3.5
2	copper	16	-
3	prepreg ($)	50	4.0
4	prepreg ($)	50	4.0

Table C.2: Stack up for the calculation of the impedance over the CIC material. The soldermask is implemented with two layers. The copper lines run in the first layer of solder resist and, therefore, the second layer of solder resist is needed to cover the copper line. The height of the additional soldermask is 16 µm.
($) Hitachi MCL-E-770G(LH) [*Hitachi*, 2017b]

D DENSE Measurements

HICANN	Error message
9	wrong hicann id 0xffffffff, fpga id 0xfffffff
29	wrong hicann id 0xffffffff, fpga id 0xfffffff
53	wrong hicann id 0xffffffff, fpga id 0xfffffff
81	wrong hicann id 0xffffffff, fpga id 0xfffffff
101	link status 0x0a
109	wrong hicann id 0xffffffff, fpga id 0xfffffff
117	link status 0x0a
137	wrong hicann id 0xffffffff, fpga id 0xfffffff
161	link status 0x0a
169	wrong hicann id 0xffffffff, fpga id 0xfffffff
177	wrong hicann id 0x14008404, fpga id 0x0
197	link status 0x0a
205	wrong hicann id 0xffffffff, fpga id 0xfffffff
213	wrong hicann id 0x10808004, fpga id 0x10c001
241	wrong hicann id 0xffffffff, fpga id 0xfffffff
277	wrong hicann id 0xffffffff, fpga id 0xfffffff
309	wrong hicann id 0xffffffff, fpga id 0xfffffff
337	wrong hicann id 0xffffffff, fpga id 0xfffffff
365	wrong hicann id 0x00000000, fpga id 0x0000000
373	wrong hicann id 0xffffffff, fpga id 0xfffffff
389	link status 0x0a
393	wrong hicann id 0x00000000, fpga id 0x0000000
401	wrong hicann id 0xffffffff, fpga id 0xfffffff
417	wrong hicann id 0x00000000, fpga id 0x0000000
433	transmission error, wrong receive values
437	wrong hicann id 0x00000000, fpga id 0x0000000

Table D.1: Error messages of the Switchram test for the DENSE001 with Automatic Repeat reQuest.

HICANN	RX_CLK [Ω]	RX_DAT [Ω]	TX_CLK [Ω]	TX_DAT [Ω]
9	x	x	113.1	114.5
29	115.8	114.0	113.8	113.1
53	115.7	114.7	113.5	114.4
81	113.2	114.2	113.1	112.0
101	116.8	x	115.3	114.0
109	116.4	114.8	113.5	114.5
117	117.3	116.8	115.7	114.5
137	116.7	115.6	115.3	114.4
161	115.6	115.0	113.7	114.0
169	116.5	115.4	113.7	114.2
177	118.0	116.6	116.2	115.2
197	116.9	115.6	114.0	115.0
205	115.4	114.4	114.5	112.9
213	117.5	116.3	115.3	116.1
241	116.8	115.5	113.7	114.5
277	116.7	115.4	115.2	113.7
309	116.3	115.5	113.6	114.7
337	114.8	113.7	113.3	112.2
365	115.0	113.8	112.9	113.4
373	116.3	115.4	114.4	113.6
389	114.0	112.9	111.4	112.6
393	114.8	114.1	113.8	112.0
401	115.5	114.6	113.5	114.7
417	115.2	114.2	112.5	113.4
433	114.0	112.6	112.5	113.4
437	114.4	113.5	112.6	113.3

Table D.2: Measurement of the termination resistance of the LVDS links on DENSE001. Open connections are marked with a "x".

Resistance between TDO and Voltage [MΩ]	HICANN									
	29	81	137	205	213	277	337	393	401	437
VDD	1.49	1.51	1.55	1.57	1.48	1.54	1.59	1.50	1.49	1.49
VDDA	1.49	1.50	1.56	1.56	1.47	1.54	1.57	1.49	1.48	1.47
DI_VCC	3.65	3.71	3.80	3.96	3.76	3.83	4.02	3.88	3.74	3.73
DI_VCCANA	1.50	1.52	1.58	1.58	1.48	1.56	1.59	1.51	1.49	1.49
DI_VCC33ANA	x	x	10.7	x	x	x	x	x	x	x
DI_VBIAS	x	x	18.8	x	x	22.4	x	22.1	20.4	x
VOL	3.09	3.21	1.80	3.55	2.77	3.42	3.27	2.87	2.22	1.79
VOH	1.80	1.84	1.95	1.94	1.76	1.92	1.93	1.83	1.74	1.77
VDDBUS	12.4	13.8	16.0	15.4	8.70	9.23	9.43	9.06	6.28	8.51
VDD25	x	x	x	x	x	14.7	x	x	x	x
VDD5	7.49	7.18	7.70	7.5	1.46	1.55	7.26	6.80	6.05	7.13
VDD12	x	x	x	x	x	x	x	x	x	x
GND	1.48	1.49	1.54	1.55	1.44	1.52	1.56	1.49	1.46	1.46
GNDA	1.48	1.49	1.51	1.55	1.45	1.53	1.56	1.49	1.45	1.45

Table D.3: Resistance measurement between JTAG TDO and the supply voltages on the DENSE001 board.
Resistance values above 30 MΩ are marked with an "x".

HICANN	Error message
5	wrong hicann id 0xffffffff, fpga id 0xffffffff
25	wrong hicann id 0xffffffff, fpga id 0xffffffff
253	transmission error, wrong receive values

Table D.4: Error messages of the Switchram test on the DENSE002 board with activated Automatic Repeat reQuest (ARQ).

HICANN	RX_CLK [Ω]	RX_DAT [Ω]	TX_CLK [Ω]	TX_DAT [Ω]
5	116.8	116.0	115.0	114.1
25	115.8	114.7	113.8	114.0
253	116.4	115.5	114.2	x

Table D.5: Measurement of the termination resistance of the LVDS links on DENSE002. Connections where no resistance was measurable are marked with an "x".

Resistance between TDO and Voltage [MΩ]	HICANN 25
VDD	1.38
VDDA	1.38
DI_VCC	3.16
DI_VCCANA	1.40
DI_VCC33ANA	x
DI_VBIAS	x
VOL	1.37
VOH	1.61
VDDBUS	4.93
VDD25	x
VDD5	6.87
VDD12	x
GND	1.38
GNDA	1.38

Table D.6: Resistance measurement between JTAG TDO and the supply voltages on the DENSE002 board. Resistances above 30 MΩ are marked with a ”x”.

Floating Gate Row Names

	Column	
Row number	Even	Odd
0	V_{reset}	n.c.
1	int_op_bias	I_{bexp}
2	V_{dllrws}	$V_{convoffx}$
3	V_{bout}	I_{convi}
4	V_{fac}	$V_{convoffi}$
5	I_{breset}	$I_{spikeamp}$
6	V_{dep}	E_l
7	I_{bstim}	I_{fire}
8	V_{thigh}	V_{syni}
9	V_{gmax3}	$I_{gladapt}$
10	V_{tlow}	V_{syntci}
11	V_{gmax0}	I_{gl}
12	V_{cira}	V_t
13	V_{gmax1}	I_{pl}
14	V_{stdf}	V_{syntcx}
15	V_{gmax2}	I_{radapt}
16	V_m	E_{synx}
17	V_{bstdf}	I_{convx}
18	n.c.	E_{syni}
19	V_{dtc}	I_{intbbx}
20	n.c.	V_{exp}
21	V_{br}	I_{intbbi}
22	n.c.	V_{synx}
23	V_{ccas}	I_{rexp}

Table D.7: Name of the floating gate cells. The function of the floating gate cell depends on the row number and if it is an even or odd column. Unused floating gate cells are labeled not connected (n.c.).

Floating Gate Tests

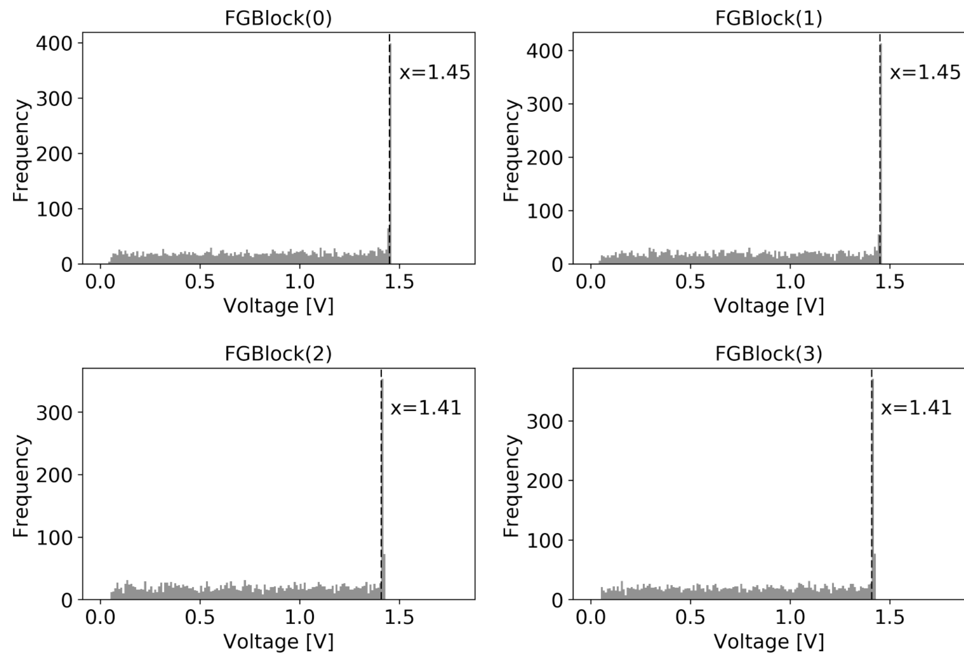


Figure D.1: Plot of the measured values for HICANN 301 on DENSE001.

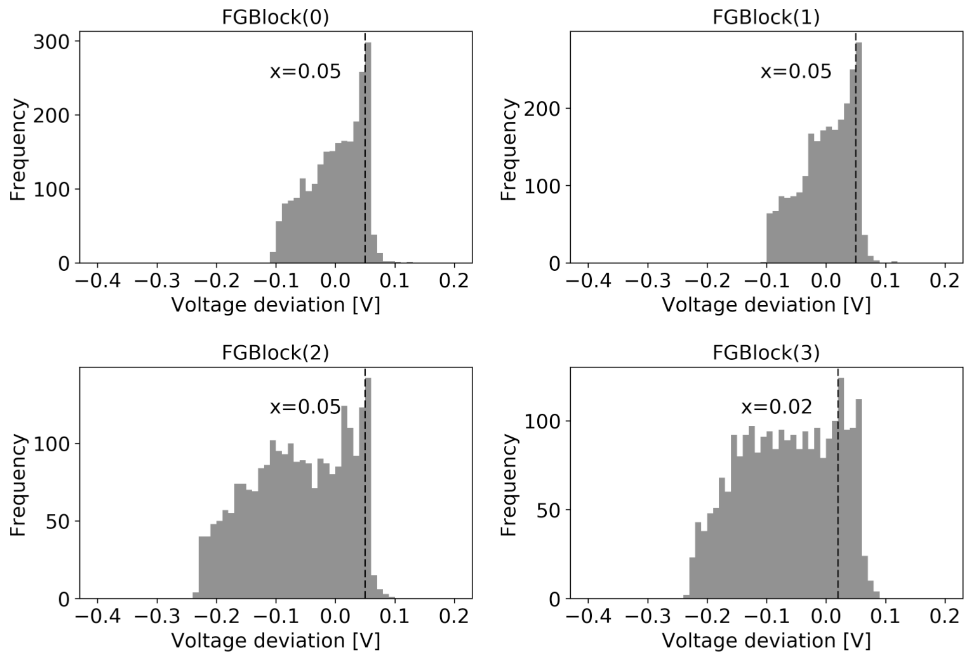


Figure D.2: Histogram of the floating gate test for HICANN 105 on DENSE001.

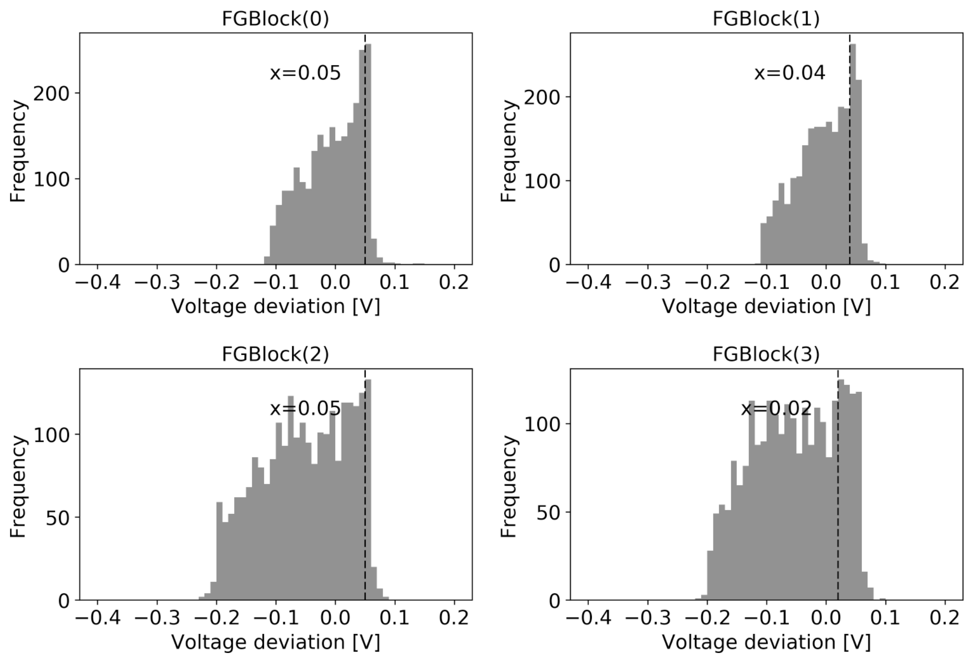


Figure D.3: Histogram of the floating gate test for HICANN 341 on DENSE001.

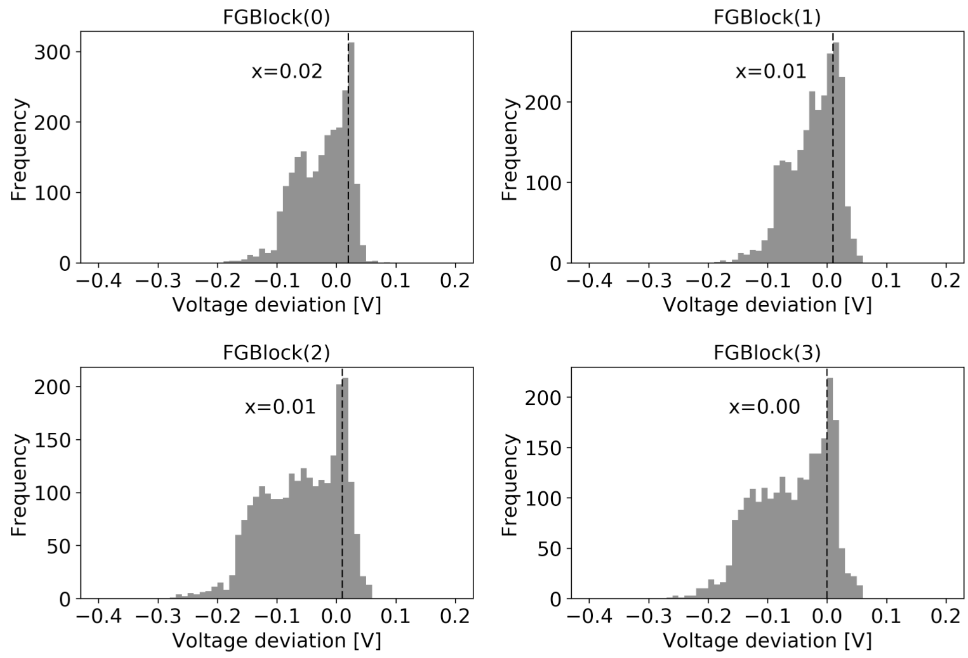


Figure D.4: Histogram of the floating gate test for HICANN 161 on HICANN002.

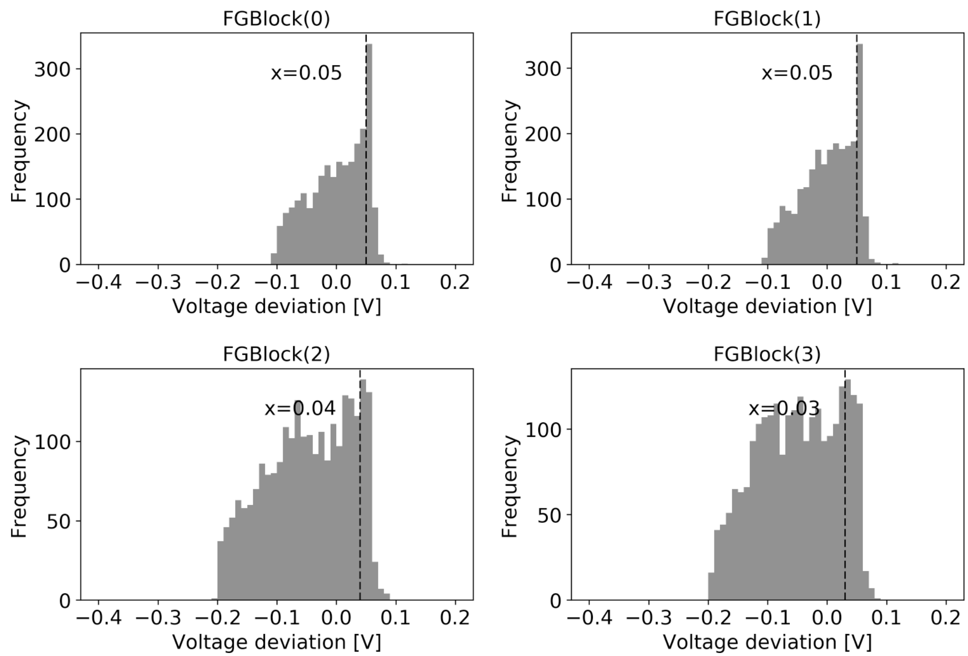


Figure D.5: Histogram of the floating gate test for HICANN 229 on HICANN002.

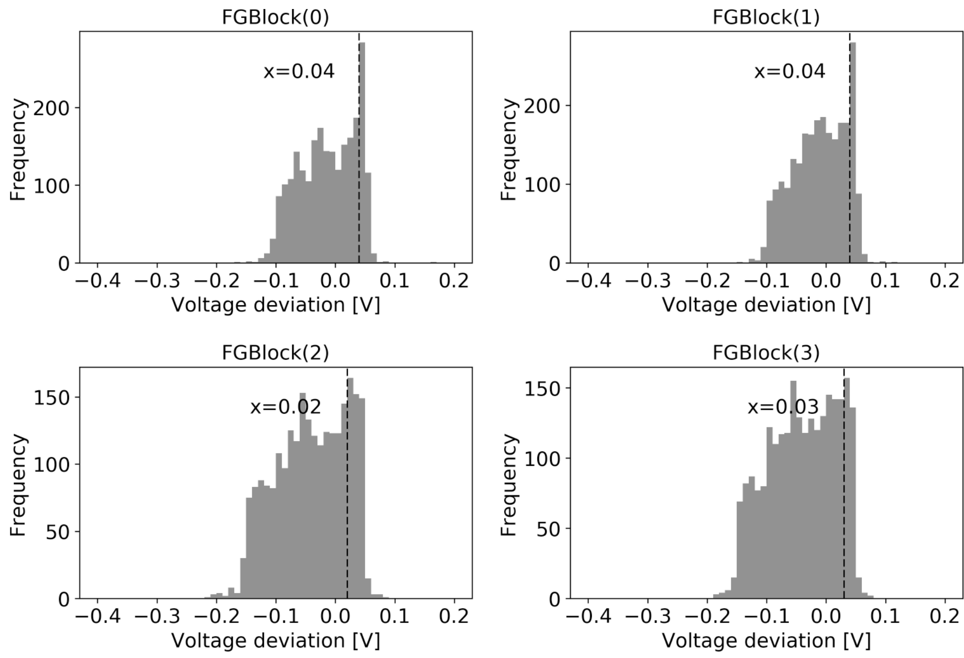


Figure D.6: Histogram of the floating gate test for HICANN 245 on HICANN002.

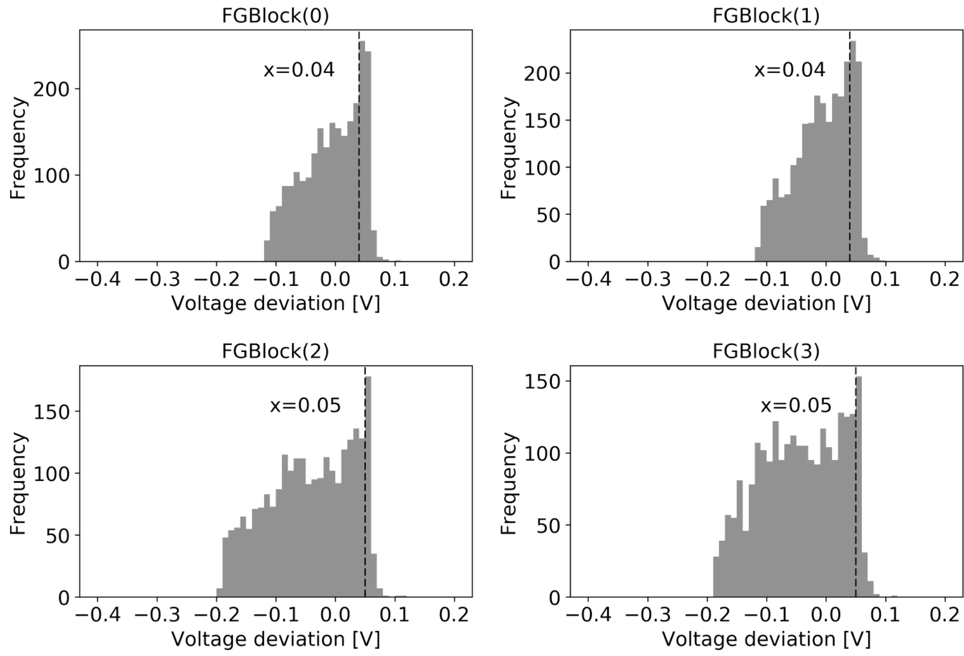


Figure D.7: Histogram of the floating gate test for HICANN 361 on HICANN002.

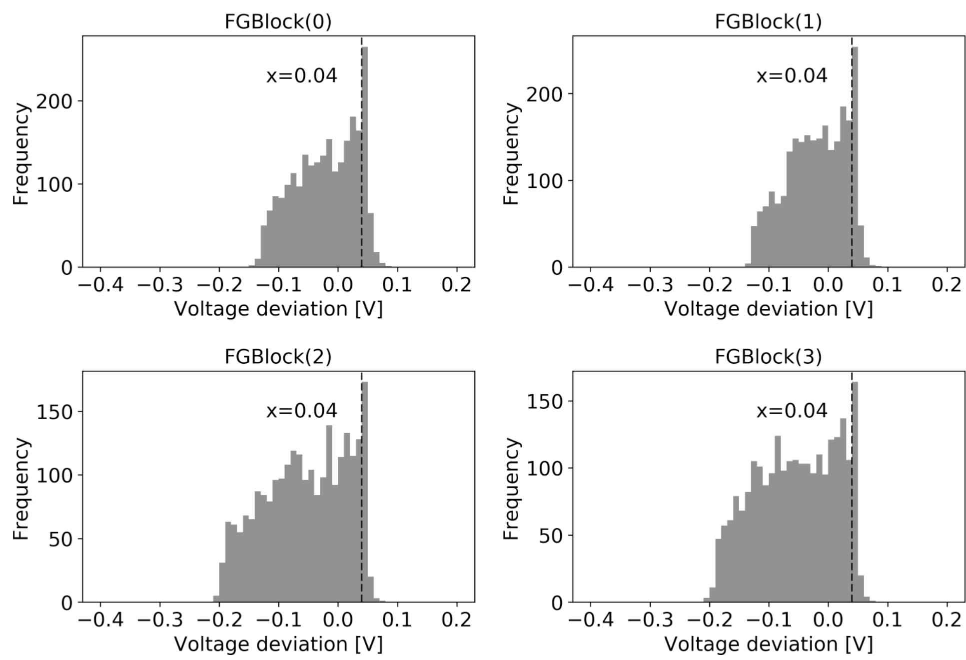


Figure D.8: Histogram of the floating gate test for HICANN 413 on HICANN002.

Calibration Software Results

The tables with the numbers of sorted out neurons is split in two tables. In the table D.8 are the HICANNs which were tested only one time. The HICANNs which are tested multiple times are in table D.9.

HICANN	Number of black-listed neurons
5	5
13	15
25	14
49	5
61	13
77	3
89	15
173	4
209	3
237	8
249	6
273	8
285	8
305	10
317	7
333	9
345	5

Table D.8: List of neurons marked as defective for the DENSE001 board.

HICANN	number of black-listed neurons			
	start	after 1st run	after 2nd run	after soldering furnace
105	74	79	71	76
133	5	6	6	9
165	11	12	15	13
181			14	7
201	15	14	14	14
217			8	8
245	5	5	7	5
253			3	11
281	11	11	6	8
289			12	11
301	3	4	2	2
313	8	9	9	10
329	6	5	6	5
341	84	82	82	77
357	5	4	4	7
385	5	5	2	5

Table D.9: List with the number of neurons sorted out during the calibration after the thermal stress tests with DENSE001.

HICANN	# black-listed neurons	HICANN (continued)	# black listed neurons
9	16	237	23
13	15	241	19
29	6	245	100
33	15	249	7
45	7	265	9
49	11	273	10
53	12	277	17
57	25	281	8
61	15	285	9
73	10	301	11
77	8	305	11
81	6	309	19
85	12	313	6
89	8	317	10
101	8	329	8
105	8	333	11
109	6	337	23
113	11	341	5
117	39	345	15
129	5	357	13
133	11	361	94
137	12	365	15
141	18	369	10
145	10	373	9
157	9	385	10
161	95	389	30
169	6	393	16
173	512	397	9
177	40	401	3
193	9	413	71
197	13	417	10
205	12	421	25
209	12	433	26
213	10	437	15
229	503	441	7
233	6		

Table D.10: Detailed list of black-listed neurons from the calibration software for the DENSE002 board.

E Cooling Calculations

In table E.1 the temperature difference between the air and the cooling surface is calculated for different heat transfer coefficients and heat dissipation values. The cooling area is assumned to be $30 \times 30\,\mathrm{cm}^2$.

Heat dissipation [W/s]	Heat transfer coefficient h [W/(m^2 K)]					
	58	100	150	200	250	290
50	6.4	3.7	2.5	1.9	1.5	1.3
100	12.8	7.4	4.9	3.7	3.0	2.5
200	25.5	14.8	9.9	7.4	5.9	5.1
300	38.3	22.2	14.8	11.1	8.9	7.7
400	51.1	29.6	19.8	14.8	11.8	10.2
500	63.9	37.0	24.7	18.5	14.8	12.8

Table E.1: Temperature difference for air cooling.

In table E.2 the temperature difference between the water and the aluminum block is calculated. It is assumed the wafer runs through a pipe with a diameter of 5 mm and a length of 140 cm. The water flow is 1.7 L/ min.

Heat dissipation [W/s]	Heat transfer coefficient h [W/(m^2 K)]					
	2300	2500	3000	3500	4000	4700
50	1.0	0.9	0.8	0.7	0.6	0.5
100	2.0	1.8	1.5	1.3	1.1	1.0
200	4.0	3.6	3.0	2.6	2.3	1.9
300	5.9	5.5	4.5	3.9	3.4	2.9
400	7.9	7.3	6.1	5.2	4.5	3.9
500	9.9	9.1	7.6	6.5	5.7	4.8

Table E.2: Temperature difference for water cooling.

F Used Hardware and Software

All test and measurements on the DENSE prototypes were conducted with a Cube Setup.

Measurement Data

All raw data that was recorded for this thesis is stored on the storage server of the Electronic Visions Group. The path to the data is "/ley/users/gguettle/phd_data".

Design Data

All the layout designs, schematics and Solidwork files are stored in git-repositories. These repositories are located on the git server of the Electronic Visions Group. The url is https://gitviz.kip.uni-heidelberg.de To access the server an BrainScaleS (BrainScaleS) account is required.

Name	Git Repository	Commit ID
DENSE	pcb-dense	3edb779ac27db75658f57833390ddd0d8e0ae08f
Double Dense Connector board (DDC)	pcb-ddc	e49e7c601b368e30ae04fb2cce72c65b6cfbec25
IBoard-to-DDC flex-rigid board (Flexadapter)	pcb-flexadapter	c0cb5e3af6592d6e9504b80a0a7386b76c56269b

Table F.1: Commit IDs of the projects

Hardware IDs

The Field Programmable Gate Array (FPGA) on the FCP board board has not a unique serial number. Therefore, the identification of the FCP board board is done with the power controller on the board. The controller has a unique number.
The FPGA is flashed with the bitfile "20160404-hmf_fpga_top.bin", its commit ID is 6a596d264890ee-dirty.

Component	ID
FCP board	0x12.38.0c.7a.8e.51.98.0c.52.00.00.00.00.00.00.00.00
iBoard	3
Analogue Readout Module	B201315

Table F.2: Hardware IDs of the used components

Operating Voltages

For the experiments the default voltage settings for a HICANN at the iBoard are used.

Voltage Name	Voltage [V]
VDDD+VDDOUT	1.78
VDDA	1.80
DI_VCC	1.77
DI_VCCANA	1.78
DI_VCC33ANA	3.29
DI_VBIAS	1.25
VOL	0.71
VOH	0.90
VDDBUS	1.20
VDD25	2.49
VDD5	5.13
VDD12	10.97

Table F.3: Operating voltages on the iBoard.

Software Stack

For the experiments and hardware control precompiled software modules were used. Before the 20th February 2017 the software module "nmpm_software/2016-10-26-1" was used. After the 20th February the module was replaced by "nmpm_software/2017-02-20-spack-2017-01-26-1".

Software Name	Commit ID
bitter	da89a7ece461ad6e97a6d7f9e9c0a5b9fccafbff
cake	b91628dab7f62eb25682c2f214391261f950e8b0
calibtic	3dc4580944110102f76c34892a82ecefe0c9eb90
cd-denmem-teststand	483ccec996ec135016099917b4e21aaaf15ea3a8
chip-teststand	d9a2f63a2a282d4bfe84dd2cf534a0ba5e248eec
ester	cc82ee2fcfb202af52a20e2bee3c778a2d8c96ce
euter	ae12b24ed5e73cff22616296a552efb365743e43
halbe	bf737de9e8cfafa02738a5878e8536257e92e944
hicann-system	923509a12240e049c20d47195e4d0cab2111efe3
lib-boost-patches	da2d3f458c419f229759d7a17c2303be25f9e289
lib-rcf	1c99a4217b5e4a8384c55707444fcbe3cd8222eb
logger	36ae203184918389d0f7a1e0a949a1ed37d9b33f
marocco	163741fe9afc57df9d0a762806c78a9392ee3406
odeint-v2	9473d7f067b9c4b56862de644898b0ac99676075
pygccxml	8ae9e19ae00c4152fa5a381eb9e663561c07345f
pyhmf	1e04ba62448012ccf3ceabd4ec207e59c4eee865
pyplusplus	5a612a7f5372d832bebacc5377e78ffd2deabcde
pythonic	e9628388d2eb0ce34db770c660ea37718d97dd3d
pyublas	85b4b20a10d16e3342b1ccc187893c2bef611899
pywrap	1744f8b7edabb5ac3b2088f9cb900e35f8234bd2
rant	4a8acd076fb9531ce61a990ef0f414935886d85d
redman	d6622274ecce7a836bfade761073221d44df0840
sctrltp	42c6d0ea49f91dc44a7aae1017cf4cd935246834
sthal	61dc4583836a601a465ec079958f41a531d08cd1
systemsim-stage2	ba16507aeb41e62795fafce072dd91fb29738502
vmodule	80efaa6cfbd16d1f607da44cedf21e7bb28aae1f
ztl	2934a12003c14e08643cf2b4b3cbe7553e860f08

Table F.4: Commit IDs of the Git repositories in the software module "nmpm_software/2016-10-26-1".

Software Name	Commit ID
bitter	da89a7ece461ad6e97a6d7f9e9c0a5b9fccafbff
cake	3f8bca050faf0cc51231ebeec74fda5db50a2f37
calibtic	443ff807cdbbaf1c831e7086be9c6611520b7710
cd-denmem-teststand	483ccec996ec135016099917b4e21aaaf15ea3a8
chip-teststand	d9a2f63a2a282d4bfe84dd2cf534a0ba5e248eec
ester	cc82ee2fcfb202af52a20e2bee3c778a2d8c96ce
euter	ae12b24ed5e73cff22616296a552efb365743e43
halbe	eaf885eca306ab25edda89f57f53f91911eb09be
hicann-system	0d377de128ee2ab5c48d925b8e318162cdfe4649
lib-boost-patches	da2d3f458c419f229759d7a17c2303be25f9e289
lib-rcf	1c99a4217b5e4a8384c55707444fcbe3cd8222eb
logger	bc17b71cce1107f856ca11a78332f76af5b06ba1
marocco	a6af6cdecbfdc201c97ed7fed371ea767c21eb1b
odeint-v2	9473d7f067b9c4b56862de644898b0ac99676075
pygccxml	8ae9e19ae00c4152fa5a381eb9e663561c07345f
pyhmf	a06569b78c0b3ff9d2f6792d52c58ddb16ac727c
pyplusplus	5a612a7f5372d832bebacc5377e78ffd2deabcde
pythonic	e9628388d2eb0ce34db770c660ea37718d97dd3d
pyublas	85b4b20a10d16e3342b1ccc187893c2bef611899
pywrap	f5ee2291d18b7a1d62ab921d8896dcc963beaa84
rant	85ce410ff59a7bc8a641c9710022576b7a39885a
redman	d6622274ecce7a836bfade761073221d44df0840
sctrltp	f880c8ff54b1bb18893bed0dfcd5c69480af8ddb
sthal	2325772f1f48237a753557be444e309abc567512
systemsim-stage2	ba16507aeb41e62795fafce072dd91fb29738502
vmodule	80efaa6cfbd16d1f607da44cedf21e7bb28aae1f
ztl	2934a12003c14e08643cf2b4b3cbe7553e860f08

Table F.5: Commit IDs of the Git repositories in the software module "nmpm_software/2017-02-20-spack-2017-01-26-1".

Acknowledgments

Finally, I want to acknowledge everyone who supported this work, especially

- Prof. Dr. Karlheinz Meier for supervision and the opportunity to work in the Electronic Visions Group;
- Prof. Dr.-Ing. Ulrich Brüning as the second referee of this work;
- Prof. Dr. Matthias Bartelmann and Prof. Dr. Peter Fischer to be part of the defense committee;
- European Union for funding our work in BrainScales and HBP;
- Dr. Johannes Schemmel for supervision and his feedback;
- the entire Electronic Visions Group;
- Julia Bing and Werner Lamade from the mechanics workshop;
- Alexander Dobler and Dr. Andreas Grübl from the electronics workshop;
- Ralf Achenbach and Markus Dorn from the ASIC Lab;
- For proof-reading: Dan Husmann, Kai Husmann, Christoph Koke, Christian Mauch, Eric Müller, Sebastian Schmitt, David Stöckel;
- my long-term office mate Dan Husmann for technical, political and all other discussions;
- my friends in the group "Clique mit Spam M&M" for constructive distraction;
- Theresa Schlegel for the final spurt support ;-)
- my family.

Statement of Originality (Erklärung):

I certify that this thesis, and the research to which it refers, are the product of my own work. Any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Ich versichere, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, August 31, 2017

....

(signature)