Characterization and Quality Control of RD53A Readout Chips and Modules for the ATLAS ITk Pixel Detector

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Michael Daas
aus
Marburg

Bonn, Mai 2021

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn.

1. Gutachter:	Prof. Dr. Jochen Dingfelder
2. Gutachter:	Prof. Dr. Klaus Desch
Tag der Promotion:	01.07.2021
Erscheinungsjahr:	2021

Contents

CHAPTER 1

INTRODUCTION

The objective of high energy particle physics is to attain a better understanding of the fundamental building blocks of our universe, the elementary particles as well as their interactions. Increasingly complex experiments have been devised over the past decades, accelerating and colliding different kinds of particles at ever increasing energies. The current peak of these developments is the Large Hadron Collider (LHC) with its two general-purpose experiments ATLAS and CMS.

The complex particle detectors of high energy physics experiments operate at the edge of technological progress, symbiotically demanding and driving developments in many different fields, from cooling and engineering to electronics- and sensor development. Currently, the pending luminosity upgrade of the LHC demands further increase in granularity, readout speed and radiation hardness of the state-of-the-art particle detectors. At the same time, the construction of the upgraded ATLAS and CMS detectors are some of the most complex and ambitious projects in particle physics history. Each detector consists of several subsystems with up to several billion readout channels each. These massive projects are undertaken by collaborations, which are distributed over hundreds of institutions, from countries all over the world. Many scientists and commercial vendors are involved in the production of every small component of the detectors. For this reason, large-scale projects like this require comprehensive Quality Assurance (QA) and Quality Control (QC) procedures in order to succeed.

This thesis covers some of the QA and QC efforts in the context of the ATLAS Phase-II upgrade that are ongoing in the Silab[1] at the Physics Institute of the University of Bonn. In the course of the High Luminosity LHC (HL-LHC) upgrade, the tracking detector of the ATLAS experiment will be completely replaced by a new, all-silicon tracking detector. This new Inner Tracker (ITk) will consist of nine layers, the innermost five of which are made up of almost 15 000 hybrid silicon pixel detector modules. The Silab is also heavily involved in the design and testing of the readout chips for the new pixel detector modules for ATLAS and CMS, which are developed by the RD53 collaboration. Furthermore, the Bonn group is going to be one of the largest module assembly- and testing sites for the ATLAS pixel detector during the production phase and is therefore involved in the development and execution of a large spectrum of QC measures and steps.

[1] Silicon Laboratory Bonn

After the basic parameters of the experiments and the fundamental principles of silicon tracking detectors are discussed in Chapter 2, the RD53A readout ASIC prototype is introduced in Chapter 3. The BDAQ53 readout system, that was co-developed as a part of this work, is introduced in Chapter 4. The following two chapters, Chapter 5 and Chapter 6, concentrate on the evaluation of the bare RD53A readout chip, verifying its general ability to fulfill the posed requirements and developing and evaluating QC routines and measurements of thousands of chips on wafer level. The last two chapters deal with RD53A as part of a hybrid detector module. General module tests and measurements are presented and the chip's performance is compared to the performance of the bare chip in Chapter 7. Finally, Chapter 8 presents and discusses measurements within the scope of the ATLAS ITk hybridization Market Survey (MS), that were accomplished in the context of this thesis.

1.1 THE LARGE HADRON COLLIDER

The Large Hadron Collider (LHC) [1] is currently the largest and most powerful particle accelerator in the world in terms of circumference, collision energy and interaction frequency. It is arguably the largest machine ever constructed. The LHC was constructed and is operated by the European Organization for Nuclear Research (CERN). Its general location is shown in Figure 1.1: The 27 km long collider ring is located in the Geneva region at the border between Switzerland and France, and was installed in the existing tunnel of the LEP[2] accelerator, between 45 m and 170 m underground.

The LHC is connected to CERN's accelerator complex, most notably the Super Proton Synchrotron (SPS), which is used as final pre-accelerator and injector for the LHC. The superconducting, symmetric proton-proton collider accelerates protons in both directions at eight straight sections around the ring to an energy of up to 6.5 TeV for each beam, resulting in a center-of-mass energy of 13 TeV at the collision points. The LHC's design collision frequency is 40 MHz. Four main experiments are located at the collision points of the accelerator: ALICE, ATLAS, CMS and LHCb. While ALICE is optimized for studying heavy ion collisions, another mode of LHC operation, and LHCb specializes in the physics of b-hadrons and measurements of CP violation, ATLAS and CMS are general-purpose particle detectors. These detectors are designed to measure and reconstruct all particles originating from the proton-proton collisions in order to test the Standard Model of Particle Physics (SM) and search for new physics phenomena.

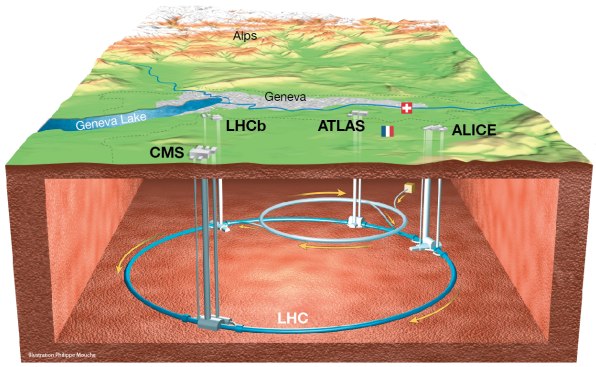


Figure 1.1: Overview of the LHC and its experiments [2]

[2] Large Electron-Positron Collider, one of the most powerful lepton accelerators in the world and the predecessor to the LHC. LEP was constructed in 1989 and decommissioned in 2000, to make way for the LHC.

The LHC and its experiments in the current form have been operated successfully for more than ten years. In this time, several ten quadrillion (10^{16}) collisions have been produced and recorded by the experiments, leading to multiple discoveries and consequences for the SM, most notably the discovery and confirmation of the Higgs boson by ATLAS and CMS in 2012, which was predicted already in 1964.

One of the key quantities that describe the performance of a particle accelerator is the *instantaneous luminosity* $\mathcal{L}$ as defined by

$$\mathcal{L} \propto \frac{N_p^2 n_b f}{4\pi F}, \qquad \text{Equation 1.1}$$

where N_p is the amount of particles per colliding bunch, n_b is the number of bunches per beam, f is the revolution frequency of the bunches in the accelerator and F is a form factor expressing the beam size and incident angle at the interaction point.

The LHC is designed for a peak luminosity of $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ for proton-proton collisions at a maximum center-of-mass energy of 13 TeV. The two general-purpose detectors ATLAS and CMS were designed to cope with this peak luminosity for the expected life span of the LHC of about 15 years.

Since the collision rate at a particle accelerator, $\frac{\mathrm{d}N}{\mathrm{d}t}$, is related to the luminosity via

$$\mathcal{L} = \frac{1}{\sigma}\frac{\mathrm{d}N}{\mathrm{d}t}, \qquad \text{Equation 1.2}$$

the total number of produced collisions in an accelerator experiment can be expressed as

$$N = \sigma \cdot \mathcal{L}_{\mathrm{int}}, \qquad \text{Equation 1.3}$$

with the *integrated luminosity* $\mathcal{L}_{\mathrm{int}} = \int \mathcal{L}\,\mathrm{d}t$ and the total proton-proton scattering cross section σ. Instead of quoting a total number of produced or recorded collisions, this quantity is typically expressed according to Equation 1.3 in terms of $\mathcal{L}_{\mathrm{int}}$ in units of inverse barn ($1\,\mathrm{b} = 100\,\mathrm{fm}^2 = 10^{-28}\,\mathrm{m}^2$). As of 2020, the LHC has delivered a total integrated luminosity of about $190\,\mathrm{fb}^{-1}$ during its first two run periods combined. The center-of-mass energy during these periods was increased from 7 TeV in 2010 to 13 TeV in 2015. Another $160\,\mathrm{fb}^{-1}$ are expected to be delivered until 2025 at the LHC's nominal design center-of-mass energy of 14 TeV. [3]

This enormous number of collected proton-proton collisions is necessary to study very rare events. For example, when multiplying the total estimated cross section of all Higgs boson production mechanisms at the LHC of about $\mathrm{O}(100\,\mathrm{pb})$ [4] with the integrated luminosity of $190\,\mathrm{fb}^{-1}$, it can be estimated that only about 10^7 collisions including the production of a Higgs boson have been produced so far. However, when taking into account acceptance and efficiency of the detector and the branching ratios of the different production- and decay channels, only a small percentage of these is actually available for analysis. Due to its statistical nature, the error of any production rate- or property measurement conducted with a particle detector scales reciprocally with the square root of the number of observed events N_{obs}:

$$\sigma_{\mathrm{stat}} \propto \frac{1}{\sqrt{N_{\mathrm{obs}}}}.$$

Therefore, for rare processes, a very large number of proton-proton collisions has to be recorded in order to achieve the required sensitivity.

1.1.1 THE HL-LHC UPGRADE

Based on the considerations in the previous section, it is estimated that the statistical gain from continuing accelerator operation at the design luminosity of $\mathcal{L}_{\text{LHC}} = 1 \cdot 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$ after the third run period, even for extended periods of time becomes very small. Halving the statistical error of measurements beyond an integrated luminosity of about $350\,\text{fb}^{-1}$ would already take more than ten years.

For this reason, the LHC will be upgraded to the HL-LHC after the third run period, increasing its instantaneous luminosity by a factor of about ten at the same center-of-mass energy to about

$$\mathcal{L}_{\text{HL-LHC}} \gtrsim 10^{35}\,\text{cm}^{-2}\,\text{s}^{-1}. \qquad [5]$$

This upgrade requires a major overhaul of the accelerator, including new 12 T superconducting magnets, new superconducting accelerator cavities and other cutting-edge technologies, developed specifically for this machine. The research and development campaigns for the accelerator upgrade have been proceeding in parallel to the LHC physics program for the last ten years. The upgrade schedule including the center-of-mass energy and the targeted integrated luminosity for each run period is shown in Figure 1.2. The HL-LHC upgrade will exploit the full potential of the largest particle physics experiments in the world and secure the machine's discovery capabilities at the energy frontier for the next two decades. [5]

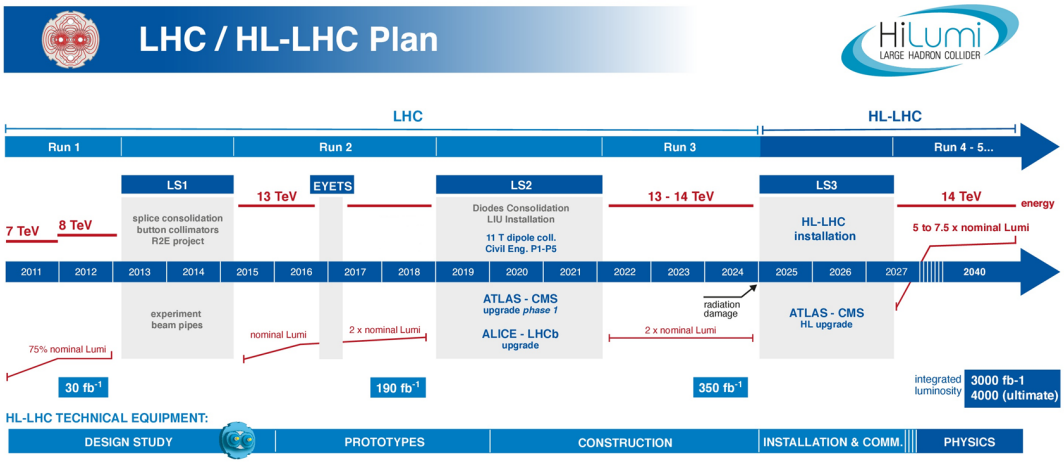


Figure 1.2: HL-LHC Project Schedule, modified from [3]. The LHC has been running successfully for the last ten years, with only a few short, planned shutdown periods for maintenance and minor upgrades. Currently, the second long shutdown is ongoing, with multiple upgrades being installed at the experiments and the accelerator itself. Another three year long run period will conclude the life span of the LHC and the machines will be upgraded to HL-LHC during a third long shutdown period.

1.2 THE ATLAS EXPERIMENT

ATLAS is the larger of the two general-purpose particle detector experiments at the LHC. It is designed to track and identify almost all types of particles. A general overview of the ATLAS detector is shown in Figure 1.3. As for most general-purpose detectors at colliders, the ATLAS detector is built in an onion-like structure around one of the four primary interaction points of the LHC, employing cylindrical *barrels* in the center and perpendicular discs, so-called *end-caps*, for each detector subsystem in the forward- and backward regions.

The cylindrical detector is roughly 44 m long and 25 m in diameter and was constructed in a cavern at the LHC interaction point 1, more than 100 m underground. The general structure of the detector consists of a tracking detector called *Inner Detector*, surrounded by a solenoid magnet system. Next, there are an electromagnetic- and a hadronic *calorimeter system*. The outermost layers consist of a *muon spectrometer* with a dedicated toroidal magnet system.

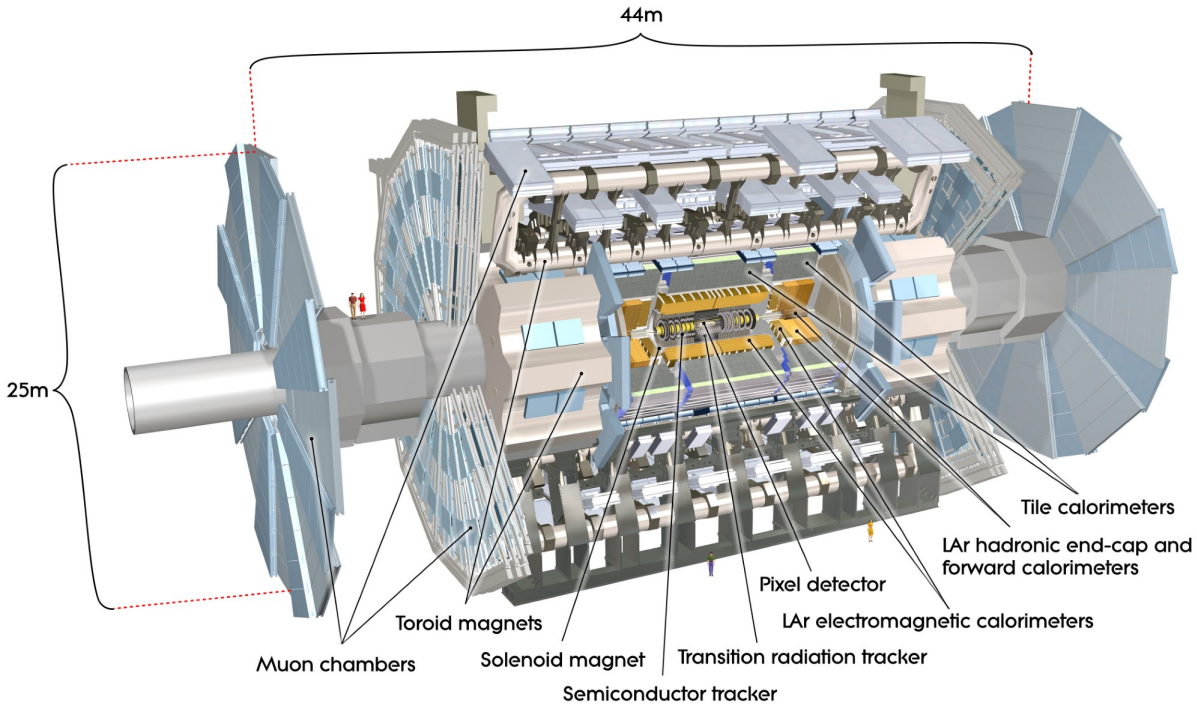


Figure 1.3: The ATLAS detector [6]. The general-purpose particle detector employs an onion-like structure, consisting of concentric barrels and perpendicular end-caps for each subsystem. The different subsystems are labeled in the figure. Multiple humans are depicted true to scale for size reference.

INNER DETECTOR

The innermost layers of the ATLAS detector consist of several different types of tracking detectors that are meant to reconstruct the trajectories of charged particles, which are curved due to the Lorentz force in a 2 T magnetic field. The direction of the curvature gives information about the particle's charge, while the degree of curvature is related to its transverse momentum. Additionally, precise tracking can allow for conclusions about whether a particle originated in the primary proton-proton interaction point or a secondary vertex. To enable all of these measurements, the Inner Detector (ID) needs to determine each particle's position at multiple points as precisely as possible. For this purpose, it consists of three sub-detector systems. [7] A schematic view of the original ATLAS ID is shown in Figure 1.4.

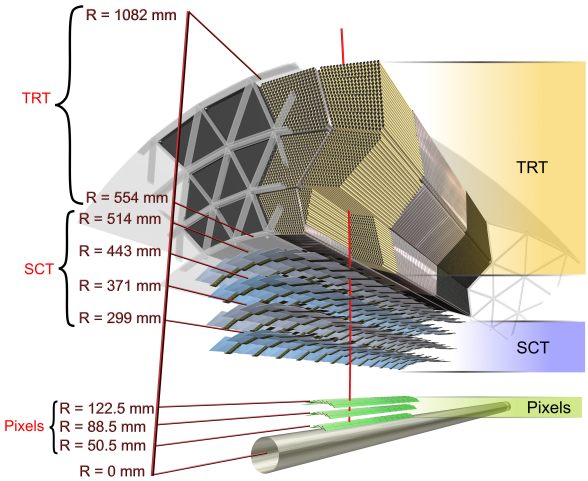


Figure 1.4: Schematic view of the original ATLAS Inner Detector (ID) without IBL [8].

The innermost subsystem of the ID is the Pixel Detector. It consists of four layers of hybrid silicon pixel detectors, the innermost of which is located only a few centimeters from the interaction point. This innermost layer is called Insertable B-layer (IBL) and was installed in a first major upgrade in 2013. Hybrid pixel detector technology offers outstanding accuracy while being able to cope with the harsh radiation environment close to the interaction point, both in terms of high occupancy tolerance and the significant radiation dose that is accumulated over the lifetime of the detector. [9, 10]

The next subdetector is the Semi-Conductor Tracker (SCT), a silicon strip detector consisting of four double layers. Silicon microstrip technology sacrifices segmentation, and therefore resolution in one dimension, for a significant reduction in complexity and material budget. This makes strip detectors the pragmatic choice for covering larger areas with lower occupancy requirements than in the innermost layers. With an active area of about 61 m^2, the SCT provides about twice as many sampled points over a larger area with roughly the same spatial resolution perpendicular to the beam axis as the pixel detector. [7]

The outermost subsystem of the Inner Detector is a straw tube detector that makes supplemental use of transition radiation, hence the name Transition Radiation Tracker (TRT). This gaseous detector uses more than 300 000 4 mm wide and almost 1.5 m long drift tubes, that detect charged particles by ionization of the xenon- or argon-based gas mixture inside them. Additionally, the voids in between the tubes are filled with material of vastly different refractive index, causing passing particles to produce transition radiation. This increases the general signal strength and further enhances the ID's particle identification capabilities. Gaseous detectors like the TRT effectively enable continuous tracking by providing an average of 36

measurement points per track at slightly lower resolution, but at the cost of a significantly lower tolerance to pile-up than silicon-based trackers. The technology was chosen in order to reduce the cost of covering the outermost part of the 7 m long and 1.2 m wide ID. [7] At the current peak instantaneous luminosity of the LHC however, the TRT is exceeding an average occupancy of 50 %, which inevitably comes with a negative impact on tracking performance and efficiency. [11]

The Inner Detector is surrounded by a superconducting solenoid magnet system responsible for the homogeneous 2 T magnetic field that penetrates the entire ID and enables Lorentz force-based momentum measurement and particle identification.

CALORIMETER SYSTEMS

Right outside the ID and its solenoid magnet system, a two-part calorimeter system is situated. It consists of an electromagnetic calorimeter stopping and measuring the total energy and location of electromagnetically interacting particles like electrons or photons. Further away from the beam axis follows the hadronic calorimeter, that is designed to stop and measure the total energy of hadronically interacting particles that have not been stopped in the electromagnetic calorimeter already.

Both parts of the calorimeter system are built as sampling calorimeters. This type of detector uses alternating layers of high-density absorber material and active elements. The incident particle produces an eletromagnetic- or hadronic shower when traversing the absorber layers. The position, shape and total energy deposition of this shower are sampled in the active layers. In the case of the ATLAS electromagnetic calorimeter, lead is used as absorber materials, while liquid argon is used as the active material of the sampling layers. The layers are folded in an accordion-like structure. The central cryostat needed to keep the employed Argon in its liquid state is additionally used to cool the Inner Detector's superconducting solenoid magnet. In the hadronic calorimeter of ATLAS, stainless steel is used predominantly as absorber material, while crystal scintillators are used as active material. [12]

Together with the charge and momentum information provided by the ID, the measurement of a particle's total energy allows for particle identification.

MUON SPECTROMETER

Due to their relatively high mass and the fact that they are not strongly interacting, muons traverse the entire ATLAS detector, including the calorimeter systems, mostly undisturbed. However, since many physics processes of interest for the ATLAS experiment involve muons and detailed information about their involvement and characteristics are vital for event reconstruction, another detector system designed specifically to measure the characteristics of muons is situated at the outer bounds of the detector. It consists of a toroidal magnetic field produced by eight superconducting air-core barrel magnets, that spans a cylindrical area of 26 m by 20 m diameter and has a varying bending power of up to 6 T m. Muon tracks are bent in this magnetic field due to Lorentz force and the curvature again gives information about the particle's momentum. To measure the direction and curvature of the bent tracks, multiple layers of tracking detectors are employed again in a barrel and two end-cap parts. Different

detector technologies are used, including Monitored Drift Tubes (MDTs) and Cathode Strip Chambers (CSCs) for precision track measurements as well as Resistive Plate Chambers (RPCs) and Thin Gap Chambers (TGCs) for independent triggering of the muon system. [13]

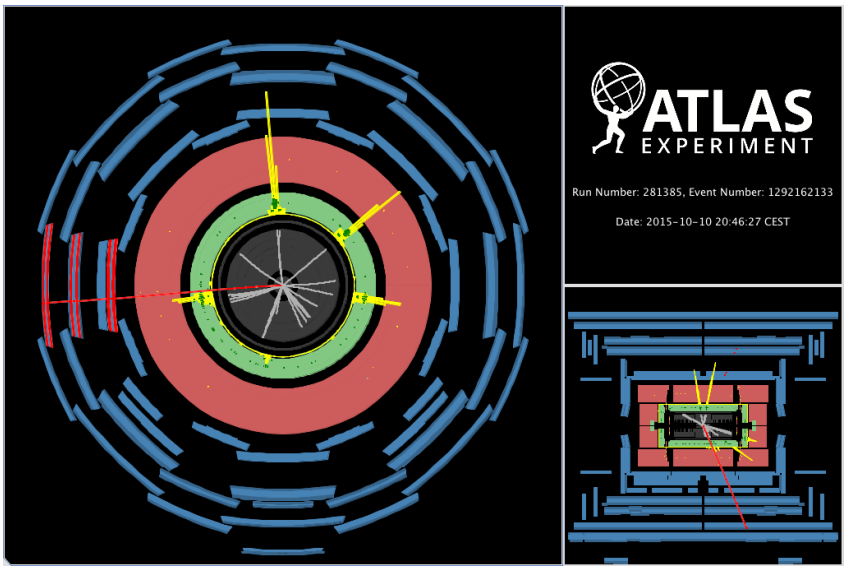


Figure 1.5: An example ATLAS event display. The ATLAS detector is shown in two different profiles, perpendicular to and along the beam axis. The subdetector systems are depicted in different colors: The ID in gray in the center, the electromagnetic calorimeter in green, the hadronic calorimeter in red and the muon system in blue. In addition, reconstructed tracks in the ID are shown in light gray, electromagnetic clusters in yellow and muon tracks in red. [14].

Figure 1.5 shows the result of all subsystems of the ATLAS detector working together. The example event display shows multiple tracks in the ID depicted in gray in the center of the cross section view. The electromagnetic calorimeter shown in green recorded five significant entries, depicted according to their energy by the yellow histograms. While the hadronic calorimeter that is depicted by the red ring only measured insignificant data in this particular event, the muon spectrometer sown in blue recognized a single muon, whose reconstructed track is shown in red.

TRIGGER AND DATA ACQUISITION SYSTEM

When operating at peak design luminosity, the LHC produces 10^9 interactions per second. Since it is far above current technology's capabilities to read out and store all of these events, a three-stage trigger system was implemented to pre-select potentially interesting events and reduce the resulting data rates to processible levels. The LVL1 trigger system makes a per-collision decision at the full rate of 40 MHz, deciding if an event is categorized as potentially interesting or as background. The resulting rate of potentially interesting events to be stored is thereby reduced to about 100 kHz, leaving the next level of trigger system a response time in the order of a few milliseconds. Here, a data reduction by another factor 10^3 is achieved by defining Regions of Interest (ROIs) to be stored instead of the full detector's raw data. A third trigger level called High Level Trigger (HLT) performs a rudimentary online analysis of the events

using several seconds of processing time. It finally decides which events are saved from the readout buffers to long-term storage for offline analysis. [15]

Even after this three-level data reduction by a total factor of 200 000, the average recorded data rates range from 320 MB/s for Run1 up to 800 MB/s for Run2 and 1 000 MB/s Run3. [16]

1.2.1 THE ATLAS PHASE-II UPGRADE

In order to meet the new challenges posed by the High Luminosity LHC (HL-LHC), all subsystems of the ATLAS detector will undergo an extensive upgrade during Long Shutdown 3 from 2025 to 2027. For the ID to cope with the new peak luminosity of up to $7.5 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, it completely removes the TRT in favor of a newly built all-silicon tracking detector named Inner Tracker (ITk). The increased average number of proton-proton interactions per bunch crossing of about 200, compared to about 40 for the current LHC, will already pose a significant challenge to the newly developed silicon detectors, and even more so for gaseous detectors.

Due to the completely redesigned trigger and DAQ systems and because of radiation damage of the existing electronics, the ATLAS calorimeter systems will undergo a complete replacement of the front- and back-end electronic systems. The muon spectrometer will remain as the only gaseous detector system in the upgraded ATLAS detector, but with upgraded readout electronics as well, to meet the new rate and occupancy requirements posed by the HL-LHC and to guarantee the necessary trigger efficiency. The trigger scheme for the ATLAS Phase-II upgrade foresees a new hardware trigger stage LVL0/LVL1 that performs real-time track reconstruction based on ROIs to reduce the LVL1 output event rate to about 200 kHz, twice the acceptance rate of the current LVL1 trigger. [17]

1.2.2 THE ATLAS ITK PROJECT

The work presented in this thesis was conducted within the scope of the ATLAS ITk project, the tracking detector of the Phase-II ATLAS experiment.

The current ID does not meet the requirements for the HL-LHC upgrade, in terms of radiation hardness, granularity, and readout speed. Since the gaseous, drift tube-based TRT of the ID is estimated to reach an average occupancy of 100 % at an instantaneous luminosity of $< 5 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, designing the upgraded tracking detector as an all-silicon detector is the obvious choice. While the currently installed hybrid pixel and microstrip detectors employ the most promising technology for application at the HL-LHC, they will have reached the limits of their capabilities with the end of Run3 as well. The radiation dose[3] will exceed the design values of $5 \cdot 10^{14}\,\mathrm{n_{eq}}\,\mathrm{cm}^{-2}$ in the pixel detector and $1.5 \cdot 10^{14}\,\mathrm{n_{eq}}\,\mathrm{cm}^{-2}$ in the SCT, bringing both subsystems to the end of their designed lifetime [7]. An increase of the mean number of pile-up events from a single proton-proton collision to almost 200 further means that the detectors may not be able to resolve tracks that are close to each other and data buffers in the readout of both the Pixel Detector and the SCT would be saturated, leading to significantly degraded event collection efficiency. [17]

For these reasons it was decided to replace the ID with the ITk, an all-silicon tracker based on multiple layers of newly developed hybrid pixel detector modules and silicon strip detectors,

[3] Non-ionizing radiation dose is typically normalized to 1 MeV neutrons per area and thus measured in $\mathrm{n_{eq}}\,\mathrm{cm}^{-2}$.

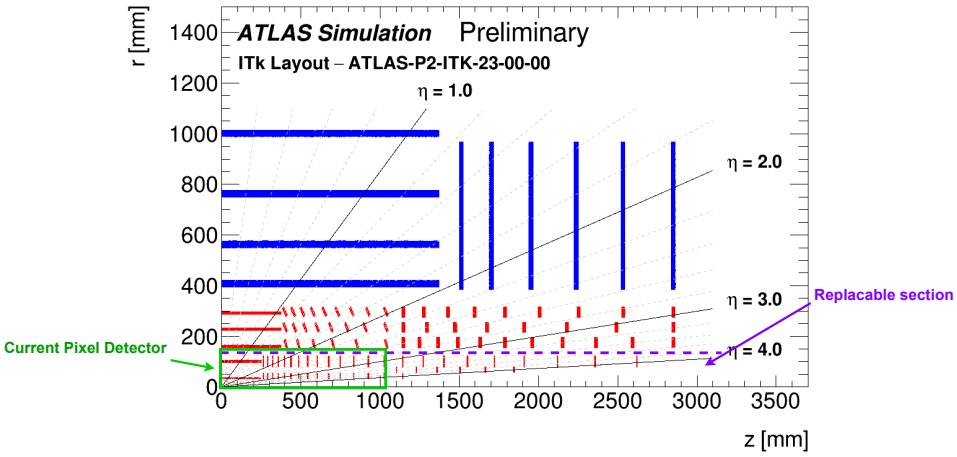


Figure 1.6: Layout simulation of the ATLAS ITk, modified from [18]. The green box indicates the size of the current ATLAS Pixel Detector, while the purple dashed line specifies the replacable inner section.

both extending to much larger radii than in the ID to replace the TRT. The current baseline layout for the ITk is shown in Figure 1.6. It consists of five layers of hybrid pixel detectors, arranged in over 2 m long barrels in the center and multiple concentric discs in the forward regions. The strip detector employs four layers of almost 3 m long barrels in the center and multiple end-caps in the forward regions as well. The total active area of silicon in the ITk will be more than 200 m^2 and the detector will employ more than five billion readout channels.

To cope with the complex spectrum of requirements, the ITk pixel detector will employ different types of hybrid pixel detector modules. As will explained in detail in Chapter 2, hybrid pixel detectors consist of a passive sensor and active circuitry, implemented in one or several readout chips. For both, mechanical and efficiency reasons, the ITk pixel detector will be made up from Triplets and Quad-Chip Modules (QCMs), which consist of three or four readout ASICs, respectively, connected to a single sensor tile. In order to decrease the cluster size and enhance tracking performance, the outer regions of the pixel detector barrels contain *inclined* modules, that are tilted towards the interaction point. Since no current detector technology could survive the expected radiation dose very close to the interaction point at a given lifetime of the experiment, that is equivalent to the expected integrated luminosity of 3 000 to 4 000 fb^{-1}, the two innermost pixel layers are designed to be easily replaceable. In order to reduce the material budget required for services, multiple modules of the ITk pixel detectors will be combined to serial powering chains to be powered with a constant current. [17]

The design of the new pixel detector readout ASIC has been ongoing for the past seven years and first prototype chips and modules have been produced and characterized within the scope of this work.

1.2.3 RD53A PRE-PRODUCTION PROGRAM

To meet the stringent requirements posed by the continuously evolving schedule of a complex project like the ATLAS Phase-II upgrade, prototyping of sensors and other components of the pixel detector modules had to be started before the final production readout ASIC was available. Due to the availability and generally good performance of this first prototype chip, an extensive pre-production program was launched based on RD53A (see Chapter 3). This prototype for the final readout ASIC was developed by the RD53 collaboration in a joint effort of ATLAS and CMS. Despite its smaller matrix size and physical dimensions compared to the final ATLAS pixel readout chip, single- and multi-chip modules are produced based on RD53A in order to investigate sensors and hybridization techniques and to develop assembly tools, flexes, carriers and other building blocks of the ITk pixel detector modules.

This program is of special relevance for this work, since many measurements that are presented in this thesis have been conducted within the scope of the hybridization MS, a pre-production campaign aimed to identify and assess potential vendors for the hybridization of the detector modules (ref. to Section 2.3.2). This MS has been conducted with RD53A-based Double Chip Modules (DCMs) due to schedule considerations.

1.3 THE CMS EXPERIMENT

The Compact Muon Solenoid (CMS) experiment is the second general-purpose particle detector at the LHC. It pursues a comparable physics program as ATLAS, with the two independent experiments aiming to confirm their respective results.

In preparation for running at the HL-LHC, all subsystems of the CMS detector are upgraded as well. A new tracker geometry, now employing multiple layers of hybrid silicon pixel, macro pixel and silicon microstrip detectors. The readout ASIC for the hybrid silicon pixel detectors of both the ATLAS and CMS Phase-II upgrades is developed in a joint effort by the RD53 collaboration. The calorimeter systems in the barrel region will only have their readout electronics upgraded, while in the forward regions completely new detectors will be implemented due to the expected radiation dose. The muon system will be extended to the forward regions, employing modern gaseous detectors. [19]

CHAPTER 2

SILICON TRACKING DETECTORS

Silicon is one of the most common elements on earth. Most sand on the earth's surface, that can be found in deserts or on beaches is made from silicon dioxide (SiO_2) and together with silicon based compounds, the element makes up more than 25 % of the mass of the earth's crust [20]. As elemental semiconductor, high purity silicon is one of the most important resources for the electronics industry.

In a typical silicon tracking detector in a particle physics experiment, silicon is used in two ways: As substrate for Integrated Circuits (ICs) to implement the necessary electronic circuits for amplification or digitization and as actual sensing volume, creating a charge signal based on traversing particles and radiation.

Based on the layout and granularity of the segmentation of the readout electrode, different types of silicon tracking detectors can be distinguished. Silicon strip detectors feature segmentation in one dimension, reducing the amount of channels necessary to cover a given area but sacrificing spatial resolution in the other direction. By combining multiple layers of strip sensors and rotating them against each other, some information parallel to the segmentation can be recovered, but strips always represent a compromise between absolute spatial resolution and complexity. When implementing a two-dimensional segmentation of the readout electrode, depending on the granularity either a pad or pixel detector is obtained. By employing modern small-featured Complementary Metal-Oxide-Semiconductor (CMOS) processes, pixel pitches in the order of a few µm can be accomplished.

The actual tracking detector is typically composed of many single entities of a few cm^2 or larger, so called modules, which are arranged in several concentric layers around the interaction point. For both technical and physical reasons, these layers are often barrel-shaped in the central region of a detector, while the forward regions tend to be covered by disc-shaped structures. Several layers ensure that multiple space points are recorded for each particle's trajectory, in order to be able to reconstruct the full track.

2.1 DETECTION PRINCIPLE

Silicon particle detectors work by reading out an electronic signal that is created by the drift of charge carriers in an electric field in the depleted volume of a junction region between

p- and n-doped silicon. The incident particle deposits energy in the sensor material due to different processes, depending on particle type and energy. These processes are discussed in Section 2.1.1. The active volume of a silicon sensor usually consists of the depleted space-charge region of a p-n-junction. Through ionization, the deposited energy creates electron-hole pairs in the semiconductor material, which drift apart in the electric field in the depletion region. As described in Section 2.1.2, the movement of the charges induces a signal in the readout electrodes. By segmenting the readout electrode in two dimensions, tracking of particles becomes possible. Since the created signal is proportional to the deposited energy, a rudimentary energy measurement is also possible with silicon detectors. The latter two principles are explained in Section 2.1.3 and Section 2.1.4, respectively.

2.1.1 PARTICLE INTERACTION WITH MATTER

Depending on particle type and energy regime, different phenomena dominate the interaction of the incident particle with the detector material. For photons, the predominant effects are the photoelectric effect, Compton scattering or pair production, depending on the photon's energy. The resulting total absorption probability in a 300 µm thick silicon sensor is shown in Figure 2.1(a). The photon is either absorbed completely in interactions based on the photoelectric effect and pair production, or is scattered at a large angle in interactions via the Compton effect. Hence, photons are absorbed with a given probability while traversing through the detector material.

In contrast, charged ionizing particles continuously lose and deposit energy in the detector material. The mean energy loss of an incident particle due to ionization per distance in an absorber material is described by the Bethe-Bloch Formula shown in Equation 2.1 [4, pp.536-541].

$$-\left\langle \frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = K\frac{Z}{A}\varrho\frac{z^2}{\beta^2}\left[\frac{1}{2}\ln\left(\frac{2m_e c_0^2(\beta\gamma)^2 T_{\max}}{I^2}\right) - \beta^2 - \frac{\delta(\beta\gamma)}{2}\right], \qquad \text{Equation 2.1}$$

where

- c_0 is the vacuum speed of light, m_e is the electron mass,
- $K = 4\pi N_A r_e^2 m_e c_0^2 = 0.307\,\mathrm{MeV\,cm^2\,mol^{-1}}$ with the electron radius $r_e = e^2/4\pi\varepsilon_0 m_e c_0^2 \approx 2.8\,\mathrm{fm}$,
- Z, A are the atomic number and nucleon number and ϱ is the mass density of the absorber material,
- z is the charge of the incident particle,
- $\beta\gamma = p/mc$ is the relativistic speed of the incident particle, with $\beta = \frac{v}{c_0}$ and $\gamma = \frac{1}{\sqrt{1-\beta^2}}$,
- $T_{\max} \approx 2m_e c_0^2(\beta\gamma)^2$ is the maximum possible energy transfer in a head-on collision,
- I is the mean ionization energy of the absorber material, for silicon $I_{\mathrm{Si}} \approx 170\,\mathrm{eV}$,

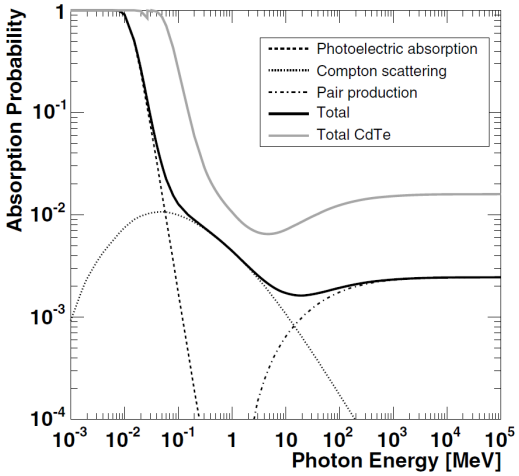
- $\frac{\delta(\beta\gamma)}{2}$ is the density correction factor that gets more important, the higher the energy of the incident particle is.

The Bethe-Bloch Formula approximates the stopping power adequately for $0.1 \lesssim \beta\gamma \lesssim 1000$. More correction factors for different energy regimes have been found, but have been omitted here, since their influence in the considered energy range is negligible.

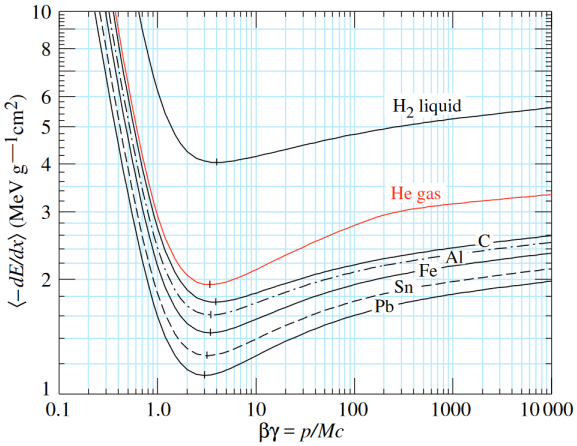
The general form of the Bethe-Bloch Formula is shown in Figure 2.1(b). For low $\beta\gamma \lesssim 1$, the mean energy loss is strongly energy dependent and falls proportionally to $\frac{1}{\beta^2}$. It reaches a minimum for so-called Minimum Ionizing Particles (MIPs) at $\beta\gamma \approx 3$. The value at this minimum, the stopping power for MIPs in silicon, can be estimated as

$$-\left\langle \frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle_{\text{MIP, Si}} \approx 3.87 \frac{\text{MeV}}{\text{cm}}. \qquad \text{[21, p.37]}$$

In the energy regime $10 \lesssim \beta\gamma \lesssim 1000$, the mean energy loss rises only moderately with increasing particle energy. For approximate calculations, the assumption of a MIP therefore holds true over a broad energy range.



(a) Total photon absorption probability in 300 µm of silicon depicted as overlap of photoelectric effect, Compton scattering and pair production, depending on the photon energy. [22, p.34]



(b) Mean energy loss of ionizing particles in matter as a function of the relativistic particle energy $\beta\gamma$, as described by the Bethe-Bloch Formula in Equation 2.1. [4, p.537]

Figure 2.1: Absorption probability and stopping power for different particle types and energies.

For highly energetic particles with $\beta\gamma \gg 1000$, radiative energy losses become the dominant effects, rising exponentially with particle energy. This is mostly relevant for electrons, positrons and highly energetic muons. For each combination of incident particle and absorber material a critical energy can be defined, where the energy loss due to *Bremsstrahlung* surpasses the contribution by ionization. Bremsstrahlung is electromagnetic radiation produced by deceleration or deflection of a charged particle, for example an electron in the Coulomb field of an atomic nucleus in an absorber. Energy loss by Bremsstrahlung can be quantified as

$$-\left(\frac{\mathrm{d}E}{\mathrm{d}x}\right)_{\text{Brems}} = \frac{E}{X_0}, \qquad \text{Equation 2.2}$$

with the energy of the electron E and the *radiation length* X_0, defined as the path length in a given absorber material, after which the incident electron's energy has been reduced to $1/e$ of its original value. For silicon, this is approximately [21, p.65]

$$X_0(\text{Si}) \approx 9.36\,\text{cm}.$$

The material budget of components in a particle detector is typically measured in units of X_0, since this provides direct overview over the detector's influence on traversing particles.

2.1.2 CHARGE MOVEMENT AND SIGNAL GENERATION

The active volume of a silicon particle detector is usually the depleted space charge region in a junction between p- and n-doped silicon. Using a typical planar n-in-p sensor as an example, a weakly p-doped bulk material is chosen and the readout electrodes are implemented as segmented, highly n-doped implantations. Due to appropriate choice of the doping concentrations of the p- and n-type silicon and by applying a reverse bias voltage, the vertical extension of the depletion region can be extended to several 100 µm, effectively filling the full volume of the sensor bulk. For the example of an n-in-p sensor, this means holding the readout electrodes at ground potential and applying a negative bias voltage to the back side, which consequently needs to be metalized, but not segmented. The energy deposited in the material by an incident particle, based on any of the effects described in the previous section, excites electrons in the space charge region, effectively raising them to the conduction band. This creates two types of mobile charge carriers, the electrons in the conduction band and the holes left by the electrons in the valence band. The average number of created electron-hole pairs per amount of deposited energy in a detector at room temperature is given by

$$N_{\text{e-h}} = \frac{E}{w}, \qquad \text{Equation 2.3}$$

with the deposited energy E and the mean required energy per electron-hole pair w. For silicon,

$$w_{\text{Si}} \approx 3.65\,\text{eV}. \qquad [21]$$

The created charges drift apart in the prevalent electric field caused by the built-in potential and the additionally applied bias voltage. According to the *Shockley-Ramo theorem*, the signal in the readout electrode is dominated by the current induced by the movement of the charges in the electric field of the electrode. This current is given by

$$i = e\vec{E}_W\vec{v}, \qquad \text{Equation 2.4}$$

with the charge carrier's charge e and velocity $\vec{v}$ and the *weighting field* $\vec{E}_W$.

The weighting field for different electrode sizes is shown in Figure 2.2. For infinitely large electrodes, it has a linear gradient in vertical direction, meaning that a moving charge carrier

will induce the same effect for any part of its drift path. For smaller electrode sizes, however, the weighting field approaches zero much faster, meaning that charges drifting in the region with vanishing weighting field will not contribute to the total signal in the electrode. Additionally, when looking at pixelated detectors with multiple electrodes, the weighting fields of neighboring electrodes overlaps, meaning that a charge created beneath one pixel cell also induces a signal in the neighboring pixels.

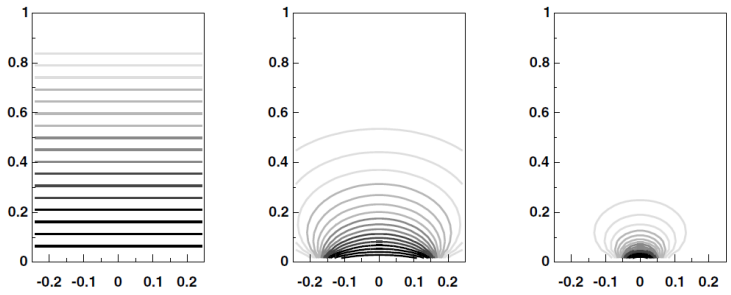


Figure 2.2: Weighting field of a single electrode with different sizes compared to the vertical bulk thickness. [22] The potential lines are shown for an infinitely large electrode (a), an electrode of 1/3 of the bulk thickness (b) and 1/10 of the bulk thickness (c).

However, the total charge induced on the electrode by a charge e moving from x_1 to x_2 in the time interval $[t_1, t_2]$ is given by

$$Q = \int_{t_1}^{t_2} i(t)\, \mathrm{d}t = e\left[\phi_W(x_1) - \phi_W(x_2)\right], \qquad \text{Equation 2.5}$$

where $\phi_W(x)$ is the weighting potential at the position x, which is generally related to the weighting field by $\vec{E}_W = -\vec{\nabla}\phi_W$. When all charge carriers have arrived at the collection node, the charge induced on this node is equal to the charge collected, while the signal on neighboring electrodes vanishes. Since charge collection in silicon sensors only takes a few nanoseconds, thanks to the high mobility of charge carriers, the signal in neighboring electrodes can usually be neglected. These considerations can be summed up as *small pixel effect*, leading to two conclusions for fine-pitched pixel detectors: Charge carriers drifting outside the weighting field towards the back side of the sensor barely contribute to the total signal and most of the signal is caused by the charges drifting very close to the collection node, where both the electric and the weighting field are strongest. The total charge collection time is governed by the carriers' drift velocity $v(x) = \mu E(x)$, with the mobility μ and the electric field strength $E(x)$. Consequently, increasing the bias voltage above full depletion, when the depleted area fills out the full depth of the sensor, is still beneficial, especially after irradiation. This is called *overdepletion*. Additionally, high-rate environments such as the LHC, where new events occur every 25 ns, demand for even shorter charge collection times for a clear distinction of events and prevention of pile-up.

A silicon pixel sensor is characterized by its IV curve, the amount of current flowing as a function of reverse bias voltage applied between the front and back side. This *leakage current*

moderately increases as more of the sensor volume is depleted. Once the depletion region reaches the opposite surface of the sensor bulk at *full depletion*, the rise of the leakage current almost saturates until at some point, the *breakdown voltage* is reached and the current increases abruptly. Both, the absolute amount of leakage current and the breakdown voltage can be a function of radiation damage (see Section 2.2). The full depletion voltage is a function of the sensor design, doping concentrations and bulk thickness and is typically in the order of a few volts up to a few 10 V. The breakdown voltage on the other hand, is mostly influenced by sensor design and can reach values of several 100 V.

2.1.3 TRACKING AND SPACIAL RESOLUTION

The spacial resolution of an idealized detector with binary readout, where no charge information is available, in x and y direction is given by [21, p.843]

$$\sigma_{x/y} = \frac{a_{x/y}}{\sqrt{12}}, \qquad \text{Equation 2.6}$$

where $a_{x/y}$ is the pixel pitch along x or y.

For small pixels, however, the charge cloud will usually be registered by multiple neighboring pixels due to *charge sharing*. This effect leads to so called *clusters* in the hit data. By taking into account charge information, if available, the center of gravity of the cluster can be determined and used as the mean hit position. The effective resolution is therefore significantly improved compared to binary detector readout. The effective resolution when including charge information is given by [21, p.844]

$$\left(\frac{\sigma_{x/y}}{a_{x/y}}\right) = \frac{1}{2\pi^2} \sum_{m=1}^{\infty} \frac{1}{m^2} \cdot e^{-4\pi^2 m^2 \left(\frac{\sigma}{a_{x/y}}\right)^2} \qquad \text{Equation 2.7}$$

It strongly depends on the size of the charge cloud relative to the pixel pitch $\frac{\sigma}{a_{x/y}}$.

2.1.4 SIGNAL AND ENERGY MEASUREMENT

A typical analog front-end electronics circuit of a silicon pixel detector is shown in Figure 2.3. For each pixel, the charge collecting implantation is connected to the left-hand side of the circuit. The first amplification stage is usually a Charge Sensitive Amplifier (CSA) with constant-current feedback. This current-integrating amplifier creates a voltage pulse (Figure 2.3 (b)) from the current induced on the collection node (Figure 2.3 (a)). The amplitude of the CSA output pulse is proportional to the total amount of collected charge, while the slope of its falling edge is a function of the feedback current. One or more amplification stages can be implemented after the CSA for gain or shaping purposes. The signal is then fed to a discriminator, that compares the signal with a fixed threshold voltage. The discriminator output switches from low to high, as soon as the input signal exceeds the threshold voltage and falls back to low, once the input falls below threshold again, thus digitizing the signal. The digital output pulse of the discriminator is shown in Figure 2.3(c).

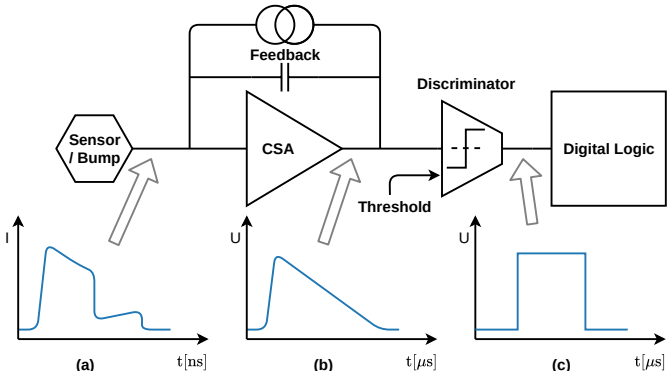


Figure 2.3: Generic pixel detector front-end electronics. The signal shape is shown as a current pulse [21, p.150] coming from the sensor (a), as a voltage pulse after the CSA with constant-current feedback (b) and after the discriminator (c).

Due to the constant feedback current, the length of the digitized pulse is proportional to the amplitude of the input signal. This length is called Time-over-Threshold (ToT) and is the typical method of energy measurement of silicon-based particle detectors. This principle is shown in detail in Figure 2.4. It shows two CSA output pulses with different amplitude U_1 and U_2, leading to pulses of different length ToT_1 and ToT_2 after the discriminator. The absolute value of the ToT is a function of the absolute threshold as well as the feedback current of the CSA. If more amplifiers are introduced into the readout chain, their shaping constants will also influence the resulting ToT. Typically, the preamplifier feedback current is adjusted to tune the ToT response to a given charge to the desired value (see Section 4.6.1). In practice, ToT is also used as a unit for measuring the deposited energy. A pule length of 5 clock cycles of the sampling clock is therefore referred to as 5 ToT.

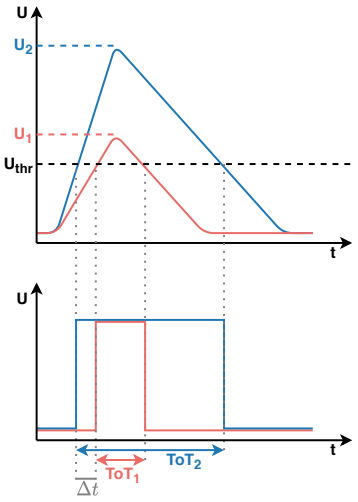


Figure 2.4: Concept of Time-over-Threshold (ToT).

Another effect of this method of digitization is that the absolute moment when the digital pulse starts is also dependent on the analog pulse's amplitude. This effect is called *timewalk* and is depicted by Δt in Figure 2.4. This timewalk leads to hits with lower charge being recognized later than hits with higher charge deposited in the sensor. This delay for small hits can become large enough, so that the affected hits are not read out within the designated trigger window anymore. The resulting effective threshold for all hits to be recorded in time is called *in-time threshold* and is an important metric for characterization of an Analog Front-End (AFE) implementation. This effect is discussed in detail in Section 4.5.6.

2.2 RADIATION DAMAGE

Since tracking detectors usually operate very close to the interaction point of a particle physics experiment, they are constantly exposed to a high flux of charged and neutral particles as well as X-ray and gamma radiation. Depending on the intensity and energy of the accelerator, this radiation has the potential of damaging the silicon structure of both the sensor and the readout ASIC. This is especially relevant in the light of the HL-LHC upgrade, since this accelerator will create one of the harshest radiation environments in the history of particle physics experiments.

Generally, two different types of radiation damage can be distinguished.

- Ionizing radiation can lead to surface damage and damage at interface layers, e.g. between Si and SiO_2. The measure typically used to assess the amount of received ionizing radiation dose is the Total Ionizing Dose (TID), measured in Gray or rad ($1\,\text{rad} = 0.01\,\text{Gy}$). The expected TID for the inner layers of the ATLAS ITk pixel detector range from 1.5 MGy up to almost 10 MGy after half of the planned lifetime of the experiment [23, p.31].

- Non-Ionizing Energy Loss (NIEL) describes damages based on energy loss effects other than ionization, e.g. collisions of the incident particle with nuclei of the lattice atoms. This can either excite a phonon in the lattice or permanently damage the crystal structure of the absorbing material, e.g. by displacement of lattice atoms. By application of the *NIEL hypothesis* [24] the bulk damage to silicon caused by different types of particles at different energies can be normalized to the equivalent damage done by neutrons with an energy of 1 MeV. The measure for damage caused by NIEL is therefore given as the 1 MeV neutron equivalent fluence $\left[\Phi_{eq}\right] = n_{eq}\,cm^{-2}$. The ATLAS ITk pixel detector is expected to receive total fluences in the order of $1.3 \cdot 10^{16}\,n_{eq}\,cm^{-2}$ by half of the intended lifetime of the experiment [23, p.31].

While radiation damage is mostly associated with the sensor properties, readout ASICs also suffer from extreme radiation, which may cause Single-Event Effects (SEEs), in addition to the effects described above. These events can lead to disruptions of the front-end operation or digital circuitry and can even cause data loss or unrecoverable states. The radiation tolerance of hybrid pixel detector readout ASICs such as `RD53A` is studied in detail and preemptive measures are taken in the chip design to increase robustness against SEEs.

2.3 MODULE DESIGN CONSIDERATIONS

Two general concepts exist for the production of a single detector module: The active detection volume and the necessary electronic circuits can be integrated on the same bulk, forming a so called Monolithic Active Pixel Sensor (MAPS), or a passive sensor and a dedicated read-out ASIC are produced independently and connected together to form a module in an additional processing step called hybridization. The resulting module is then called a hybrid pixel detector. Both concepts have distinct advantages and drawbacks, which will be illustrated in the following sections [25].

2.3.1 Monolithic Pixel Detectors

In recent years, MAPS have increased in popularity since their performance has caught up and even surpassed other concepts, with new production technologies becoming available to the physics community. Designs for this kind of detector have gained tremendously from new process features like full CMOS technologies with large metal stacks or the use of highly resistive substrates. This modern type of monolithic detector is often referred to as Depleted Monolithic Active Pixel Sensor (DMAPS), due to the fact that a fully depleted volume can be used for charge collection. An example for the implementation of a DMAPS is shown in Figure 2.5.

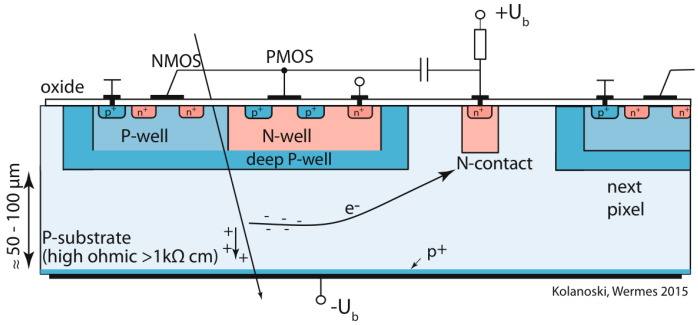


Figure 2.5: Example for a DMAPS detector. [22] Read out electronics are integrated in the same bulk as the active volume by using multiple nested p- and n-wells to achieve full CMOS capabilities.

The foremost advantage of monolithic detector concepts is the simplicity of their production. Once a design has been thoroughly tested and qualified, producing large quantities of these detectors, as is necessary for most modern experiments in high-energy particle physics, becomes comparably simple. The production cost in general is very well defined and low compared to other concepts, due to the fact that only a single or at most a handful of vendors is involved in the process. The production yield is generally high and the cost per unit can even be reduced further by making use of highly available standard CMOS processes.

Threshold and noise performance, detection efficiency as well as power consumption of MAPS are comparable to other concepts. A shortcoming of MAPSs for their application very close to the interaction point of future LHC experiments, however, is their poorer ability to withstand large doses of ionizing, as well as non-ionizing radiation.

2.3.2 Hybrid Pixel Detectors

Hybrid pixel detectors offer more flexibility in the choice of manufacturing processes, as the technologies for the readout ASIC and the sensor can be chosen independently. For example, the readout chip can be manufactured in a modern, small feature-size CMOS process that enables implementation of the complex circuitry in the given size of the pixel area and with the lowest possible current consumption, while the much simpler sensor can be produced using a significantly cheaper process with a larger featuresize. This approach offers a lot of benefits for the design and production, but comes with an additional, expensive and error-

prone manufacturing step, the process of interconnecting the two ASICs pixel by pixel via high-density fine-pitch bump-bonding.

A schematic view of a hybrid pixel detector module is shown in Figure 2.6. The pixelated readout chip is shown on the bottom with the connections to the sensor on top. Tiny solder balls, so-called bump-bonds, are placed on either of the chips before the sensor is placed on top of the readout chip with its pixel output connections pointing down. The two ASICs are aligned and the whole stack is heated up until the solder balls melt and form a mechanical and electrical connection between the two chips. The connection to the detector services and Data Acquisition (DAQ) system is typically realized with wire-bonds on the bottom edge of the readout chip, that overlaps with the sensor.

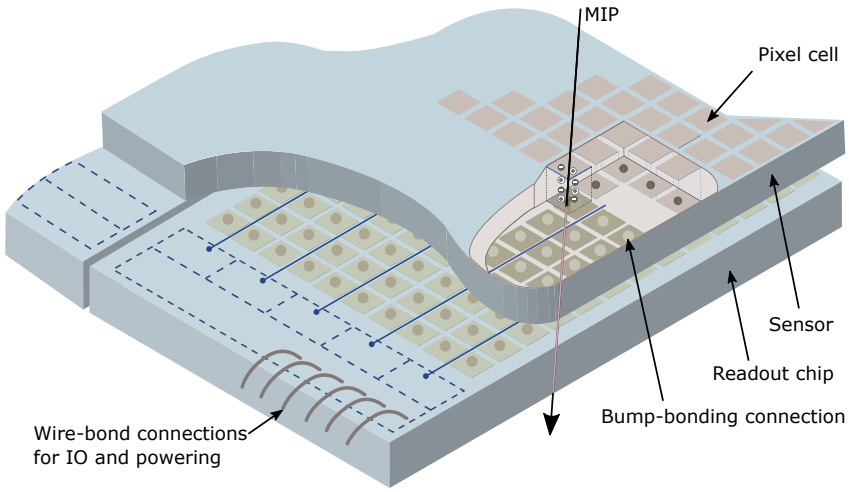


Figure 2.6: Schematic view of a hybrid pixel detector with square pixels, modified from [26, p.47]. A MIP is shown traversing through a single pixel cell in the sensor. The readout chip is connected via bump bonding. Additionally, the wire-bond connections of the readout chip are depicted.

Since the manufacturing yield can vastly differ based on design complexity and is a function of the area of a single ASIC entity, sensors can typically be manufactured larger than the readout chips of a hybrid particle detector. A commonly used approach to overcome this problem is to use so called multi-chip modules, where several smaller readout chips are bump-bonded to a larger sensor tile to form a single module. For the current generations of detector modules, two, three or four readout chips are connected to a single sensor, forming Double Chip Modules (DCMs), Triplets or Quad-Chip Modules (QCMs), respectively [23]. Older generations of hybrid particle detectors, for example the current ATLAS pixel detector, used up to 16 readout ASICs on a single sensor tile [9].

Despite the drawbacks of hybrid pixel detectors, namely the additional manufacturing complexity, the increased material budget and the higher cost compared to monolithic detector designs, hybrid detectors are still used close to the interaction points, in high-luminosity and -radiation environments, e.g. in the HL-LHC, since the ability to choose independent processes for the readout chip and the sensor allows for a radiation hardness that is not yet achievable with monolithic designs. The prospective hybridization process for the ATLAS ITk pixel detector modules is described and analyzed in-depth in Chapter 8.

CHAPTER 3

THE RD53A READOUT ASIC PROTOTYPE

RD53A is a prototype readout ASIC for hybrid pixel detector modules. It was developed by the RD53 collaboration, a joint research and development collaboration between the ATLAS and CMS experiments. The goal of this collaboration is to develop the methods and foundations for a new-generation readout ASIC for hybrid silicon pixel detectors, that is able to cope with the conditions and requirements of HL-LHC. This should be accomplished using 65 nm CMOS technology. The resulting readout chips are to be used in the upgrades of the tracking detectors of both ATLAS and CMS.

The RD53 collaboration was established in 2013, with the expressed goal to evaluate and qualify the chosen 65 nm CMOS process in terms of radiation hardness, to develop the methods and foundations to design large and complex ASICs for application in the ATLAS and CMS tracking detectors, and to develop and characterize building blocks that can directly be used in the designs of these final chips [27].

After a number of years of developing the methods and framework for the design, several demonstrator and prototype chips for specific features, as well as two small prototype chips called FE65-P2 [28] and Chipix [29], a large-scale prototype readout chip was produced in order to demonstrate the suitability of the chosen CMOS technology for the HL-LHC upgrades of ATLAS and CMS in a large format IC. This readout chip is called RD53A and is shown in Figure 3.1. It incorporates two different digital buffer architectures and three different flavors of analog front-end electronics design that should be tested and evaluated. A decision needs to be made for both the CMS and ATLAS production chips, which will only use one analog

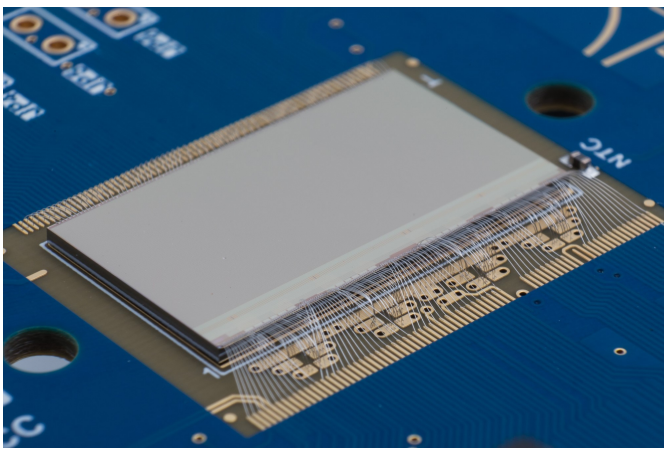


Figure 3.1: Bare RD53A chip wire-bonded to a Single-Chip Card (SCC)

design each. RD53A already incorporates most of the key features of the production chip, such as 50 µm pixel pitch, high hit rate capability including high speed Input/Output (I/O) and serial powering capabilities. This chip was never intended to be used by the experiments as a production chip, but instead is supposed to be used to conduct module pre-production steps, such as market surveys and development of methods and tooling for production.

In August 2017, the RD53A design was submitted to the foundry and the silicon arrived at the first testing sites in early December 2017. The chip was shown to be working mostly within specifications in the following months. By the end of 2018 the ATLAS collaboration conducted a review to decide which Analog Front-End (AFE) design would be used in the production chip for the ATLAS Phase-II upgrade. The decision to use the Differential front-end was reached in early 2019 [30]. A summary of the review process can be found in Section 5.7. The decision was independent of the choice the CMS collaboration was going to make a few months later. The CMS experiment will use the Linear front-end design in its pixel detector readout chip [31]. Hence, the production chips for the two experiments will not only differ slightly in physical matrix size, but also in the fundamental design of the analog electronics.

In 2018, the scope of the RD53 collaboration was extended by request of both ATLAS and CMS to include the design and fabrication of the actual production chips for both experiments [32].

3.1 DESIGN

The physical layout of RD53A is shown in Figure 3.2. The chip is 20 mm wide, since the final production chips for both ATLAS and CMS were going to be at least 20 mm wide and so the main pad frame could be reused in the design of these chips. Due to the available space in the wafer reticle, the chip height was limited to 11.6 mm, which is about half of the expected height for the production chips for the two experiments [33].

The main pad frame, which is used to wire-bond the chip to a SCC or flexible PCB (flex), is located along the bottom edge of the chip. Above this pad frame, there is the peripheral circuitry of the chip, including the eight Shunt Low-Dropout Regulators (sLDOs), a 12-bit Analog-to-Digital Converter (ADC), multiple bias Digital-to-Analog Converters (DACs), the chip's Clock/Data Recovery (CDR) and Phase-Locked Loop (PLL) circuits, the Power-on Reset (POR) circuit, a block with multiple different sensors, as well as the global analog and digital logic of the chip, called Analog Chip Bottom (ACB) and Digital Chip Bottom (DCB), respectively. All these blocks combined take up about 1.7 mm of height along the full width of the chip. Above the bottom logic block follows the main sensitive pixel matrix of the chip, consisting of 400 columns and 192 rows of 50 µm × 50 µm wide pixels. On the top edge of the chip, there are more wire-bond pads for various testing and debugging functions and signals, which are supposed to be removed in further iterations of the chip [33].

3.1.1 ANALOG ISLANDS IN A DIGITAL SEA

The pixel matrix itself can be broken down into 1200 *cores*, each containing 8 × 8 pixels. Design and verification is done for one core, which is then stepped and repeated to reach the desired total matrix size.

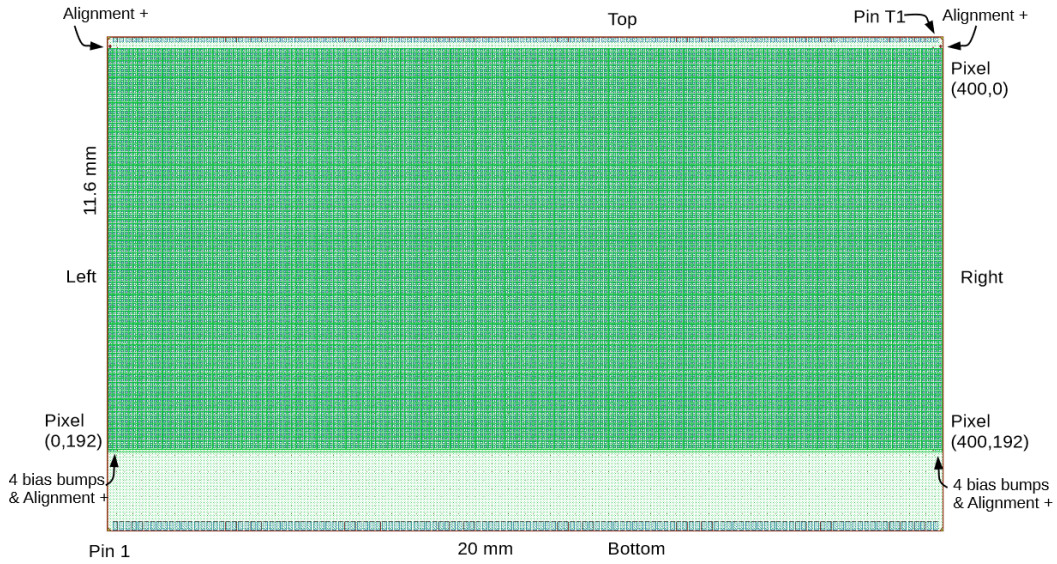


Figure 3.2: Top view of the RD53A design [33]. The chip is 20 mm or 400 columns wide and 11.6 mm or 192 rows high. The *bottom* of the chip is defined as the edge containing the main pad frame. The pixels are counted from the top left (0,0) to the bottom right corner (400,192).

Within a core, the RD53 collaboration followed the "analog islands in a digital sea" design philosophy. Here, the analog circuitry for multiple pixels is combined into a single island that sits in the midst of a sea of synthesized digital logic. This way, the efforts for shielding the analog electronics can be minimized while taking full advantage of modern synthesis algorithms for digital logic, which would be much harder to design and especially to verify manually.

Figure 3.3 shows the general design concept of RD53A, where the AFEs of four pixels are combined into one island with an area of 70 µm × 70 µm. This island also houses four bump pads with a 50 µm × 50 µm pitch, which is used to connect the readout-pixel input with the corresponding output of the sensor pixel. The synthesized digital logic around the islands cannot be attributed to one pixel anymore in a simple way, as depicted in the figure. Instead, the placement of logic gates is optimized automatically by a synthesis tool. It has been shown that this difference in neighboring circuitry does not influence the characteristics of the adjacent AFE circuitry [33]. The synthesis of one core is executed as a whole and the full design of a core is then repeated over the matrix.

The main motivation and most important benefit of this design approach is verification. Many methods and experience in verifying synthesized digital logic already existed and were used and refined in the verification process for RD53A. Apart from the digital verification, it was made sure that a single core is small enough to be verified using transistor-level analog simulation within a reasonable amount of time [33, 34]. Furthermore, the chosen design strategy offers advantages for power distribution and homogeneous biasing over the full pixel matrix, which is a significant concern, regarding the fact that the power regulators are located exclusively in the analog bottom logic block (ACB) and the power consumption of a single

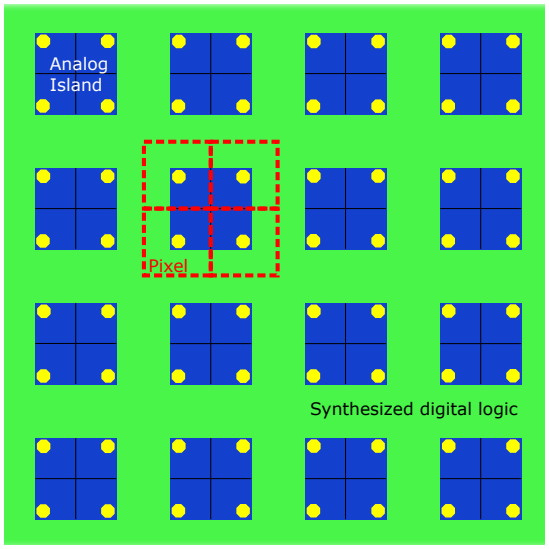


Figure 3.3: Schematic view of the "analog islands in a digital sea" design strategy. One core consists of 64 pixels. Four pixels (red) share an analog island (blue) which houses the implementation of the four AFEs as well as the bond pad (yellow). The digital circuitry of all pixels in a core is synthesized as a whole (green).

pixel can add up to a maximum of 14 µA when combining the analog and digital domains [33].

3.1.2 ANALOG FRONT-END FLAVORS

Since RD53A was designed as a prototype chip, one of its main purposes, next to the general proof of concept, was the evaluation of different digital and analog architectures and designs. For this purpose, the pixel matrix is subdivided to implement two different digital buffer architectures, Distributed Buffer Architecture (DBA) and Central Buffer Architecture (CBA), which are explained in detail in Section 3.1.3. This division is illustrated in Figure 3.4.

Furthermore, the matrix is divided vertically into three regions with different implementation of the AFE circuit design. The three AFEs are called Synchronous Analog Front-End (SYNC), Linear Analog Front-End (LIN) and Differential Analog Front-End (DIFF) and are described in the following subsections [35]. While the SYNC and LIN have been prototyped before in Chipix and the DIFF was evaluated in the FE65-P2 submission, the purpose of RD53A was to implement all three of them on the same chip, in combination with both buffer architectures, to evaluate and compare them and enable the experiments to make a well-founded decision for one of the AFE flavors to be used in their respective production chips (see Section 5.7).

The main purpose of the AFE of the pixel is the amplification, shaping and digitization of the input signal coming from the sensor pixel, which is bump-bonded directly to the input stage of this circuit. The greatest challenges involve stable low-threshold and low-noise operation of the amplifier stages, creating a homogeneous threshold distribution over the matrix and coping with the sensor leakage current, which increases over the lifetime of the detector as radiation damage of the sensors increases. The three AFEs of RD53A implement different methods to accomplish this.

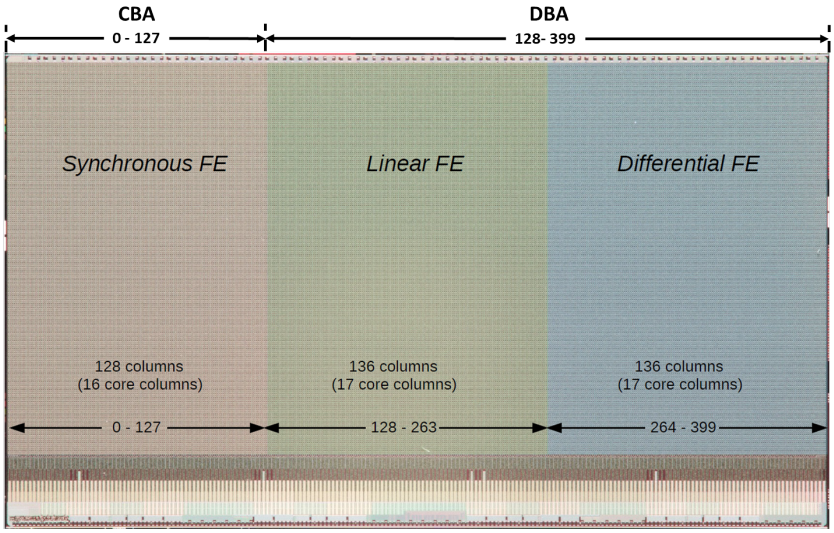


Figure 3.4: Location of the three analog front-end flavors on the RD53A chip, modified from [33, p.18]. The number and naming scheme of the corresponding columns is indicated in the figure, along with the number of core columns (columns of 8x8 pixel cores). At the top of the matrix, the division of the two buffer architectures is shown.

SYNCHRONOUS FRONT-END

The SYNC is shown in a simplified schematic view in Figure 3.5. It implements a single-stage CSA with Krummenacher feedback as pre-amplifier. The Krummenacher circuit in the feedback loop of the pre-amplifier serves to discharge the feedback capacitor and the Krummenacher (feedback) current can be adjusted to control the return-to-baseline of the CSA, enabling manual adjustment of the ToT response. At the same time, the Krummenacher circuit is intended to compensate sensor leakage current, preventing it from being amplified together with the signal and being fed to the subsequent amplification stages [36]. In addition to the adjustable Krummenacher current, two different feedback capacitors have been implemented and can be included into the circuit manually by means of the pixel configuration register, which allows to choose between three different gain settings for the pre-amplifier.

Digitization is achieved by a simple differential amplifier discriminator, that is AC-coupled to the pre-amplifier in order to ignore fluctuations in its output baseline. The discriminator threshold is controlled by a global voltage called VTH_SYNC, that is the same for all pixels of the SYNC. Differences in the actual threshold of the pixels resulting from this global voltage lead to a non-negligible *threshold dispersion* over the pixel matrix. This dispersion is usually compensated by means of trim DACs, which is the case for both the LIN and DIFF. The SYNC, on the other hand, implements a mechanism called Auto-Zero (AZ) and is based on a latch, that is coupled to both the inverting and non-inverting output of the discriminator via two capacitors. The latch needs to acquire the baseline regularly, then compares the input signal to the stored baseline and generates the discriminator output. This way, the additional preparation step of tuning trim DACs in every pixel can be omitted for the cost of having to send an AZ pulse at least every 100 µs during operation.

The additional latch stage in the discriminator can also be used as an in-pixel oscillator with a frequency of up to 800 MHz in an alternative mode, enabling faster ToT counting and thus much better energy resolution. [33, 37]

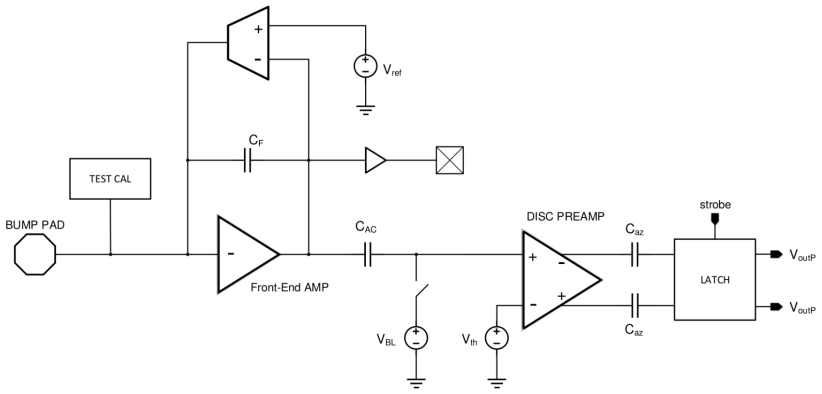


Figure 3.5: Simplified schematic view of the SYNC [33, p.22]. The single-stage CSA with Krummenacher feedback is directly connected to the bump pad, while its output is AC-coupled to the discriminator. The comparator is implemented from a pre-amplifier that is AC-coupled to a latch that is responsible for comparing the input signal to the baseline that is acquired regularly.

LINEAR FRONT-END

The LIN takes a more traditional approach to analog signal processing and digitization. As shown in Figure 3.6, it is based again on a single-stage CSA for amplification and pulse shaping. Like the SYNC, the LIN uses a Krummenacher feedback network for leakage current compensation, which can be tuned by different peripheral DACs, e.g. `REF_KRUM_LIN` and `KRUM_CURR_LIN` controlling the Krummenacher reference voltage $V_{\text{REF\_KRUM}}$ and the recovery current I_K, respectively. As for the SYNC, the feedback capacitance of the LIN, and with it the gain of the amplifier, can be selected from two different settings by using the `GAIN_SEL` bit of the pixel register (see Table A.2) to use C_{F1} either alone or in parallel with C_{F2}.

The output of the amplification stage is DC-coupled to a threshold discriminator that compares the signal to a static threshold voltage, provided by a global threshold DAC called `Vthreshold_LIN`. Similar to the SYNC, the discriminator is also responsible for the charge measurement by means of a ToT counter.

Like most pixel detector readout ASICs, including the predecessors of `RD53A` currently in use in the ATLAS detector, the LIN implements a 4-bit Local Threshold DAC (TDAC) for manual fine tuning of the threshold voltage for each pixel individually and reducing the threshold dispersion over the pixel matrix. For the LIN, this TDAC is binary weighted and a setting of 15 corresponds to the lowest possible local threshold value, while 0 sets the highest possible value. Tuning the pixels' TDAC to the optimal value to minimize threshold dispersion over the matrix has to be performed manually by the DAQ system but should be stable for extended periods of time, once achieved. [33, 38]

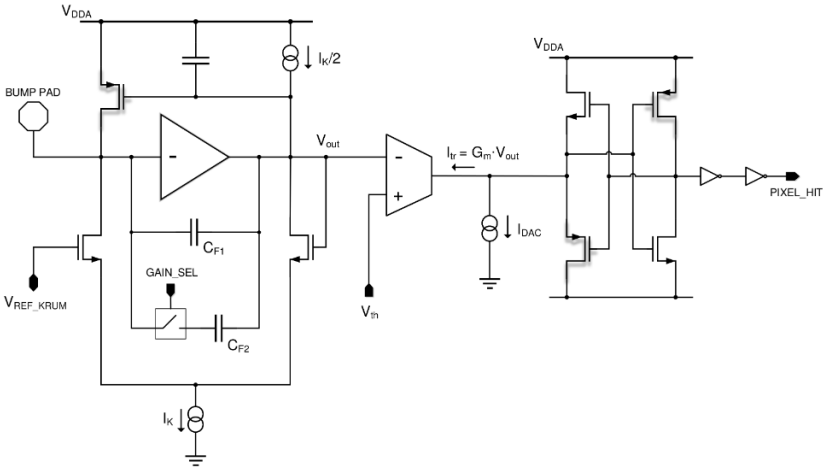


Figure 3.6: Simplified schematic view of the LIN [33, p.18]. The main amplification stage (left half) is directly coupled to the sensor bump pad. The discriminator (right half) is DC-coupled to the output of the CSA and is responsible for pulse digitization and ToT counting.

DIFFERENTIAL FRONT-END

In contrast to the other two AFEs, the DIFF only implements analog circuitry within the analog island, leaving the digital pulse processing and buffering to the digital core. This allows for further optimization regarding a low minimum stable threshold and low noise, as well as low threshold dispersion.

Figure 3.7 shows a simplified block diagram of the DIFF. The analog front-end is implemented with three stages. The first stage is a CSA with traditional capacitive feedback and continuous reset. Like in the other two flavors, the gain of the pre-amplifier can be adjusted by acting on a global configuration bit to either include C_F in the feedback loop or disconnect it. Since the DIFF is missing a Krummenacher circuit, high sensor leakage currents of more than 10 nA could potentially saturate the DC-coupled pre-amplifier. For these extreme cases, the DIFF implements an LCC circuit in the feedback loop of the pre-amplifier, that can be connected to the circuit if enabled. However, based on expectations at the time of designing the AFE, the continuous feedback alone should be able to prevent saturation during detector operation [33]. The connection of the LCC circuit as well as the inclusion of C_F are controlled by the global register `CONF_FE_DIFF`.

Both, the input and output signals of the pre-amplifier are fed to the differential inputs of the second amplification stage, that provides an additional gain and applies the threshold to the signal. This pre-comparator is implemented with two branches, that are offset by two independent global threshold voltages, `VTH1_DIFF` and `VTH2_DIFF`. Both of these voltages can be trimmed by the four TDAC bits of the DIFF, while the fifth TDAC sign bit selects which branch is to be adjusted, effectively allowing threshold tuning over a five bit range. The possible TDAC values therefore range from -15 for the lowest to +15 for the highest local threshold. This design is optimized for low threshold and low noise operation.

The last stage is a traditional comparator stage, that compares the threshold-adjusted signals

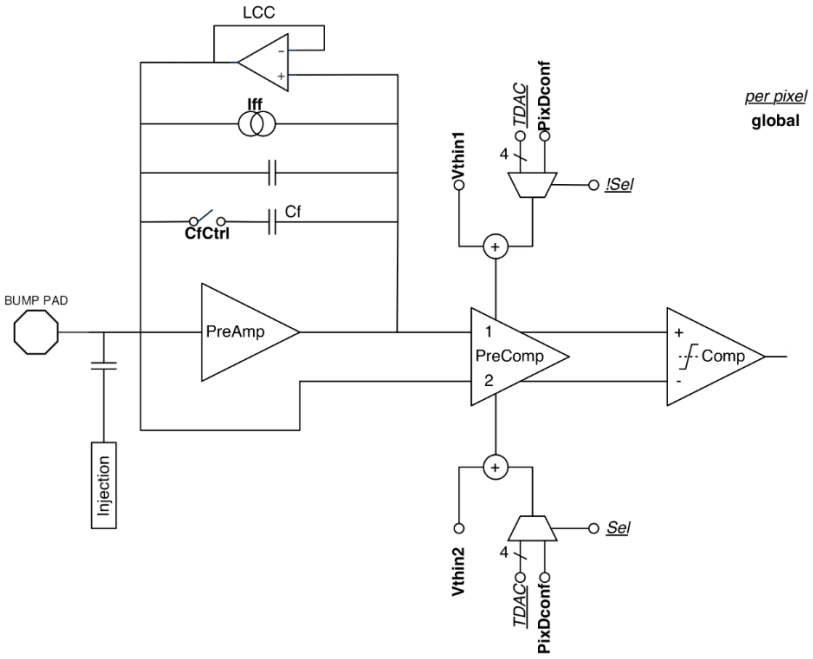


Figure 3.7: Simplified schematic view of the DIFF, modified from [33, p.20]. The first stage (PreAmp) is implemented as a traditional CSA with capacitive feedback and continuous reset. Additionally, the feedback loop includes an optional Leakage Current Compensation (LCC) circuit. The second stage (PreComp) provides additional gain and applies the threshold to the differential input signal. Effective 5-bit local threshold tuning is implemented by only applying the TDAC bits two one of the two threshold voltages. The third stage (Comp) is formed by a traditional continuous time comparator stage.

from the first amplification stages and outputs a digital pulse whose length depends on the ToT and thus on the charge collected by the sensor. [28, 33]

3.1.3 DIGITAL PIXEL LOGIC

The digital pixel logic is almost identical for all three AFEs and is synthesized as a whole for a pixel core. Figure 3.8 shows every pixel's signal processing chain.

The input to the circuit on the left can either originate from the pixel's AFE or from an injection pulse called `CAL_EDGE` (see Section 3.1.4). The whole digital blocks of the entire core column can be enabled or disabled by means of the digital enable bit (`EN_CORE_COL_[FE]`). The signal is then split into two independent branches: Gated by the `HitOr Enable` bit (see Table A.2), the signal is linked with the other pixels in its HitOr network by a logical OR gate, implementing the chip's HitOr feature. The other branch constitutes the main digital pixel logic. The general `Enable` bit is used to mask the pixel off if necessary. The *Hit Discriminator & Counter* block evaluates the binary input signal to decide whether an in-time hit is detected. If this is the case, the ToT value is determined and stored per pixel. Since the `RD53A` ToT counter

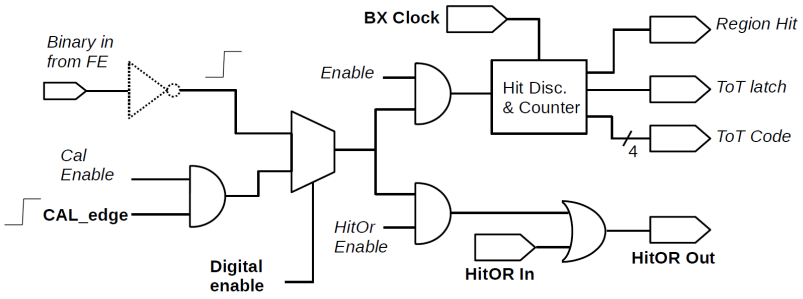


Figure 3.8: Single-pixel digital signal processing chain [33, p.28]. The AFE signal or an injection pulse (CAL_EDGE) is gated by the digital enable bit before being divided into the hit discrimination and counting path, gated by the pixel enable bit (see Table A.2) and the HitOr path gated by the pixel HitOr enable bit. The dashed inverter is only implemented for the DIFF.

size is 4 bit, any ToT value of 15 or greater is encoded as 0b1111[1] in RD53A. In addition, when a hit is recognized in the *Hit Discriminator & Counter* block, a latency timer is started. When this timer reaches its programmed latency value and a trigger is present, a trigger tag is generated and stored together with the ToT information for later readout. If no trigger is present, the ToT value is erased and the latency counter is reset.

BUFFER ARCHITECTURES

RD53A implements two different digital buffer architectures, Distributed Buffer Architecture (DBA) and Central Buffer Architecture (CBA), dividing the pixel matrix vertically. They differ with regard to the distribution of ToT buffers and latency counters between pixels within a defined region. In both architectures, the latency counters are global for the region, while ToT buffers are either distributed between the individual pixels (DBA) or also implemented as common region memory (CBA). Moreover, the region size differs between the two architectures, with four pixels in a region for DBA and 16 pixels per region for CBA.

While the CBA efficiently suppresses ToT $= 0$ values, simply by not having to store information for pixels without a hit at all, it needs to store an additional hit map to attribute the hit to the correct pixel. The DBA on the other hand requires no hit map, but needs to store ToT $= 0$ for pixels without a hit. This makes the DBA more efficient for higher region occupancy with more pixels per region being typically hit, favoring smaller region designs. The CBA on the other hand will be more efficient for lower region occupancy, enabling larger regions to be defined. Furthermore, resource efficiency of both architectures depends on the register size of both the latency counters as well as the ToT buffers. Smaller ToT registers would mean a clear advantage for DBA, while larger ToT registers clearly favor CBA. In the case of RD53A with 9 bit timers and 4 bit ToT registers, the performance of both architectures is similar. [33]

[1] Binary numbers are denoted with a leading 0b, hexadecimal numbers with a leading 0x. Numbers without any leading identifier are always decimal numbers.

3.1.4 CHIP BOTTOM

The bottom part of RD53A below the active pixel matrix is called chip bottom and contains numerous global blocks of the chip. It is subdivided into a digital part (DCB), which contains elements for data processing, start-up and reset logic, and global registers and an analog part (ACB) housing the chip's CDR and PLL circuits, the monitoring block including a 12 bit ADC, the sLDO regulators and a calibration block to generate the calibration injection voltages (see Section 3.1.5).

CDR AND PLL

RD53A needs different clock signals for different functions and signals. Apart from the 40 MHz bunch crossing clock (BX_CLK), which is provided externally by the accelerator, a 160 MHz clock (CMD_CLK), a 640 MHz and a 1.28 GHz clock (SER_CLK) are generated on-chip by a custom-designed Clock/Data Recovery (CDR). The CDR block utilizes a reference clock that is generated by a Phase-Locked Loop (PLL), locking to the input command stream. Since the 40 MHz BX_CLK, that is the main clock distributed to the pixel matrix, is generated by dividing the 160 MHz CMD_CLK, there are four different possibilities for its phase shift. The correct option is chosen by means of a special "sync" pattern that needs to be sent to the chip on start-up or after the command stream was interrupted. The internally generated clocks can be delayed in steps of 1.6 ns by means of a delay clock (DEL_CLK), in order to synchronize the chip operation with the accelerator bunch crossings. [39]

sLDO REGULATORS

The intended powering mode of the pixel detector modules in the ATLAS ITk is serial powering, where multiple modules are connected to a single power supply rail in series, reducing the amount of necessary power cables significantly. To enable even power distribution between the modules of a serial powering chain, the chain is supplied with a constant current defined by the total power consumption of all connected modules and some additional headroom. To generate the necessary constant digital and analog voltages, each RD53A chip contains two Shunt Low-Dropout Regulators (sLDOs). These custom-designed regulators are capable of supplying a constant output voltage of 1.2 V for both, the analog and digital voltage rails with a typical total power consumption of 750 mA with a typical input voltage of just 1.5 V and a current headroom in the order of 10 to 20 %. By shunting a variable current directly to ground, differences in the power consumption of one module can be decoupled from the rest of the serial powering chain. [33, 40, 41]

The sLDOs developed within the scope of the RD53 readout chips are extensively characterized individually. [41–43]

ON-CHIP ADC

For measuring analog voltages and currents for self-monitoring purposes, RD53A features a 12 bit Successive Approximation Register (SAR) ADC in the Analog Chip Bottom (ACB). Together with a two-stage multiplexer, this monitoring ADC can be used to evaluate other

DACs on the chip. It is also used to read out the four on-chip temperature sensors located at different positions on RD53A.

When the ADC's reference voltage VREF_ADC is trimmed to the nominal 900 mV, the ADC's Integrated Non-Linearity (INL) has been shown to be below 1 LSB. [44]

3.1.5 OTHER DISTINCTIVE FEATURES

This section describes additional features of RD53A that are relevant to the measurements and results reported in later sections. For an extensive list and detailed explanation of all features of the chip, refer to the RD53A Manual [33].

HITOR NETWORKS

Important distinctive features of RD53A are the four HitOr networks to access the unprocessed pulses of all pixels, which can be used for self-trigger like operation or for measurements with increased charge resolution using the Time-to-Digital Converter (TDC) method (see Section 4.3). In contrast to its predecessor the FE-I4 [45], where the whole pixel matrix was combined into one output signal, RD53A features four independent HitOr networks, each of which is fed by one quarter of the pixels. By connecting only non-adjacent pixels to the same network, the analysis of hit clusters in the readout system by means of the raw analog signals becomes possible. The HitOr output and the associated enable bit is part of the digital pixel logic and shown in Figure 3.8. [33]

CALIBRATION INJECTION CIRCUIT

Each pixel of RD53A contains a calibration injection circuit that is able to generate an artificial charge signal at the input of the AFE. This charge signal is used extensively in the characterization and general operation of the readout chip. It serves as general input for self tests (e.g. Analog Scan, see Section 4.5) as well as reference charge for measurement (Threshold Scan) and tuning (Global / Local Threshold Tuning) of the pixels threshold voltages as well as their ToT response (ToT Tuning).

To compensate for differences in the local ground level of different pixels, the calibration injection circuit is based on two differential voltages called VCAL_MED and VCAL_HIGH, that are distributed from the ACB to the top for each core column. Each pixel includes a switch to toggle between the differential voltages and thereby generate a voltage step that is fed to an injection capacitance connected to the AFE input. Using the Injection Enable bit of the pixel register (see Table A.2), each in-pixel injection circuit can be enabled or disabled separately. This is necessary, since the injection lines should not be loaded with too many pixels simultaneously, in order to respect both timing and amplitude driving capabilities (see Section 4.3.2). [33]

The absolute charge generated by the injection circuit as a function of the voltage difference between VCAL_MED and VCAL_HIGH (in short: ΔVCAL) has been calibrated on a handful of single-chip assemblies with different types of sensors, by first determining the mean value of the in-pixel injection capacitance and then measuring the amplitude of the injection pulse,

which is available via the pad frame of RD53A. These results have been verified by using charge spectra of different sources to deduce a linear relation between the arbitrary setting parameter ΔVCAL and the resulting charge threshold value in units of e^-. [46]

TRIGGER COMMAND

A *trigger* is an external command supplied to the chip to initiate a readout of the hits stored in the buffers of the matrix. Upon receiving a trigger command, a global trigger table in the DCB is filled with the current value of the bunch crossing counter, giving absolute timing information to the associated hits, and a user definable trigger tag. The matching trigger ID is distributed to the matrix, where all pixels respond with data, if they have data stored in their local buffers and the according latency counter has already reached the programmed value. The trigger table can store 32 triggers before it needs to be emptied by reading out the data. For this reason, BDAQ53 sends 32 trigger commands for each injection pulse by default and reads out the data as fast as possible. This way, a hit is harder to miss if the latency of the chip and the DAQ are not perfectly synchronized.

TEMPERATURE SENSORS

To provide a general idea about the environmental conditions of the chip, RD53A offers four diode-based temperature sensors, located at different points in the chip's periphery, including close to some of the sLDOs, and in the opposite corner. By applying two different currents to these sensors with a dedicated DAC and measuring the resulting voltage using the monitoring ADC, the chip substrate temperature can be deduced from a calibration curve. In practice, these temperature sensors have been shown to have an accuracy of $\pm 4\,°C$ under ideal conditions. [33, 44]

RING OSCILLATORS

For monitoring the radiation damage of an RD53A chip, it features eight independent ring oscillators, each of which was designed for a frequency around 1 GHz and drives a 12 bit counter. While an enable signal (supplied via GLOBAL_PULSE_ROUTE) is set to high, the ring oscillators drive their respective counters, and by reading out the counters, a rather precise frequency measurement is achieved. The frequency of the ring oscillators depends on the TID received, as well as the supply voltage (VDDD) and the temperature. By keeping both, the temperature and the supply voltage constant, e.g. during an irradiation campaign, the received TID can be monitored by regularly reading out the ring oscillators. [33]

3.2 CONNECTIVITY

For communication with the outside world, RD53A features a bottom pad frame containing 198 wire-bond pads. Being a prototype chip, RD53A additionally includes a top pad frame with an additional 97 wire-bond pads for debugging and monitoring of various signals and voltages.

Data is formatted with a custom format and encoded with the Xilinx Aurora protocol, before being sent out by means of four independent 1.28 Gbit/s serializers.

Usually, an SCC is used to communicate with single, bare chips, while assembled modules are typically connected to the DAQ by means of a module flex and the appropriate adapter.

3.2.1 DATA LINK

The four independent 1.28 Gbit/s output serializers and Current-Mode Logic (CML) drivers are located in the chip bottom region. They use the 640 MHz clock generated by the CDR to serialize the Aurora encoded output data, using Double Data Rate (DDR). The serializers' speed can be reduced by a factor of 2 to 8 and every serializer block can be completely disabled, allowing the chip to output data with a variable rate between 160 Mbit/s (1 lane at $\frac{1.28\,\text{Gbit/s}}{8} = 160\,\text{Mbit/s}$) and 5.12 Gbit/s (4 lanes at 1.28 Gbit/s). [33]

The performance of the serializers and CML drivers has been evaluated in depth. [47]

OUTPUT DATA FORMAT

Different data frames containing either hit data, requested register data or idle patterns are time-multiplexed into the outgoing data stream. These Aurora frames are always 64 bit long, while the actual data words or event headers consist of 32 bit. Depending on the amount of hits, an Aurora frame can therefore either contain one or two data words. When there is no data to send out, `RD53A` outputs either idle patterns or Aurora framing words which can be used for alignment as well as receiving PLL and CDR circuits.

If the user requests a register read-back, the register data is inserted into the data stream every N Aurora frames, where N is user-programmable to reserve the required bandwidth for data. Even if no register read-back is requested, the register frame will be sent out. In this case, it is filled automatically with register data from configurable addresses.

An event header frame consists of a 7 bit header, that is used to distinguish the word from normal data words, the trigger ID and trigger tag of the event, followed by the 15 bit Bunch Crossing ID (BCID), denoting the relative timing of the hit with respect to the issued trigger command. Multiple data words can follow an event header, depending on the number of pixels reporting a hit. Every data word contains hit data for a pixel region, with the first 16 bit used for the region address followed by an 4 bit ToT value for each of the four pixels in the region (see Section A.1). [33]

XILINX AURORA 64B/66B ENCODING

After the 64 bit Aurora frames are assembled, two additional bits are prepended. This Aurora sync header can either be 01 to indicate an Aurora data frame or 10, denoting either a user data frame or an Aurora command frame. These two independent output channels enable fixed bandwidth allocation for different frame types and simplify load balancing between multiple activated Aurora lanes. The original 64 bit of the Aurora frame are then scrambled before they are sent to the serializers. By these means, the Aurora cores take care of DC balancing and optional error correction using Cyclic Redundancy Check (CRC) during data transmission. [33, 48]

3.2.2 SINGLE-CHIP CARD

The RD53A Single-Chip Card (SCC) is a 10 cm × 10 cm Printed Circuit Board (PCB) that a single RD53A chip can be glued and wire-bonded to. It allows to access all external functions of the chip, including all four data output lanes, all four HitOr lanes and all other signals available on either the bottom or top pad frame. Additionally, the card provides easy connection to a laboratory power supply by means of common Molex connectors. The powering mode of the chip can be selected via jumpers between Shunt Low-Dropout Regulator (sLDO) mode, pure Low-Dropout Regulator (LDO) mode and direct powering, bypassing the regulators completely. Additionally, the power source for the CDR and PLL circuits can be selected individually between VDDA, VDDD and an externally supplied voltage. Jumpers are also used to set the chip ID to the desired value, as well as for trimming IREF to the nominal 4 µA. More headers and optional LEMO connectors give access to various analog voltages and debug signals.

For routing the four high-speed data lanes to the DAQ, the RD53 collaboration decided to use commercial DisplayPort (DP) connectors and cables, as this standard provides four shielded high-speed lanes, rated for much more than the necessary bandwidth of 1.28 Gbit/s, in a single cable and both cables and connectors are highly available and affordable. The CMD signal is routed from the DAQ to the chip on an auxiliary lane of the primary DP connection, while a secondary DP connector carries all four HitOr outputs of the chip, as well as the connection to the NTC located close to the lower right corner of the chip. [49]

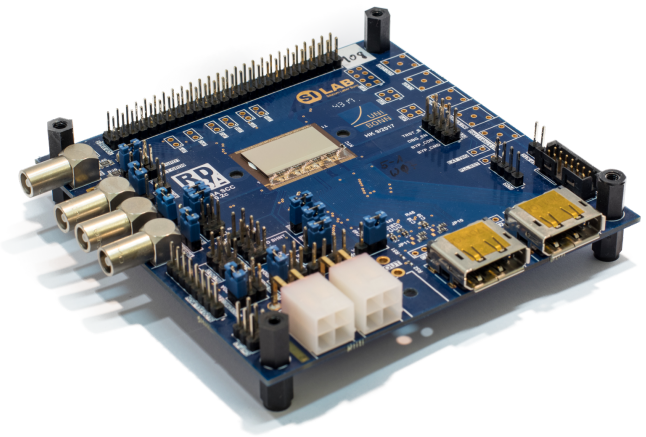


Figure 3.9: An RD53A Single-Chip Module (SCM) wire-bonded to a Single-Chip Card (SCC). The front edge houses two Molex power connectors and two DP connectors for data output and HitOr.

Figure 3.9 shows a typical RD53A SCC with a single-chip module mounted in the foreseen position. The position was chosen to be compatible with the chip position on the FE-I4 SCC for compatibility reasons. The chip is glued onto a massive copper area throughout all layers of the PCB to provide some passive cooling and enable active coolers to be attached from the bottom. For irradiation campaigns, a special "light" version of the SCC was designed, that

contains the lowest possible amount of copper, including a cutout behind the chip. The SCC provides pads for all bottom wire-bond pads of RD53A as well as all top pads. The latter are routed directly to the header on the top edge of the SCC.

RD53B

Based on the testing results of RD53A, the RD53B framework was developed by the RD53 collaboration. This generalized framework facilitates flexible design alterations between the two production chip versions for ATLAS and CMS, while benefiting from a common code and design base, simplifying the verification process significantly.

The production chip for the ATLAS ITk pixel detector is called ITkPix, while the production chip for the CMS experiment was named CROC. At the time of writing this thesis, the first version of the ATLAS chip, ITkPix-V1, had just arrived at the first testing sites. It is the direct successor to RD53A and the first chip based on the RD53B framework.

CHAPTER 4

BDAQ53 - A READOUT SYSTEM FOR RD53 ASICS

BDAQ53 is a readout system and verification framework for the family of pixel detector readout ASICs designed by the RD53 collaboration. Distinctive features of these newly developed chips, such as a new command interface and data format, or the specified data transfer rate of up to 1.28 Gbit/s, require the development of adequate readout systems for testing and characterization. Multiple systems are being developed for this purpose, some of which will eventually evolve into the final detector readout integrated into ATLAS. [50]

BDAQ53 is developed in the Silab at the Physics Institute of the University of Bonn. Development started in 2016 based on the previous generation of readout systems, USBPix[51] and PyBAR [52], and is driven by a core developer team consisting of Tomasz Hemperek, David-Leon Pohl, Michael Daas, Yannick Dieter, Mark Standke and Marco Vogt. Several students, as well as internal and external users have contributed various features and functionality and the open-source nature of the project encourages contributions and improvements.

The design of BDAQ53 is focused on versatility and rapid customization, since intended usage scenarios include tabletop chip- and module characterization measurements, permanent setups e.g. for wafer probing, module testing or serial powering tests, as well as measurements involving inflexible environments like irradiation- or test beam campaigns. This is achieved by using a modular approach to hardware and software, using a commercial Field-Programmable Gate Array (FPGA) board, and employing mostly Python [53] and Verilog [54] as programming languages. BDAQ53 has been used successfully for many measurements, irradiation- and test beam campaigns and is utilized in multiple stationary setups, including the RD53A wafer probing setup described in Chapter 6. The majority of the results presented in this thesis were obtained using BDAQ53.

4.1 HARDWARE

BDAQ53 is based on a Xilinx Kintex7 FPGA [55] that is enclosed in a commercially available daughter board, the Enclustra Mercury+ KX2 [56]. A custom base board for the KX2 was designed to conveniently interface with the FPGA, as well as the DAQ PC and the readout chip.

The custom base board gives access to various debugging signals and is highly customizable via simple firmware modifications. Multiple LEDs are used to inform about link status on all of the used data lanes. The setup is compact and lightweight and uses only commercially available cables and components. Alternatively, BDAQ53 also supports a fully commercial hardware platform, albeit with limited range of functions.

4.1.1 BDAQ53 CUSTOM HARDWARE

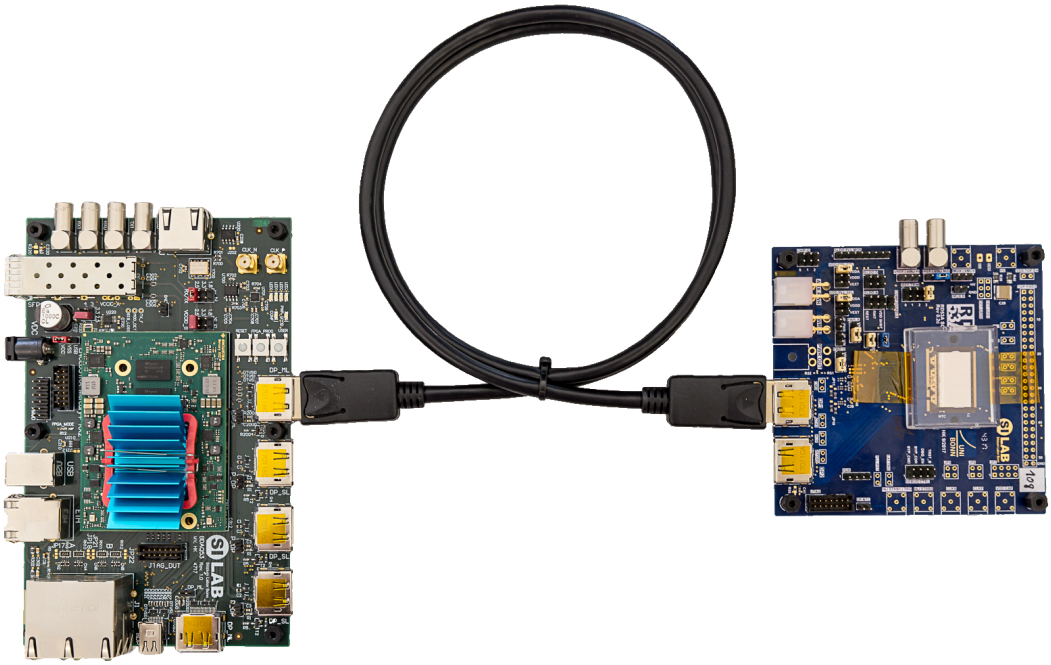


Figure 4.1: BDAQ53 base board (left) with FPGA daughter board, connected to an RD53A SCC (right) via DisplayPort (DP). [50]

A minimal BDAQ53 setup is shown in Figure 4.1. The BDAQ53 board with mounted KX2 is shown on the left. It features five full size DisplayPort (DP) connectors as well as a Mini DP connector to interface with multiple readout chips. The first DP connector offers up to four data lanes, which can be used to read out a single readout chip at full data rate or multiple chips at reduced rate across a single cable, depending on the routing on the SCC or flex. The three additional connectors on the long edge of the PCB connect a single data lane each to the FPGA. All four primary DP connectors additionally carry the CMD signal generated on the FPGA. An additional DP and a Mini DP connector on the short side of the BDAQ53 board can be used to route up to four HitOr lines each to the FPGA for trigger generation or charge measurement (see Section 4.3.3). Additionally, these secondary connectors carry auxiliary signals, for example to read out the NTC on a connected SCC or flex or to communicate with add-on cards via I^2C. A standard RJ45 connector as well as an SFP+ port are used to connect to the DAQ PC via 1 Gbit/s or optionally 10 Gbit/s Ethernet. The FPGA can be programmed with the provided

firmware, using the integrated programmer that is interfaced via a USB connector. The board is powered via this USB port or via an additional power port. Several voltage regulators included in the KX2 board provide different voltage levels to power the FPGA itself as well as all other active components. Multiple other connectors including four RJ45, four LEMO and other connectors can be used to access various signals that can be variably routed via firmware.

The Mercury+ KX2 board was chosen as FPGA board for BDAQ53, because of its long-term availability. It provides the base board with access to eight Multi-Gigabit Transceivers (MGTs) of the Kintex7 FPGA, which are used for the seven available data lanes as well as the 10 Gbit/s Ethernet connection.

In addition to the custom-designed base board, BDAQ53 also supports other hardware platforms based on a Kintex7 FPGA, notably the commercially available Xilinx KC705 development board. Due to hardware limitations, this board by itself only supports a single data lane. By using custom adapters designed within the RD53 community, the FMC connector of this board can be employed to extend the number of supported data lanes. [50]

For the work documented in this thesis, the BDAQ53 base board was mostly used to conduct the measurements shown in later chapters.

4.1.2 STRUCTURE OF FIRMWARE MODULES

The functional firmware and hardware blocks of BDAQ53 are shown in Figure 4.2. The firmware is written in Verilog [54] and follows a modular approach. As depicted in Figure 4.2, the I/O part was kept completely separate from the core firmware. The latter part contains all functional blocks necessary for operating and interfacing with the readout chip, while the former part handles communication with the DAQ PC and contains the PLL. This approach simplifies portability between hardware platforms and also enables interfacing with the firmware completely independent of hardware components, when run in a simulator.

The basil [57] framework is a general and modular system testing and control framework, that provides generic FPGA firmware modules as well as drivers for many different laboratory appliances. It is developed in the Silab Bonn as well. In the context of BDAQ53, basil provides base classes for many frequently used firmware modules like FIFO buffers, I^2C controllers and others. These parts of the firmware are instantiated from basil, while features specific to the RD53 readout chips, like the command encoder are part of BDAQ53. [50]

Based on the physical Ethernet interface between the BDAQ53 board and the DAQ PC, two different protocols are used for communication between BDAQ53 firmware and software. For communication with the basil bus and configuration of the different firmware modules, UDP is used because of its low overhead and latency. Since error correction is advantageous for data words that are being sent from the FPGA to the software, TCP is used here. The firmware tags every Aurora data word from the chip additionally with a custom header as well as the channel ID, in order to distinguish the data of multiple connected readout chips.

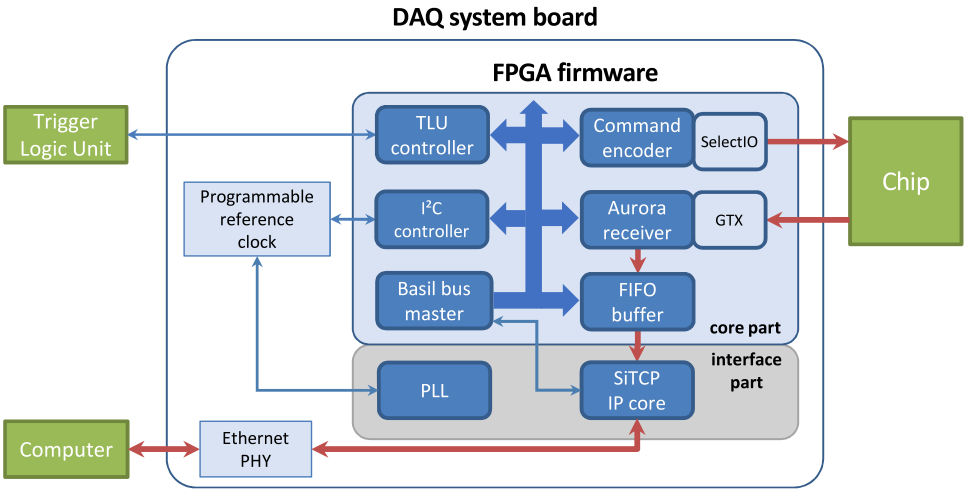


Figure 4.2: Block diagram of BDAQ53 firmware modules [50]. Red arrows indicate data-, blue arrows control flow.

4.2 SOFTWARE

The software framework of BDAQ53 is written mostly in Python [53], with a focus on versatility and rapid customization. As for hardware and firmware, a modular approach was chosen for multiple reasons. First, when using the readout system, only modules necessary for the given application need to be instantiated, which decreases overhead and setup work.

Additionally, a modular approach simplifies maintenance and testing. Fully self-contained modules can be assigned to maintainers and verified by automatic tests. This concept of Continuous Integration (CI) is used throughout the framework, with automatic tests of most of the software, based on both simulation and actual hardware, running automatically with every published change.

BDAQ53 software and firmware are open source and hosted in a public repository[1].

4.2.1 STRUCTURE AND MODULES

Most of the relevant classes of the BDAQ53 software framework, as well as their relations are shown in Figure 4.3. The most important classes include a *Chip* object that is specific to the connected readout chip and is instantiated once for every connected Device Under Test (DUT). It handles all chip specific functions like register writing and reading, configuration, calibration constants, and more. Additionally, there is a *BDAQ53* object that is instantiated once for every readout board in the setup. It handles communication between firmware and software, error handling and all other BDAQ53-specific functions. Finally, there are multiple *Scan* scripts that encapsulate a single measurement or tuning routine. They inherit from a base class called *scan_base* that provides algorithm-agnostic methods, e.g. for data handling. The scan script itself only provides the algorithm for a specific measurement, that is independent of the specific readout hardware or type of DUT in the setup. These scripts utilize other classes for specific

[1] http://gitlab.cern.ch/silab/bdaq53/

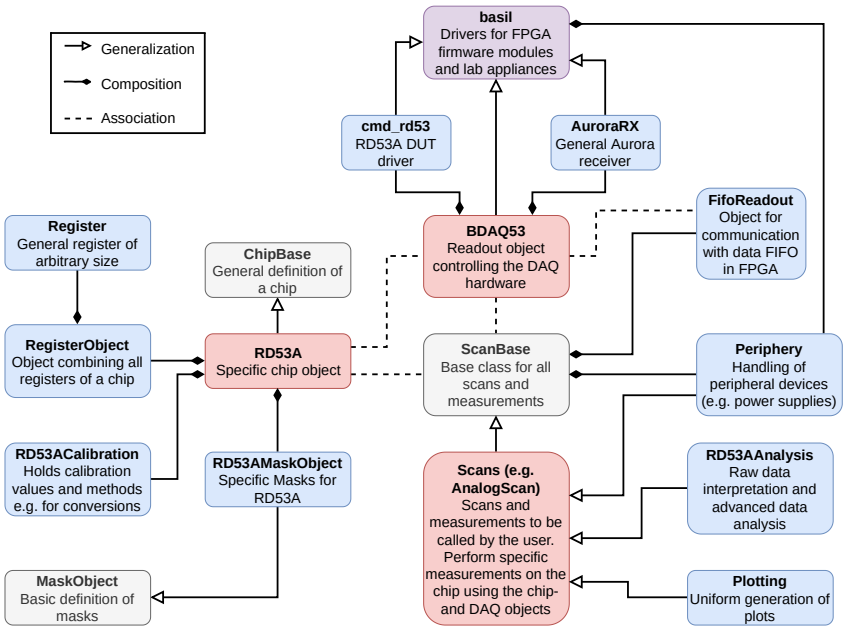


Figure 4.3: Class diagram of the BDAQ53 software framework [50]. Main classes are depicted in red, base classes in white. The basil framework (purple) provides base classes for hardware interfaces.

functions, like an *Analysis* class that interprets the raw data words from the readout chip, a *Plotting* class that contains all functions necessary for data visualization, and others.

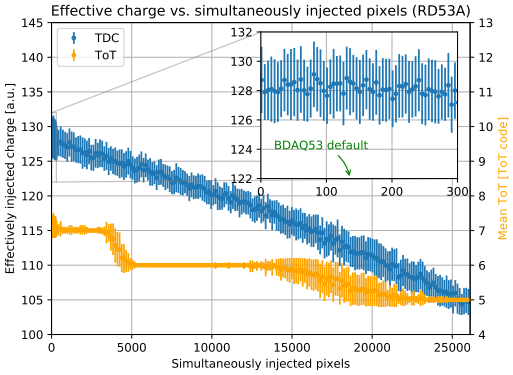
4.3 BASIC FEATURES

In this section, some selected basic features are explained in detail, since they are essential for the measurements shown in later chapters, or because they were developed within the scope of this work.

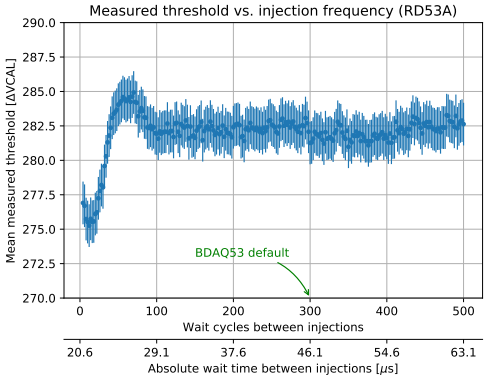
4.3.1 MASK SHIFTING

Readout ASICs of the RD53 family all feature a charge injection circuit (refer to Section 3.1.5 for RD53A) that is used in many scans within BDAQ53 to produce analog hits independent of external particle sources. While parts of this circuit are implemented in every pixel individually, other parts are shared per core, per core column or even globally. It is technically possible to inject a charge into all pixels at the exact same time, but since the injection DACs VCAL_MED and VCAL_HIGH are not capable of driving the load of all pixels at once, the accuracy of the effectively injected charge will suffer. This effect is shown in Figure 4.4(a). The number of simultaneously enabled pixels on the x-axis is plotted versus the effectively injected charge measured with the TDC method on the left y-axis. The charge clearly decreases with the amount of simultaneously injected pixels. The effect on the measured ToT is shown in yellow with respect to the right y-axis. Due to the much lower charge resolution, the effect begins later for ToT, but is still visible. This is a clear sign of the injection circuit not being able to inject the

requested charge into many pixels at once due to overload.



(a) Effects of injection circuit overload, measured on an RD53A chip. An increasing number of pixels is enabled at once (using all columns of the LIN) and a charge corresponding to a ToT of 7 is injected into all enabled pixels simultaneously. If too many pixels are injected simultaneously, the injection circuit cannot supply the requested charge and the effectively injected charge decreases.



(b) Effect of increasing injection frequency, measured on an RD53A chip. The mean measured threshold from a Threshold Scan is shown as a function of the delay time between injections using the charge injection circuit. If the injection frequency is too high, the effectively injected charge decreases, since the injection circuit has not settled yet and a lower mean threshold is measured.

Figure 4.4: Limitations of the RD53A injection circuit.

To avoid this effect when testing the full pixel matrix, only a small subset of pixels must be injected at a time, while shifting over the enabled subsets. This process is called *mask shifting*. The absolute number of pixels the injection circuit can supply simultaneously depends on the absolute requested charge, the location of the specific pixels and other factors. Therefore, the decision was made to only enable one pixel per column by default in BDAQ53, leading to 136 simultaneously enabled pixels and 192 shifting steps necessary to sweep over the full column height of RD53A. Since only one of the three AFE flavors is enabled at a time and all columns of a flavor can be safely enabled simultaneously, scanning a full RD53A chip requires a total of $3 \cdot 192 = 576$ mask shifting steps.

4.3.2 ANALOG INJECTION METHOD

Many scans rely on the calibration injection feature of RD53A and its successors. The implementation of this feature in BDAQ53 basically follows the suggested method [33], but implements some additional functions.

Before any injections can be performed, a *setup* function has to be called, which sets up the appropriate registers: INJECTION_SELECT is set to analog injection with a specifiable fine delay, VCAL_MED and VCAL_HIGH are set to the predefined values to define the amount of charge to inject and GLOBAL_PULSE_ROUTE is set up for sending AZ commands. As a last step, the CAL_EDGE is primed, i.e. set to low, so a rising edge can be generated to perform the actual injection.

As a first step of the actual injection, an AZ pulse is issued a certain timespan before every injection, if the SYNC is enabled. With the default settings for hit latency and injection frequency, this results in an AZ frequency well below 100 µs. As a workaround for an implementation issue in the digital core logic of the SYNC (see Section 5.3.1), an ECR command is issued before every injection and corresponding trigger commands, if the SYNC is enabled. This prevents the hit storage logic of the SYNC from getting stuck.

Independent of the enabled front-end flavor, the CAL_EDGE signal is issued next, producing the actual charge injection. After a configurable latency that depends on the chip's LATENCE_CONF setting, a trigger command with a configurable trigger pattern is issued. By default, this trigger pattern corresponds to 32 consecutive triggers. As a last step, the CAL_EDGE signal is set to low again, priming it for the next injection.

Since the injection circuit is inherently inert, the injection frequency needs to be watched closely. If the injection signal is issued too soon after priming CAL_EDGE, the DAC does not have enough time to settle and the resulting injection voltage might differ from the set value. This effect is shown in Figure 4.4(b). The mean measured threshold from a Threshold Scan (see Section 4.5.4), which is very sensitive to the absolute injected charge, is shown as a function of the number of wait cycles (a special synchronization pattern sent to the chip) inserted in between priming and injection, or the absolute time between injections. The measured threshold first increases slightly when the critical wait time of about 100 wait cycles is reached due to overshoot of the injection DAC. When the injection speed is increased further, the measured threshold decreases significantly. To stay clear of this issue, BDAQ53 by defaults injects 300 wait cycles between priming and injection.

4.3.3 Self-Trigger Mode

When operating a silicon pixel detector module or assembly with actual particle sources like radioactive sources or in a particle accelerator beam, trigger commands need to be generated to initiate the readout of the recorded hits. In the experiment, this is handled by an elaborate triggering scheme, but for laboratory and characterization measurements there are only two options: Either the chip is able to generate triggers on-chip (so called *self-trigger mode*) or the DAQ system needs to generate the trigger commands, usually by means of fast external devices like scintillators close to the DUT. Since RD53A does not implement a native self-trigger mode, an effective and simple triggering scheme was implemented in BDAQ53. This scheme makes use of the chip's HitOr networks to generate triggers within a specialized firmware module in the FPGA, effectively enabling self-triggered operation for RD53A without the need for a complex setup with external detectors for trigger generation. The only additional requirement is a second DP connection routing the chip's HitOr output to the BDAQ53 base board. RD53B will feature an internal self-trigger mode.

4.3.4 TDC Method

The TDC Method [26] is a DAQ-based charge measurement method for silicon pixel detectors and can be used in conjunction with or as an alternative to the chip's internal ToT measurement. It uses the chip's HitOr feature to sample the raw pulses coming from the front-end electronics

with a Time-to-Digital Converter (TDC) implemented in the DAQ's FPGA, using a much higher clock frequency than the 40 MHz BX_CLK used internally by the readout chip. In case of BDAQ53, a 640 MHz clock is used, enabling a 16 times higher resolution of 1.5625 ns instead of 25 ns for ToT. Implementation of this method in BDAQ53 enabled the measurement of high-resolution charge spectra that were used for precise calibration of the chip's injection DACs [46].

4.3.5 MULTI-CHIP READOUT

When working with multi-chip modules (DCMs, Triplets or QCMs), it is convenient to be able to control, configure and read out all involved readout ASICs simultaneously or at least without the need to change any hardware connections. BDAQ53 is well suited for this task: The primary *multi-lane* DP connector of the BDAQ53 base board supports connecting up to four chips in parallel, though the readout bandwidth in this case is limited to a single Aurora lane per chip. Furthermore, handling of multiple chips with individual configuration files and data handling has been implemented into the BDAQ53 software framework [58].

When preparing for QC of large numbers of such modules, however, connecting even more than one full QCM to the readout system is preferable. For this use case, a multiplexer add-on card for the BDAQ53 base board was developed. This card allows one to physically connect up to four QCMs to a single BDAQ53 setup, one of which (four readout ASICs) is electrically connected for readout. CMD signals are routed to all four modules in parallel, enabling parallel operation and configuration of all four modules, while allowing for serial readout and characterization of one module at a time with a single BDAQ53 board and readout PC.

Due to the Ethernet interface between the BDAQ53 hardware and the readout PC and based on the sufficient computing power of modern PCs, multiple BDAQ53 setups can be connected and controlled by a single DAQ PC. This allows for easy and cost effective scaling of the multi-chip readout capabilities of BDAQ53.

4.4 CALIBRATION ROUTINES

Before any meaningful results can be achieved with a readout chip, various values have to be calibrated. Some of these calibrations are conducted once for a limited sample size and the resulting value is used as a default value unless the specific DUT is calibrated. For RD53A, this is the case for the ToT-, ADC-, and temperature sensor calibration. On the other hand, the reference voltages have to be adjusted individually for every chip.

4.4.1 CALIBRATE VREF AND CALIBRATE VREF_ADC

The reference bandgap voltage outputs for both the analog and digital sLDO regulators can be trimmed by means of a register setting (VOLTAGE_TRIM). In order to optimize the analog (VDDA) and digital (VDDD) supply voltages, the Calibrate VREF routine can be used to scan through the different settings for VOLTAGE_TRIM automatically and select the optimal trim bit setting. The settings are scanned by starting from the highest output value and the scan stops before a critical voltage is reached that would disrupt communication.

The same routine is also available for trimming the reference voltage for the Successive Approximation Register (SAR) ADC of the chip, VREF_ADC. Since VREF_ADC is also the reference voltage for the injection DACs VCAL_MED and VCAL_HIGH, it is important to trim this reference voltage to be as close as possible to the nominal voltage.

4.4.2 Calibrate ADC

Apart from the reference voltage VREF_ADC, the on-chip ADC also has direct trim bits to trim the ADC's response to a given input voltage. The Calibrate ADC routine can be used to trim these settings automatically.

4.4.3 Calibrate Temperature Sensors

The temperature is measured with the four on-chip temperature sensors by applying a linear calibration function to the voltage of these sensors, that is measured with the on-chip ADC. The Calibrate Temperature Sensors routine can be used to empirically determine the constants of this calibration function by cooling or heating the whole assembly to a known temperature and using the NTC on the SCC or flex as a refrence.

4.5 Scans

This section describes and explains the scan procedures and algorithms that were used for the characterization measurements of RD53A and modules shown in later chapters.

4.5.1 Register Test

A Register Test is a basic test designed to verify if all registers of the DUT are working as expected. For this purpose, every register that is defined as writable and is not essential for operation is set to an alternating 0b01 pattern of the same length as the register and is then read back. After the return value is checked to be exactly the correct pattern, the process is repeated with an alternating 0b10 pattern, making sure that every bit has been checked for its ability to store either 0 or 1. After all registers are checked, the chip is reset to the default values to make sure that subsequent scans do not start with an undefined initial setting.

Since register defects are rare and working registers are crucial for chip operation, the result of the Register Test is boolean, yielding True if all registers passed the test and False if any of them failed.

4.5.2 Digital Scan

The Digital Scan is designed to test the digital matrix functionality, as well as the global powering, configuration, hit processing and communication of the DUT. Digital injections are induced in every pixel, using the mask shifting algorithm explained in Section 4.3.1. The generation of digital hits is independent of the functionality of the AFE of a given pixel, since the artificial hit information is injected into the circuit in parallel to the AFE's comparator. In

order to gain some statistical relevance, for each mask step, the injection command is issued 100 times by default.

The result is a hitmap that shows exactly 100 hits for all working pixels, less than 100 hits (usually 0) for digitally dead pixels and more than 100 pixels for noisy pixels. Since the AFEs of the pixels are disabled during a `Digital Scan`, noisy pixels in this stage a very rare, as the source of the noise must be independent of the analog circuitry.

Since no additional hit processing is necessary for digital injections, the timing usually is not as critical as for hits involving the full analog signal processing chain. Therefore, the relative BCID histogram will usually show all recorded hits in a single bin. Also, since no ToT variation is possible with digital injections, the ToT code histogram also reports all hits with the same ToT.

4.5.3 Analog Scan

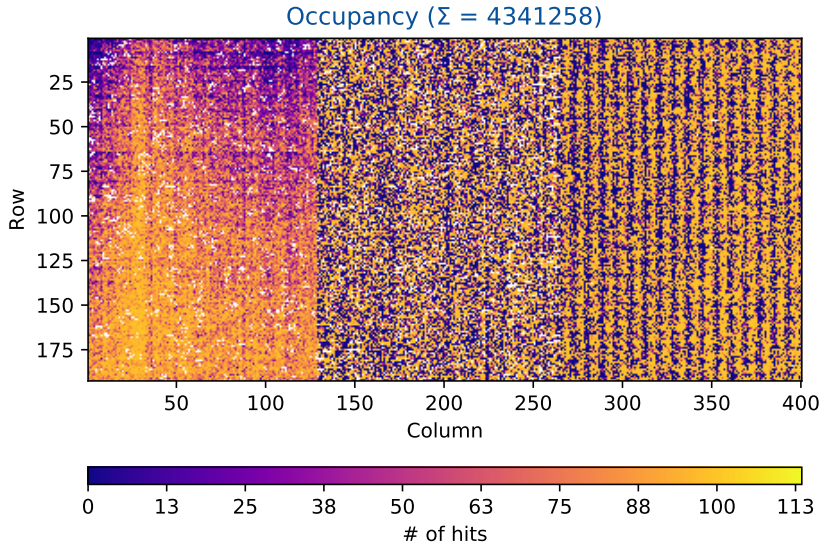


Figure 4.5: An example hitmap from an `Analog Scan`. The injected charge was chosen close to the mean value of the untuned threshold to provoke a suboptimal result. The color scale ranges from pixels with no recorded hits (blue) to noisy pixels with more than the expected 100 hits (yellow). White pixels were disabled completely during the scan.

The `Analog Scan` works the same way as the `Digital Scan`, but uses the calibration injection circuit to produce a hit signal at the input of the AFE stage. The same mask shifting algorithm is used in order to stay within the specified limits of the injection circuit.

The resulting hitmap again shows exactly 100 hits per pixel by default. In case of the `Analog Scan`, noisy pixels showing more than 100 hits, or less efficient pixels showing less than 100 hits are much more common. An explicitly bad example hitmap is shown in Figure 4.5, in order to illustrate most of the possible features. In this case, an `Analog Scan` was performed where the injection charge was chosen close to the mean value of the untuned charge detection threshold of the matrix. This leads to some pixels, shown in blue, which did not recognize

any hits at all. Most pixels recognized between 0 and 100 hits. On the other hand, noisy pixels recognizing more than the 100 injected hits can be identified in bright yellow. The hitmap allows to easily identify features like the threshold gradient of the SYNC visible in the figure (refer to Section 5.3.2) or the inhomogeneity of the DIFF (refer to Section 5.5.1).

When evaluating the timing information in the relative BCID histogram, each AFE should be evaluated separately, since the performance and absolute timing can differ. To make sure that the resulting hit timing is not influenced by the propagation time of the injection pulse, an `Injection Delay Scan` should be performed beforehand. In this case, all hits shown in the histogram should be recognized within about one BCID (25 ns). Taking into account that the absolute hit times are not synchronized precisely enough with the trigger commands, the relative BCID histogram should show the hits in one to two neighboring bins.

If the ToT response was tuned prior to running the `Analog Scan` and the injection charge was set to the tuning target value, the ToT histogram should show a peak at the tuned value (see Section 4.6). This can be used to assess the result of the `ToT Tuning`.

4.5.4 Threshold Scan

In order to find the effective charge detection threshold of any given pixel, as well as the mean threshold and threshold dispersion over an area of pixels, the `Threshold Scan` effectively consists of multiple `Analog Scans` with increasing target charges. The target charge (`VCAL_HIGH` at constant `VCAL_MED`) is varied over a configurable range with configurable step size and all pixels are injected with all selected charges. When plotting the response of each pixel as a function of the injected charge, a so-called S-curve is formed. This curve is defined by Equation 4.1 and shown in Figure 4.6.

$$N_{\text{hits}}(Q_{\text{inj}}) = \frac{A}{2} \cdot \texttt{erf}\left(\frac{Q_{\text{inj}} - \mu}{\sigma \cdot \sqrt{2}}\right) + \frac{A}{2} \qquad \text{Equation 4.1}$$

When sweeping the injected charge over the effective charge collection threshold, an ideal pixel without any noise would respond with an ideal step function, not recognizing any hits as long as the charge is below threshold and immediately registering all hits once the threshold is crossed. Due to electronic and other noise sources in the real system, the actual response curve is given by the Gaussian error function `erf` shown in Equation 4.1, where the threshold μ is defined as the charge for which 50 % of the injected hits are recognized and the noise σ is given by the slope of the response curve at this point.

By fitting this function to the response data for every pixel individually, a threshold- and noise value for each pixel is obtained. These values are then histogrammed and the mean value for both can be found by calculating the mean of all individual values. For redundancy and robustness against single non-converging fits, the mean threshold and noise values can also be found by fitting a Gaussian distribution to the resulting histograms.

4.5.5 Injection Delay Scan

Since the analog charge injection signal is created at one point in the ACB and has to be propagated over the physical size of the chip, the precise timing of the charge injection will

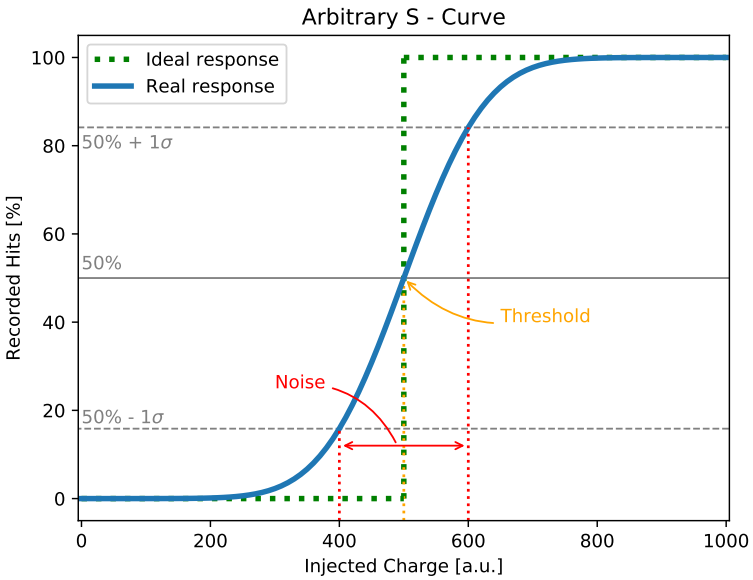


Figure 4.6: An arbitrary S-curve as defined by Equation 4.1. The response function of an ideal pixel without any noise is given by the dotted green curve, while a real pixel will show the solid blue response curve. The mean threshold is defined as the charge, for which 50 % of the injected hits are recognized. the mean noise can be reconstructed from the crossing of the measured S-curve with the 50 % $\pm$ 1σ lines.

be dependent on the pixel's position in the matrix. To minimize this position dependence for timing-critical investigations, a fine delay can be applied to the injection command in steps of 1.563 ns (using the 640 MHz `DEL_CLK`).

To find the optimal fine delay value for each pixel, the `Injection Delay Scan` injects a defined charge at different fine delay settings. In the hit data, it then looks for the transitions between bins in the relative BCID distribution and saves the optimal fine delay setting to move the effective injection time to the center of the desired bin.

The resulting injection delay map is used as an input for subsequent, timing-critical scans like the `In-Time Threshold Scan`.

4.5.6 In-Time Threshold Scan

In testing and characterization measurements, `BDAQ53` reads out 32 consecutive BCIDs for every charge injection command by default. This is done to improve robustness, as it increases the probability to record a hit, even if the relative delay settings of the charge injection, hit generation and the readout system do not match exactly.

In actual LHC operation however, only a single trigger command per bunch crossing in the accelerator is generated. Therefore, hits with a delay outside the accepted range, both below and above, will not be recorded. To represent this behavior during characterization and to assess the timing performance, an in-time threshold is defined in addition to the hardware threshold as the measured threshold when only hits within a certain delay range are accepted.

Typically, the accepted range is one BCID bin wide and its absolute position is determined by the characteristics of the AFE and various delay settings within different parts of the chip. Due to timewalk (see Section 2.1.4), hits with lower charge are typically recorded later than hits with higher charge. For this reason, the in-time threshold is usually higher than the hardware threshold. The difference between the in-time- and the hardware threshold is called overdrive and is a characteristic feature of the AFE.

The `In-Time Threshold Scan` uses the same basic algorithm as the normal `Threshold Scan`, but applies the injection fine delay map that needs to be generated beforehand by means of a `Injection Delay Scan` to the injection command. This ensures that only the intrinsic timing performance of the signal processing chain is investigated and any influence of the charge injection circuit is eliminated. In the analysis step, the hardware threshold is calculated by taking all hits into account. In a second step, all out-of-time hits are rejected and the in-time threshold is measured. The scan then outputs the measured absolute in-time threshold as well as the calculated overdrive.

4.5.7 Noise Occupancy Scan

Noise occupancy is a measure for the amount of hits generated from noise per issued readout trigger. If a typical noise occupancy of 10^{-6} hits/trigger is required, any pixel that responds at least once within a million issued trigger commands without any real signal due to either an actual particle in a connected sensor or charge injection is considered *noisy*.

The `Noise Occupancy Scan` is designed to identify noisy pixels for two main reasons: First, the amount of noisy pixels at a given tuned threshold is an important measure for AFE characterization and needs to be measured. Secondly, for operating a pixel detector module in a test beam or with radioactive sources, noisy pixels should be disabled in order to avoid flooding the readout chain and buffers with noise hits.

Both is accomplished by the `Noise Occupancy Scan`, by sending a user-defined number of triggers without any charge injection and recording all hits originating from these triggers. Another setting called *minimum occupancy* is defined to set up a threshold for pixels to be classified as noisy. It is recommended to send more triggers than the desired noise occupancy limit. For example, by default the `Noise Occupancy Scan` is set up for a target noise occupancy of 10^{-6} hits/trigger by sending 10^7 triggers and classifying only pixels with more than 10 hits as noisy. This is done to improve the statistical significance of the results, since pixels with a noise occupancy lower than the desired limit could by chance also generate a hit during the scan period and would be classified as noisy even though their actual noise occupancy is not as high.

The scan outputs the number of pixels that were classified as noisy with the given settings for the number of triggers and minimum occupancy, as well as a disable map that disables all noisy pixels for subsequent scans.

4.5.8 StucksPixel Scan

A pixel is defined as *stuck* if the output signal of its comparator is always high independent of the input signal of the comparator. This behavior can appear due to very high noise levels

in the AFE of the pixel or as a result of misconfiguration. Since only a rising or falling edge in the comparator signal can be recognized as a hit by the digital logic, these pixels simply appear as dead pixels not recognizing any of the expected hits in the scan output data. The number of stuck pixels under most conditions is negligible and therefore often not quantified. When running a *self-triggered* source scan (see Section 4.3.3), however, stuck pixels present a serious problem. Since in self-trigger mode triggers for reading out the matrix are generated based on the logical OR of the analog output signals of all pixels on the HitOr network, a single stuck pixel will render the whole HitOr network inoperable. The logical OR output will simply always be high and no triggers can be generated for the whole part of the matrix connected to the particular HitOr network.

To enable self-triggered operation, stuck pixels need to be identified and disabled beforehand. The `Stuck Pixel Scan` takes advantage of the chip's implementation of the calibration injection circuit. By setting `CAL_EDGE` to high and acting on `INJECTION_SELECT` to switch from digital to analog injection, pixels with stuck high comparator output will produce a rising edge on their hit output and a hit can be recognized. By repeating this step several times and shifting over the subset of enabled pixels, stuck pixels can be identified and disabled. Similarly to the `Noise Occupancy Scan`, the `Stuck Pixel Scan` outputs the number of identified stuck pixels as well as a modified disable map.

4.5.9 CROSSTALK SCAN

Crosstalk can be a feature of both, the readout ASIC and the sensor. It describes the phenomenon that a signal in one pixel is also recognized in one or several of its neighbors, degrading vertex resolution and accuracy of the charge measurement. This behavior can have multiple causes, with the most significant source being capacitive coupling between the pixel implantations in the sensor or via the substrate in the readout chip. This effect is not to be mistaken with *charge sharing*, which is caused by diffusion of the charge cloud in the sensor [22]. Since crosstalk can be strongly influenced by the implementation of the readout chips Analog Front-End (AFE), it is an important metric for front-end chip characterization.

During the `Crosstalk Scan`, a large charge is injected into selectable patterns of neighboring pixels around an investigated pixel. By sweeping over the total injected charge, a crosstalk threshold is determined, which in turn enables the calculation of the amount of crosstalk relative to the absolute threshold of the readout pixel. Typically, the entire dynamic range of the injection DACs `VCAL_MED` and `VCAL_HIGH` is used and patterns with multiple neighboring pixels need to be selected to achieve a crosstalk charge above the (tuned) threshold of the investigated pixel. Different injection patterns include a *ring* pattern, injecting in all surrounding pixels to maximize the measured charge, a *cross* pattern that injects charge into all direct neighbors of the investigated pixels and different combinations of only *left*, *right*, *top* or *bottom* neighbors, to investigate the direction, column or row dependency of the crosstalk. [46]

The `Crosstalk Scan` can be used to determine the relative crosstalk of assemblies, but the absolute charge created by crosstalk in a bare readout chip without sensor bump-bonded to it is usually not large enough to be recognized in the investigated pixel. Comparing the crosstalk levels of different AFE implementations is still possible, however, by using an assembly with a sensor that implements the same sensor pixel flavor for all AFE flavors. In this case, any

differences in the crosstalk performance must originate from the implementation of the AFE.

4.5.10 Bump–Connectivity Scans

When assessing the quality of the pixel detector module hybridization process (see Chapter 8), using charge generated in the sensor by a radioactive source is the obvious choice. However, this method comes with two significant drawbacks: First, gathering a significant amount of hits in all pixels of a module can take a long time, depending on the activity and collimation of the utilized source. Secondly, working with radioactive sources complicates the setup, since appropriate shielding and handling of the source are necessary. Furthermore, since bump-bond connectivity is often investigated as a function of thermal stress (see Chapter 8), using radioactive sources in combination with a climate chamber can make the setup even more inflexible.

For these reasons, two different algorithms have been developed in the scope of `BDAQ53` to assess the bump-bond quality without the need of radioactive sources: a *Crosstalk*-based algorithm and an algorithm based on *Threshold and Noise shift* between different sensor working points. [58]

Crosstalk-based Algorithm

Pixels with disconnected bumps can be identified by utilizing the `Crosstalk Scan` described in Section 4.5.9. By using the *ring* injection pattern, i.e. injecting charge in all neighbors of the investigated pixel at the same time, the absolute difference in crosstalk levels between connected and disconnected pixels is large enough to facilitate a classification of the bump-bond quality. While this algorithm works rather fast, especially compared to running a full source scan, its accuracy is limited by the absolute level of crosstalk in the given sensor, which is designed to be as low as possible, as well as the AFE's sensitivity to crosstalk, which is designed to be as low as possible, too. Due to the dependency of the charge generation on the connectivity of the neighboring pixels, the crosstalk-based bump connectivity scan cannot definitely identify the bump connectivity of a pixel that is surrounded by disconnected pixels. These pixels have to be classified as *unknown*. Hence, this algorithm is not suited to positively identify larger clusters of disconnected pixels, which historically has been shown to be a common failure mode in large bump-bonded structures.

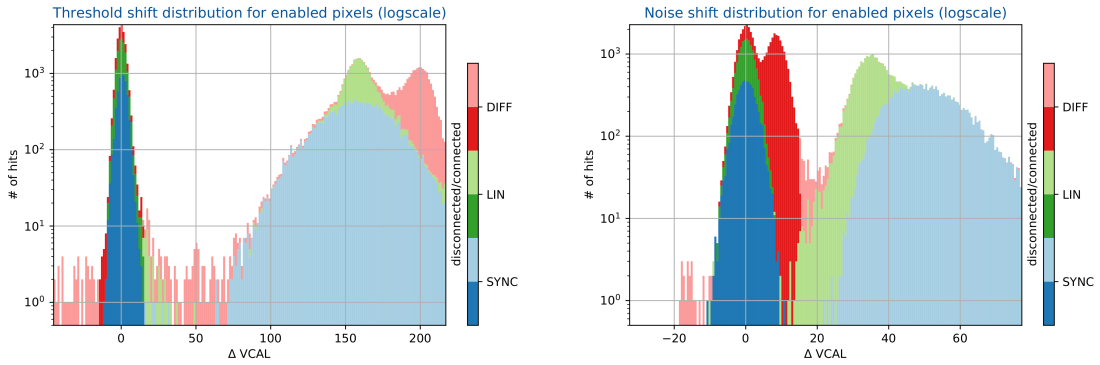
The scan automatically analyzes the resulting occupancy map with regard to bump connectivity and creates a bump connectivity map, classifying each pixel as either *connected*, *disconnected* or *unknown*.

Using `RD53A`-based single-chip assemblies, this algorithm is expected to overestimate the amount of disconnected pixels by up to 6 %, even after optimizing multiple chip parameters for high sensitivity to crosstalk and tuning each AFE flavor the the lowest possible threshold individually. [58]

Noise / Threshold Shift Algorithm

Operating the sensor of a hybrid pixel detector module at different bias voltages can have a significant influence on the measured charge collection threshold and noise. Especially when

applying a small forward bias voltage of around +1 V, threshold and noise levels that are determined by running a `Threshold Scan` are significantly higher compared to running the sensor at normal reverse bias voltage levels. Naturally, this effect can only be observed in pixels with good bump-bonds, since it originates in the sensor.



(a) Threshold shift between a reverse bias voltage of −40 V and a forward bias voltage of +1 V.

(b) Noise shift between a reverse bias voltage of −40 V and a forward bias voltage of +1 V.

Figure 4.7: Result of a threshold/noise shift based `Bump Connectivity Scan` for a special `RD53A`-based single-chip assembly with a significant amount of disconnected bumps. The shift in threshold and noise is shown for each AFE flavor individually (SYNC in blue, LIN in green and DIFF in red). Disconnected pixels are depicted with bright color, while connected pixels are indicated by faint color. The classification of pixels is based on source scan data.

The noise/threshold shift based bump connectivity scan therefore consists of two `Threshold Scans` at a typical bias voltage and at a small forward bias. Connected pixels will typically exhibit a significantly different threshold and noise in the two scans. The result of this method is shown in Figure 4.7. By applying a predefined cut to the measured distributions of both threshold and noise shifts between the two working points, every pixel is classified as *connected* or *disconnected*. The scan automatically creates a bump-connectivity map based on this classification.

Due to the continuous distribution of shift values, especially in the DIFF, the amount of falsely classified pixels will depend slightly on the defined cut value. However, by using the same `RD53A`-based single-chip assemblies as for the crosstalk-based algorithm, the noise/threshold shift based bump-connectivity scan was shown to be slightly more accurate. It overestimates the amount of disconnected pixels by less than 3.5 % [58]. Additionally, since this algorithm works for each pixel individually, larger clusters of disconnected pixels can be identified with the same accuracy as solitary disconnected pixels.

MERGED-BUMP SCAN

While the `Disconnected Bump Scans` described previously concentrate on identifying pixels with no connection between readout chip and sensor, the `Merged Bump Scan` was implemented to identify the other common failure mode of hybridization processes: Merged bumps are bump-bonds that connect more than one sensor pixel to one pixel of the readout chip. This

can be a result of multiple bump-bonds melting together during the hybridization process (see Section 8.1).

The `Merged Bump Scan` works similarly to the crosstalk-based `Disconnected Bump Scan` but only injects a small charge just above the tuned threshold into a single neighboring pixel. This way, if the investigated pixels recognizes a hit, it must have a low-ohmic connection to its neighbor and both pixels are categorized as *merged*.

4.6 TUNING ALGORITHMS

Chip *tuning* refers to algorithms for optimizing specific characteristics of a readout ASIC by testing out different settings for the appropriate registers. The three most commonly tuned chip characteristics are the global and local threshold and the ToT response.

Since a complete tuning of a readout chip typically involves multiple different routines, so called *meta-scripts* are available that combine the different script calls necessary for a complete tuning. For example, the `Threshold and ToT Meta Tuning` routine first executes a `Global Threshold Tuning`, then runs multiple iterations of `ToT Tuning` and `Local Threshold Tuning` (if applicable) and finally includes the necessary `Stuck Pixel Scan` and `Noise Occupancy Scan`. As a last, optional step, the meta script also calls the `Threshold Scan` to visualize the results of the performed tuning.

4.6.1 ToT Tuning

`ToT Tuning` means optimizing the ToT response to return a given ToT code for a defined injection charge. A typical target response is for example ToT code 7 for a charge of 10 ke$^-$. The ToT response is adjusted by changing the shaping amplifier's feedback current. A higher feedback current leads to a faster return-to-baseline and therefore to a lower ToT, while a lower feedback current generates a slower return-to-baseline and therefore higher ToT values (see Section 2.1.4). The name and size of the register controlling the feedback current is specific for the AFE and is predefined in the `ToT Tuning` script. For `RD53A`, these registers are `IBIAS_KRUM_SYNC` for SYNC, `KRUM_CURR_LIN` for LIN and `VFF_DIFF` for DIFF.

The `ToT Tuning` injects the target charge into every pixel and measures the mean ToT response. The corresponding feedback current register is then scanned using Binary Search until the optimal setting is found. In the case of `RD53A`, this process is repeated for each AFE individually.

The output is a chip configuration file containing all the settings that were used during the scan, including the optimized settings for the feedback current registers. To assess the success of the `ToT Tuning`, a simple `Analog Scan` with the injection charge set to the tuning target can be performed. The resulting ToT histogram should show a significant peak at the target ToT code.

4.6.2 Global Threshold Tuning

Typically, in silicon pixel detectors each pixel's charge detection threshold is defined by a global reference voltage DAC called Global Threshold DAC (GDAC), and a pixel-specific local offset,

called Local Threshold DAC (TDAC). The `Global Threshold Tuning` tries to optimize the Global Threshold DAC (GDAC), so that the mean global threshold of the matrix is as close to the desired target value as possible. To achieve this, the algorithm again uses Binary Search to scan over the relevant GDAC register settings while injecting the defined target charge.

Since the GDAC is again implemented per AFE in `RD53A`, the `Global Threshold Tuning` optimizes all three AFEs independently, by tuning `VTH_SYNC` for SYNC, `Vthreshold_LIN` for LIN and `VTH1_DIFF` for DIFF[2].

The result of a `Global Threshold Tuning` is again a chip configuration file containing all chip settings used during the tuning procedure, as well as the optimized GDAC register settings.

4.6.3 Local Threshold Tuning

Tuning a chip's local threshold means optimizing every pixel's Local Threshold DAC (TDAC) individually to minimize the threshold dispersion and achieve a uniform response to incident particles with a given charge. This is once again implemented by using a modified Binary Search algorithm with predefined step sizes on each pixel individually, while injecting the target charge multiple times in each iteration. The algorithm has been optimized empirically based on `RD53A` and for each AFE flavor individually, defining unique start values and step sizes for the LIN and DIFF. As explained in Section 3.1.2, in `RD53A` manual TDAC tuning is only necessary for the LIN and DIFF, while the SYNC does not have TDACs but implements an automatic threshold adjustment mechanism.

The result of a `Local Threshold Tuning` is a map of TDAC values for every single pixel, that is saved together with all other masks that were applied during the tuning in a *masks* output file, which can be loaded in subsequent scans.

The pixels' response to different TDAC settings can be influenced by other registers as well, including the feedback current setting used to tune the ToT response during `ToT Tuning`. On the other hand, the TDAC setting also influences the pixels' ToT response slightly. Therefore, the tuning of both the ToT response and charge collection threshold of a pixel matrix needs to be an iterative process.

Since a suboptimal `Local Threshold Tuning` can have a negative influence on the assessment of the performance of an AFE flavor, for example by not reaching the AFE's lowest possible threshold or overestimating the threshold dispersion or noise, it is important for an unbiased front-end characterization that the `Local Threshold Tuning` uses the given AFE flavor to its full potential. To ensure this, multiple different tuning algorithms have been evaluated, most of which converged to the performance levels presented in Chapter 5. This leads to the assumption that the shown performance is indeed dominated by the performance of the chip and not a feature of the algorithm.

Furthermore, using `RD53A` as an example, the results of the presented `Local Threshold Tuning` algorithm match the performance figures achieved by using a completely independent

[2] Even though the DIFF implements a differential Global Threshold DAC (GDAC) consisting of `VTH1_DIFF` and `VTH2_DIFF` (see Section 3.1.2), best results can be achieved for low absolute thresholds when keeping `VTH2_DIFF` constant at 50 while changing `VTH1_DIFF` according to the target threshold.

readout system and tuning algorithm quite well [44]. This leads again to the conclusion, that these figures in fact represent the AFE's performance and the `Local Threshold Tuning` is well suited for usage in readout ASIC characterization.

4.6.4 Noise Tuning

`Noise Tuning` or *Threshold Baseline Tuning* [59] is another tuning algorithm for silicon pixel detectors that was implemented in `BDAQ53` for use with RD53 readout chips. This algorithm uses a predefined maximum noise occupancy (as discussed in Section 4.5.7) and tunes all pixels to the lowest possible threshold, while not exceeding the noise occupancy limit. This is achieved by first successively lowering the Global Threshold DAC (GDAC) until a certain amount of noise hits is recognized. After the lowest possible GDAC value is found, the Local Threshold DAC (TDAC) of all pixels is optimized by first setting it to the highest possible local threshold setting and in each iteration decreasing the TDACs of pixels not yet exceeding the noise occupancy limit, thus decreasing their local threshold and potentially increasing their noise occupancy. This iterative process is repeated until all pixels operate as close to the noise occupancy limit as possible. A defined amount of the noisiest pixels of the matrix can be automatically disabled, as during a `Noise Occupancy Scan`, to further decrease the mean operating threshold. Once the number of disabled pixels reaches a user-defined limit, the algorithm stops and the chip is tuned to the lowest possible threshold for the given noise occupancy limit.

Since the `Noise Tuning` optimizes both the GDAC and the TDACs, the result consists of both a configuration output file containing the optimal GDAC setting for each AFE and a *masks* output file containing a map of the optimized TDAC settings.

4.7 Measurement Routines

Measurement in this context refers to a script that uses the `BDAQ53` framework and infrastructure to perform measurements with a defined scope on a DUT. These scripts are typically used during test beam- or irradiation campaigns and allow for e.g. operating `BDAQ53` together with additional devices like a beam telescope [60].

4.7.1 External Trigger Scan / Source Scan

The `External Trigger Scan` configures the setup to accept triggers from an external source, e.g. an EUDET Trigger Logic Unit (TLU) [61]. No injection- or trigger commands are generated in the `BDAQ53` software. All data coming from the connected DUTs is saved and analyzed after a user-defined stop criterion is met or the scan is aborted manually.

A special version of the `External Trigger Scan` is the `Source Scan`. This script enables the effective self-trigger operation described in Section 4.3. A special firmware module is enabled, that generates triggers in real time, based on the HitOr output of the DUT. This feature is used extensively to enable simple setups, for example to collect hits originating from a radioactive source without the need for additional detectors for trigger generation.

4.8 SUMMARY OF BDAQ53 SCANS

Table 4.1 summarizes all BDAQ53 scans that are relevant for the results shown in this thesis.

Calibration routines	
Calibrate VREF(_ADC)	Calibrates the references for VDDD, VDDA or VREF_ADC.
Calibrate ADC	Calibrates the ADC response by means of the ADC trimbits.
Calibrate Temperature Sensors	Calibrates the on-chip temperature sensors with respect to a reference NTC in a climate chamber.
Scans	
Register Test	Tests every register of the DUT for its ability to store arbitrary values.
Digital Scan	Injects artificial, digital hits into every pixel to test functionality of the digital signal processing.
Analog Scan	Injects artificial, analog hits into every pixel to test functionality of the full signal processing chain.
Threshold Scan	Scans over different artificial injection charges to determine each pixel's effective charge detection threshold.
Injection Delay Scan	Scans the delay between injection of artificial charges and recognition of hits for each pixel, to apply the appropriate fine delay.
In-Time Thr. Scan	Measures the total charge detection threshold as well as the effective threshold for in-time hits and calculates the overdrive.
Noise Occupancy Scan	Measures noise hits and disables noisy pixels above the limit.
Stuck-Pixel Scan	Finds and disables pixels whose signal processing chain is stuck in one state and not reacting to new hits anymore.
Crosstalk Scan	Measures the degree of crosstalk between neighboring pixels based on artificial charge injection.
Bump-Conn. Scan (Crosstalk)	Uses differences in crosstalk to identify poorly connected bump-bonds.
Bump-Conn. Scan (Thr. Shift)	Identifies poorly connected bumps based on the shift of threshold and noise with reverse- and forward bias of the sensor.
Merged Bump Scan	Identifies merged bumps based on their increased crosstalk.
Tuning routines	
ToT Tuning	Tunes the mean ToT response of the matrix to a specific value at a defined injection charge.
Global Thr. Tuning	Tunes the mean charge detection threshold of the matrix to a given value.
Local Thr. Tuning	Tunes the TDAC of each pixel to a given threshold.
Noise Tuning	Tunes both the global and local thresholds of all pixels to the lowest possible value based on a given noise occupancy.
Measurement routines	
External Trigger/ Source Scan	Allows BDAQ53 to acquire hits with RD53A based on an arbitrary trigger source, e.g. effective self-triggered operation.

Table 4.1: Summary of BDAQ53 routines.

CHAPTER 5

CHARACTERIZATION OF RD53A

The performance of a bare readout ASIC comprises both the digital and the analog performance. While digital performance is usually rated as "all-or-nothing", meaning hit storage, processing and transmission will either work within specification or not at all, the analog signal processing chain has to be carefully verified to fulfill the requirements. This is also where the three flavors of RD53A differentiate themselves and an in-depth analysis is required.

To complicate matters even further, different requirements can be defined to reflect the different regions in the detector, the prospective detector modules will be installed in. In the *Inner Layer*, the requirements regarding minimum stable in-time threshold and return-to-baseline time of the CSA are harder to reach, but higher current consumption per pixel is permitted due to the fact that not too many modules will be operated at these settings. For the *Outer Layer*, the maximum total current consumption is the limiting factor, while lower demands are posed on the minimum stable in-time threshold and return-to-baseline time of the CSA. This can be further motivated by the varying requirements in the different regions of the detector. For example, operation at extremely low thresholds is especially important in the innermost layers, where the size of the sensor's depletion region and therefore the absolute signal strength suffers most severely due to radiation damage over the lifetime of the detector. The same argument holds true for the regions at the far ends of the barrel parts of the detector, where extreme particle incident angles are expected. This effect decreases the effective charge independently of the size of the depletion region. The pixels' current consumption limits on the other hand are governed by the available cooling power in the different detector regions. For the inner layers, special effort is made to provide the necessary power- and cooling budgets for enabling the pixels' operation at the required performance. Due to the increasing distances and multiplicity of modules in the outer layers, the inevitably decreasing cooling performance demands for lower power budgets, while at the same time, the performance requirements, e.g. for in-time threshold and return-to-baseline, can be relaxed here.

The requirements used as a basis for the AFE review conducted by the RD53 collaboration in December 2018 are shown in Table 5.1.

Detailed guidelines on how to tune the AFEs to the desired performance and power consumption have been worked out by the respective IC designers and are summarized in Appendix B. The conclusions drawn from this review as well as the independent reviews held by the ATLAS

	Min. stable thr. w/o sensor	Min. stable thr. with sensor	Current consumption per pixel	CSA ret. to basel. @ 6 ke$^-$
Inner Layer	1 000 e$^-$	1 200 e$^-$	(5.5 µA), 5.0 µA or 4.5 µA	133 ns ≈ 5.3 ToT
Outer Layer	1 300 e$^-$	1 500 e$^-$	(4.0 µA), 3.5 µA or 3.0 µA	225 ns ≈ 10 ToT

Table 5.1: Requirements for RD53A AFE operation for the AFE review conducted in December 2018. The requirements are divided into two categories, *Inner Layer* and *Outer Layer*, depending on the future location of the prospective module in the ITk.

and CMS experiments are discussed in Section 5.7. The results presented in this chapter have been obtained with settings optimized for the *Inner Layer* specification, with a per-pixel current consumption of about 5 µA. This is typically the most demanding situation for the readout ASIC, since performance and power consumption do not scale linearly.

A typical characterization procedure for a pixel detector readout ASIC consists of the following steps:

1. **First start-up**: The DUT is tested and powered on for the first time. Power consumtion, communication and general functionality are tested.
2. **General analog characterization**: All voltage- and current references are calibrated and verified to work within the expected ranges. The injection circuit is tested and calibrated.
3. **Verification of full functionality**: The chip's general ability to detect, store and read out hits is tested.
4. **In-depth analog characterization of the AFE**: The characteristics and performance of the AFE electronics are tested and measured. In the case of RD53A, this step has to be conducted independently for all three AFE flavors. Important characteristics are the threshold dispersion and noise after tuning to typical operating thresholds, the minimum stable threshold, the overdrive and the ToT response.

The following sections show the results and details of these steps for RD53A, before summing up the findings and recapitulating the Analog Front-End reviews held by the RD53, ATLAS and CMS collaborations, respectively.

5.1 First Start-Up and Analog Characterization

When the RD53A chip arrived at the first testing sites in December 2017, it could be operated within a couple of hours. The chip was first tested in Direct Powering mode, supplying the nominal 1.2 V to the VDDA and VDDD rails directly. First tests in LDO mode showed that a larger than expected percentage of chips exhibited problems to establish communication. After a detailed investigation of the problem, it turned out to be caused by an unfortunate combination of minor features. First, the output voltage of the regulator reference Band Gap is slightly lower than expected from simulations. This leads to lower than nominal core voltages VDDA and VDDD at the default trim setting. This effect is investigated in detail in Section 6.4.

Additionally, the chip's CDR and PLL circuits, which are internally powered from the VDDA rail, work less reliably in the low supply voltage corner than expected from simulations; they need a minimum supply voltage of about 1.15 V to work properly [39].

Depending on process variations, the combination of these two effects can lead to a condition where the chip's core voltages at start-up and default setting of the regulator trim bits are too low for the CDR and PLL circuits to work and communication cannot be established. Unfortunately, the regulator voltage trim bits are implemented as an on-chip register, VOLTAGE_TRIM, and general communication with the chip needs to be established before the core voltages can be increased by acting on VOLTAGE_TRIM.

A simple workaround for this issue was found by adding a resistor from VIN_A to VDDA, in order to increase VDDA by a few 10 to 100 mV, depending on the resistor's value. This could easily be implemented on the existing Single-Chip Card (SCC) design and the option was added to the layouts for the various flexes that were developed for RD53A-based modules.

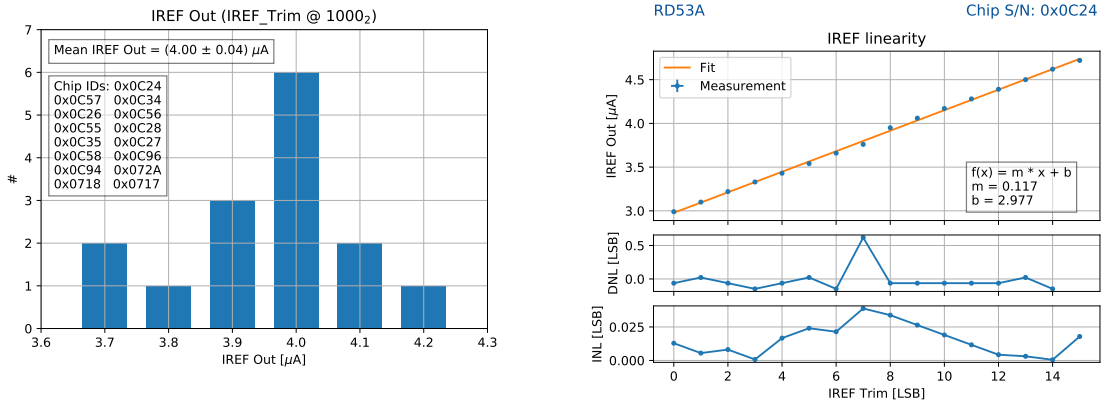
General characterization of the chip design includes confirmation of the linearity of various DACs and trim settings. The most important ones for further characterization measurements are the global reference current IREF that most bias DACs are based on, and the chip's differential injection DAC that is heavily used by most of the scan algorithms described in Chapter 4.

5.1.1 Global Reference Current

IREF is RD53A's global master reference current. The output of all of the chip's bias DACs is scaled with regard to this reference. Therefore, it is important for chip operation and especially for comparison measurements, that IREF is as close to the nominal value of 4 µA as possible. To ensure this, four trim bits are available to adjust IREF. This is done by wire-bonding these four IREF_TRIM bond pads either to VDDA for a high value or ground for a low value, effectively adjusting IREF in 16 discrete steps. To simplify testing and trimming in laboratory environments, the SCC design foresees to trim IREF manually using simple jumpers. All IREF_TRIM wire-bonds are always bonded and the signals are routed to a header and pulled to ground by means of a pull-down resistor.

Two measurements have been performed to verify the nominal value of IREF and the functionality of the trimming mechanism. First, the nominal reference current was measured for a small sample of chips with IREF_TRIM set to the default value of 0b1000. The result is shown in Figure 5.1(a). The mean output reference current is almost exactly 4.00 µA and the variance of 40 nA is rather low. This measurement was later repeated with a much larger statistic sample size based on wafer probing data. The results are shown in Section 6.4.

Secondly, the performance of IREF_TRIM was investigated by measuring the produced IREF current for any trim bit setting. An example measurement for one chip is shown in Figure 5.1(b). The graph shows a good linearity, allowing to trim IREF from 3 µA to about 4.8 µA. This range provides comfortable margins to both sides of the nominal value of 4 µA. The step size is determined by the linear fit to be about 117 $\frac{\text{nA}}{\text{LSB}}$, which allows for trimming IREF with sufficient accuracy.



(a) Variance of IREF output at default trim bit setting for multiple chips.

(b) Linearity of IREF trim bits measured on an RD53A chip.

Figure 5.1: Analysis of IREF.

5.1.2 Injection Circuit

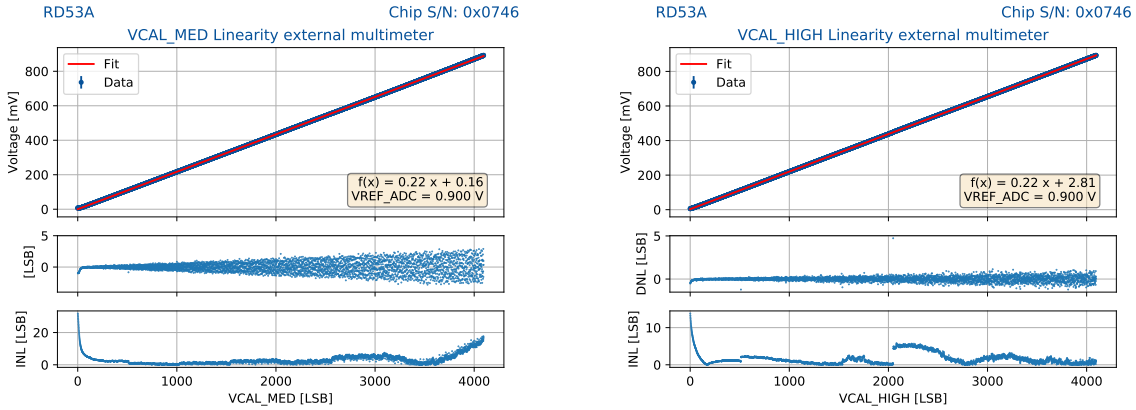
To obtain a good understanding of the effective charge injection as a function of the differential injection DAC, which is essential for most of the characterization methods described in Chapter 4, the injection circuit was examined in the very beginning of the RD53A characterization campaign. The value of the injected charge depends on two characteristic properties: The value of the injection capacitance implemented in each pixel and the linearity of the global differential injection DAC, VCAL_MED and VCAL_HIGH. The effective injection capacitance has been measured indirectly by means of the absolute charge calibration presented in [46]. The linearity measurement of the injection DACs is shown in the following section as an example for DAC linearity measurements that have been performed for all important DACs of RD53A.

Injection DAC Linearity

The quality of a voltage or current DAC implementation is characterized by measuring the output for every possible setting and assessing the quality of a straight line fit to the data points. Additionally, the Differential Non-Linearity (DNL) can be defined as the deviation of the actual output from the ideal output at a given setting. The Integrated Non-Linearity (INL) is defined as the integral of the DNL along the transfer function.

Figure 5.2 shows the DAC linearity measurements for the two global injection DACs of RD53A, VCAL_MED and VCAL_HIGH. Both DACs exhibit an acceptable linearity with the largest deviations at both, the low and high end of the range. This is a common effect for wide-spanning voltage DACs due to potential issues with the involved transistors' operating points when the output voltage approaches one of the the supply rails' potentials. For this reason, the injection charge is usually determined during chip characterization by setting VCAL_MED to 500 and adjusting VCAL_HIGH within the range $[500, 3500]$, according to the desired charge, thus leading to a charge that is as precisely determined as possible.

The slope of the straight line fit of about $(0.22 \pm 0.04)\ \frac{\text{mV}}{\text{LSB}}$ can be used together with the injec-



(a) Linearity of the injection DAC VCAL_MED measured on an RD53A chip.

(b) Linearity of the injection DAC VCAL_HIGH measured on an RD53A chip.

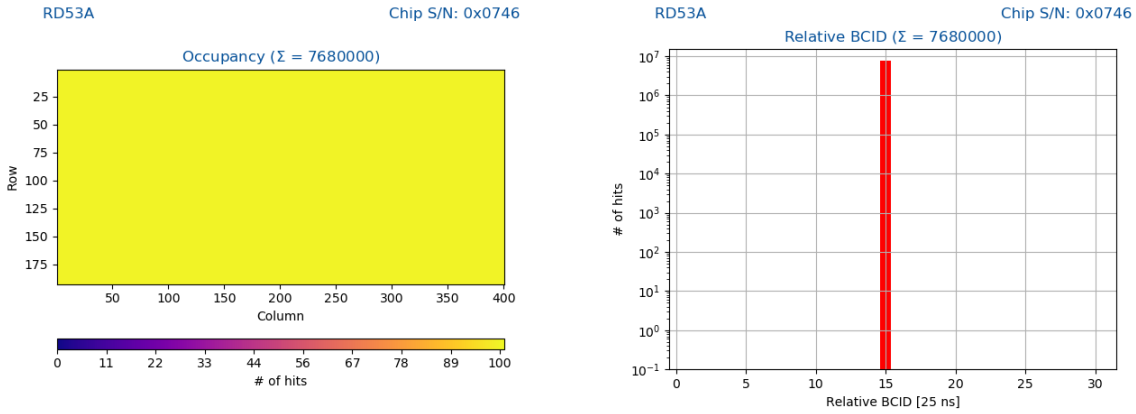
Figure 5.2: Analysis of the global differential injection DAC.

tion capacitance measurement to estimate the effectively injected charge in units of Coulomb or electrons. However, in practice and in most of the measurements presented in this thesis, a direct transfer function between the difference of the two differential injection DACs, ΔVCAL = VCAL_HIGH - VCAL_MED and the effectively injected charge is used, that was derived from the calibration of a complete SCM with a variable X-ray source, since this method includes all parasitic capacitances and additional effects of the full system [46].

5.2 VERIFICATION OF GENERAL CHIP FUNCTIONALITY

A first step in the verification of the chip's general functionality is to make sure that all global and in-pixel registers work as expected. This is done using the Register Test. In a good chip, all tested registers will work as expected, i.e. reading back the same values they were set to. Digital performance is then verified by running a Digital Scan (see Section 4.5) and checking that no hits are lost. This confirms that the digital hit processing circuitry of all pixels as well as the peripheral digital logic and the complete Aurora chain up to the readout system work as expected. An exemplary result of a Digital Scan is shown in Figure 5.3. Here, all three AFEs can be scanned at once, since the BDAQ53 mask shifting algorithm takes care of enabling one AFE after the other and not injecting too many pixels at once (see Section 4.3.1). All pixels report exactly 100 hits resulting from 100 injections per pixel, and all hits are recognized with a well-defined delay after the trigger command.

Next, an Analog Scan is run to verify basic analog functionality of all pixels. This scan usually provides a good overview over the amount of generally working pixels. By injecting a large charge well above threshold, the general functionality of pixels can be evaluated without influence of the analog performance of the respective AFE. The result of an Analog Scan is shown in Figure 5.4. Here, only the hitmap is evaluated, since both the ToT response and the hit timing are not tuned yet and are therefore evaluated per front-end later on. Again, all pixels recognized the 100 injected analog hits, verifying the general functionality of all pixels analog



(a) Hitmap showing the number of recorded hits for each pixel in the RD53A matrix.

(b) Histogram of the relative BCID of all recorded hits.

Figure 5.3: Result of a Digital Scan of all three AFEs.

and digital circuitry.

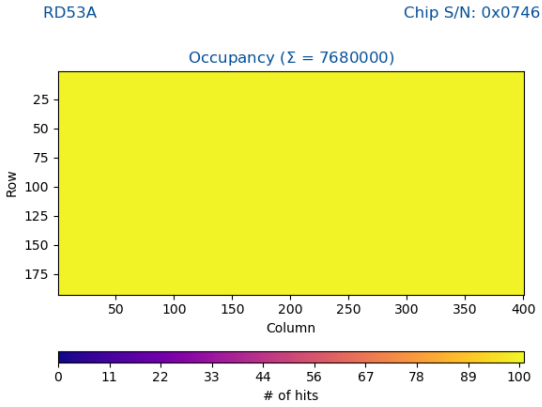


Figure 5.4: Result of an Analog Scan of all three AFEs. The hitmap shows the number of recorded hits for all pixels of the matrix.

After verifying basic functionality, the bare chip performance is dominated by the performance of the Analog Front-End (AFE). Therefore, the performance evaluation for RD53A is usually conducted for each AFE independently. The minimum stable threshold, threshold dispersion and noise are evaluated by means of a Global and Local Threshold Tuning and a subsequent Threshold Scan. The ToT response of the front-end is first tuned (ToT Tuning) and then evaluated by means of an Analog Scan injecting exactly the target charge. Furthermore, the timing performance of the AFE is evaluated by means of an In-Time Threshold Scan and comparing the resulting in-time threshold to the basic threshold determined before.

5.3 SYNCHRONOUS FRONT-END

The Synchronous Analog Front-End (SYNC) is the leftmost AFE flavor of RD53A, occupying columns 0 to 128. With only 16 core columns, its matrix is slightly smaller than that of the other two flavors, which both span over 17 core columns. It implements a synchronous design with a single stage CSA with Krummenacher feedback feeding a synchronous comparator (see Section 3.1.2). In contrast to the other two AFEs, the SYNC implements an automatic baseline acquisition scheme (Auto-Zero (AZ)), that requires a regular pulse being sent to the pixels instead of a manual Local Threshold DAC (TDAC). This feature poses an additional requirement for the DAQ system.

5.3.1 DISCOVERED ISSUES

During testing it was discovered that the digital logic of the SYNC can completely freeze from time to time, so that no more hits are generated and can be read out. An `Analog-` or `Digital Scan` sometimes runs without issues, other times only the first hits are reconstructed. After some testing and debugging, it was discovered that the SYNC can generate wrong hits where the ToT value 0xFFFF is stored. Since the digital bottom logic was never tested with this nonsensical input pattern, it can get stuck once this error occurs, preventing any hits to be output once a trigger command is received.

A DAQ-based workaround for this issue was found by sending regular `ECR` commands, resetting all the hit buffers and counters in the digital logic. In injection-based scans this command is sent right before every block of 100 injections, so that no actual hits are lost. For operation with external trigger sources, the command was integrated into the regular AZ procedure during which the SYNC has a small dead time anyways. Unfortunately, this `ECR` command also resets the global BCID counter, which removes information from the hits that is used in the analysis class of `BDAQ53` for an additional validity check of the data. Consequently, this optional check was disabled for the SYNC.

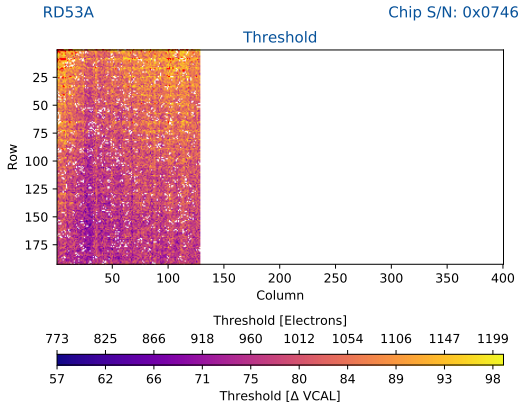
5.3.2 THRESHOLD PERFORMANCE

For an evaluation of the threshold- and noise performance of an AFE, the matrix is first tuned to a mean threshold of about 1 000 e^- and the ToT response is tuned to 5.3 ToT at 6 ke^-. Subsequently, a threshold scan is performed that records a response curve (S-curve, see Section 4.5.4) and extracts the threshold and noise for each pixel.

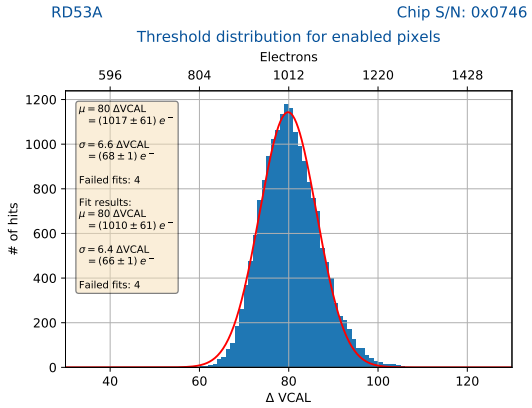
THRESHOLD DISPERSION

Due to the AZ feature described in Section 3.1.2, manual per-pixel threshold tuning of the SYNC is not necessary. The threshold dispersion is minimized automatically by sending regular AZ pulses. In injection based scans, this pulse is sent before every injection by `BDAQ53`, resulting in an AZ period of about 50 µs, well below the maximum recommended period of 100 µs.

The resulting threshold map and -dispersion is shown in Figure 5.5. White pixels in the threshold map shown in Figure 5.5(a) were disabled by the `Noise Occupancy Scan` or the `Stuck Pixel Scan` run in the course of the tuning procedure. The resulting mean threshold



(a) Threshold map of the SYNC with optimal AZ frequency.



(b) Threshold distribution of the SYNC with optimal AZ frequency.

Figure 5.5: Threshold result of a `Threshold Scan` of the SYNC with optimal AZ frequency after tuning the GDAC to a mean threshold of about 1 000 e^-.

shown in Figure 5.5(b) is very close to the target value of 1 000 e^-. The threshold dispersion of $(66 \pm 1)\,e^-$ is comparatively high. This result is deteriorated by the gradient of the threshold exhibited in Figure 5.5(a), which is most likely caused by the additional delay over the matrix from bottom to top in the AZ signal.

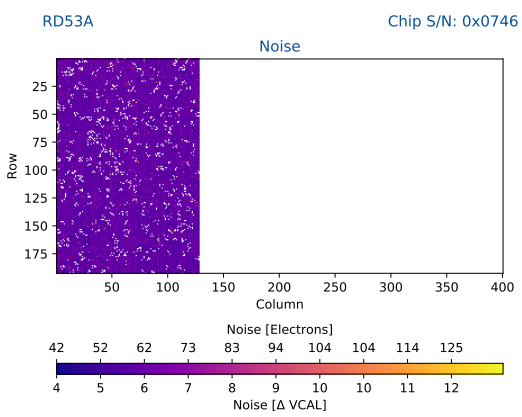
NOISE

The noise performance of the SYNC, that is extracted from the same threshold scan as the threshold distribution discussed in the previous section, is shown in Figure 5.6. The noise of a CSA is typically measured in units of Equivalent Noise Charge (ENC), i.e. the number of electrons at the input which would lead to an output signal equivalent to the measured noise amplitude.

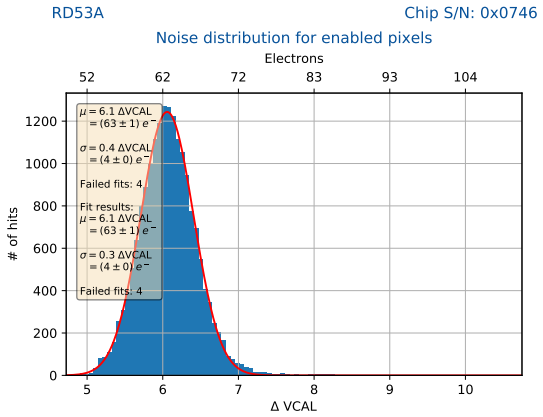
The noise map of the SYNC in Figure 5.6(a) exhibits no gradient over the matrix, indicating that the noise performance is not influenced by the effect discussed previously. The noise distribution is shown in Figure 5.6(b). The mean ENC of the SYNC of about $(63 \pm 1)\,e^-$ is typically the highest of the three AFEs of `RD53A`.

MINIMUM STABLE THRESHOLD

The minimum stable threshold is defined as the minimum threshold the AFE can be operated at, with a noise occupancy not larger than 10^{-6} after disabling not more than 1 % of noisy pixels. While these conditions are found iteratively by the `Noise Tuning` algorithm (see Section 4.6.4) for the LIN and DIFF, where also the TDACs have to be optimized, the procedure is simpler for the SYNC. Here, the global threshold is simply lowered until the conditions are met. The resulting threshold is shown in Figure 5.7. Figure 5.7(a) shows the timing distribution of all recorded hits. This histogram can be used to estimate the effective noise occupancy, since noise hits have no timing correlation with the trigger pulses generated by the threshold scan and



(a) Noise map of the SYNC with optimal AZ frequency.

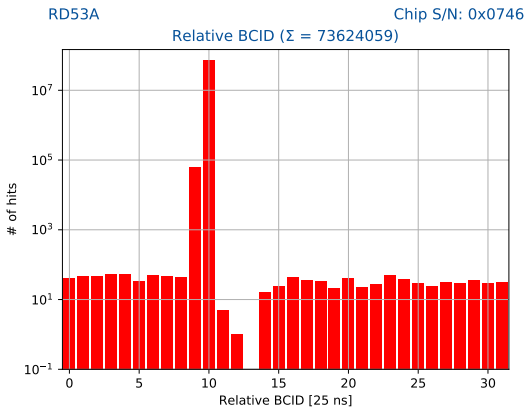


(b) Noise distribution of the SYNC with optimal AZ frequency.

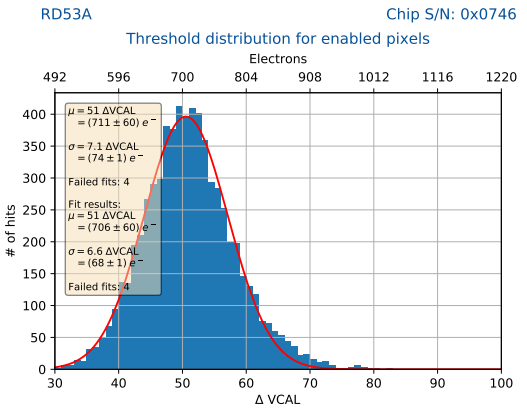
Figure 5.6: Noise result of a `Threshold Scan` of the SYNC with optimal AZ frequency after tuning the GDAC to a mean threshold of about 1 000 e^-.

are homogeneously distributed over all bins. In contrast, hits caused by the deliberately sent injections are typically found in one or two bins, as shown in Section 5.2. Therefore, the noise floor can be estimated at about $3 \cdot 10^1$ hits in this scan, while the actual scan used about $5 \cdot 10^7$ hits. The difference of $> 10^6$ between these two figures gives a rough estimate of the effective noise occupancy.

The resulting threshold distribution is shown in Figure 5.7(b). The mean minimum stable threshold of the SYNC is $(706 \pm 60)\,e^-$ with a dispersion of $(68 \pm 1)\,e^-$, identical to the usual performance shown in the previous sections.



(a) Relative BCID distribution of the `Threshold Scan` at the minimum stable threshold of the SYNC. The separation of the deliberate hits from the noise floor is larger than 10^6.



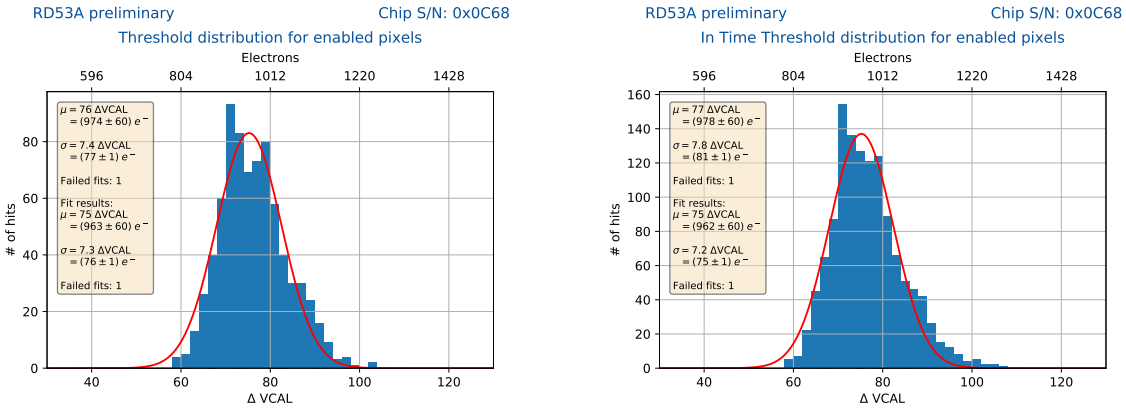
(b) Threshold distribution of the SYNC operated at the minimum stable threshold.

Figure 5.7: Minimum stable threshold of the SYNC.

OVERDRIVE

The overdrive is defined as the difference between the measured threshold when all hits are taken into account, regardless of their relative BCID and their in-time threshold when only a single relative BCID bin is accepted.

The result of an `In-Time Threshold Scan` (see Section 4.5.6) of columns 56 to 64 in the SYNC after tuning the global threshold to about 1 000 e^- and with sending AZ pulses in the optimal frequency range is shown in Figure 5.8. Due to the synchronous nature of its design, the SYNC inherently has no overdrive. The in-time threshold in Figure 5.8(b) is the same as the normal threshold, that is shown in Figure 5.8(a). This is different for the other two AFEs, which have a non-zero overdrive.



(a) Threshold distribution of the selected range of the SYNC taking all hits into account, regardless of their relative BCID.

(b) In-Time threshold distribution of the selected range of the SYNC taking only in-time hits into account.

Figure 5.8: Overdrive estimation for the SYNC. Columns 56 to 64 were selected for a narrow timing distribution.

5.3.3 TOT RESPONSE

The ToT response is evaluated by means of an `Analog Scan` injecting the target value of 6 ke^-. The resulting ToT response histogram for the SYNC is shown in Figure 5.9. The distribution peaks at 5 ToT and the mean of the distribution is 5.2 ToT with a standard deviation of 1.0 ToT, close to the target value of 5.3 ToT. All in all, the ToT computation and tuning of the SYNC works as expected and satisfies the requirements, however, the dispersion of the ToT values is large compared to the predecessor chip ATLAS `FE-I4`, which featured an in-pixel DAC for ToT tuning. [62, p.54]

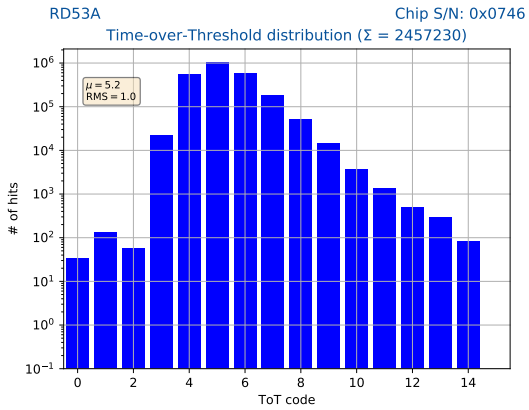


Figure 5.9: ToT response of the SYNC to an injection of 6 ke⁻ after tuning to 5.3 ToT at 6 ke⁻.

5.4 LINEAR FRONT-END

The Linear Analog Front-End (LIN) is the center AFE flavor of RD53A, located in columns 128 to 264. It implements a well-established design with a single-stage CSA with Krummenacher feedback, a standard comparator and ToT counter and a manual 4 bit Local Threshold DAC (TDAC) (see Section 3.1.2).

From a DAQ standpoint, the LIN is the easiest of RD53A's AFEs to operate, since it requires no regular pulses, special mask- or extensive register setting optimization.

5.4.1 DISCOVERED ISSUES

First measurements of the LIN's performance showed that the untuned threshold dispersion of the AFE is slightly higher than expected from simulations. While this effect should have been easy to compensate by means of tuning, another issue was found when first threshold tuning routines were developed and tested. The LIN implements a threshold fine-adjust DAC, LDAC_LIN that adjusts the step size of a single TDAC LSB, allowing the user to trade tuning accuracy for range. However, as shown in Figure 5.10, this DAC saturates after about 180 of 1 023 LSB, capping the TDAC step size at about 330 e^-.

While this feature does not necessarily hurt the LIN's threshold performance directly, a possible impact cannot be excluded. In addition, the AFE's versatility and robustness definitely suffer from this anomaly, since a good matching between the global and local threshold tuning algorithms is required.

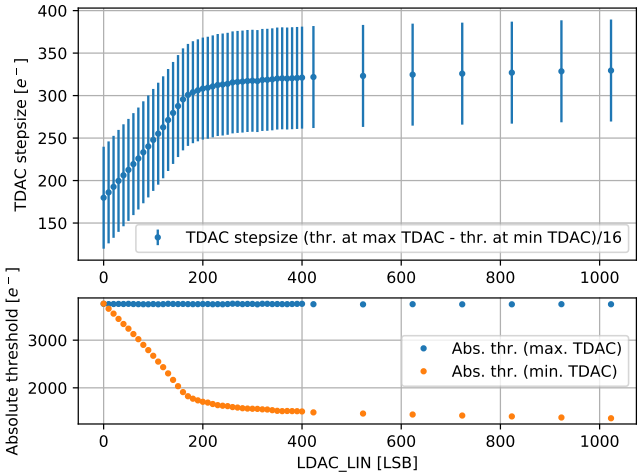


Figure 5.10: TDAC step size of the LIN as a function of the defining parameter LDAC_LIN. The step size is defined as the difference between the effective threshold at the maximum TDAC setting and the minimum, divided by the amount of available TDAC settings. Contrary to the expectation, the TDAC step size saturates after less than 20 % of the total range.

5.4.2 THRESHOLD PERFORMANCE

In contrast to the SYNC, the LIN implements standard 4 bit Local Threshold DACs (TDACs), which need to be tuned manually in order to minimize the threshold dispersion over the matrix. The range is inverted with regard to the resulting threshold, meaning that a TDAC setting of 15 corresponds to the lowest possible local threshold while a value of 0 sets the highest possible threshold. The quality of this tuning can be assessed by means of the TDAC distribution, shown in Figure 5.11. The distribution resembles a Gaussian function centered around the mean value, 7.7 in this case, while 7.5 would be the optimal value for the LIN. All TDAC values are used and the lowest and highest bins are not overly full, indicating that the step size for the TDAC was well chosen. Based on these observations, the resulting threshold performance can be expected to reflect the optimal performance of the AFE.

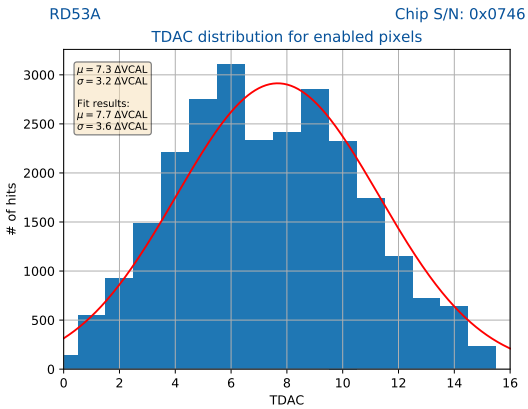


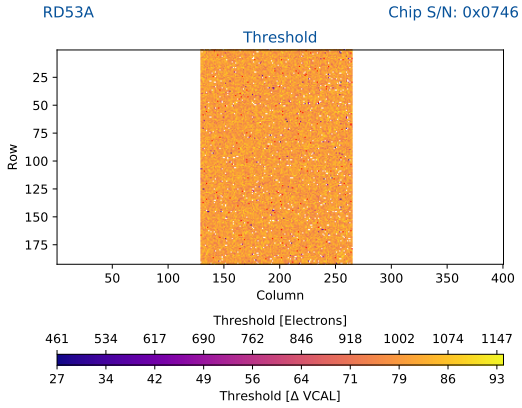
Figure 5.11: TDAC distribution of the LIN after tuning the GDAC and TDACs to a mean threshold of about 1 000 e^-.

THRESHOLD DISPERSION

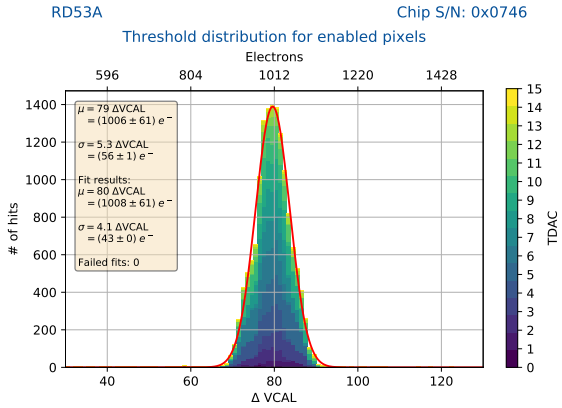
Figure 5.12 shows the threshold performance of the LIN after tuning the full AFE to a target threshold of 1 000 e^-. The threshold map in Figure 5.12(a) shows a homogeneous threshold over the full AFE with no visible gradients. As shown in the threshold distribution plot in Figure 5.12(b), the resulting mean threshold is close to the target threshold of 1 000 e^- and the measured threshold dispersion of $(43 \pm 1)\,e^-$ is low, especially compared to the result of the SYNC shown in the previous section. Additionally, in comparison to the SYNC, considerably fewer pixels had to be disabled due to noise or being stuck.

NOISE

The noise of the LIN is homogeneously distributed, as shown in Figure 5.13(a). While this is comparable to the SYNC, the mean ENC of the matrix, shown in Figure 5.13(b), is $(56 \pm 1)\,e^-$ and thus again considerably lower than that of the SYNC.

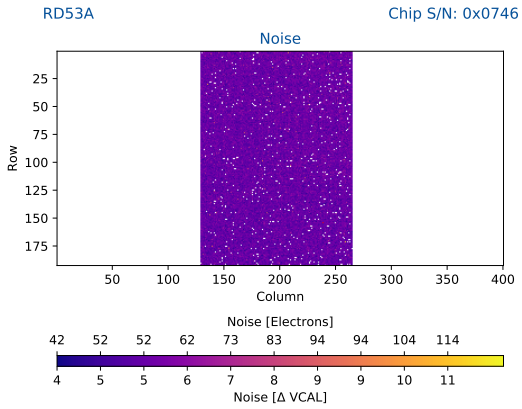


(a) Threshold map of the LIN after tuning.

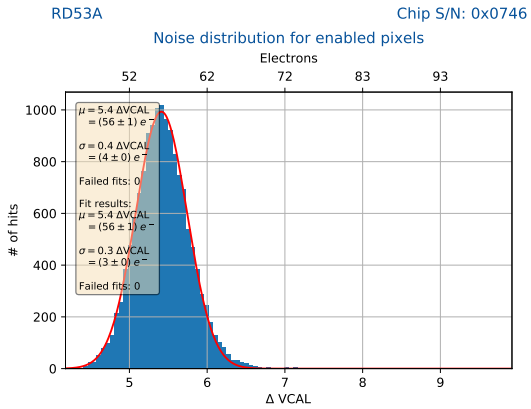


(b) Threshold distribution of the LIN after tuning. The color scale additionally shows the TDAC distribution within each bin of the histogram.

Figure 5.12: Threshold result of a `Threshold Scan` of the LIN after tuning the GDAC and TDACs to a mean threshold of about 1 000 e⁻.



(a) Noise map of the LIN after tuning.

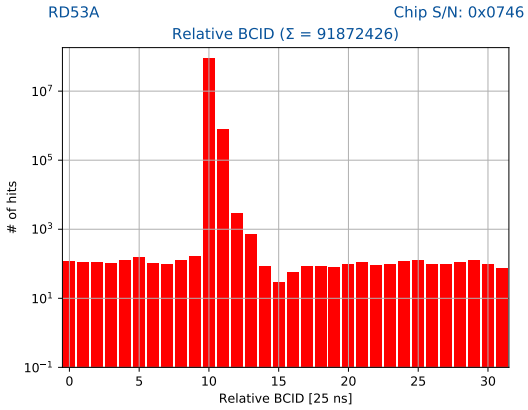


(b) Noise distribution of the LIN after tuning.

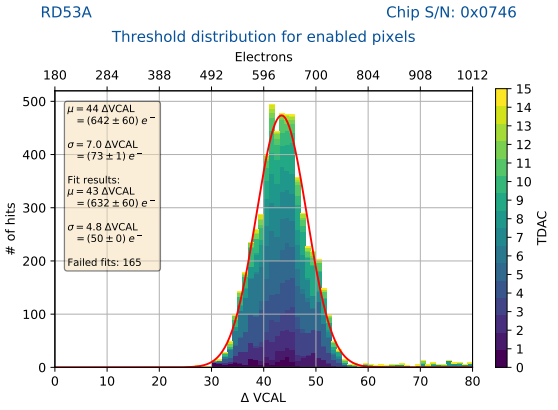
Figure 5.13: Noise result of a `Threshold Scan` of the LIN after tuning the GDAC and TDACs to a mean threshold of about 1 000 e⁻.

MINIMUM STABLE THRESHOLD

The minimum stable threshold of the LIN as a result of the `Noise Tuning` algorithm is shown in Figure 5.14. The separation between noise floor and deliberate hits is about 10^6, as can be seen in Figure 5.14(a), while the resulting distribution of the minimum stable threshold is shown in Figure 5.14(b). The mean minimum stable threshold of $(632 \pm 60)\,e^-$ is slightly lower than for the SYNC. The dispersion of $(50 \pm 1)\,e^-$ is comparable to the normal performance. The TDAC distribution as a measure for the quality of the tuning result can be estimated from the color scale in the histogram.



(a) Relative BCID distribution of the `Threshold Scan` at the minimum stable threshold of the LIN. The separation of the deliberate hits from the noise floor is about 10^6.



(b) Threshold distribution of the LIN operated at the minimum stable threshold.

Figure 5.14: Minimum stable threshold of the LIN.

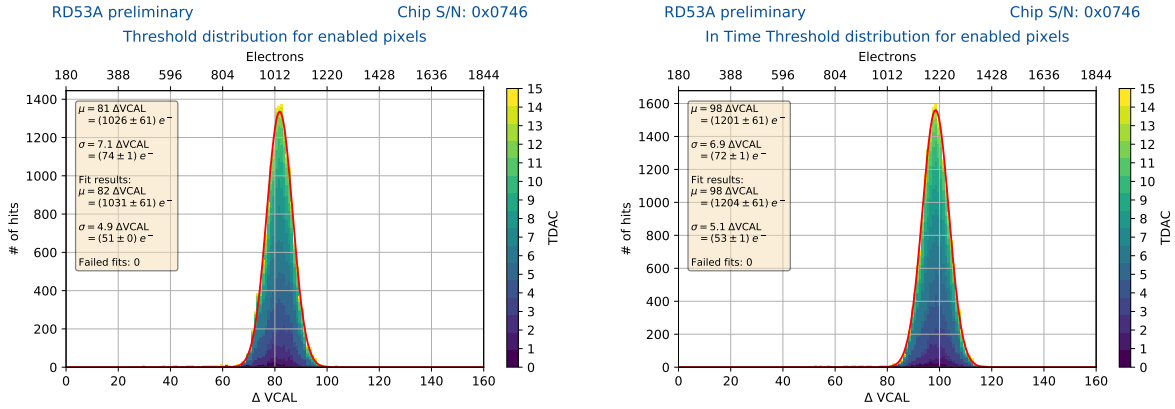
OVERDRIVE

The result of an `In-Time Threshold Scan` (see Section 4.5.6) after a tuning to a threshold of about 1 000 e^- for a single core column (using columns 128 to 136) is shown in Figure 5.15. The normal threshold, taking all hits into account is shown in Figure 5.15(a), while the resulting in-time threshold, using only hits with a relative BCID of 10, is shown in Figure 5.15(b). The overdrive is calculated to be:

$$\begin{aligned} \text{Overdrive} &= \mu_{\text{in-time}} - \mu \\ &= (1\,204 \pm 61)\,\text{e}^- - (1\,031 \pm 61)\,\text{e}^- \\ &= (173 \pm 86)\,\text{e}^-. \end{aligned}$$

5.4.3 TOT RESPONSE

As was done for the SYNC, the ToT response of the LIN was tuned to 5.3 ToT at 6 ke^-. The ToT distribution resulting from an analog scan with an injection charge set to the same value as the tuning target is shown in Figure 5.16. Most of the hits reported a ToT of 5, while the distribution drops rapidly to both sides of the peak. Note that the number of hits in Figure 5.16 is plotted on a logarithmic scale on the y-axis. The mean of the distribution is found to be 5.2 ToT with $\sigma = 1.3$ ToT, very close to the target value of 5.3 ToT. Again, calculation, storage and output of the ToT of recorded hits work as required for the LIN, while the dispersion is as large as expected without an in-pixel DAC for ToT tuning.



(a) Threshold distribution of the selected range of the LIN taking all hits into account, regardless of their relative BCID.

(b) In-Time threshold distribution of the selected range of the LIN taking only in-time hits into account.

Figure 5.15: Overdrive estimation for the LIN. Columns 128 to 136 were selected for a narrow timing distribution.

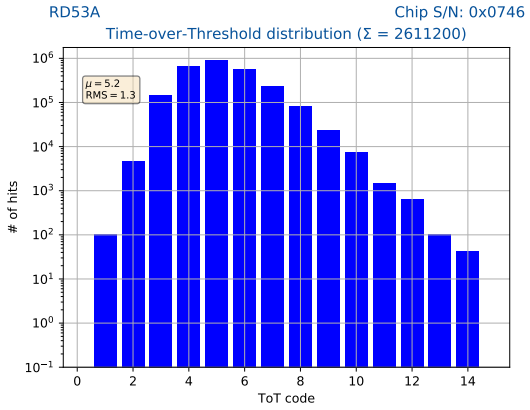


Figure 5.16: ToT response of the LIN to an injection of 6 ke⁻ after tuning to 5.3 ToT at 6 ke⁻.

5.5 DIFFERENTIAL FRONT-END

The last and rightmost AFE flavor of RD53A is the DIFF, spanning columns 264 to 400. It implements a two-stage design with a charge sensitive preamplifier with capacitive feedback and a differential pre-comparator with two global threshold voltages and an effective 5 bit TDAC (see Section 3.1.2).

Operation of the DIFF is somewhat more complex due to the use of two correlated GDACs. Due to an implementation issue described in the following Section 5.5.1, it requires different sets of global register settings and pixel masks, depending on the intended usage.

5.5.1 DISCOVERED ISSUES

Early on during testing of the DIFF, it became clear that the total threshold dispersion without tuning was much larger than expected. Both the ToT and untuned threshold distribution exhibited a suspiciously constant pattern repeating in every core. Upon further investigation, it became apparent that a missing buffer element between the discriminator output of the AFE and the input of the digital pixel logic causes a significant capacitive load on the discriminator output. The extent of this parasitic capacitance depends on the length of the connecting routing, which is synthesized automatically. The effective load per pixel in pF is shown for a single core in Figure 5.17 [63].

ROW \ COL	0	1	2	3	4	5	6	7
0	0,032	0,006	0,001	0,001	0,015	0,045	0,036	0,054
1	0,032	0,007	0,002	0,001	0,005	0,031	0,039	0,058
2	0,03	0,006	0,032	0,002	0,004	0,034	0,041	0,061
3	0,037	0,005	0,004	0,002	0,003	0,042	0,047	0,061
4	0,066	0,049	0,041	0,001	0,002	0,045	0,04	0,069
5	0,06	0,048	0,042	0,002	0,003	0,033	0,054	0,064
6	0,028	0,006	0,014	0,003	0,015	0,032	0,041	0,058
7	0,024	0,007	0,005	0,003	0,003	0,038	0,043	0,059

Figure 5.17: *Good pixel mask* for a single core of the DIFF. Green pixels are enabled, while red pixels are disabled when the good pixel mask is selected. Each cell contains the effective discriminator output load in pF.

This parasitic capacitance severely limits the bandwidth of the affected pixels, distorting the signal edges, which leads to degradation of the timing and ToT performances. At the same time, the output signal is attenuated, which leads to a significantly higher threshold that is measured by means of a `Threshold Scan` and can cause hits being missed completely.

As a workaround, it was decided to define a *good pixel mask*, as highlighted by the green pixels in Figure 5.17, and only evaluate the pixels that are least affected by this issue when characterizing and comparing the AFE's performance. Additionally, a set of parameters was defined, that equalizes the performance of all pixels of a core as much as possible for use in scenarios where the highest possible statistical sample size is required, for example when evaluating a sensor bump-bonded to an RD53A chip.

For the measurements in this chapter, the good pixel mask is used in order to compare the performance of the DIFF to the other two AFEs in RD53A, since this bug would be easily fixed in an eventual successor chip. For measurements involving the whole chip or assemblies based on RD53A, for example those presented in Chapters 7 or 8, the full matrix is used when possible, making use of the optimized parameter set to equalize the pixels performance.

5.5.2 THRESHOLD PERFORMANCE

Like the LIN, the DIFF implements in-pixel TDACs to minimize dispersion of the pixel threshold distribution manually. In contrast to the LIN, however, the DIFF uses a 5 bit TDAC, offering a better granularity and trimming range. In contrast to the LIN, the TDAC values of the DIFF are arranged symmetrically around 0, with −15 corresponding to the lowest possible local threshold and 15 to the highest.

The TDAC distribution resulting from a tuning of the DIFF to a target threshold of 1 000 e^- is shown in Figure 5.18. The distribution is well centered around 0, while all values seem to be in use with no excessive occupancy of the overflow bins.

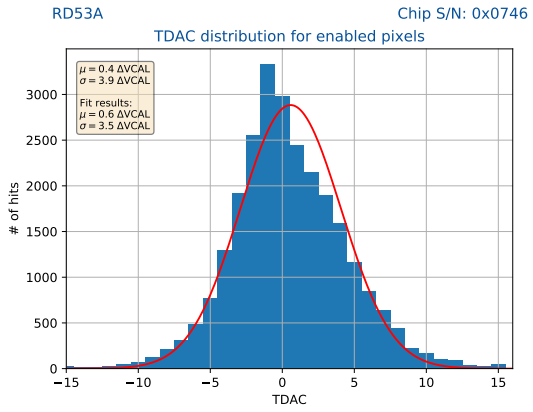


Figure 5.18: TDAC distribution of the DIFF after tuning the GDAC and TDACs to a mean threshold of about 1 000 e^-.

THRESHOLD DISPERSION

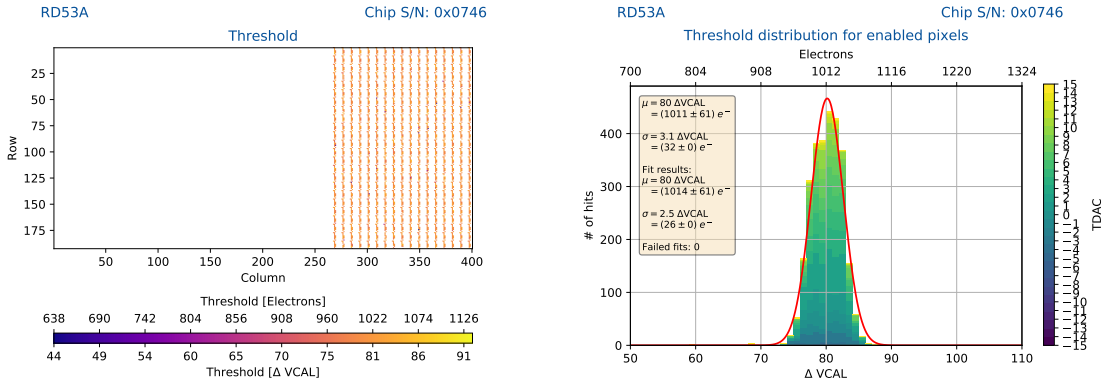
The threshold performance of the DIFF is shown in Figure 5.19. The homogeneity of the threshold values resulting from the tuning for the enabled pixels is very good, as can be seen in Figure 5.19(a). The threshold distribution in Figure 5.19(b) shows a mean threshold of $(1\,014 \pm 61)\,e^-$, very close to the tuning target of 1 000 e^-, while the dispersion of $(26 \pm 1)\,e^-$ is the lowest of the three AFEs.

NOISE

Figure 5.20 shows the noise performance of the DIFF. The noise is homogeneously distributed over the enabled pixels, as shown in Figure 5.20(a). While this was the case for the other two AFE flavors as well, the mean ENC of the DIFF of $(39 \pm 1)\,e^-$ is significantly lower.

MINIMUM STABLE THRESHOLD

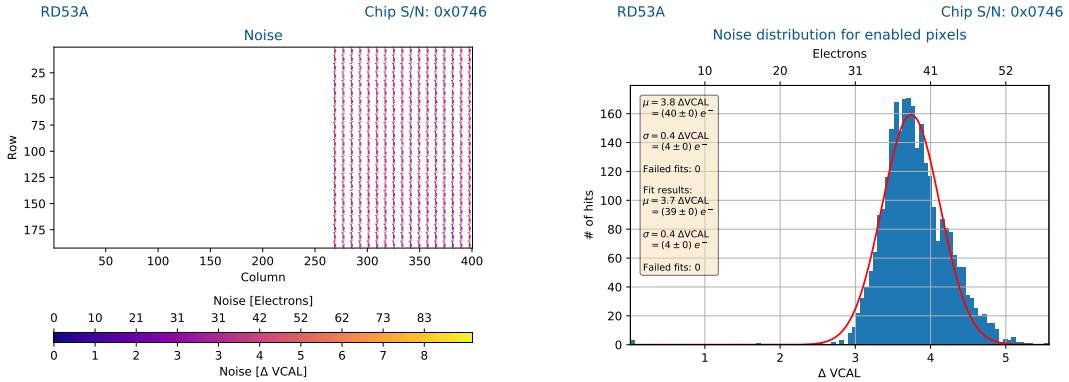
The DIFF reaches the lowest minimum stable threshold of the three AFE flavors, as shown in Figure 5.21. The hit timing distribution in Figure 5.21(a) shows a separation of deliberate hits from the noise floor well above 10^6. Again, the number of hits is plotted on a logarithmic



(a) Threshold map of the DIFF after tuning.

(b) Threshold distribution of the DIFF after tuning. The color scale additionally shows the TDAC distribution within each bin of the histogram.

Figure 5.19: Threshold result of a `Threshold Scan` of the DIFF after tuning the GDAC and TDACs to a mean threshold of about 1 000 e⁻.



(a) Noise map of the DIFF after tuning.

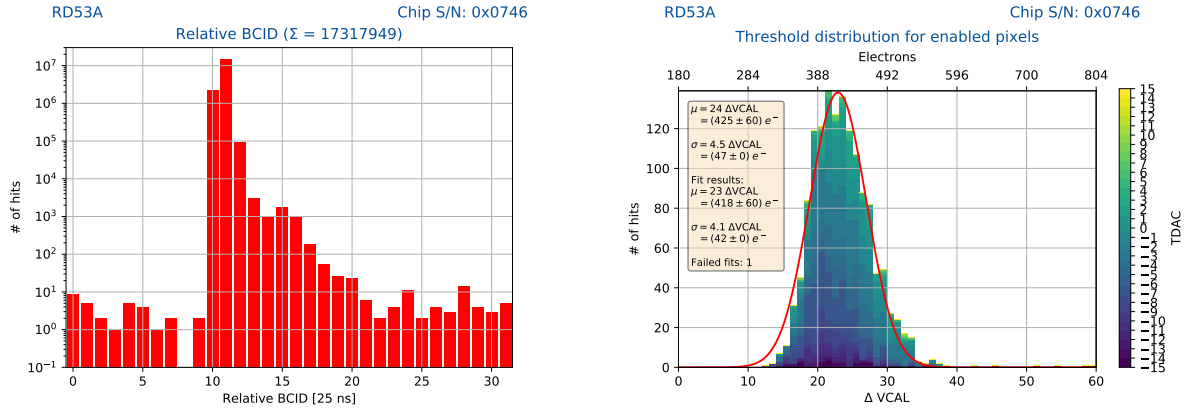
(b) Noise distribution of the DIFF after tuning.

Figure 5.20: Noise result of a `Threshold Scan` of the DIFF after tuning the GDAC and TDACs to a mean threshold of about 1 000 e⁻.

scale. The threshold distribution in Figure 5.21(b) shows a mean minimum stable threshold of the *good pixels* of the DIFF of $(418 \pm 60)\,e^-$, significantly lower than the other two flavors. The threshold dispersion at this low threshold of $(42 \pm 1)\,e^-$ is slightly worse than the DIFF's regular performance, but still very small.

OVERDRIVE

Figure 5.22 shows the result of an `In-Time Threshold Scan` after a tuning to a threshold of about 1 000 e⁻ for the *good pixels* of a single core column (using columns 306 to 314). The normal threshold, taking all hits into account, is shown in Figure 5.22(a), while the resulting in-time threshold, using only hits with a relative BCID of 11, is shown in Figure 5.22(b). The overdrive



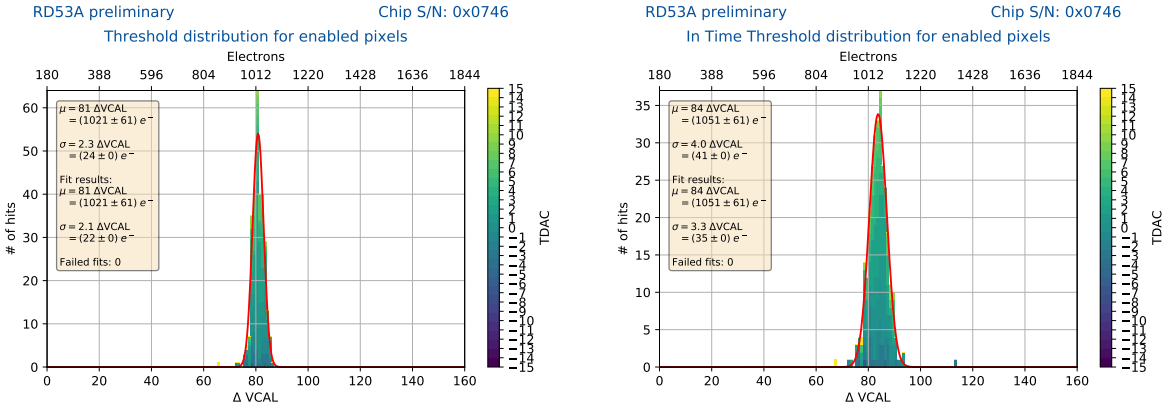
(a) Relative BCID distribution of the `Threshold Scan` at the minimum stable threshold of the DIFF. The separation of the deliberate hits from the noise floor is lower than 10^6.

(b) Threshold distribution of the *good pixels* of the DIFF operated at the minimum stable threshold.

Figure 5.21: Minimum stable threshold of the DIFF.

us calculated to be:

$$\begin{aligned} \text{Overdrive} &= \mu_{\text{in-time}} - \mu \\ &= (1\,051 \pm 61)\,\text{e}^- - (1\,021 \pm 61)\,\text{e}^- \\ &= (30 \pm 86)\,\text{e}^- \end{aligned}$$



(a) Threshold distribution of the selected range of the DIFF taking all hits into account, regardless of their relative BCID.

(b) In-Time threshold distribution of the selected range of the DIFF taking only in-time hits into account.

Figure 5.22: Overdrive estimation for the DIFF. Columns 306 to 314 were selected for a narrow timing distribution.

5.5.3 ToT Response

The ToT response of the *good pixels* of the DIFF is shown in Figure 5.23. The distribution peaks at 5 ToT, steeply falling off to both sides of the peak. The mean of the distribution is 5.4 ToT, very close to the tuning target of 5.3 ToT, while the width of the distribution is not too wide at 1.3 ToT. Like for the other flavors, acquisition, storage and tuning of ToT values in the *good pixels* of the DIFF work as required, with comparably large dispersion.

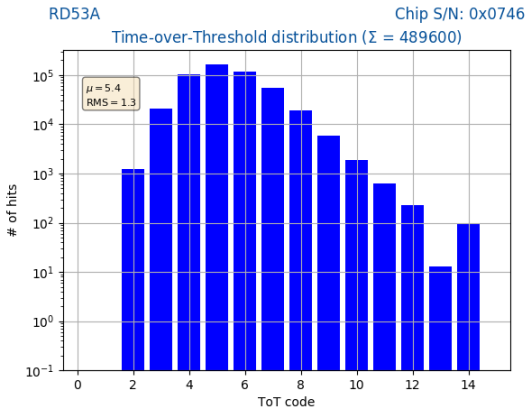


Figure 5.23: ToT response of the *good pixels* of the DIFF to an injection of 6 ke$^-$ after tuning to 5.3 ToT at 6 ke$^-$.

5.6 Summary

In summary, all three AFE flavors are capable of satisfying the requirements that were posed by the RD53 collaboration for their design. However, a few unplanned design features have been identified during testing. They degrade the performance of the respective AFE or require a workaround in the DAQ system. For the SYNC, the digital logic can get stuck due to wrongly created hits, necessitating a reset of the digital logic to be issued regularly. The LIN's TDAC range saturates earlier than expected, degrading the tuned threshold performance of the AFE. Due to a missing buffer in the implementation of the DIFF, the comparator output of this AFE flavor is loaded with an additional, routing-dependent parasitic capacitance, which degrades the timing and ToT performances of the AFE.

On the other hand, the SYNC's Auto-Zero (AZ) feature required some additional implementation work in the DAQ system, but once implemented successfully eliminates the need for manual threshold tuning. The threshold dispersion of the SYNC is comparable to the LIN, although it exhibits a gradient in the threshold map. While the noise of the SYNC is typically the highest of the three AFE flavors of `RD53A`, it is still rather low when viewed against a global reference. It is able to operate at a rather low minimum stable threshold, again comparable to the LIN, but generally struggles with a large amount of noisy pixels. The timing performance of the SYNC is excellent and the best of the three AFEs.

The LIN is a new implementation of the classic AFE design of older generations of pixel

detector readout chip front-ends. It works reliably with respect to threshold and noise performances, while not offering any remarkable performance features. The timing performance of the LIN is inferior to the other two AFE flavors of RD53A, but still fulfills the requirements.

The performance of the DIFF is hard to assess in RD53A, due to the missing buffer degrading timing performance. Despite this handicap, the threshold and noise performances of the DIFF are excellent and it outperforms the other two flavors. The minimum stable threshold of about 400 e^- is impressive and well below the requirements. The timing performance, while hard to assess in this prototype, is also very promising. It has a lower overdrive than the LIN, but higher than the SYNC.

Table 5.2 shows a summary of the numeric results presented in the previous sections. It shows that, despite their vastly different design approaches, all three AFE flavors of RD53A are capable of fulfilling the posed requirements. The dominant source of uncertainties shown in the previous sections is the charge conversion function from ΔVCAL to e^-, as described in Section 3.1.5.

	SYNC	LIN	DIFF (*good pixels*)
Threshold dispersion at 1 000 e^-	(66 ± 1) e^-	(43 ± 1) e^-	(26 ± 1) e^-
Mean ENC at 1 000 e^-	(63 ± 4) e^-	(56 ± 3) e^-	(39 ± 4) e^-
Minimum stable threshold	(706 ± 68) e^-	(632 ± 60) e^-	(418 ± 60) e^-
Overdrive	0 e^-	(173 ± 86) e^-	(30 ± 86) e^-
Mean ToT (6 ke^-) after tuning	(5.2 ± 1.0) ToT	(5.2 ± 1.3) ToT	(5.4 ± 1.3) ToT

Table 5.2: Summary of AFE characterization results.

Compared to the predecessor chip, ATLAS FE-I4, all three AFEs of RD53A significantly reduced the mean noise and also the minimum stable threshold is considerably lower. A more differentiated grading has to be performed when considering the mean threshold dispersion. The SYNC's dispersion is a little larger than the FE-I4's, while the LIN and DIFF undercut the FE-I4 by a small and comfortable margin, respectively. However, this result is shifted even further in RD53A's favor when considering that the mean dispersion was analyzed at a mean threshold of 3 ke^- for the FE-I4 while it is measured at 1 ke^- for RD53A. The mean overdrive of all three AFEs of RD53A is two to ten times lower than the FE-I4's in case of the LIN and DIFF and not present at all for the SYNC. Only in terms of ToT dispersion the FE-I4 is at an advantage compared to RD53A due to its per-pixel ToT tuning capability. [62]

5.7 ANALOG FRONT-END REVIEW

In December 2018, an Analog Front-End review was held by the RD53 collaboration in order to evaluate whether all three AFEs of RD53A satisfy the requirements, collect a list of necessary changes in the case that one of them is selected to be used in one of the successor chips and to give a recommendation to both ATLAS and CMS, as to which AFE is best suited for their experiments. The decision was based on the collective effort and measurement results obtained by multiple testing sites. These results were the general performance metrics of the three AFEs, as presented in the previous sections, as well as additional measurements based on SCMs and

irradiated front-end chips.

While the reviewers noted the advantages of the SYNC, especially the low overdrive and the fact that no manual threshold tuning is required, they also noted the subpar crosstalk performance (refer to [46]). Furthermore, the reviewers feared that the comparably large influence of leakage current on the AFE's noise performance (see Section 7.3) would pose an unpredictable risk when being integrated into multi-chip modules, especially after high irradiation.

The reviewers generally endorsed the performance of the LIN and the relative simplicity of its design. The proposed solution for the saturation of the LIN's TDACs (see Section 5.4.1) was met with approval by the reviewers, while the comparably large overdrive represents a concern. However, the reviewers proposed to decrease the preamplifier feedback capacitance, trading linearity of the response function for speed of the circuit. Since both of the proposed changes are quite easy to implement in a successor chip, they did not degrade the generally positive impression the LIN left with the reviewers.

With regard to the DIFF, the reviewers complimented the designers on the excellent low-threshold performance, while naturally referring to the impact of the parasitic capacitance at the comparator output as a limitation with a low-risk solution. Like for the SYNC, the reviewers were concerned about an effect of even comparably small amounts of leakage current on the DIFF's threshold and ToT distributions. The reviewers noted that the included LCC circuit has not been tested enough to understand the severity of this effect.

Based on these considerations, the RD53 collaboration gave the recommendation to use the Linear Analog Front-End (LIN) for both the ATLAS and CMS experiments, since it satisfies all requirements and bears the lowest risk for integration in a production chip. They also recommended to continue investigating the other two AFE implementations and their limitations. [64]

Decision for the ATLAS Experiment

An independent review was held by the ATLAS collaboration in March 2019, to decide which AFE should be used in the ATLAS ITk production Front-End ASIC. The SYNC was excluded due to the large number of noisy pixels and the fundamental changes to the Detector Control System (DCS) that would be necessary to efficiently use the SYNC's AZ mechanism. The risk of the proposed fixes for both LIN and DIFF were considered acceptable. Based on the expected superior performance of the DIFF compared to the LIN, including all proposed fixes, the Differential Analog Front-End (DIFF) was selected by the ATLAS collaboration to be used in the ITk's pixel detector module readout chip `ITkPix`. [30]

Decision for the CMS Experiment

The CMS collaboration also held an independent review in May 2019 prepared by a task force to identify the AFE flavor best suited for their tracker pixel module readout chip. Since this task force did not identify any additional arguments to exclude one of the three AFEs or strongly favor one of them, CMS decided to follow the recommendation of the RD53 collaboration and use the Linear Analog Front-End (LIN) in their production chip `CROC`. [31]

5.8 CONCLUSION

The results presented in this chapter have been obtained over the course of two years and have been validated by several research groups around the world involved in lively exchange and discussions at regular meetings. The findings and performance comparison of the three AFE flavors of RD53A have proven to constitute a solid baseline for the multiple reviews that have been held in order to identify strengths and weaknesses of the different flavors and to make decisions as to which Analog Front-End should be used in the various successor chips of RD53A.

Specific circuits of RD53A are independently characterized in depth. This includes the sLDO regulators [43] and the PLL and CDR circuits of RD53A [39], as well as their respective design iterations for RD53B. Additionally, irradiation studies and a full characterization of the performance of RD53A after different types, doses and dose rates have been conducted [39, 65].

Valuable input could also be provided to the RD53 collaboration for the design of the RD53B framework, which has been finished and verified. The first chip based on RD53B and the first direct successor to RD53A, ITkPix-V1, is becoming available to the first testing sites by the time of writing this thesis.

CHAPTER 6

WAFER-LEVEL TESTS OF RD53A

Every fabrication process and integrated circuit is subject to defects and constant process variations. It is not uncommon to see several ten percent of produced ASICs not working at all or showing malfunctions. This yield, i.e. the number of working ASICs relative to the produced amount, is an important figure in semiconductor device manufacturing. It is typically a function of both the total area of an IC and the complexity of its design. Wafer probing in this context refers to testing the manufactured integrated circuits on wafer level, before further processing.

6.1 QUALITY CONTROL OF BARE FRONT-END ASICS ON WAFER LEVEL

Quality Control (QC) during production of larger quantities of detector modules is an important aspect to monitor and control the whole production process. This is especially important for hybrid detector modules, where QC becomes a multi-stage process. The readout ASICs and sensors need to be tested individually before the hybridization step, to control the yield of fully functioning modules. Both, the individual ASICs and the hybridization process itself are too expensive to match up readout chips and sensors randomly. Instead, only known to be good readout ASICs and sensors should be used for the hybridization in the first place.

A common method in industry for QC of ASICs during the production process is *wafer probing*, that is testing the chips individually on the silicon wafer they are produced on. This method typically requires a *probe card*, a special PCB equipped with microscopic needles or *probes* that are pressed onto the wafer to contact a single ASIC and a *probe station*, a large machine that is capable of moving the wafer with micrometer precision relative to the probe card in order to contact one ASIC after the other.

In the scope of this thesis, a wafer probing setup and routine for `RD53A` mass testing have been developed and successfully used to test 8 900 `RD53A` chips from 100 of the 112 wafers that have been produced to date. From the first engineering batch of 12 wafers, 11 were diced directly after production to distribute chips to the testing sites as soon as possible, while one wafer was used to develop the wafer probing setup.

6.1.1 QC REQUIREMENTS

The requirements to classify an RD53A chip as *working* (green), *malfunctioning* (yellow) or *not working* (red) have been defined beforehand based on the characterization results of single chips, presented in Chapter 5. A series of electrical and digital tests is performed on each ASIC to check for fulfillment of these requirements. This section gives a short introduction to the performed tests during wafer probing and the criteria used for classification of the chips based on their test results.

ELECTRICAL TESTS

Electrical tests require no communication with the chip and can be conducted on all tested chips. The cuts applied to the results of these tests in order to categorize the chips as either *working* (green), *malfunctioning* (yellow) or *not working* (red) are shown in Table 6.1.

Evaluated value	**Category**		
	Green	**Yellow**	**Red**
Power-on			
Power-on behavior	No open or short	-	Open or short
IREF trimming			
IREF after trimming [µA]	> 3.925 and < 4.075	> 3.9 and < 4.1	< 3.9 or > 4.1
Regulator IV curves			
Digital regulator slope [Ω]	> 0.75 and < 0.95	> 0.5 and < 1.2	< 0.5 or > 1.2
Digital regulator offset [V]	> 0.85 and < 1.05	> 0.6 and < 1.3	< 0.6 or > 1.3
Analog regulator slope [Ω]	> 0.75 and < 0.95	> 0.5 and < 1.2	< 0.5 or > 1.2
Analog regulator offset [V]	> 0.85 and < 1.05	> 0.6 and < 1.3	< 0.6 or > 1.3

Table 6.1: Cuts for the categorization of chips based on the results of electrical tests during wafer probing.

- **Power-on:** The chip is powered on in Shunt mode and should not exhibit an open or short. The power consumption should be within defined limits. Further testing is performed in LDO mode, to be able to also identify chips that can be used for lab testing.
- **IREF trimming:** The global reference current IREF is measured for every possible trim bit setting. The value should be trimmable as close to the nominal 4 µA as possible.
- **Regulator IV curves:** The input behavior of the sLDO regulators is measured and a fit is performed to the linear part of the curve. The slope and offset values of both regulators should be within the defined limits.

LOW-LEVEL DIGITAL TESTS

Low-level digital tests first test the ability to communicate with the chip, while some other tests already require working communication and cannot be performed otherwise. The cuts applied

to the results of these tests in order to categorize the chips are shown in Table 6.2.

Evaluated value	Category		
	Green	Yellow	Red
Communication test			
# of tries to establish communication	1	≤ 5	> 5
VREF trimming			
VDDA after trimming [V]	> 1.1 and < 1.3	> 1.0 and < 1.4	< 1.0 or > 1.4
VDDD after trimming [V]	> 1.1 and < 1.3	> 1.0 and < 1.4	< 1.0 or > 1.4
VREF_ADC trimming			
VREF_ADC after trimming [V]	> 0.85 and < 0.95	> 0.7 and < 1.1	< 0.7 or > 1.1
Injection DAC calibration			
VCAL_MED slope $[\frac{\text{mV}}{\text{LSB}}]$	> 0.18 and < 0.23	> 0.15 and < 0.26	< 0.15 or > 0.26
VCAL_MED offset [mV]	> 5 and < 70	> 0 and < 75	< 0 or > 75
VCAL_HIGH slope $[\frac{\text{mV}}{\text{LSB}}]$	> 0.18 and < 0.23	> 0.15 and < 0.26	< 0.15 or > 0.26
VCAL_HIGH offset [mV]	> 5 and < 70	> 0 and < 75	< 0 or > 75
Total power consumption (configured)			
Total input current after configuration [A]	> 0.45 and < 0.80	> 0.35 and < 0.90	< 0.35 or > 0.90

Table 6.2: Cuts for the categorization of chips based on the results of low-level digital tests during wafer probing.

- **Communication test:** The chip's power is cycled and a simple check is performed to assess whether Aurora communication with the chip could be established. If there is no communication, the chip is power cycled and the test is performed again. This is repeated up to five times.
- **VREF trimming:** For both regulators, the trim bits are scanned at the same time, beginning at the maximum value 31. The resulting VDDA and VDDD are measured. The trim bits for both regulators are reduced by one, until one of the measured voltages reaches 1 V or all trim bits have been scanned. The optimal trim bit setting to obtain VDDA / VDDD closest to the nominal 1.2 V is saved to be used in future chip configuration files. They are used for the rest of the testing procedure. The value of the optimal VDDA / VDDD is evaluated.
- **VREF_ADC trimming:** The same procedure from before is repeated for the ADC- and injection DAC reference voltage VREF_ADC.
- **Injection DAC calibration:** The output voltage of both injeciton DACs VCAL_MED and VCAL_HIGH is measured for five discrete values spanning from 500 to 3 500. The slope

and offset obtained from a linear fit to these data points are evaluated and can be used as calibration values for future chip configuration files.

- **Total power consumption (configured):** The total current consumption of the chip that is powered with a constant voltage is measured, after a defined configuration is applied.

HIGH-LEVEL DIGITAL SCANS

High-level digital scans require working communication with the chip. Therefore, these tests can only be conducted on chips with working communication. The cuts applied to the results of these tests in order to categorize the chips are shown in Table 6.3.

Evaluated value	Category		
	Green	Yellow	Red
`Register Test`			
# of wrong registers	0	-	> 0
`Digital Scan`			
Amount of good pixels [%]	> 99	> 68	$<= 68$
`Analog Scan`			
Amount of good pixels [%]	> 99	> 68	$<= 68$
`Noise Occupancy Scan`			
Amount of noisy pixels [%]	< 1.0	≤ 5.0	> 5.0
`Stuck Pixel Scan`			
Amount of stuck pixels [%]	< 0.1	≤ 1.0	> 1.0
`Threshold Scan`			
Mean threshold [Δ`VCAL`]	> 179 and < 181	> 170 and < 190	< 170 or > 190
Threshold dispersion [Δ`VCAL`]	< 40	< 50	> 50
Amount of untuned pixels [%]	< 5	< 10	> 10

Table 6.3: Cuts for the categorization of chips based on the results of high-level digital tests during wafer probing.

- **`Register Test`:** All global registers that can be changed without leaving the chip in an undefined state, are written with a 0b01 and a 0b10 and read back afterwards (see Section 4.5.1). The read value is compared to the written value.

- **`Digital Scan`:** All pixels are injected with a digital hit 100 times. Analog pixel circuitry is disabled. A good pixel reads back exactly 100 hits. A pixel reporting 0 hits is defined as *dead*, a pixel reading back any amount of hits between 0 and 100 is defined as *bad* and a

pixel reading back more than 100 hits is defined as *noisy* (see Section 4.5.2). The relative amount of good pixels compared to enabled pixels is evaluated.

- **`Analog Scan`:** All pixels are injected 100 times with an analog hit with a charge of 800 ΔVCAL . A good pixel reads back exactly 100 hits. A pixel reporting 0 hits is defined as *dead*, a pixel reading back any amount of hits between 0 and 100 is defined as *bad* and a pixel reading back more than 100 hits is defined as *noisy* (see Section 4.5.3). The relative amount of good pixels compared to enabled pixels is evaluated.

- **`Threshold Tuning`:** Global and local threshold of all AFEs are tuned to a target value of 180 ΔVCAL (see Section 4.6). The tuning result is evaluated in the next scans.

- **`Noise Occupancy Scan`:** 10^7 random trigger commands are sent to the chip and all resulting hits are recorded. Every pixel that recorded more than ten hits is considered *noisy* and is disabled (see Section 4.5.7). The relative amount of noisy pixels compared to enabled pixels is evaluated.

- **`Stuck Pixel Scan`:** A pixel is defined as *stuck* if it is so noisy, that its comparator output is effectively stuck high. These pixels would not be classified as *noisy* by a normal `Noise Occupancy Scan`, since they never generate hits. A hit from stuck pixels can be provoked by toggling between analog and digital injection (see Section 4.5.8). The relative amount of stuck pixels compared to enabled pixels is evaluated.

- **`Threshold Scan`:** A `Threshold Scan` is performed over all three AFEs simultaneously to evaluate the result of the `Threshold Tuning` (see Section 4.6). The mean threshold of all pixels is evaluated as well as the overall threshold dispersion and the amount of untuned pixels.

6.2 WAFER PROBING SETUP

Within the scope of this work, a setup for wafer probing of `RD53A` has been established, tested and used to probe a large portion of the available wafers. The setup consists of a semi-automatic 300 mm probe station, situated in a clean room to avoid contamination of the chips with dust and other particles that could complicate further processing. The Cascade PA300 probe station used here offers an absolute precision of $\pm$2.0 µm in both x- and y direction and the repeatability of movements in x-, y- and z direction is $\pm$1.0 µm. [66] An inside view of the probe station with an `RD53A` wafer loaded onto the *chuck* is shown in Figure 6.1. The wafer is held in place by the use of vacuum. After an initial alignment of the wafer in Θ, the rotational dimension of the chuck with respect to the probe card, the wafer can be moved in position, so the first chip's pads align with the contact needles of the probe card. The probe station features a movable microscope with 100$\times$ magnification to help with this process. To help find the correct height, so that all needles contact the chip pads with the required force, the probe card features two edge sensors, which are described in Section 6.2.1. After this first, manual process, the probe station is capable of contacting all 89 chips on the wafer automatically, using a map of relative chip positions that was created during the initial setup process.

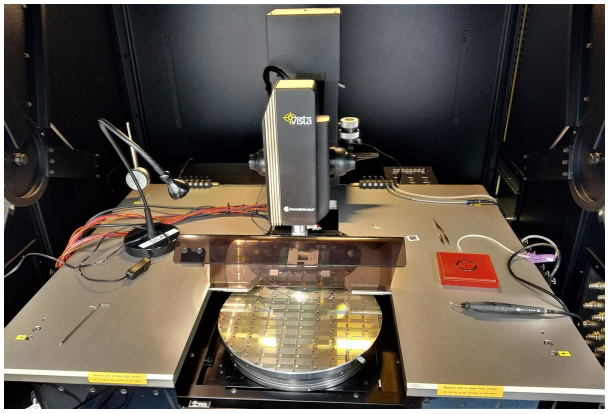


Figure 6.1: Inside view of the Cascade PA300 probe station.

Apart from the probe station, the wafer probing setup consists of a BDAQ53 installation as described in Chapter 4. Instead of the SCC or a flex, the BDAQ53 board is connected to the *probe card* via two DP cables. Two low-voltage power supplies and a Source Measure Unit (SMU) are used to power the DUT and conduct analog measurements. Everything is controlled by a DAQ PC running the BDAQ53 software framework. By employing the basil framework [57], BDAQ53 is able to control the power supplies, SMU as well as the probe station.

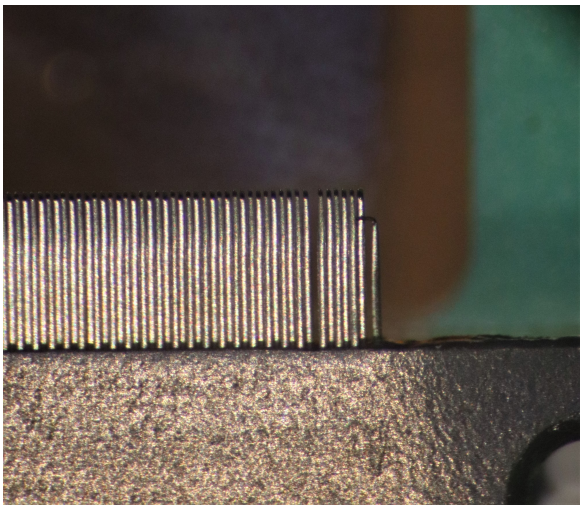
6.2.1 PROBE CARD

The RD53A probe card is shown in Figure 6.2(a). Its design is based on the RD53A SCC (refer to Section 3.2.2), but instead of wire-bonding, contact with the chip is established via 198 microscopic needles made from tungsten. A close up view of some of these needles is shown in Figure 6.2(b). They are about 2 cm long, though more than half of this length is embedded in a mounting block. The needles have an L-shape, bending down towards the chip in a 90° angle on one end.

At both ends of the needle array, the probe card features an *edge sensor*, a simple switch made up of the outermost needle contacting the DUT and another needle with an L-shape in the horizontal plane, held in contact with the outermost contact needle by spring tension. Once the DUT reaches the appropriate contact height, the contact between these two needles is opened due to the contact needle being pressed upwards. This contact opening can be read out by the DAQ via I^2C and additionally makes an LED light up as a signal to the operator that the correct contact height has been reached.

The RD53A probe card gives access to all the signals of the bottom row wire-bond pads of RD53A, while the top row is not accessible. Compared to the SCC, several active components, like I^2C controlled switches and multiplexers (MUXs) have been added to the design to allow for fully automatic testing and measurement of all available signals and voltages. In contrast to the SCC, IREF can be trimmed by software using I^2C switches instead of physical jumpers. In an additional modification, a switch was added to select between LDO and Shunt modes programmatically.

(a) Top view of the RD53A probe card. The probing needles are located at the bottom edge of the central cutout. A white ceramic stiffener is mounted on the top side of the PCB to improve rigidity.



(b) Closeup view of the probing needles viewed from the bottom. The two rightmost needles form one of the two edge sensors.

Figure 6.2: RD53A probe card.

6.2.2 PROBING ROUTINE

The wafer probing routine was developed within the scope of this work to assess if a chip fulfills the QC requirements detailed in Section 6.1.1. It was optimized for full testing of as many chips as possible, since also malfunctioning chips can be used for certain laboratory measurements, depending on the failure mode. The most important limiting factor during the development of the testing routine was the total time, the measurement routine takes to fully test a chip. For efficient mass testing, and also as preparation for wafer probing as part of the FE chip QC during mass production of modules for the ITk, the routine was developed with the goal of keeping a 24 h frequency, including changing wafers and making contact. This goal was reached with the routine taking less than 15 minutes to test a fully working chip, while *malfunctioning* or *not working* chips can be discarded even faster, depending on the failure mode.

A flow diagram of the probing routine that is executed for every chip on a wafer automatically is shown in Figure 6.3. It starts in the top left corner with making contact with a new chip, which is handled by the probe station on receiving a control command issued by BDAQ53. Before any voltage is applied to the chip, the output of the two edge sensors is checked. Only if both sensors report good contact, the chip is powered on in Shunt mode. If any of the two independent power supply channels for the analog and digital input draws significantly more current than expected, the power supplies are immediately powered off in order to protect the delicate needles and the chip is discarded. Otherwise, in order to assess the possibility of implementing a chip as part of a serial powering chain, the input IV curves of the two sLDO

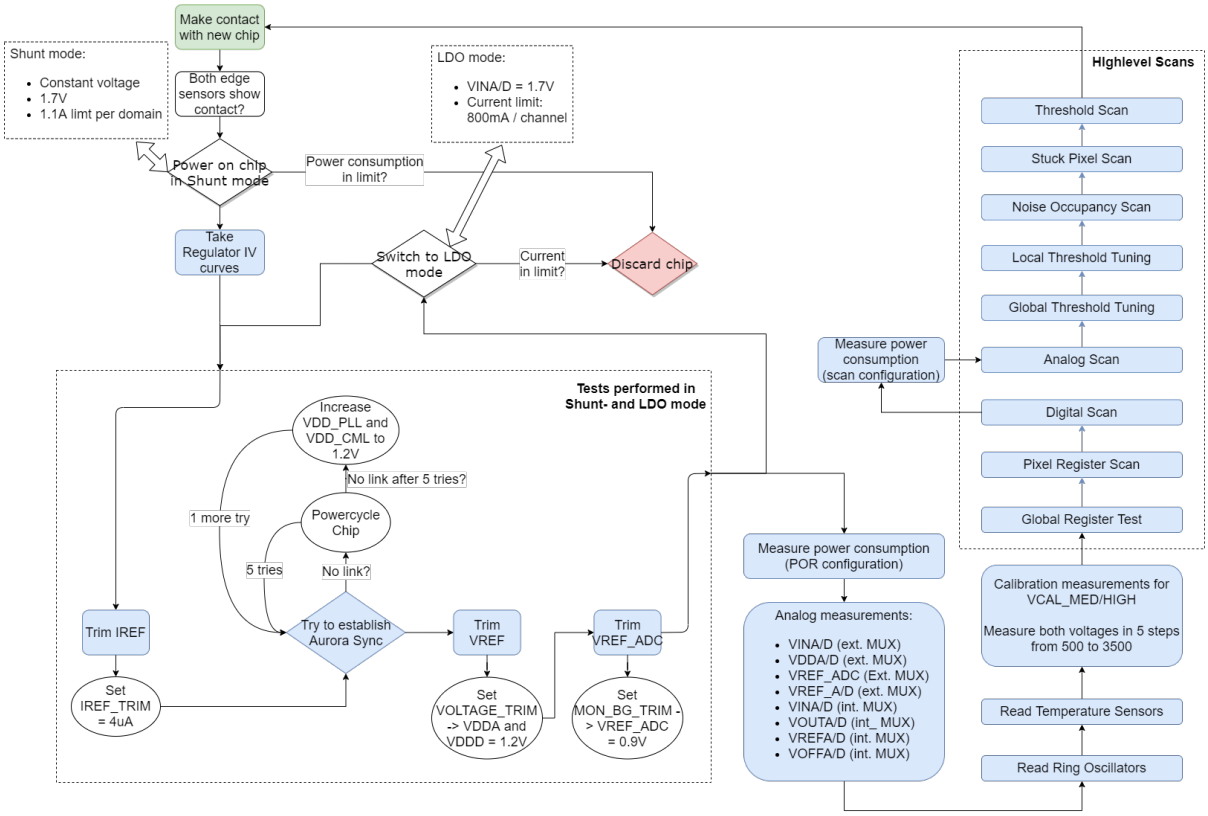


Figure 6.3: Flowchart of the RD53A wafer probing routine.

regulator domains are measured. Next, a collection of trimming and communication tests is conducted, that is highlighted by the dashed box in the bottom left of Figure 6.3. First, all 16 possible trim bit settings for IREF are selected and the resulting IREF is measured for each. All raw data is saved to disc, while the routine selects the optimal trim bit setting for the rest of the testing routine. The PLL and CML circuits of the chip are powered manually, using an additional 2-channel low-voltage power supply and the AUX_PWR connector, to avoid losing a large percentage of chips due to the start-up issue discussed in Section 5.1. However, to test if the chip would communicate under realistic conditions, both auxiliary voltages are set to the measured VDDA at start-up conditions. The routine tries to establish communication up to five times, power cycling the chip in between every iteration. If no communication could be established after five tries with start-up conditions, VDD_PLL and VDD_CML are set to the nominal 1.2 V and the routine tries again five times to establish Aurora sync. In this case, the chip is marked as *malfunctioning*, but testing is continued, since it can potentially still be used for testing by applying the workaround discussed in Section 5.1. Once communication with the DUT has been established, the trim registers for VREF_A, VREF_D and VREF_ADC can be trimmed. Like for IREF, the resulting VDDA, VDDD and VREF_ADC are saved to disc for every trim bit setting and the optimal trim bit is selected automatically for the rest of the routine. This block of tests is repeated once again in LDO mode. The data for both modes is saved for reference, but only the results in LDO mode are assessed for chip classification.

After trimming is complete and the optimal trim bits have been selected for all reference voltages and currents, a set of analog measurements is conducted, including the power consumption at default configuration and the resulting voltages trimmed in the previous step. The on-chip ADC is used to read out the on-chip temperature sensors and the ring oscillators for radiation dose measurement are read out once, to ensure their general functionality.

As a last analog measurement, both calibration injection DACs `VCAL_MED` and `VCAL_HIGH` are characterized by measuring the output voltage for five different points over the full range, fitting a straight line to the data points and extracting the slope and offset.

Finally, a set of high level scans are executed, starting with a `Register Test`, a `Digital Scan` and an `Analog Scan` of the full chip. After the configuration step of the `Digital Scan`, the power consumption of both domains of the chip is measured again. Following a standard `Global` and `Local Threshold Tuning` including a `Noise Occupancy` and `Stuck Pixel Scan`, a `Threshold Scan` is executed to generate assessable results. In earlier versions of the routine, the `Threshold Tuning` was omitted. The `Threshold Scans` of these wafers are evaluated against different requirements from the ones specified in Section 6.1.1.

In an asynchronous analysis step, the data of all chips of the probed wafer is evaluated against the requirements and the chips are categorized as *working* (green), *malfunctioning* (yellow) or *not working* (red), according to the criteria defined in Section 6.1.1.

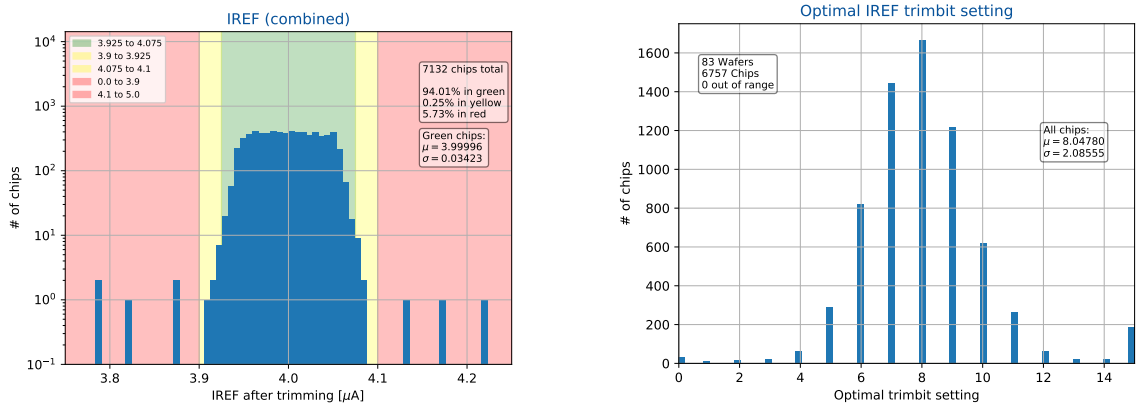
6.3 RESULTS

This section summarizes the results from 83 wafers probed in Bonn and at CERN. Data from an additional 16 wafers, that are affiliated with CMS and were probed at another testing site using the same setup and routine, was unfortunately not available for analysis within this thesis. Furthermore, the last production lot of 25 wafers has been probed with an upgraded routine, that optimizes the testing time while concurrently adding more tests. This was done in order to prepare for and to closer resemble the conditions during wafer testing of the final production chips. However, the more comprehensive test routine and especially the implemented time savings during the initial communication test (refer to Section 6.2.2) lead to significantly more chips being characterized as *malfunctioning* compared to the original routine. This is a consequence of the communication link instabilities of `RD53A` described in Section 5.1. Therefore, only the 58 wafers that were probed with the original routine are included in the yield analysis of this section. For the other wafers, about 38 % of green chips are characterized as yellow instead.

In the following sections, detailed analyses of the results of each of the performed tests are presented. While the graphs show automatically calculated percentages up to two decimal digits, the total sample size of less than 10 000 chips that is evaluated here only justifies one significant decimal figure.

6.3.1 IREF TRIMMING

For the `IREF` trimming test, every possible setting for the `IREF` trim bits is configured by means of a software-controlled digital switch on the probe card. The resulting `IREF` is then measured



(a) Histogram of the trimmed IREF values of all tested chips. The background colors depict the regions used to classify the chips. The yield for this test is shown in the top right corner.

(b) Histogram of the optimal IREF trim bit setting of all tested chips.

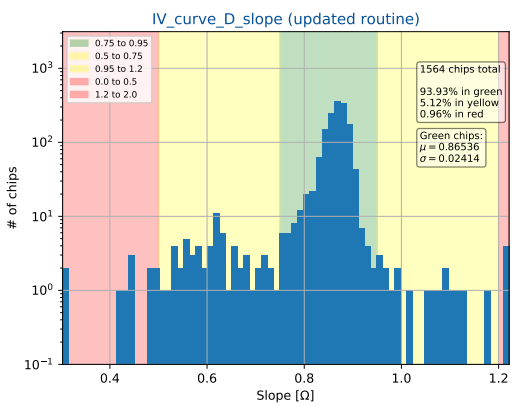
Figure 6.4: Results of the IREF trimming test.

directly, using the external SMU. After all values have been acquired, the optimal trim bit setting is extracted by finding the measurement that is closest to the nominal IREF value of 4 µA. This value is evaluated with regard to the requirements presented in Section 6.1.1. The result of this test for all tested chips is shown in Figure 6.4(a). 94.0 % of the tested chips can be trimmed to (4.000 ± 0.075) µA and are classified as *working*, while about 5.7 % only reach a value that deviates by more than 100 nA from the nominal value or could not be trimmed at all. These chips are categorized as *not working*. The extracted optimal trim bit setting is then set up for the rest of the testing routine for this chip, to make sure the chip is tested with the correct IREF.

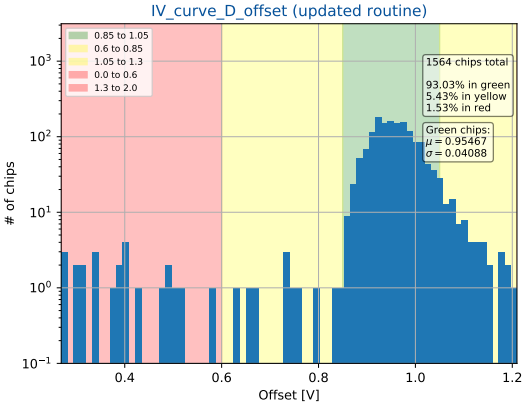
Figure 6.4(b) shows the distribution of the extracted optimal IREF trim bit settings for all tested chips. The distribution exhibits a Guassian shape with a mean value close 8. This leads to the conclusion that the chip's IREF circuit behaves as expected from simulations and the trimming range is well chosen.

6.3.2 REGULATOR IV CURVES

To test whether a chip is fit for use in a serial powering chain, both its regulators need to work within a given range of parameters. To assess this, the regulator input IV curve is measured by powering the chip in Shunt mode and varying the input current at the power supply from 1.0 A to 0.5 A per channel in decreasing order and measuring the resulting output voltage. Two examples for this measurement are shown in Figure D.1. For working points in this range, the regulators should behave linearly, with a slope of 900 mΩ and an offset of 1 V. Figure D.1(a) depicts an example IV curve of well-behaving regulators. However, sometimes one or both of the regulators of a chip behave irregularly, with a strongly deviating slope or offset or even with a non-linear trend. An example for a chip with a malfunctioning analog regulator is shown in Figure D.1(b). These malfunctioning chips are not suitable for use in serial powering chains but



(a) Histogram of the slope values extracted from the digital regulator IV curves of chips tested with the updated routine. The background colors depict the regions used to classify the chips. The yield for this test is shown in the top right corner.



(b) Histogram of the offset values extracted from the digital regulator IV curves of chips tested with the updated routine. The background colors depict the regions used to classify the chips. The yield for this test is shown in the top right corner.

Figure 6.5: Results of the IV curve test for the digital regulator.

may still be appropriate for other testing purposes.

The distribution of the extracted mean slope values of the digital regulator IV curves of all chips that were tested with the updated routine are shown in Figure 6.5(a), while the corresponding offset values are shown in Figure 6.5(b). For both histograms, more than 93 % of chips are categorized as *working*. The mean slope of digital regulators of the tested chips is $(865 \pm 24)\,\mathrm{m\Omega}$ and the corresponding mean offset is $(966 \pm 41)\,\mathrm{mV}$, both close to the expected values. Chips that are categorized as working in this test should be suitable for serial powering tests without limitation.

However, when considering the corresponding histograms for the analog regulators of the same sample of chips, shown in Figure D.2, it is apparent that the analog regulators do not work within specification much more often than their digital counterparts. Only about 73 % of the tested chips are categorized as *working* with regard to their analog regulator's slope and offset. The most plausible cause for this phenomenon is, that all control signals for both regulators are generated in the chip's digital logic, which is powered by the digital power domain and therefore depends on the digital regulator. Since there is no level shifter for these signals, their amplitude directly depends on the digital supply voltage and a possible level mismatch could cause the analog regulator to fail more often than the digital regulator.

All in all, the regulator IV curve tests attain their goal of identifying chips that are suitable for serial powering tests. Since operation in `Shunt` mode is of low relevance for most other use cases, the regulator IV curve tests are ignored for the overall yield figures.

6.3.3 COMMUNICATION TEST

The communication test consists of assessing the Aurora sync between chip and readout system multiple times, power cycling the chip in between, until communication is established. This

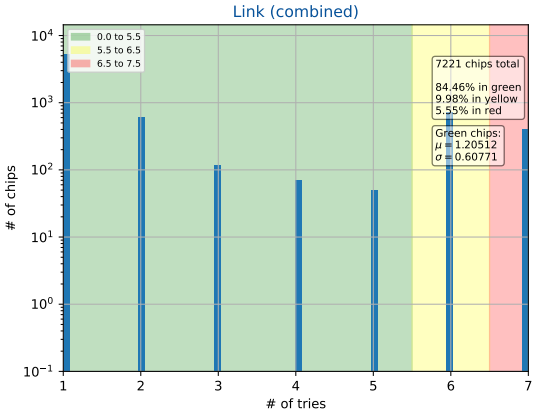


Figure 6.6: Results of the communication test of all tested chips. The background color depicts the regions used to classify the chips. The yield for this test is shown in the top right corner.

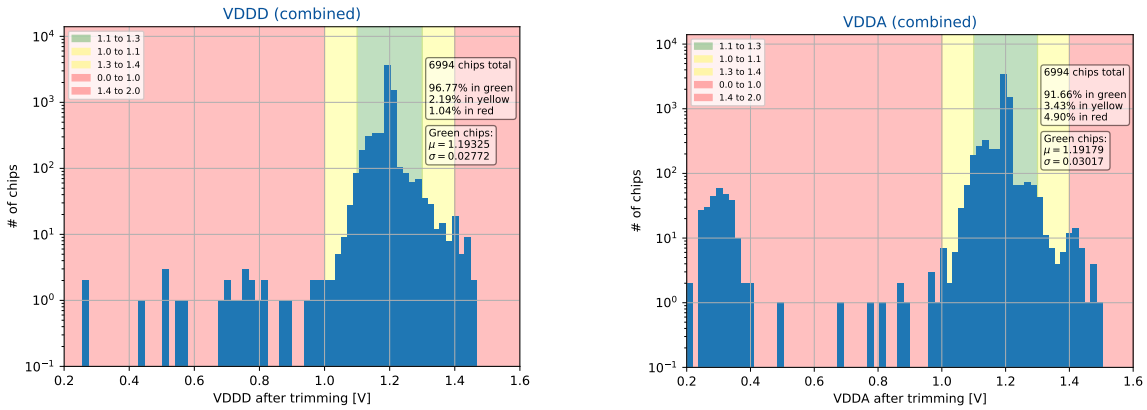
is first tried up to five times with the supply voltages for the PLL and CML circuits VDD_PLL and VDD_CML set to the chip's default (untrimmed) analog supply voltage VDDA, to check if the chip would be able to communicate under normal conditions without any workaround being implemented. If communication starts working within these five tries, the chip is categorized as *working*. Otherwise, VDDA is set to the nominal 1.2 V and Aurora sync is tested again. If communication with the chip only works in this sixth try, the chip is classified as *malfunctioning*, since it can still be used in scenarios where a workaround like the pull-up of VDDA described in Section 5.1 can be implemented. Chips that do not communicate with the readout system, even when powered with the nominal supply voltage, are categorized as *not working*.

The results of this test for all tested chips are shown in Figure 6.6. A fraction of 84.5 % of all tested chips manage to establish communication at the default supply voltage, even though a significant fraction of these chips requires one or more power cycles. About 10 % of the tested chips are generally working, but require a voltage boost of VDDA to establish Aurora sync. About 5.6 % of chips do not communicate at all, even with nominal supply voltage.

6.3.4 VREF TRIMMING

Similar to the IREF trimming test, the capability of each chip to adjust its digital and analog supply voltages, VDDD and VDDA, is assessed by applying every trim bit setting and measuring the resulting voltages. The two supply voltages are tested at the same time to make sure the on-chip power rails do not deviate too much. The range is scanned from the highest to the lowest voltage setting, since a too low voltage will result in loss of communication with the chip.

The results for VDDD and VDDA are shown in Figure 6.7(a) and Figure 6.7(b), respectively. For VDDD, 96.8 % of all tested chips can be trimmed to the nominal value of 1.2 V with an accuracy of ± 100 mV. The distribution resembles a Gaussian with a mean value of 1.19 V and a few outliers that are classified as *not working*. For VDDA, only 91.7 % of tested chips can be trimmed to (1.2 ± 0.1) V. The main peak of the distribution again resembles a Gaussian distribution with



(a) Histogram of the trimmed VDDD values of all tested chips.

(b) Histogram of the trimmed VDDA values of all tested chips.

Figure 6.7: Results of the reference voltage trimming tests. The background colors depict the regions used to classify the chips. The individual yield for each test is shown in the top right corner.

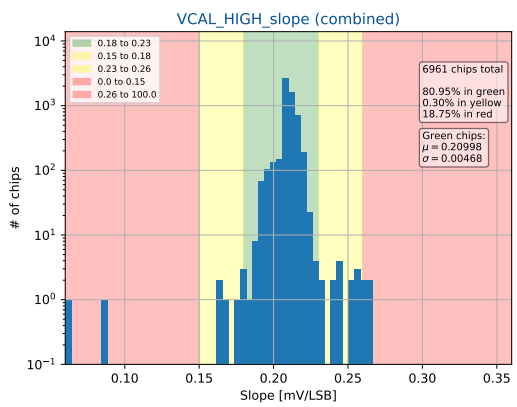
a mean value of 1.19 V. However, a significant secondary peak can be observed at very low voltages, indicating defective analog regulators. These chips are mostly responsible for about 4.9 % of tested chips in total that are categorized as *not working* in this test.

Furthermore, the distribution of the optimal VREF_D and VREF_A trim bits can be analyzed. These results are shown in Figure D.4(a) and Figure D.4(b), respectively. Both distributions show the expected shape, but the mean values of both are shifted towards the higher end of the spectrum, with the mean optimal trim bit setting for the digital supply voltage being about 20 and for the analog supply voltage being about 21. This is another indication for the unexpected, erroneous behavior of RD53A's voltage regulators, discussed previously in Section 5.1.

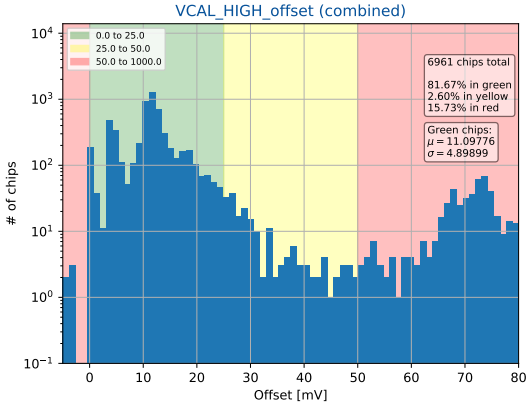
Additionally, the chip's reference voltage for the monitoring ADC and the calibration injection circuit, VREF_ADC is also trimmed and assessed. The result of this test as well as the distribution of optimal trim bits is shown in Figure D.3 and Figure D.4(c), respectively. 92.2 % of tested chips can be trimmed to a nominal value of (0.9 ± 0.1) V, with the mean of the distribution located at 0.89 V. The mean optimal trim bit setting for VREF_ADC is about 12, much closer to the center of the range than for VREF_D and VREF_A.

6.3.5 INJECTION DAC CALIBRATION

Next, the output voltage of the calibration injection DACs, VCAL_MED and VCAL_HIGH is measured for several settings distributed over the DACs' range. The original wafer probing routine only took measurements at settings of 500 and 3500, while the updated routine uses five measurement points at 500, 1250, 2000, 2750, and 3500. Here, the results of both routines are combined. A straight line is fitted to the measured points and the slope and offset values are extracted. For VCAL_HIGH, the distribution of the slope values of all tested chips is shown in Figure 6.8(a) and the offset values are shown in Figure 6.8(b). The corresponding results for VCAL_MED are shown in Figure D.5. A fraction of 81.0 % of all tested chips exhibit a slope of



(a) Histogram of the extracted VCAL_HIGH slope values of all tested chips.



(b) Histogram of the extracted VCAL_HIGH offset values of all tested chips.

Figure 6.8: Results of the calibration measurement for VCAL_HIGH. The background colors depict the regions used to classify the chips. The yield for this test is shown in the top right corner.

VCAL_HIGH in the expected range of 0.18 to 0.23 $\frac{\text{mV}}{\text{LSB}}$ and are therefore categorized as *working*. The mean slope of all tested chips is $(0.210 \pm 0.005)\ \frac{\text{mV}}{\text{LSB}}$, confirming the measured slope on single chips of $(0.22 \pm 0.04)\ \frac{\text{mV}}{\text{LSB}}$, shown in Section 5.1.2 with great statistical accuracy.

Furthermore, a mean offset of (11 ± 5) mV is observed, with 81.7 % of chips exhibiting an offset between 0 and 25 mV and being categorized as *working*.

The same conclusions regarding both yield and slope and offset of the DAC can be drawn for VCAL_MED, based on the results in Figure D.5(a) and Figure D.5(b).

6.3.6 TOTAL POWER CONSUMPTION

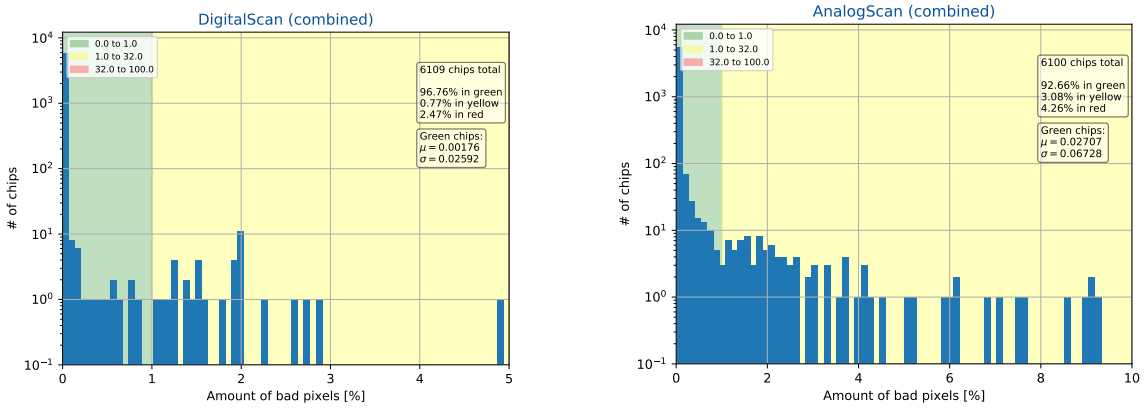
For the total power consumption test, the output current of both channels of the main power supply is measured and added together. Since the chip is powered with a constant voltage in LDO mode, the current consumption is directly proportional to the total power consumption of the chip. The measurement is taken after the configuration step of the Digital Scan. This ensures that for each chip, the same configuration is loaded during the power consumption measurement. Unusually low power consumption is an indication for some kind of malfunction of the chip, which will usually also be caught by other test stages. A higher than usual power consumption, on the other hand, can cause problems in later tests, for example if the chip is to be used in a multi-chip module.

The results of the power consumption test are shown in Figure D.6. The power consumption of all tested chips is well confined between 0.45 A and 0.8 A, with 99.6 % of chips falling into this category. The mean combined digital and analog current drain of all tested chips is (600 ± 16) mA in the given configuration.

6.3.7 Register Test

The first high level scan based on BDAQ53 that is executed during wafer testing is a Register Test as described in Section 4.5. The result of this test of all chips that were probed with the updated routine is shown in Figure D.7. For every chip the Register Test was successfully run on, all 85 tested global registers worked perfectly. However, the Register Test, like all of the following high level tests, is only executed and included in the statistical evaluation, if Aurora sync could be established and the scan could be executed in the first place.

6.3.8 Digital and Analog Scan



(a) Histogram of the Digital Scan results of all tested chips.

(b) Histogram of the Analog Scan results of all tested chips.

Figure 6.9: Results of the Digital and Analog Scans of all tested chips. The x-axis shows the relative amount of bad pixels. The background colors depict the regions used to classify the chips. The yield for this test is shown in the top right corner.

For the Digital and Analog Scan (see Section 4.5), each pixel is categorized as good if it recorded all injected hits and no more, and as bad otherwise. As criterion for the chip performance, the relative amount of bad pixels is then evaluated with regard to the requirements stated in Section 6.1.1. In both cases, 1 % of bad pixels or less is categorized as *working*, while chips with up to 32 % of bad pixels are still categorized as *malfunctioning*. This limit is chosen, since chips with a digitally or analog dead SYNC can still be used for some testing purposes.

For the Digital Scan, whose result is shown in Figure 6.9(a), 96.8 % of tested chips exhibit more than 99 % of digitally working pixels, while the majority of the remaining chips is categorized as *not working*.

The result of the Analog Scan test is shown in Figure 6.9(b). 92.7 % of all tested chips have less than 1 % of dead pixels and are categorized as *working*. In contrast to the Digital Scan, 3.0 % of tested chips are categorized as *malfunctioning*. This is caused mostly by chips with 1 to 3 % of bad, mostly noisy pixels. Finally, 4.3 % of all tested chips are classified as *not working*. The most common failure mode here is chips that report not a single analog hit, hinting at an erroneous calibration injection circuit on these chips.

6.3.9 Threshold Tuning

As a final testing step, a full `Threshold Tuning` is executed (see Section 4.6). While the quality of the tuning result itself is evaluated with the `Threshold Scan` in the following section, the `Noise Occupancy Scan` and `Stuck Pixel Scan`, that are run as part of the tuning routine, are assessed separately.

While the rest of the `Threshold Tuning` test was only added with the updated probing routine, the `Noise Occupancy Scan` was also included in the original routine and is therefore evaluated based on all tested chips. The result of the `Noise Occupancy Scan` test is shown in Figure D.8. Here, 89.6 % of all tested chips exhibit less than 1 % of noisy pixels and are categorized as *working*. The tail of the distribution falls off approximately exponentially, leading to 9.3 % of chips being classified as *malfunctioning* with 1 to 5 % of noisy pixels and 1.1 % of chips being characterized as *not working*.

Figure D.9 shows the result of the `Stuck Pixel Scan`, that was added with the updated routine. Of all chips tested with this new routine, 91.7 % exhibit less than 0.1 % stuck pixels and are therefore classified as *working*. A fraction of 6.7 % of tested chips have 0.1 to 1 % stuck pixels and are categorized as *malfunctioning*. The remaining 1.6 % of chips have more than 1 % of their pixels classified as stuck, which can have a negative impact on testing accuracy when using these chips in modules. They are therefore categorized as *not working*.

6.3.10 Threshold Scan

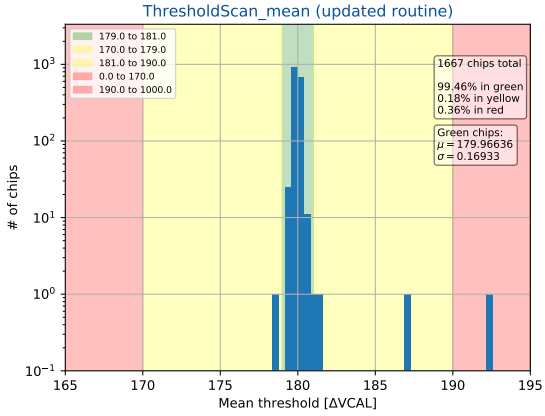


Figure 6.10: Results of the mean threshold from a `Threshold Scan` of all chips tested with the updated probing routine. The background color depicts the regions used to classify the chips. The yield for this test is shown in the top right corner.

The `Threshold Scan` is used to evaluate the quality of the `Global` and `Local Threshold Tuning`, by assessing the resulting mean threshold of the full chip, the threshold dispersion over the full matrix and the relative amount of untunable pixels separately.

The result of the mean threshold test for all chips tested with the updated routine is shown in Figure 6.10. A fraction of 99.5 % of tested chips can be tuned very well to a mean threshold of $(180.0 \pm 0.2)\,\Delta\mathrm{VCAL}$ (about 2 000 e^-), even though all three AFEs are activated simultaneously.

While the original probing routine did not include Threshold Tuning, a Threshold Scan with all three AFEs set to a default value corresponding to a threshold around 330 ΔVCAL (about 3 500 e^-) was always performed, but only the mean value was evaluated for categorization. The result of this test on all chips that were probed with the original probing routine is shown in Figure D.11. The spread is much larger than for the updated routine, since neither the global nor the local threshold of any AFE are tuned. About 80.6 % of chips are categorized as *working* and another 15.8 % are classified as *malfunctioning*. However, since the Threshold Scan is the final test stage, and most chips with extreme results in this test have already failed in an earlier stage, a change of the yield of this test has only minimal influence on the total yield figures.

For the updated routine including global and local Threshold Tuning, however, the total threshold dispersion and the amount of untuned pixels are evaluated as well. Figure D.10(a) shows the result of the dispersion test based on the same Threshold Scan as before. A fraction of 98.9 % of all tested chips exhibit a mean dispersion over the full matrix of less than 5 ΔVCAL . The mean dispersion of all tested chips is (3.7 ± 0.2) ΔVCAL , or about 40 e^-. This result is especially impressive when considering that all three AFEs of RD53A are evaluated simultaneously.

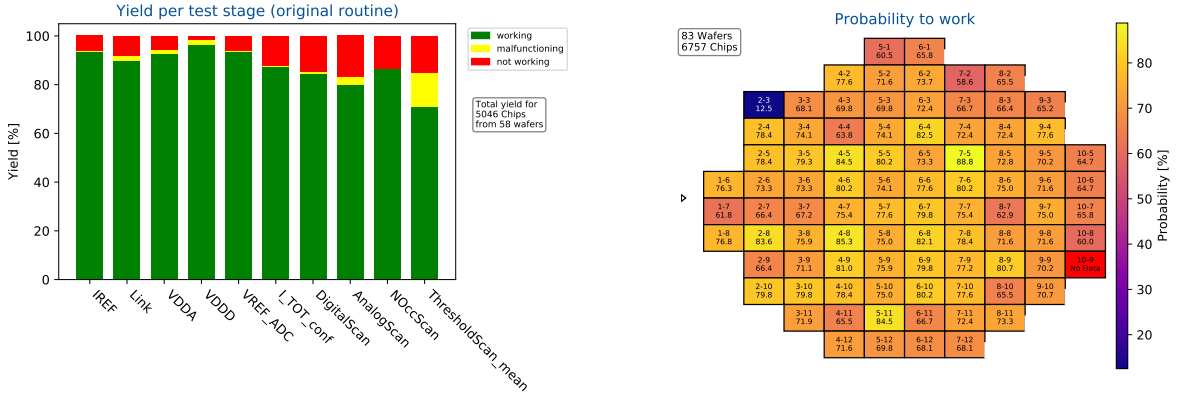
Finally, the result of the untuned pixel tests, again based on the same Threshold Scan as before is shown in Figure D.10(b). Pixels are defined as untuned based on a Threshold Scan after running a Threshold Tuning, if their resulting threshold is either exactly 0.0 ΔVCAL or deviates from the target values by more than a factor of five. Too many untuned pixels can have a negative impact on testing accuracy. A fraction of 93.2 % of all chips tested with the updated routine have less than 5 % of their pixels classified as untuned and are therefor categorized as *working*. Another 5.9 % of the tested chips are categorized as *malfunctioning*, since they exhibit 5 to 10 % untuned pixels. Less than 1 % of the tested chips are categorized as *not working* since they exhibit more than 10 % untuned pixels.

6.3.11 SUMMARY AND YIELD ANALYSIS

For the original probing routine, when combining all of the individual tests presented in the previous sections, a total yield of 64.3 % *working* chips, 16.7 % *malfunctioning* and the remaining 19.1 % *not working* chips can be observed. While chips of the red category may still work to some extend, the label *not working* in this context can also be expressed as *not suitable for module building*. These chips have at least one characteristic weakness that disqualifies them for use in the ATLAS ITk RD53A pre-production program. *Malfunctioning* chips of the yellow category may still be suitable for certain use cases, but require a closer look at their respective issues. Green chips, categorized as *working*, should be suitable for all use cases in the scope of the RD53A pre-production program.

An exception to this rule is the use of chips for serial powering tests, since the characteristic chip attributes for this use case, the regulator IV curves, are excluded from the categorization. This decision was made due to the rather low yield of chips suitable for serial powering and the fact, that serial powering tests are only one testing scenario among many others and the characteristics of the chip's sLDO regulators has little to no influence on other tests.

To ensure that no systematic errors are overlooked in the analysis of RD53A chips, the overall



(a) Summary of yields of the different test stages. The figures for all evaluated test stages are shown.

(b) Probability of a good chip as a function of its position on the wafer. The probability shown is the total number of green and yellow chips in this position divided by the number of tested wafers.

Figure 6.11: Summary of wafer probing results.

yield can be analyzed more closely. Figure 6.11(a) illustrates the total yield for every test stage that is evaluated for the final yield figures. It shows that no single test stage exhibits a drastically lower yield. A lower yield can be observed for higher level tests and scans, starting with the total power consumption test or rather the `Digital Scan`, since these tests already require working Aurora sync. Therefore, a slightly lower yield is expected for these stages. The large amount of chips categorized as *malfunctioning* in the `Threshold Scan` mean threshold test is caused by the fact, that the original probing routine did not include a `Threshold Tuning`, but rather just a `Threshold Scan` with default values. This result therefor reflects the full range of untuned default values under the influence of process variations.

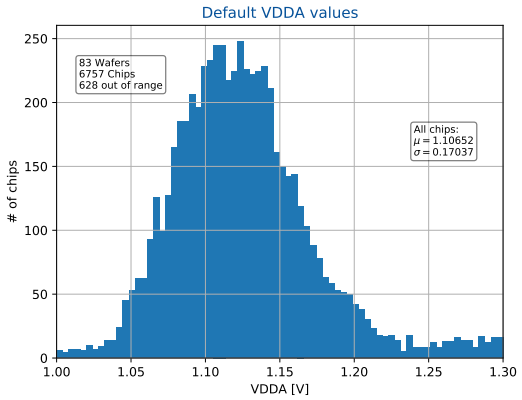
In Figure 6.11(b), the mean probability for a chip to be categorized as *working* is depicted as a function of the chip's position on the wafer. The distribution is rather homogeneous over the wafer, with probabilities ranging from about 60 % to almost 90 %. Exceptions are the two chips at positions 10-9 and 2-3, which have a probability of 0.0 % and 12.5 %, respectively. These chips are located very close to the wafer edge and are typically not fully processed. While they are included in the wafer map for the sake of completeness, their results are usually excluded from any yield analysis.

No significant gradient or hotspots can be identified on the wafer map, indicating that the mean probability of a chip to work is not correlated with its position on the wafer. A very faint radial gradient can be observed, with the chips at the center of the circular wafer having a marginally higher yield than the chips located at the wafer edges. However, this is a common pattern in semiconductor processing and not a systematic error of the presented wafer probing setup.

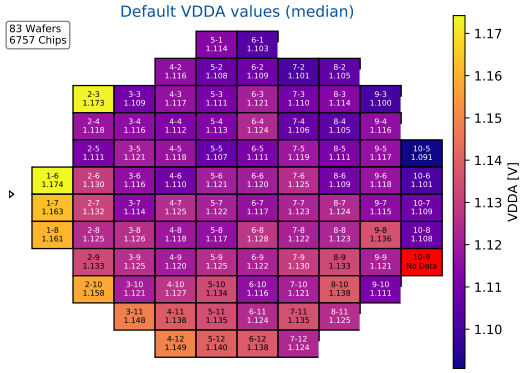
6.4 Special Analyses

The large number of RD53A chips tested during probing of almost 100 wafers allows for a statistical analysis of more interesting phenomena that would otherwise remain hidden in the statistical uncertainties. One of several examples for such an advanced analysis is the deeper investigation of process variations. The following section describes the analysis of the variation of the default supply voltages of RD53A as a consequence of such a statistical process variation.

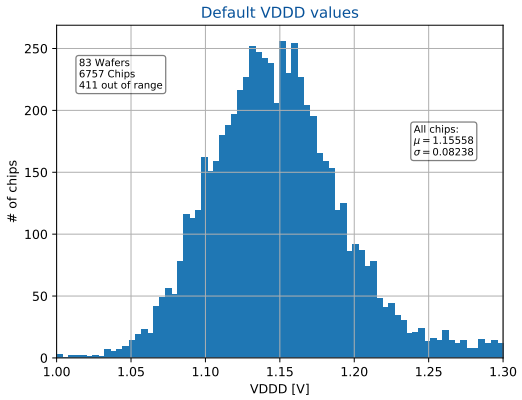
6.4.1 Analysis of Supply Voltages



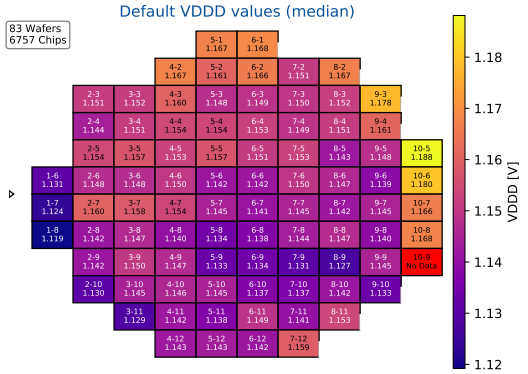
(a) Histogram of the untrimmed default VDDA values of all probed chips.



(b) Untrimmed VDDA values by position on the wafer. The notch of the wafer is located on the left hand side of the figure. In each chip position, the median of the untrimmed default VDDA values is depicted as numeric value and via the color scale.



(c) Histogram of the untrimmed default VDDD values of all probed chips.



(d) Untrimmed VDDD values by position on the wafer. The notch of the wafer is located on the left hand side of the figure. In each chip position, the median of the untrimmed default VDDD values is depicted as numeric value and via the color scale.

Figure 6.12: Untrimmed VDDA and VDDD values of all probed chips.

Figure 6.12(a) shows the distribution of the untrimmed VDDA values of all tested chips. Here, untrimmed refers to the value at a trim bit setting of 16, which is given directly after powering on the chip and which is responsible for the initial establishing of communication between the DAQ system and the chip. The distribution peaks at 1.11 V, marginally lower than the nominal value of 1.20 V, which was the expected mean value from design simulations. Despite the unprecedented sample size for RD53A, the significance of this observation is still low with a standard deviation of 0.17 V. However, the lower mean value definitely has some implications for operating the chip, as discussed in Section 5.1. Additionally, the width of the distribution is surprisingly large compared to simulations. The reason for this is apparent when plotting the mean untrimmed value as a function of the chip's position on the wafer, as is shown in Figure 6.12(b). The mean untrimmed value varies from about 1.09 V in the upper right to about 1.17 V in the lower left region of the wafer. This effect is probably caused by typical variations in doping concentrations during production, which can show a linear or radial gradient over the wafer, depending on the exact production mechanisms involved. This is even more plausible, since in LDO mode, as is the case during this measurement, VDDA is determined by the regulator's Band Gap circuit, which is very sensitive to changes in doping concentration.

The corresponding plots for the digital supply voltage VDDD are shown in Figures 6.12(c) and 6.12(d). Curiously, while the wafer map of the mean VDDD values shows the same gradient, its sign is inverted. Further investigation revealed that the layout of the analog and digital regulator circuits are in the physical wafer mask of RD53A, explaining the opposite influence of the same production variations on the two Band Gap circuits.

6.5 CONCLUSION

An extensive testing routine has been developed and executed on a large number of RD53A chips at wafer level. The test results have been evaluated in order to classify chips into three categories: *Working*, *malfunctioning* and *not working*. This categorization is heavily relied upon for chip selection during module production in the context of the ATLAS ITk RD53A pre-production program.

Overall, almost 9 000 RD53A chips have been tested with a specialized, semi-automatic wafer probing setup that has been developed in this thesis. Detailed analyses of the results of more than 7 000 of these chips have been presented. About 80 % of these chips can be used for module production, bench testing, irradiation campaigns and other activities.

Furthermore, the large sample size has made detailed analyses possible and provided insights that improved the understanding of the chip and the associated CMOS process. This further facilitated valuable feedback for the design process of the successors of RD53A on the way to producing the pixel detector readout ASIC for the ATLAS Inner Tracker (ITk).

Last but not least, the setup served as preparation for wafer probing within the Quality Control (QC) efforts during module production for the ATLAS Phase-II upgrade. At the time of writing, a probe card for the updated pad frame of RD53B has become available and the setup is being adapted to ITkPix-V1.

CHAPTER 7

PIXEL DETECTOR MODULES BASED ON RD53A

Different types of assemblies and modules based on RD53A are produced within the context of the ATLAS ITk pre-production program (refer to Section 1.2.3). In this thesis, the first available RD53A Single-Chip Modules (SCMs) were characterized and the results are presented in this chapter. Furthermore, RD53A-based Double Chip Modules (DCMs) were investigated in the context of the hybridization MS, which is the subject of Chapter 8. Other types of modules using RD53A are being produced to evaluate other aspects of the final product as well as the production work flow. Examples include special Quad-Chip Modules (QCMs) for serial powering tests [43] as well as QCMs with full ITk-sized sensors that are used to simulate the production flow during the mass production stage.

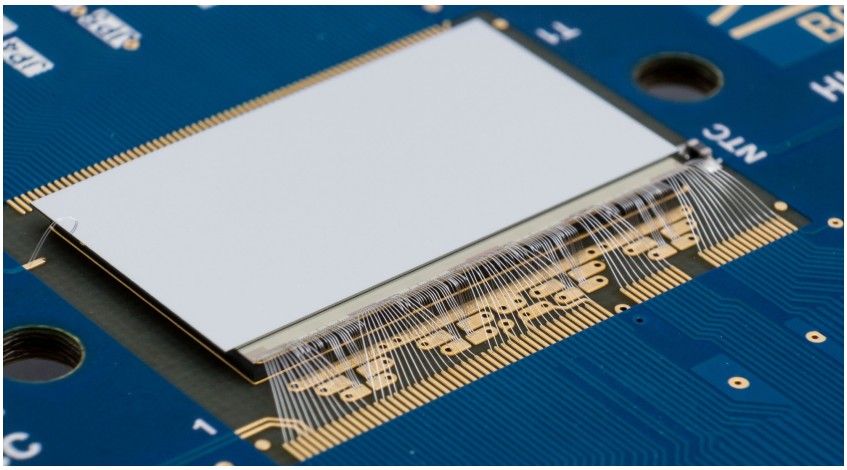


Figure 7.1: An RD53A Single-Chip Module (SCM) on an SCC.

RD53A SCMs consist of a single RD53A readout chip that can be optionally thinned down from the back side to about 150 µm and a single sensor tile of roughly the same dimensions as the readout chip. Sensors are typically thinned down to 100 µm or 150 µm. Figure 7.1 shows a typical RD53A SCM on a Single-Chip Card (SCC). The thinned sensor is bump-bonded to a full-thickness ($\approx$ 775 µm) RD53A chip. The assembly is then glued to a standard RD53A SCC.

The objective of the investigation of these SCMs is to characterize RD53A as part of a hybrid

pixel detector. Consequently, the focus of this chapter is the comparison of the modules' performance to the bare chip's performance established in Chapter 5. The presented measurements therefore concentrate on readout chip performance instead of sensor performance.

7.1 SENSORS

In the context of the ATLAS ITk pre-production program, conventional planar and 3D sensors, as well as passive CMOS sensors are being investigated.

Generally, the sensor pixel size can vary, but the pitch of its bump-bond pads has to match the 50 µm × 50 µm pitch of RD53A. The most basic sensor design approach translates the bump-bond pitch directly to a 50 µm × 50 µm sensor pixel pitch. Alternatively, the sensor design can make use of rectangular pixels, for example with dimensions of 25 µm × 100 µm, with two sensor pixels on top of each other corresponding to two horizontally neighboring pixels in the readout chip. This approach requires an additional redistribution layer in the sensor design to map the pixels bump-bond connections to RD53A's 50 µm × 50 µm pitch as well as additional processing of the raw data in software to reflect the correct geometry in the data analysis.

The RD53A SCMs investigated in this thesis almost exclusively employ conventional planar n-in-p silicon sensors from the HLL[1] SOI-3 and SOI-4 productions [67, p.31 ff.].

7.2 CONNECTIVITY

Different approaches exist to connect a module to the DAQ system. For basic testing purposes, the same SCC (refer to Section 3.2.2) that was used for mounting and testing bare readout ASICs can be employed. In addition to the bottom-row wire-bonds, which are needed for communication with the chip, the sensor back side needs to be contacted as well in order to apply a reverse bias voltage to the sensor. This is realized on the SCC by multiple additional wire-bonds in the top left corner of the module, connecting the sensor back side to a dedicated high voltage (HV) connector. This can be seen on the far left of Figure 7.1. The top-row wire-bond pads of RD53A are not accessible in assemblies and modules, since the sensor slightly overlaps the top edge of the readout chip.

In the context of this thesis, the SCMs presented in this chapter were mounted and tested on SCCs exclusively.

For use in the ATLAS ITk, however, a dedicated, minimal flexible PCB (flex) is used to connect the module to the DCS and DAQ systems. This flex is typically considered to be part of the complete module, since it is permanently glued to the sensor back side during module assembly. In the context of the RD53A pre-production program, different types of flexes have been developed. Different flexes are used for RD53A-based Double- and Quad Chip Modules, due to the different requirements regarding connections and physical size. Additionally, a dedicated flex has been designed for use with special QCMs for serial powering chain tests [43].

[1] Halbleiterlabor - Semiconductor Laboratory of the Max-Planck-Gesellschaft, Munich, Germany

7.3 RESULTS

Performance characteristics of RD53A SCMs are evaluated for two main reasons. The performance of the three different AFE flavors needs to be confirmed with a physical sensor attached to the readout chip, in order to be able to make a well-founded decision. Bare chip performance is a good indicator for comparing the three flavors to each other but in the actual experiment, the readout chip needs to work together with a sensor as a complete module and large performance limitations of the AFE connected to a sensor compared to the bare chip performance should not stay unnoticed. Secondly, since the SCMs were the first RD53A-based modules available, they are used to assess whether RD53A is suitable as a basis for the characterization of other building blocks and processing steps, such as sensors or the module hybridization within the scope of the pre-production program.

7.3.1 SENSOR IV SCAN

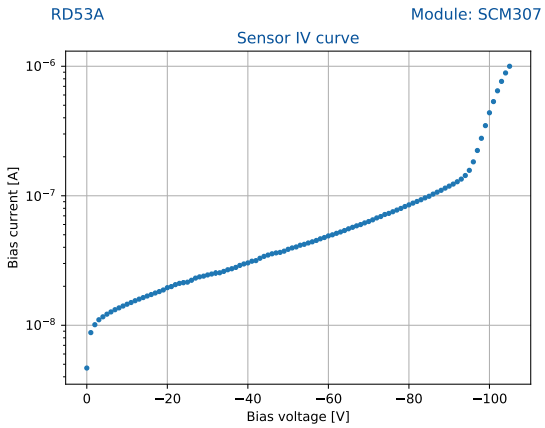


Figure 7.2: IV curve of the tested SCM 307 with sensor W10-8-50x50 from the HLL SOI-3 production, measured at room temperature.

The performance metrics of RD53A in a module should be determined under realistic conditions, including significant leakage current, as can be expected during operation of the modules in the experiment. Therefore, the first measurement conducted when characterizing a module is the sensor IV curve, illustrating the leakage current as a function of the applied reverse bias voltage. The IV curve of a typical RD53A SCM with a 150 µm thick sensor from the HLL SOI-3 production with $50 \times 50\,\mu m^2$ pixels is shown in Figure 7.2. On a logarithmic scale, the curve shows a linear behavior between a bias of a view volts and about 95 V, where the sensor breaks down. Full depletion for this module type is typically achieved around 30 to 40 V, though the typical curvature and saturation of the curve is not visible in this example. This is probably caused by larger leakage current contributions covering this feature of bulk leakage current. For all further measurements presented in this chapter, a bias voltage of 80 V was applied, leading to a total leakage current of about 100 nA while fully depleting the sensor volume.

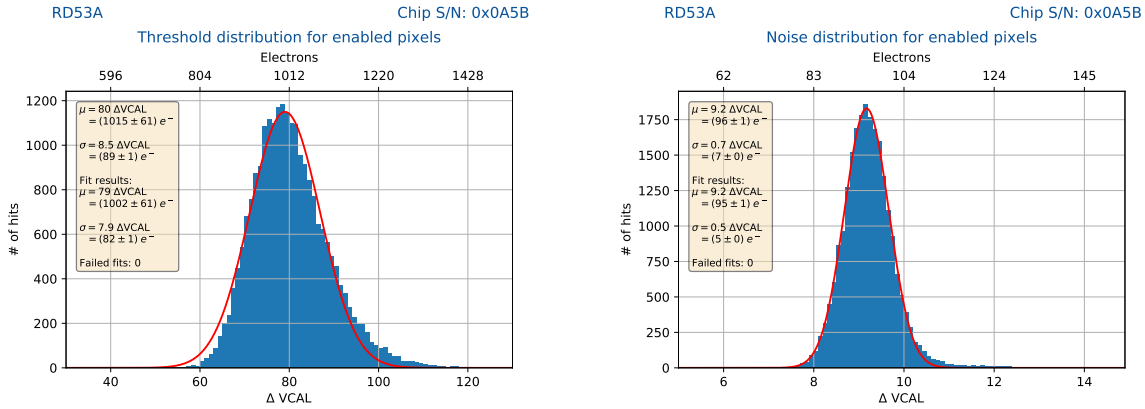
While for the SYNC and LIN, the Krummenacher feedback circuit should automatically

prevent the AFE input from saturating, the DIFF implements a manual LCC circuit. According to the RD53A manual [33, p.21], enabling the LCC circuit is only required for leakage currents in the order of 10 nA per pixel. However, enabling the circuit for lower leakage currents should not have any negative effects and its general functionality needs to be verified. Therefore, it was enabled and set to LCC_DIFF = 20 for the following measurements. In order to test the leakage current compensation, highly irradiated sensors would be required, which were not available at the time of testing.

7.3.2 ANALOG FRONT-END PERFORMANCE WITH SENSOR

To get a better picture of the performance of the three different AFE flavors of RD53A in a more realistic scenario as part of a module, the main performance metrics shown and discussed in Chapter 5 are measured again on a module. In the following measurements, a bias voltage of 80 V was applied to the sensor, leading to a total leakage current of about 100 nA. All tests were conducted at room temperature.

SYNCHRONOUS FRONT-END



(a) Threshold distribution of the SYNC in an SCM with optimal AZ frequency.

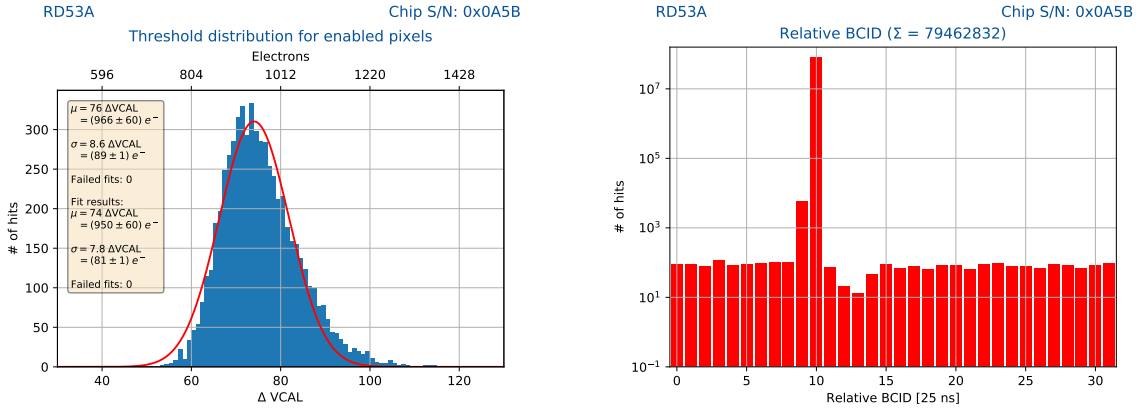
(b) Noise distribution of the SYNC in an SCM with optimal AZ frequency.

Figure 7.3: Threshold and noise results of the SYNC in an SCM after tuning to a mean threshold of about 1 000 e⁻.

To compare the threshold and noise performance of the AFEs to the bare chip performance presented in Chapter 5, the same procedure was applied. The AFE flavors are tuned to a threshold of about 1 000 e⁻ separately, and a threshold scan is performed, yielding the resulting mean threshold, threshold distribution and mean ENC. These results can be compared to the corresponding values of the bare chip to investigate the influence of the sensor pixel's additional capacitance and leakage current.

The results of this threshold scan for the SYNC after tuning the global threshold to 1 000 e⁻ is shown in Figure 7.3. The associated maps are shown in Figure D.12. AZ commands were sent with a period of about 80 µs. The resulting mean threshold almost perfectly matches the

tuning target and the threshold distribution looks as expected from previous measurements, including the gradient from top to bottom in the threshold map that is discussed in Section 5.3.2. The mean threshold dispersion of the SYNC in a module at a threshold of 1 000 e^- is about $(82 \pm 1)\,e^-$, about 16 e^- larger than for a bare chip. The ENC of the SYNC in a module is about $(95 \pm 5)\,e^-$, about 32 e^- higher than on a bare chip. This is generally expected, since the sensor pixel adds a fluctuating contribution to the input capacitance, the main factor for threshold dispersion as well as noise. Besides, even low levels of leakage current as in this case can have a negative influence on the AFE's performance, despite of the Krummenacher circuit that is implemented in the SYNC for leakage current compensation.



(a) Minimum stable threshold of the SYNC in an SCM.

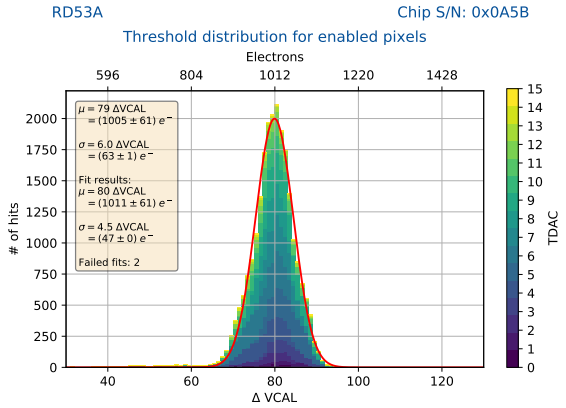
(b) Relative BCID distribution at the minimum stable threshold of the SYNC in an SCM.

Figure 7.4: Minimum stable threshold results of the SYNC in an SCM.

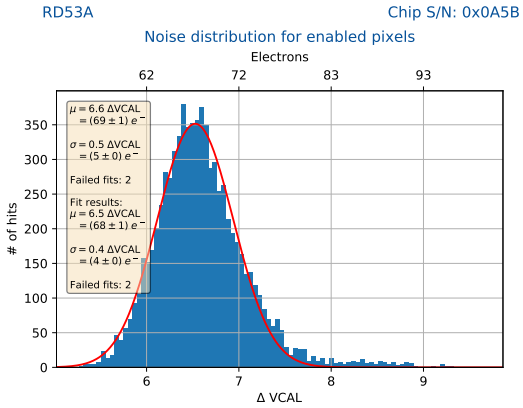
The minimum stable threshold of the SYNC in a module is about $(966 \pm 60)\,e^-$, as shown in Figure 7.4. This is significantly higher than the minimum stable threshold of $(706 \pm 60)\,e^-$ that could be reached with a bare chip. This is another effect of the increased noise on the module compared to a bare chip.

LINEAR FRONT-END

The result of a threshold scan of the LIN on the SCM after tuning to about 1 000 e^- is shown in Figure 7.5. The mean of the threshold distribution in Figure 7.5(a) is very close to the tuning target and the associated threshold map in Figure D.13(a) shows no gradient or irregularities. The threshold dispersion of the LIN on the module is $(47 \pm 1)\,e^-$, only 4 e^- higher than for the bare chip. The mean ENC of the LIN in a module is $(68 \pm 4)\,e^-$, as shown in Figure 7.5(b). This is 12 e^- more than on a bare chip. Additionally, the distribution exhibits a long tail towards higher values, indicating a growing amount of noisy pixels. These noisy pixels are randomly distributed, as can be seen in the associated noise map in Figure D.13(b). Compared to the SYNC, the results of the LIN in a module show the same tendencies, but the effects are less pronounced. This leads to the conclusion, that the LIN is able to cope with the aggravated conditions in an actual hybrid pixel detector module better than the SYNC.

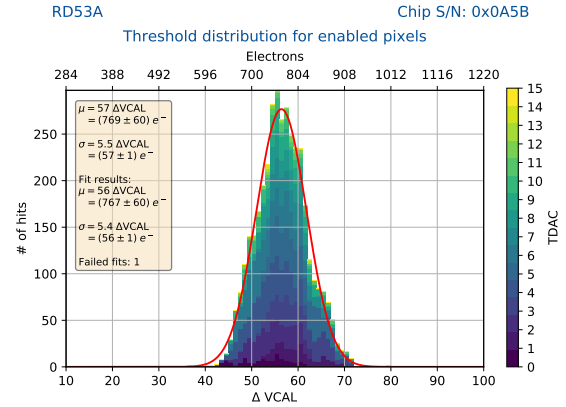


(a) Threshold distribution of the LIN in an SCM after tuning.

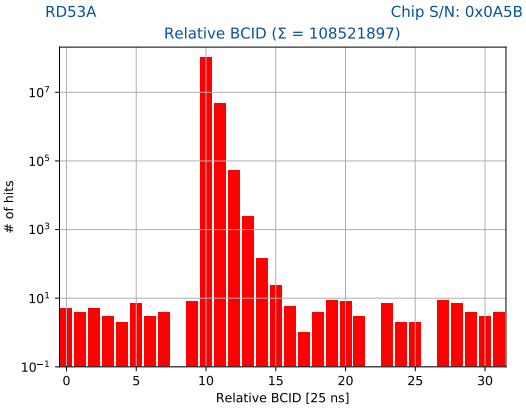


(b) Noise distribution of the LIN in an SCM after tuning.

Figure 7.5: Threshold and noise results of the LIN in an SCM after tuning to a mean threshold of about 1 000 e⁻.



(a) Minimum stable threshold of the LIN in an SCM.



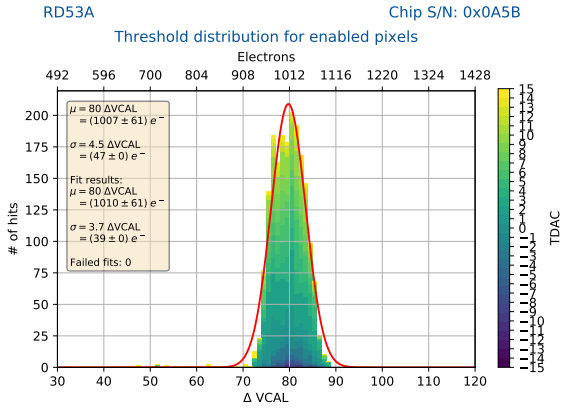
(b) Relative BCID distribution at the minimum stable threshold of the LIN in an SCM.

Figure 7.6: Minimum stable threshold results of the LIN in an SCM.

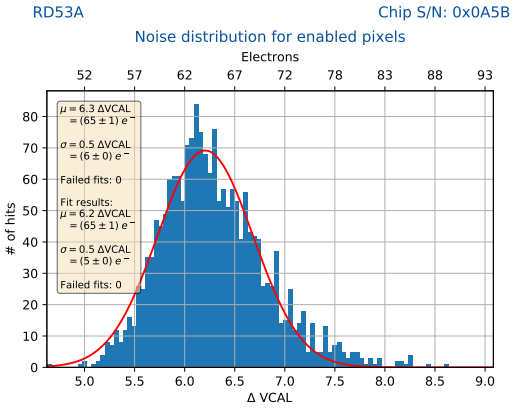
Figure 7.6 shows the minimum stable threshold of the LIN in an SCM. The `Noise Tuning` algorithm resulted in a threshold of (767 ± 60) e⁻, 135 e⁻ higher than the minimum stable threshold achievable on a bare chip. Again, these results show that the influence of the sensor on the AFE's performance is less pronounced than for the SYNC.

DIFFERENTIAL FRONT-END

The threshold distribution for the DIFF in an SCM after tuning both global and local threshold to about 1 000 e⁻ is shown in Figure 7.7(a), with the associated threshold map shown in Figure D.14(a). The target threshold was reached accurately and the mean threshold dispersion is (39 ± 1) e⁻, an increase of 13 e⁻ compared to the bare chip's dispersion of (26 ± 1) e⁻. The mean



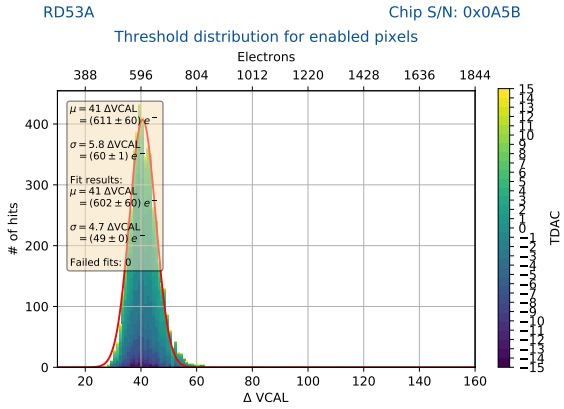
(a) Threshold distribution of the DIFF in an SCM after tuning.



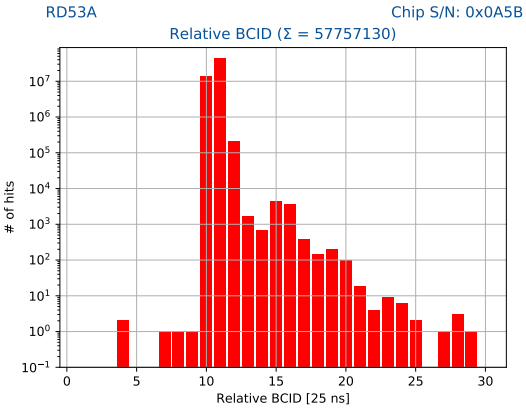
(b) Noise distribution of the DIFF in an SCM after tuning.

Figure 7.7: Threshold and noise results of the DIFF in an SCM after tuning to a mean threshold of about 1 000 e⁻.

ENC of the DIFF in a module at a threshold of 1 000 e⁻, shown in Figure 7.7(b), increased by 26 e⁻ compared to a bare chip, to (65 ± 5) e⁻. The noise map is presented in Figure D.14(b) and shows no remarkable features. The active leakage current compensation circuit of the DIFF was enabled and set to `LCC_DIFF` = 20 for this measurement. While the sensor's influence on the threshold dispersion of the DIFF is comparably low, the increase of the noise is rather large.



(a) Minimum stable threshold of the DIFF in an SCM.



(b) Relative BCID distribution at the minimum stable threshold of the DIFF in an SCM.

Figure 7.8: Minimum stable threshold results of the DIFF in an SCM.

The minimum stable threshold of the DIFF in an SCM is shown in Figure 7.8. Like in the previous measurement, the DIFF's active leakage current compensation was enabled and set to `LCC_DIFF` = 20. Compared to the bare chip, the minimum stable threshold increased by about 184 e⁻ to (602 ± 60) e⁻. As for the mean ENC, the performance decrease of the DIFF ranges in-between the LIN and the SYNC.

AFE PERFORMANCE SUMMARY

A comprehensive comparison of the performance figures discussed in the previous sections is shown in Table 7.1. It illustrates the changes in key performance parameters as a consequence of an increased sensor capacitance and leakage current compared to the bare front-end chip.

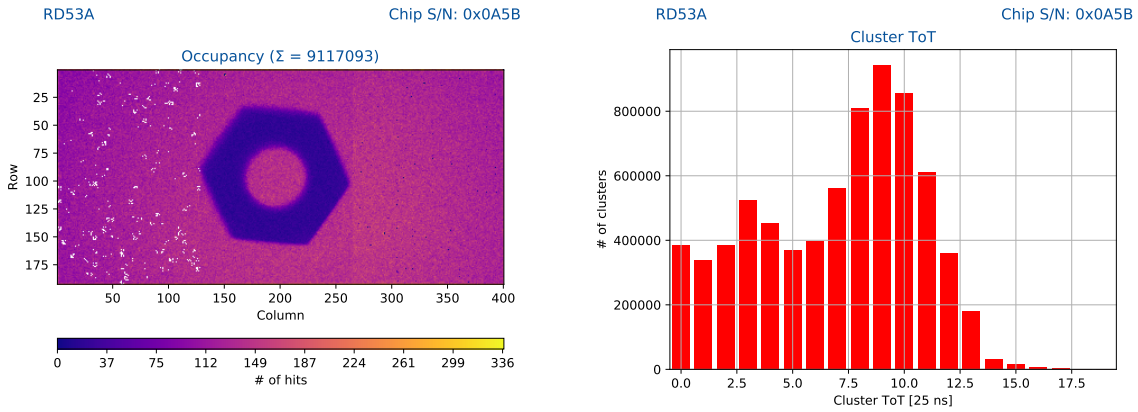
	SYNC	LIN	DIFF (*good pixels*)
Threshold dispersion at 1 000 e^-	$(82 \pm 1)\,e^-$	$(47 \pm 1)\,e^-$	$(39 \pm 1)\,e^-$
Change of threshold dispersion	$16\,e^-$	$4\,e^-$	$13\,e^-$
Mean ENC at 1 000 e^-	$(95 \pm 5)\,e^-$	$(68 \pm 4)\,e^-$	$(65 \pm 5)\,e^-$
Change of mean ENC	$32\,e^-$	$12\,e^-$	$26\,e^-$
Minimum stable threshold	$(966 \pm 60)\,e^-$	$(767 \pm 60)\,e^-$	$(602 \pm 60)\,e^-$
Change of min. stable threshold	$260\,e^-$	$135\,e^-$	$184\,e^-$

Table 7.1: Summary of AFE characterization results on an SCM.

Based on these results, the conclusion can be drawn that the LIN is best able to cope with the changed conditions. The increase of the threshold dispersion, mean ENC, as well as the minimum stable threshold in an SCM compared to the bare chip results are the lowest of the three AFEs of RD53A. However, due to the much better performance in a bare chip, the DIFF is still the best performing AFE in a module, when comparing the absolute values. The SYNC, on the other hand, suffers most from the increased capacitance and leakage current, increasing its already high metrics by the largest amount of the three flavors. Further investigation, especially with irradiated sensors generating several orders of magnitude more leakage current would be necessary to draw a complete picture of the three front-ends' performance in a long-running high-energy particle physics experiment. However, due to scheduling reasons and the urgency for ATLAS and CMS to make a decision for one of the AFEs, in order to continue with the design of the production chips, these time consuming measurements were never conducted and the decision was based on expectations extrapolated from simulations and the results presented in this section.

7.3.3 SOURCE SCANS

To illustrate the general functionality of an RD53A SCM as a particle detector, the module is irradiated with a ^{241}Am X-ray source. The module is operated in self-trigger mode, meaning the necessary trigger commands are generated based on the HitOr output (see Section 4.3.3) and all three AFEs of the module were tuned to a mean threshold of about 3 500 e^- and a ToT response of 10 ToT at 16.5 ke^-, in order to optimally represent the main peak of the ^{241}Am spectrum at 59.6 keV. To verify the correct pixel mapping and general functionality, an isolated M3 screw nut was placed on the sensor back side. The resulting hitmap is shown in Figure 7.9(a). The figure demonstrates a good homogeneity of the hit distribution, allowing conclusions about the homogeneity of the threshold distribution across the three AFEs to be drawn. The shape of the screw nut is visible in the center of the hit map, demonstrating that the pixel mapping is correct and the module can be operated as a particle detector. Additionally, this is one of the first measurements demonstrating the operation of RD53A in self-trigger mode. The clustered ToT distribution of this source scan is shown in Figure 7.9(b). It exhibits a prominent peak at



(a) Hitmap of a `Source Scan` of an `RD53A` SCM.

(b) Clustered ToT distribution of the same `Source Scan`.

Figure 7.9: `Source Scan` of an `RD53A` SCM irradiated with a ^{241}Am X-ray source. A screw nut was placed on top of the sensor to illustrate the module's functionality.

9 ToT, slightly lower than the expected 10 ToT. This is expected due to clustering- and threshold effects. The recorded spectrum nicely represents the expected spectrum of ^{241}Am, given the limited accuracy of the ToT tuning of `RD53A`.

In a nutshell, this measurement represents a simple but comprehensive validation of `RD53A`'s capability to be used as readout ASIC in a hybrid pixel detector module. It can be tuned to an almost perfectly homogeneous response over the full matrix and the charge spectrum is plausible. Self-triggered operation of `RD53A` together with `BDAQ53` has been confirmed to be a viable option for measurements.

7.3.4 TEST BEAM CAMPAIGNS

Several test beam campaigns with `RD53A` modules have been conducted within the scope of this thesis. The first campaign was organized in July 2018, two months after the first `RD53A`-based SCMs were produced, at the CERN SPS North Area test beam facility. Here, a secondary beam from the SPS, consisting of electrons, hadrons and muons with a variable momentum of 10 to 400 GeV/c is available for the users. The goal of this first test beam was first of all to again verify the viability of `RD53A` as readout ASIC for any modules that would be produced within the scope of the ATLAS ITk pre-production program. The second important goal of this test beam campaign was to verify and finish the development of `BDAQ53`, to ensure the DAQ system was ready for measurements in test beams in conjunction with a beam telescope, including integration into larger readout frameworks like EUDAQ [68].

The result of an overview measurement from this first test beam campaign demonstrating `BDAQ53`'s capabilities is presented in Figure 7.10. In this simple measurement, an `RD53A`-based SCM with a 150 µm thick HLL sensor with a pixel pitch of 25 × 100 µm^2 (`W8-2-25x100`) was operated in self-trigger mode and irradiated in the beam. However, in this measurement, a special hitbus mask, that is the mask defining which pixels' output is accepted to create a trigger, is applied. the mask was chosen to resemble the official `RD53A` logo. This creates an

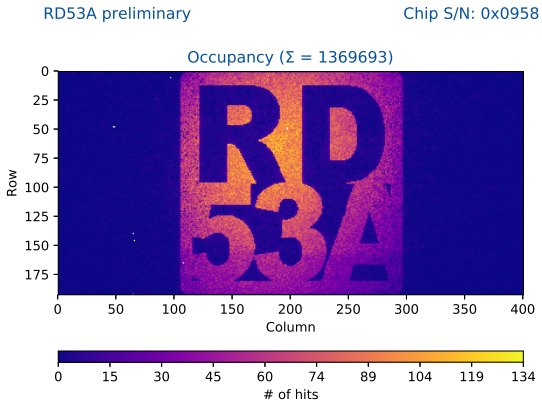


Figure 7.10: Hitmap of a first measurement in the 2018 CERN SPS test beam campaign. The RD53A SCM is operated in self-trigger mode and triggers are only accepted from certain pixels resembling the RD53A logo.

overlay of the shape of the logo with the actual SPS beam profile, impressively demonstrating BDAQ53's simple masking capabilities and ability to cope with the high hit rates of a test beam measurement.

Furthermore, during this first test beam campaign, several RD53A SCMs were operated through BDAQ53 together with ACONITE [69], a 6-plane EUDET-type beam telescope, to perform the first particle detection efficiency measurements with an RD53A-base module. Based on these measurements and developments during the first test beam campaign, both, BDAQ53 and RD53A could be established as viable options for multiple Market Surveys and further characterization measurements within the ATLAS ITk pre-production program.

7.4 CONCLUSION

The measurements and results presented in this chapter have been incorporated into several decision processes, including the Analog Front-End (AFE) reviews summarized in Section 5.7. The performance results of the AFEs showed that the three flavors differentiated themselves further by exhibiting different susceptibility to an increased sensor capacitance and leakage current. All three flavors showed a small deterioration in all key performance parameters, with the LIN showing the lowest disparity between module and bare chip performance. However, all three AFE flavors still nominally fulfill the posed requirements with ease. Additionally, different measurements demonstrate the functionality and viability of RD53A as a readout ASIC for hybrid particle detector modules. This is important groundwork for the ATLAS ITk pre-production program that is started mostly based on RD53A due to scheduling reasons. RD53A can be tuned to a uniform and homogeneous threshold over the full matrix and has successfully been used with radioactive sources and in test beams to evaluate sensors. This is an important fact to confirm and reassure, since RD53A-based modules will be used in both, the official ITk sensor Market Survey (MS) as well as the official hybridization MS. The latter is presented in detail in the following Chapter 8.

CHAPTER 8

QUALIFICATION OF THE HYBRIDIZATION PROCESS

During the production of the ATLAS ITk, almost 15 000 bare hybrid pixel detector modules will be produced by one or several industrial vendors. To ensure a smooth and timely production process, these vendors have to be selected carefully with regard to their ability to produce and deliver the required quality and quantity of modules within the given scheduling constraints. The quality of the vendor's processing and hybridization techniques and the resulting performance of the produced modules are particularly important and have to be assessed for all considered vendors in order to enable a well-founded decision.

This is the scope and objective of the hybridization Market Survey (MS), which is conducted as a two-step process. As a first step in 2018, potential suppliers were identified, based on a technical questionnaire as well as detailed discussions between the procurement offices of CERN, ATLAS and CMS and the vendors. Based on the results of this pre-evaluation, seven companies were invited to tender and apply for a prototype production order.

Of these seven companies selected for step two of the MS, two subsequently withdrew their offers as flip-chip vendor, due to financial and production process concerns. The remaining five vendors are assessed in detail in the second step of the MS. In the following, they are referred to by their original vendor codes from step one, `Vendor A`, `Vendor D`, `Vendor G`, `Vendor L` and `Vendor M`. Of these vendors, `Vendor D` and `Vendor G` had to form a consortium in order to fulfill all the required processing steps, while the other three vendors were able to perform all processing in-house. However, since some of the employed sensor foundries supplied sensors already including under-bump metallization (UBM), and two of the flip-chip vendors also served as sensor foundries, the resulting matrix of possible combinations becomes quite complex. In the end, nine different combinations of vendors were accepted and assessed, as shown in Table 8.1.

In step two of the hybridization MS, each of the five remaining vendors was tasked with processing five sensor wafers and five front-end chip wafers as well as the delivery of 20 flip-chipped Double Chip Modules (DCMs). The delivered modules were then characterized according to a predefined qualification procedure and acceptance criteria [70].

The Silab Bonn is one of the four qualified assembly and testing sites for the hybridization MS. The QA and QC measurements of 33 Double Chip Modules (DCMs) have been conducted for the MS within the scope of this work. This chapter first gives a general introduction to

Comb.	Flip-chip Vendor	Bump Vendor	UBM Vendor	Sensor Vendor
C1	Vendor A	Vendor A	Foundry 2	Foundry 2
C2	Vendor A	Vendor A	Vendor A	Foundry 3
C3	Vendor A	Vendor A	Foundry 6	Foundry 6
C4	Vendor D	Vendor F	Vendor D	Foundry 4
C5	Vendor D	Vendor L	Vendor D	Foundry 4
C6	Vendor G	Vendor F	Vendor G	Foundry 1
C7	Vendor L	Vendor L	Vendor L	Foundry 6
C8	Vendor M	Vendor M	Vendor M	Foundry 3
C9	Vendor M	Vendor M	Vendor M	Foundry 6

Table 8.1: Summary of vendor combinations for the hybridization MS.

the hybridization processes of hybrid pixel detectors and the RD53A-based DCMs used for the hybridization MS. Next, the MS requirements and the test methodology employed to characterize the DCMs are defined. Finally, this chapter describes the results acquired for the hybridization MS within the scope of this work and the conclusions that can be drawn from them.

8.1 HYBRIDIZATION

Hybridization, the construction of a hybrid pixel detector module from the two basic components, readout ASICs and passive sensor tile, is a complex process with multiple stages. First, both wafers undergo additional processing steps to reach the desired thickness and connectivity and to enable the connection of both chips. For the sensor wafer, these processing steps include thinning and backside metallization, if appropriate, and under-bump metallization (UBM). In addition to these processing steps, the bump-bonds are typically deposited on the thinned front-end chip wafer. Since the hybridization of silicon pixel particle detectors is a comparatively complex but low-volume process, compared to industry standards, die-to-die bonding is usually employed instead of die-to-wafer or even wafer-to-wafer bonding, which would not easily allow for the production of multi-chip modules. Hence, both, the sensor- and front-end chip wafers are diced in order to obtain the bare dies, which are then bonded together in a final step called *flip-chip* process.

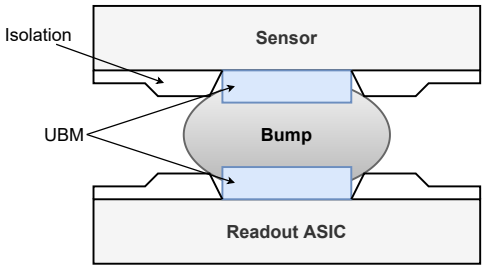


Figure 8.1: Schematic of a single bump-bond connection.

Figure 8.1 shows the schematic view of a single, finalized bump-bond connection. Within the respective opening of the passivation layer on both, the sensor tile and the front-end chip, the UBM is applied to enable a firm mechanical and electrical bond of the bump with both chips. The bump material is then deposited on one or both of the two chips with a fine-pitched,

photo-lithographic mask.

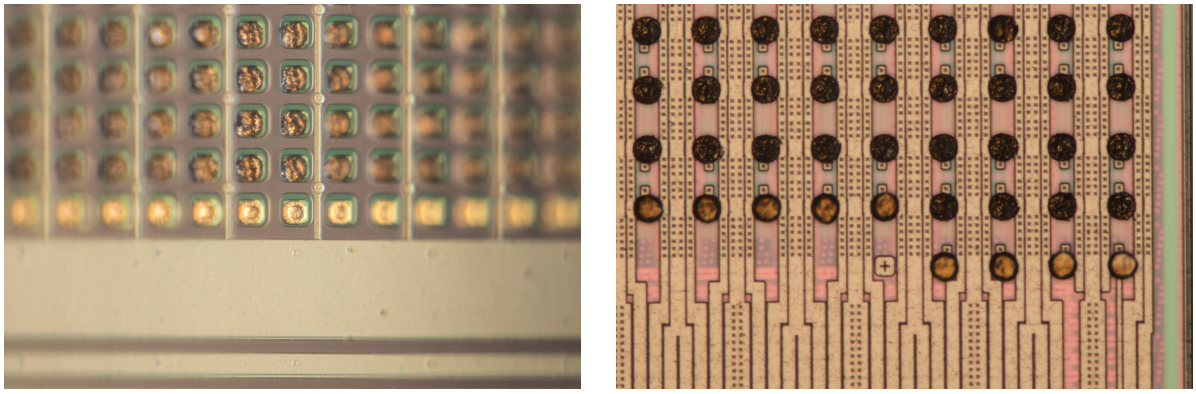
For a typical pixel detector module, a large number of bumps is deposited onto the pixel matrix with a comparatively fine pitch and high density over a large area. While high-density hybrid interconnection in general is a commonly used process in industry and research [71], the specific requirements of hybrid pixel detectors for use in high-energy particle physics experiments pose considerable challenges for the vendors involved in the process. For example, the RD53A-based DCMs depicted in this chapter feature more than 150 000 pixels with a pitch of 50 µm × 50 µm over the total matrix area of about 4 cm^2. Since both chips are connected with a bump for each single pixel, the resulting bump density is about $4 \cdot 10^4$ bumps $/cm^2$ over the full module area. This challenge will increase even further with the final production modules for the ATLAS ITk upgrade, which will be Quad-Chip Modules (QCMs) with ITkPix-V1, that features twice the matrix size of RD53A, quadrupling the total bump-bond area while maintaining the same bump-bond density.

8.1.1 PREPARATORY STEPS

In general, there are three major preparatory steps necessary before the flip-chip process. First, both wafers are thinned down to the desired final thickness. For the foundries and technologies typically used for the readout ASICs for hybrid silicon pixel particle detectors, the wafers are processed with a thickness of about 775 µm. For the final experiment, a much lower volume is desired to decrease the detector's material budget and increase the efficiency. Both, front-end chips and sensors are typically thinned down to a final thickness of 150 to 100 µm. This entails two significant challenges for the vendors: First, they need to have access to the necessary equipment and expertise to handle up to 300 mm wafers. Secondly, the thinned wafers are fragile and need to be handled with extreme care during the following production steps. One strategy to tackle this task is to use a handling wafer made from glass substrate, that is attached to the thinned ASIC wafer, requiring more specialized know-how and equipment. [72]

Secondly, an under-bump metallization (UBM) layer is deposited on the bump pads on both, the readout- and the sensor wafer. Since the usual pad material in the employed CMOS processes is typically aluminum, which is not well suited for soldering connections, another metal layer needs to be applied to the bump pads on both, the ASICs and sensor tiles to form a suitable foundation for the bump-bond connection. Different UBM materials can be used, depending on the targeted bump material and temperature limitations of the sensor material, if applicable. For solder bump interconnections, copper or an alloy of copper, nickel and gold are typical UBM materials on the sensor side. For example, the ATLAS IBL pixel detector modules used a sputtered titanium-tungsten (Ti:W) adhesion layer and an electro-deposited copper layer as UBM [73, p.18]. The UBM layer on the DCMs used for the measurements presented in this chapter is shown in Figure 8.2(a). For low-temperature processes involving indium bumps, a pure nickel UBM can be used, but has to be guarded from oxidization by a thin gold layer. [72–74]

In a third step, the bump-bonds are deposited on one of the wafers, typically the front-end chips. As a basis, a UBM stack similar to the options described before is applied to the bump-bond pads on the front-end chip wafer. The wafer is spin-coated with a photosensitive material and a fine-pitched pattern is applied via mask-based photo lithography to form the basic mold



(a) Under-bump metallization on a sensor tile of an RD53A DCM (Bottom row, gold).

(b) Deposited bumps on an RD53A front-end chip (Bottom row, light brown).

Figure 8.2: Under-bump metallization on a sensor tile and deposited bumps on an RD53A chip. The photos were taken on a disassembled module after flip-chip. Only the bottom row of pads and bumps, respectively, show the condition before flip-chip, while the upper rows comprise the broken bump-bonds on both sides.

pattern for the bump-bonds. After the UBM is applied to the wafer, the bump material is deposited for example through electroplating. To form spherical bumps from the deposited material, a reflow step is employed. The resulting bumps on an RD53A readout chip are shown in Figure 8.2(b). Different materials are available for the solder bumps, ranging from classical eutectic solder alloys like lead-tin (PbSn) or tin-silver (SnAg) to pure indium (In), which is an industry standard for low-temperature hybrid interconnection processes. In case of the ATLAS pixel detector modules both, classical solder bumps made from PbSn as well as In were used, while for the production of the ATLAS IBL, SnAg solder bumps were employed exclusively. [72–75]

Due to the concept of multi-chip modules and the differing sizes of sensor tile and readout chips, hybrid pixel detectors are typically flip-chipped in a die-to-die interconnection process instead of more cost effective variants like die-to-wafer or wafer-to-wafer interconnection. Therefore, the next step is *dicing*, that is separating the single dies on the processed wafers. Regarding the posed constraints for the total module envelope, and in order to avoid damaging the dies, the dicing has to be performed as precisely as possible.

8.1.2 FLIP-CHIP PROCESS

The actual die-to-die interconnection, or *flip-chip* process, starts with the precise alignment of the dies using special alignment markers on both, the sensor tile and the front-end chip. The typically required alignment accuracy is 2 to 3 µm [72]. However, in the optimal case, the chips will self-align during the reflow process due to the surface tension of the molten solder, if applicable. The chips are then locked into position by means of a highly viscous, flux-based tracking agent.

Depending on the interconnection method and bump material, two different flip-chip techniques are commonly used for hybrid pixel particle detectors. Processes with solder-based

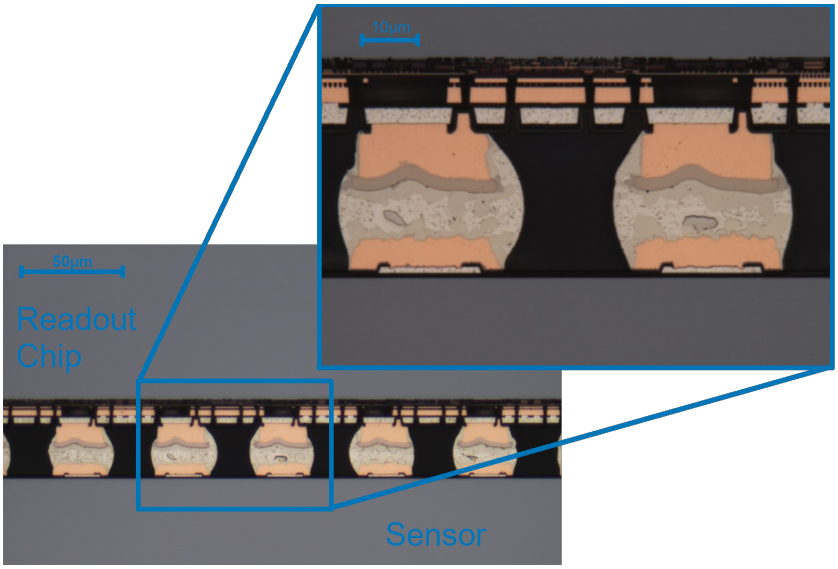


Figure 8.3: Cross section polish of a flip-chipped RD53A DCM (Combination C2) [76].

bumps (PbSn or SnAg) form a mechanical and electrical interconnection when evenly heating the whole package of readout chips with bumps, precisely fixed to the sensor tile in a second reflow process until the bumps melt at a temperature of about 250 °C. While this classical bump-bonding technique is well proven and tested for comparatively thick dies, special measures have to be taken to avoid bending of thinned chips due the high temperatures during reflow and the unavoidable mismatch in coefficient of thermal expansion (CTE) between the copper layers in the integrated electronics on top of the dies and the silicon bulk material underneath. For a commonly used option to mitigate this effect, a glass substrate wafer is temporarily bonded to the back side of the processed wafer and diced together with the actual dies. The resulting chip stack is strong enough to survive the reflow process. However, the temporary bonding material has to be temperature stable to enable laser-induced debonding only after the flip-chip process is completed.

A cross section polish of a successfully hybridized RD53A DCM is shown in Figure 8.3. This module of vendor combination C2 was fully processed by Vendor A, including UBM, bump deposition, and flip-chip processing and uses Ti:W-sputtered Cu UBM and SnAg solder bumps. The copper of the UBM is clearly visible on both dies and while it has partially started to dissolve, it is still solid enough to form a rigid mechanical interconnection and good electrical contact to both, readout chip and sensor. The solder bump itself has completely melted and wetted the copper pillar from the sides, providing additional support and contact surface.

Another promising method, especially for the hybridization of highly thinned ASICs, is thermo-compression bonding with In bumps. For this method, indium bumps are deposited on top of a nickel-based UBM layer on both, the sensor- and the front-end chip wafer, by electroplating. Like for the solder bump-bonding process described above, both wafers are diced and the individual chips are flip-chipped in a die-to-die bonding process. In contrast to the solder-based hybridization process, however, thermo-compression is used to form a mechanical and electrical connection between the two In bumps, requiring a high bonding

pressure in the order of several MPa, but significantly lower temperatures of a few 10 °C. In theory, this decreases the risk of CTE mismatch-induced bond failures, since the applied temperature is significantly lower and the chips are additionally stabilized during the bonding process. However, the beneficial self-alignment of the chips due to the surface tension of molten bumps does not apply here, since the In bumps never fully melt. [77]

Of the five flip-chip vendors assessed during the hybridization MS, only `Vendor M` uses thermo-compression with indium bumps in both combinations C8 and C9. All other vendors and combinations C1-C7 use SnAg solder bump-bonding, albeit with slight variations in the form and composition of the UBM and bump-bonds. Besides, `Vendors D`, `G` and `L` provide additional support to the assembly during the flip-chip process to minimize warping due to CTE mismatch.

8.2 RD53A DOUBLE CHIP MODULES

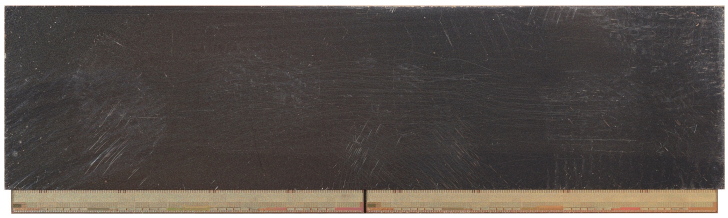


Figure 8.4: A bare `RD53A` DCM (Combination C6). The metallization on the sensor backside (top) appears dark. The metallization layer is scratched due to an invasive metrology measurement. The bottom part of the two `RD53A` readout chips with the wire-bond pads can be seen at the bottom.

The stack of sensor tile and readout chips after the flip-chip process is typically referred to as *bare module*. A bare `RD53A` DCM of vendor combination C6 is shown in Figure 8.4. It consists of a sensor tile of about 10 mm × 40 mm that is pictured on top in the figure. Due to the sensor backside metallization, it appears dark in the picture. Along the bottom edge, the wire-bond pads of the two `RD53A` front-end ASICs can be seen. The total module envelope is determined by the sensor size on three sides and the overlap of the readout chips on the fourth. The `RD53A`-based DCMs for the hybridization MS were built using probed `RD53A` chips, in order to improve the yield and thereby the amount of assessable modules. Since the focus of the hybridization MS is the assessment of the hybridization quality, the maximum number of pixels per module should be evaluated, in order to maximize the statistical sample size. For this purpose, the three AFE flavors of both `RD53A` chips were tuned to behave as similarly as possible, including the use of the full matrix optimization set of parameters for the DIFF (ref. to Section 5.5). However, differences in threshold, occupancy and mean ToT between the flavors can still be visible in the results. Additionally, in order to compensate for the larger gap between the two readout chips in the center columns of the sensor, some sensor designs include larger pixels in this position, while other designs simply leave this gap as inactive area. All of these facts have to be taken into account for the assessment of the bump-bond quality.

8.2.1 DCM Flex

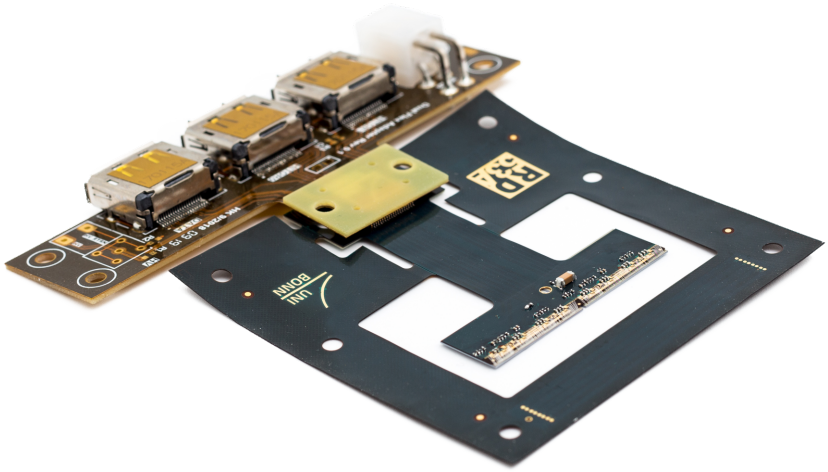


Figure 8.5: A fully assembled RD53A DCM. The flex (dark green) is glued to the sensor backside and the readout chips are connected to the flex via wire-bonds. The flex is connected to the readout adapter (brown) for connection to the readout system.

Before a module can be evaluated, it needs to be *assembled* in order to provide connectivity of the bare module to the DAQ system. This assembly process consists of two main steps: Attaching a flexible PCB (flex) to the bare module by means of a rigid epoxy adhesive and wire-bonding the two readout ASICs to the flex in order to accomplish electrical connectivity. The resulting assembly is referred to as a *dressed module* or simply *module*. A fully dressed module is shown in Figure 8.5. The flex is depicted in dark green in the figure and consists of a frame for handling and stabilization that holds no traces or components. The main part in the center of the frame's cutout includes some passive components on top of the module area and routes all the signals from the wire-bond pads on the bottom edge to two high-density connectors at the end of the so-called *pig tail*. Mated to the flex via these two connectors is a detachable *readout adapter*, depicted in brown. The readout adapter is another flexible PCB with an additional stiffener, that distributes the traces from the high-density connectors to the usual DP and power connectors for interfacing with the readout system.

The RD53A DCM flex routes only the primary Aurora lanes of both readout chips to the primary DP connector, which is also responsible for distributing the CMD signal to the ASICs. Additionally, all four HitOr networks of both readout chips are routed to two dedicated DP connectors to enable efficient testing with radioactive sources using BDAQ53's self-trigger functionality. The supply voltage for the chips is applied via a standard four-pin connector and routed through the same high-density connectors to the flex, for powering in LDO- or Shunt mode. For biasing the sensor, the flex features an additional HV line that is especially isolated to facilitate the comparatively high voltages involved. It is wire-bonded to the sensor backside through the HV hole in the center of the flex and includes a large filter capacitance and two 100 kΩ HV resistors, one on the flex and one on the readout adapter. Last but not least, the

RD53A DCM flex contains an NTC located directly on top of the module that can be employed to monitor the module temperature during operation.

Since the flex itself consists of multiple copper layers and is firmly and rigidly attached to the full sensor backside, it adds another layer to the module stack that can cause mechanical stress during thermal cycling due to mismatched CTE. Since the modules in the final experiment will also be attached to an even more complex flex, it seems appropriate to evaluate the hybridization quality in the MS with a similar flex attached. As for the final ITk module flex, special measures have been taken in order to minimize the copper content of the RD53A DCM flex. As defined beforehand, the flex should have a total thickness of about 220 µm in the module area, consisting of two 20 µm thick copper layers with a filling factor of 100 % and 50 %, respectively. In between, the flex contains a 150 µm thick dielectric layer and on both the top and the bottom, a cover layer with a thickness of 30 µm is foreseen. [70]

For routing reasons, the RD53A DCM flex used for the measurements presented in this chapter instead features three copper layers with a thickness of about 12 µm each, adding up to about the required copper content, albeit with a different spatial distribution. Consequentially, the dielectric layer is split into two layers of 75 µm thickness. The total thickness of the DCM flex of 218 µm almost exactly matches the specification.

8.2.2 Overview of Tested Modules

A comprehensive overview of all RD53A DCMs tested in the light of the hybridization MS within the scope of this thesis is shown in Table 8.2. In total, 33 out of the 78 modules delivered to Bonn by the five different vendors have been assembled and tested, in order to optimize the sample size for each vendor combination as much as possible within the given schedule constraints. Vendor combinations C1 and C2 were tested with three and four modules, respectively, combination C3 with 3 modules. Combinations C4 and C5 were tested with one module each. Combination C6 was tested with four modules, C7 with four modules, and C8 and C9 with two modules each. Each combination of vendors was therefore assessed based on at least two modules, except for combinations C4 and C5. For these combinations involving Vendor D, only two modules were delivered to Bonn, while the rest was tested at other testing sites.

However, despite using preselected readout chips, some modules still showed malfunctions that complicated or completely prevented the assessment of the hybridization quality. Furthermore, some modules were accidentally broken during the investigation process. This mainly concerns combination C6, since these were the first modules to be delivered to Bonn in June 2020, but also some modules of other combinations. Nonetheless, the number of tested modules per combination listed in the previous paragraph includes fully tested modules only.

8.3 Testing Methodology and Requirements

The testing methodology and acceptance criteria for the hybridization quality of the modules delivered withing the scope of the hybridization MS were determined beforehand and are defined in the *Technical Specification for the Hybridization of Hybrid Pixel Detector Modules for the ATLAS ITk* [70].

Module	Comb. (Tab.8.1)	Status
DCM1	C6	Fully tested
DCM2	C6	Fully tested (low statistics)
DCM3	C6	Not tested (defective FE chip)
DCM4	C6	Not tested (defective flex)
DCM5	C6	Fully tested
DCM6	C6	Not tested (defective FE chip)
DCM7	C6	Not tested (broken during HV investigation)
DCM8	C1	Fully tested
DCM9	C1	Fully tested
DCM10	C6	Not tested (broken during HV investigation)
DCM11	C6	Fully tested
DCM13	C1	Fully tested
DCM14	C3	Fully tested
DCM15	C2	Fully tested
DCM16	C3	Test not finished (sensor broken after TC)
DCM17	C3	Not tested (defective FE chip)
DCM18	C6	Not tested (defective FE chip)
DCM19	C7	Fully tested
DCM20	C4	Fully tested (defective FE chip)
DCM21	C7	Fully tested
DCM22	C5	Fully tested (low statistics)
DCM23	C3	Fully tested
DCM24	C3	Fully tested
DCM25	C9	Fully tested
DCM26	C9	Fully tested
DCM27	C2	Fully tested
DCM28	C7	Fully tested
DCM29	C7	Fully tested
DCM30	C8	Fully tested
DCM31	C8	Fully tested
DCM32	C2	Fully tested (low statistics)
DCM33	C2	Fully tested

Table 8.2: Summary of modules tested for the hybridization MS.

The test procedure and hybridization assessment methodology consists of seven steps and is depicted in Figure 8.6. After a first **visual inspection**, the dimensions of the module are measured in an optical **metrology** step. The bare module is then contacted with a simple probe station to measure and assess the **sensor IV curve**. Next, the module is **assembled** by attaching a flex to the bare module with a defined adhesive, gluing- and curing procedure. After the glue has cured, the module is wire-bonded to the flex. Subsequently, the module undergoes a first **electrical test** stage consisting of a standard test and tuning procedure, a second IV curve measurement and a `Source Scan`. The main goal of the electrical test procedure is to assess

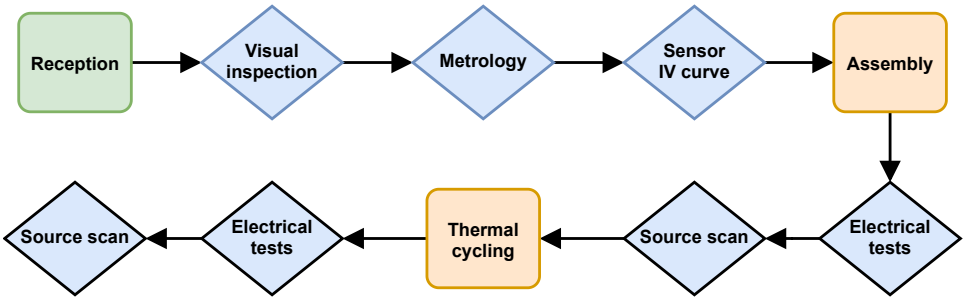


Figure 8.6: A flowchart depicting the test procedure for the hybridization MS measurements.

the hybridization quality by identifying disconnected or merged bump-bond interconnections. In order to accelerate the aging process expected in the later experiment for testing purposes, the module is then **thermally cycled**. This step consists of 20 cycles from −40 °C to 60 °C in rapid succession, causing significant mechanical stress for the bump-bond interconnections. Finally, a second, identical **electrical testing** step serves to identify bump-bond failures that formed during thermal cycling. The highlighted steps and the corresponding acceptance criteria are explained in detail in the following sections.

VISUAL INSPECTION

Visual inspection of the module is performed either manually using a microscope or by means of a high-resolution photograph. The purpose of this step is first of all to assess the shipping procedure of the vendor and assess whether the module arrived at the testing sites without damage. By evaluating the cutting edges of the dies, the quality of the wafer dicing at the vendor can be assessed.

METROLOGY

The metrology stage serves to determine the physical dimensions of the module, including the *envelope*, consisting of width and height, and the thickness or, consequentially, the planarity of the module. The envelope is assessed as a measure for the quality of the vendor's dicing process. The dicing edge is required to have an accuracy of ±20 µm for sensors and ±10 µm for readout ASICs. Additionally, break offs at the dicing edge are only tolerated up to a total size of 20 µm and as long as the IC is not electrically damaged. The height of the assembly is a good measure for the accuracy of the wafer thinning process of the vendor. The final thickness of readout ASICs is required to be 150^{+25}_{-10} µm, while sensors from the different foundries are typically available with a thickness of 100 µm or 150 µm. Significant variations in total module thickness can be a hint for complications during the flip-chip stage.

While several manual or automatic methods exist for performing the metrology measurements of a module, the measurements presented in this thesis were conducted manually with a measuring microscope with an accuracy of approximately 2.2 µm in x and y (horizontally), and 5 µm in z (vertically) [78]. Unfortunately, this microscope does not include a vacuum chuck to firmly attach the module to the base. As a consequence, planarity measurements could show

warping of the whole module and do not necessarily indicate problems during the flip-chip process. Additionally, since the height measurement with this device is based on focusing on the different layers and the base of the microscope chuck is made from glass, a reliable measurement of the absolute module height is hard to achieve with this device.

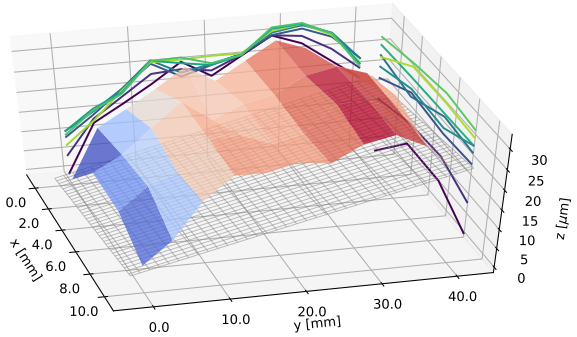


Figure 8.7: Example metrology measurement of a bare RD53A DCM.

Figure 8.7 shows an example metrology measurement of an RD53A DCM. The envelope of this module was measured as 41 139(2) µm by 11 805(6) µm, where the numbers in parentheses depict the variance of the width measurement over the module height, and vice versa. This module therefore fulfills the requirements of $(41\,100 \pm 50)$ µm in width. For the hybridization MS, no requirement for the total module height after the flip-chip process is posed. While the planarity does not need to fulfill a requirement either, it can still be interesting for the investigation of the hybridization quality. Due to the nature of the measurement method, the planarity is only assessed from the sensor backside. To get a comparable measurement of the planarity, a virtual plane is fitted to the lowest four measurement points. This fit plane is depicted in the figure as light gray mesh. The maximum warp is then defined as the maximum distance of any measurement point to this plane. For this module, this method results in a warp of about 32 µm which does not indicate any problems. Finally, the difference in z between the readout chip's surface and the sensor backside is measured at multiple points along the bottom edge. Knowing the approximate thickness of the sensor tile, the mean bump thickness can be estimated from this measurement. For the example described here, the mean thickness of sensor and bumps amounts to 132(4) µm. With an approximate sensor thickness of 100 µm, this leaves about 32 µm for the bump thickness after the flip-chip process, well within the expected range.

Bare Module IV Curve

The sensor IV curve (ref. to Section 7.3.1) is measured on the bare module right after reception to make sure the module is operational and to ensure the sensor has not been damaged during the hybridization process. Additionally, it provides a comparison measurement before the assembly process. Since the bare module does not feature any easy-to-use interfaces yet, a simple manual probe station is used to contact the sensor backside and an analog ground pad on one of the readout chips to apply the reverse bias voltage to the sensor. Two parameters of

interest are extracted from this measurement: The sensor leakage current at operation voltage normalized to 20 °C and the breakdown voltage of the sensor.

ASSEMBLY

The assembly procedure consists of attaching the flex to the bare module and electrically connecting the readout chips to the flex via wire-bonding. For the attachment procedure, an epoxy-based adhesive similar to *Araldite 2011* [70] is dispensed manually in a regular pattern on the bottom side of the flex. Bare module and flex are aligned using a special jig and pressed together with a small weight applied to the top part of the alignment jig for even distribution. The package is shortly heated to a moderately elevated temperature to allow the adhesive to spread evenly and then cured at room temperature. This way, an evenly distributed glue layer with a thickness of 30 to 50 µm is achieved.

Finally, the module is electrically connected to the corresponding pads on the flex via wire bonds. This includes setting the correct trim bits for `IREF`, as measured during wafer probing (see Section 6.3.1).

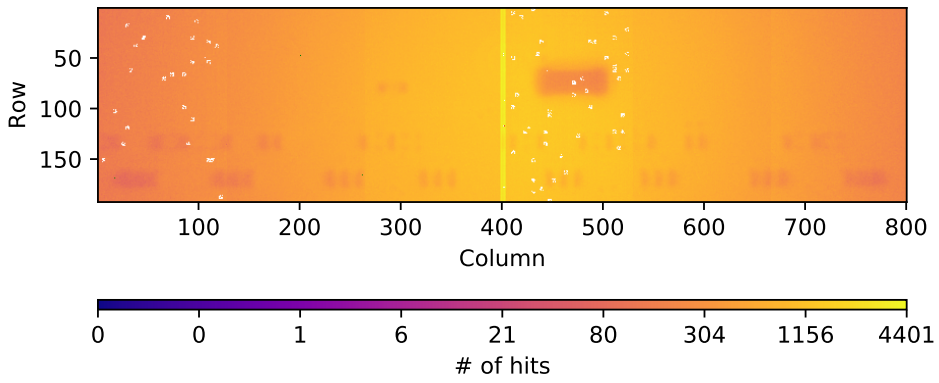
FIRST ELECTRICAL TEST

Electrical tests include connecting the assembled module to a `BDAQ53` setup (see Chapter 4) via the readout adapter. These tests are conducted in a shielded and insulated test box on an active cooling plate at 20 °C to prevent the modules from overheating and establish a comparable setting independent of the environmental conditions. The set of tests is less focused on characterization of the readout chip, like the tests presented in chapters 5 through 7, but on the assessment of the bump-bond interconnection quality. Therefore, the front-end chips are tuned to give as similar results as possible between the three AFE flavors. After an initial `Digital-` and `Analog Scan` to confirm the chips' general functionality, the sensor IV curve is measured again through the flex and readout adapter. The normalized leakage current and breakdown voltage of the sensor are extracted from the measurement. No significant deviations from the bare module measurement are expected, apart from the influence of the two serial bias resistors in the HV line. An appropriate operation voltage is determined from the IV curve and applied for the following measurements.

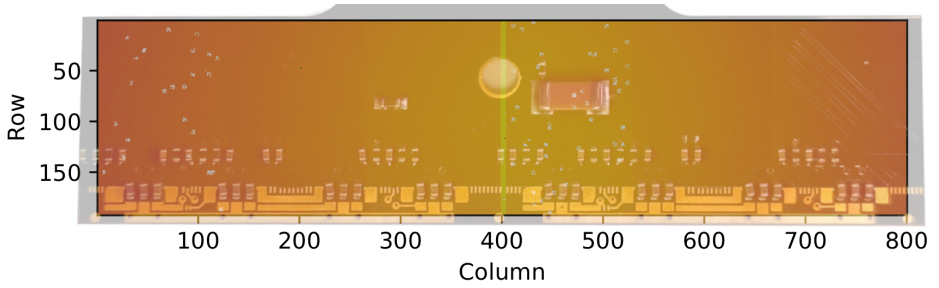
Next, the global and local thresholds of all three AFE flavors of both front-end chips are tuned to about 3 000 e^- and a `Noise Occupancy-` and `Stuck Pixel Scan` are performed to disable misbehaving pixels in the front-end chips. These pixels can of course not be assessed regarding their bump-bonding interconnection, but usually only a very small percentage of pixels has to be disabled. The exact tuning parameters and the resulting register settings for the measurements presented in this chapter are listed in Section C.2.

Finally, a ^{241}Am X-ray source is placed on top of the module and both readout chips are operated in parallel in self-trigger mode for one hour, acquiring on average about 10^7 hits in total. This corresponds to at least 150 hits/pixel in the outermost pixels, which are illuminated the least by the circularly collimated source. An automated algorithm is then applied to estimate the amount of poorly connected and entirely disconnected bump-bonds based on the acquired hits in each pixel, the according energy distribution based on ToT and other

indicators. The results of this algorithm are always confirmed visually by inspecting the hitmap of acquired data during the `Source Scan`. Figure 8.8(a) shows such an occupancy map of a fully connected `RD53A` DCM. The map shows six columns in the center that exhibit about twice the number of the other pixels. This is explained by larger edge pixels in the sensor design. Furthermore, some less efficient areas with a very specific pattern can be observed, that are caused by shadowing effects due to the passive components on the flex. Figure 8.8(b) shows the same occupancy map superimposed with a top-view image of the flex.



(a) Occupancy map, logarithmic color scale.



(b) Flex overlay with 50% opacity.

Figure 8.8: Occupancy map of a `Source Scan` of an `RD53A` DCM with a ^{241}Am X-ray source. White pixels are disabled. The superimposed flex in (b) illustrates the position of the mounted components, which cause shadowing effects.

For some of the tested modules, a `Bump Connectivity Scan` based on the noise / threshold shift algorithm (see Section 4.5.10) is performed in addition to a `Source Scan`. By comparison of the results of this scan and the `Source Scan`, the `Bump Connectivity Scan` was additionally verified [79]. For some modules presented in this chapter, performing a `Source Scan` is impossible, for example due to large amounts of stuck pixels. In these cases, only the `Bump Connectivity Scan` is used to assess the bump-bond interconnection quality.

Thermal Cycling

During the lifespan of the final experiment, the detector will be exposed to changing thermal conditions due to the operation- and maintenance cycles. These changing conditions cause

significant thermal and mechanical stress for the bump-bond interconnections. To artificially accelerate this aging process, the modules are rapidly thermo-cycled between −40 °C and 60 °C 20 times. A *Weiss LabEvent T/210/70/5* temperature chamber is used for this procedure [80]. The modules are not powered during thermal cycling (TC) and the readout adapters are removed. However, the modules stay on their aluminum carrier plate during the cycling procedure. A temperature sensor is attached firmly to the carrier plate of one of the tested samples, as close to the actual module as possible, and is used as reference for the cycling procedure. The temperature chamber cools and heats the air inside the chamber at a rate of approximately $-5\,\frac{\mathrm{K}}{\mathrm{min}}$ and $+8\,\frac{\mathrm{K}}{\mathrm{min}}$, respectively [80]. Once the target air temperature is reached, the algorithm waits until the temperature sensor on the module has reached the desired temperature and keeps it stable for an additional two minutes, before the cycle is continued. The resulting temperature progression during a thermal cycling run is shown in Figure 8.9. At the given maximum heating and cooling rate of the temperature chamber, a single cycle takes about 38 min, resulting in a total run time of 12.5 h for a complete TC run. Since the employed temperature chamber features an additional air dryer, condensation during passive TC is no issue.

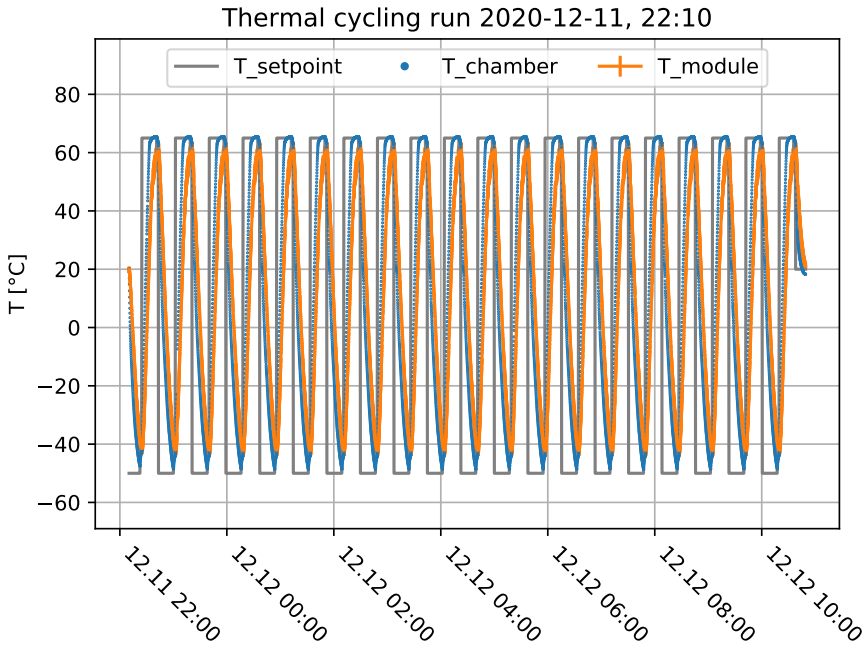


Figure 8.9: Temperature development during a thermal cycling run. The gray curve depicts the setpoint of the temperature chamber, the blue points show the air temperature as reported by the chamber and the orange points represent the temperature measured close to the actual module.

SECOND ELECTRICAL TEST

After the thermal cycling, another full electrical test is performed, including the same set of tests as the first electrical test stage. On the one hand, this serves to assess if the modules survived the cycling process. On the other hand, it is used to identify bump-bond interconnections that broke due to the thermal and mechanical stress during the cycling process. The acceptance criteria described in the following section are applied to the number of disconnected bumps *after* thermal cycling.

8.3.1 ACCEPTANCE CRITERIA

The acceptance criteria for the RD53A DCMs produced and delivered in the scope of the hybridization MS were defined before production and testing [70]. Not all criteria could be verified in the scope of this work, due to lack of information or devices. The most relevant acceptance criteria for the hybridization MS that have been surveyed for the tested modules are defined in the following list, while a comprehensive enumeration of all criteria is included in Section C.1 of Appendix C.

- Width of the sensor tile: $(41.100 \pm 0.050)\,\mathrm{mm}$.
- Leakage current $< 1.5\,\mu\mathrm{A\,cm^{-2}}$ at $V = V_{\mathrm{depl}} + 50\,\mathrm{V}$ and $20\,^{\circ}\mathrm{C}$.
- Breakdown voltage $> V_{\mathrm{depl}} + 80\,\mathrm{V}$.
- No ASICs failing after assembly.
- Number of total individual bump failures per module $< 1\,200$.
- Failure cluster size (number of individual bump failures per ASIC, which are grouped in a cluster) < 50.

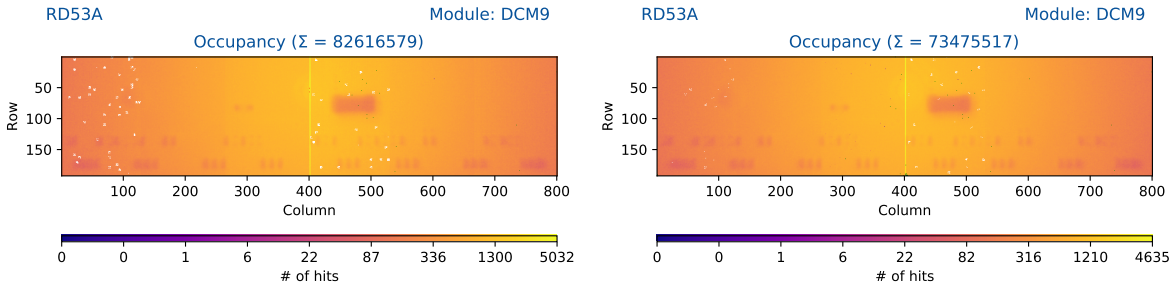
Module dimensions and sensor characteristics before and after assembly have been verified to be compliant with the acceptance criteria in most cases. Diverging cases could be retraced and understood and are followed up by the MS for the ATLAS ITk pixel sensors, that is conducted in parallel. Due to this overlap, the evaluation of the hybridization quality and vendor combinations in the following section primarily focuses on the acceptance criteria for the number of failing bump-bond interconnections.

8.4 RESULTS

In this section, the results acquired for all modules tested for the hybridization MS in the scope of this work are presented. The results are structured by vendor combination to give an overview of the vendors' performance. For each combination, only the most remarkable results are discussed in detail. Additional results and complementary plots are presented in Section D.3 of Appendix D.

COMBINATION C1

Three modules were assembled for vendor combination C1, DCM8, DCM9 and DCM13, and all three of them could be tested without complications. Figure 8.10(a) shows the occupancy map of the Source Scan of DCM9 before TC. While a few noisy pixels had to be disabled, especially in the two SYNC AFEs, that cannot be assessed regarding their bump-bond connectivity, most other pixels exhibit an appropriate number of hits. No clusters of disconnected pixels can be observed. The automatic algorithm counts 18 poorly connected or completely disconnected pixels and close investigation of the occupancy map confirms this first, good impression.



(a) DCM9 source scan occupancy map before TC.

(b) DCM9 source scan occupancy map after TC.

Figure 8.10: Occupancy map of a Source Scan of DCM9 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.

After TC, the situation does not change dramatically, as can be seen in Figure 8.10(b). Here, 47 pixels are automatically determined as poorly or entirely disconnected. A summary of the results of all three modules tested for combination C1 is shown in Table 8.3, while the corresponding occupancy maps for DCM8 and DCM13 are shown in Figures D.15 and D.16, respectively.

	Before TC			After TC		
Module	Poor	Disconnected	Unknown	Poor	Disconnected	Unknown
DCM8	14	466	1 060	44	428	748
DCM9	2	16	579	6	41	167
DCM13	0	3	609	1	3	388

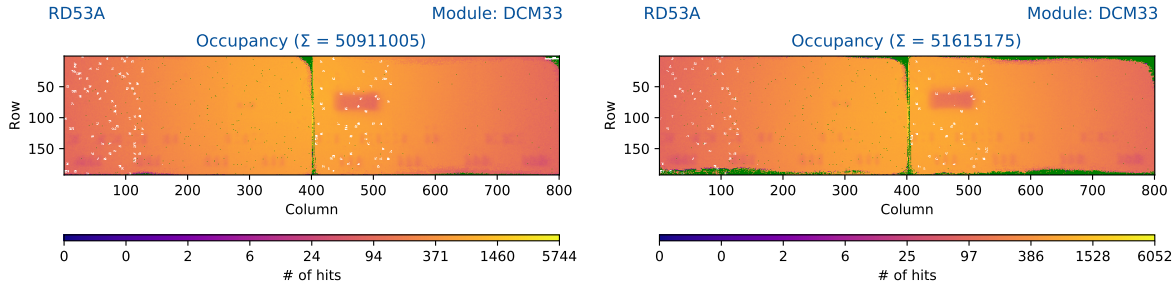
Table 8.3: Summary of modules tested for vendor combination C1.

To sum up, all three tested samples of vendor combination C1 show much less than the maximum allowed 1 200 failed bump-bond connections, even after TC. Vendor combination C1 can therefore be recommended as supplier for the ATLAS ITk modules without reservations.

COMBINATION C2

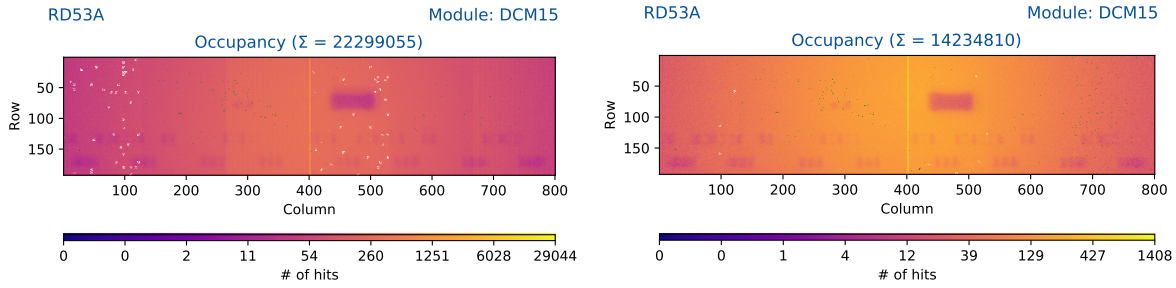
For vendor combination C2, involving Vendor A in combination with a different sensor type and UBM vendor compared to combination C1, four modules were tested in total, DCM15, DCM27, DCM32 and DCM33. Three of these modules could be tested without complications, but

DCM32 exhibited a very low total occupancy after the Source Scan, limiting the statistical significance of this test. Due to limited time available for extended testing, the evaluation of the bump-bond connectivity for this module was carried out based on Bump Connectivity Scans, which generally has a higher margin of error than the Source Scan-based analysis. However, two of the tested samples, namely DCM27 and DCM33, exhibited large areas of disconnected pixels along the module edges after TC. Figure 8.11(b) illustrates this effect on DCM33. A closer investigation reveals, that the problematic areas already misbehaved before TC, suggesting that a general problem in the hybridization process is responsible for this effect. The according plots for DCM27 in Figure D.17 exhibit comparable features, albeit at smaller scale.



(a) DCM33 source scan occupancy map before TC. (b) DCM33 source scan occupancy map after TC.

Figure 8.11: Occupancy map of a Source Scan of DCM33 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.



(a) DCM15 source scan occupancy map before TC. (b) DCM15 source scan occupancy map after TC.

Figure 8.12: Occupancy map of a Source Scan of DCM15 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.

However, regarding the results of DCM15 in Figure 8.11(b) no significant disconnections can be observed. DCM15 was even cycled a second time for an addition 20 thermal cycles beyond the requirement without exhibiting significant amounts of poorly- or entirely disconnected bump-bonds. In extension, keeping in mind the higher margin of error due to the evaluation based on Bump Connectivity Scans, this also holds true for DCM32. For an additional cross check, the result of the low-occupancy Source Scan is shown in Figure D.18. Even though the total occupancy is too low for a statistically significant result, no systematic disconnections like for DCM27 and DCM33 can be observed here. The quantitative results of all three modules are

summarized in Table 8.4.

	Before TC			After TC		
Module	Poor	Disconnected	Unknown	Poor	Disconnected	Unknown
DCM15	4	109	508	14	196	43
DCM27	703	229	121	652	916	170
DCM32	0	1 477	885	0	1 504	885
DCM33	5 103	1 451	1 348	3 996	7 207	1 012

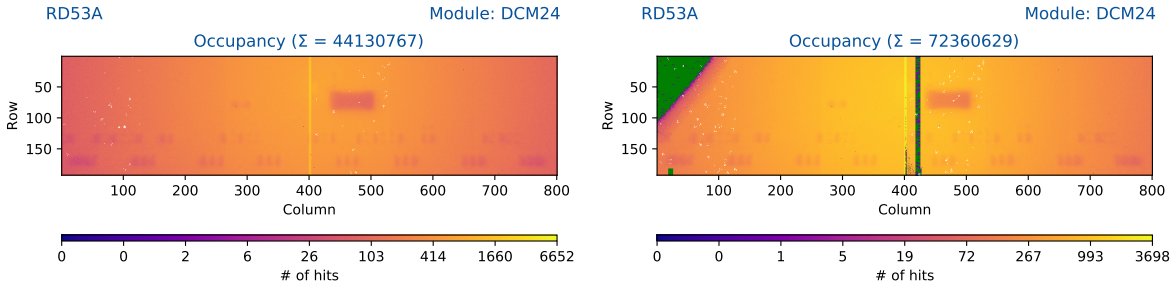
Table 8.4: Summary of modules tested for vendor combination C2.

Upon further analysis, it turns out that, even though all modules are produced by the same combination of vendors, the sensors supplied by Foundry 3 were thinned to two different measures. Curiously, the two good modules, DCM15 and DCM32 both used 100 µm thin sensors, while the problematic modules, DCM27 and DCM33 both used 150 µm thick sensors. This correlation suggests a general problem in the hybridization process of Vendor A for 150 µm thick sensors from Foundry 3. Vendor A was informed of the results and will investigate the problems further.

All in all, vendor combination C2 can only be recommended for the production of ITk modules for 100 µm thin sensors, until the problems observed for 150 µm thick sensors can be resolved.

COMBINATION C3

For the last vendor combination involving Vendor A, combination C3, five modules have been assembled. Unfortunately, two of those modules could not be assessed completely due to reasons unrelated to the hybridization. The remaining three modules, DCM14, DCM23 and DCM24 could be tested without major complications.



(a) DCM24 source scan occupancy map before TC. (b) DCM24 source scan occupancy map after TC.

Figure 8.13: Occupancy map of a Source Scan of DCM24 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.

While one of the tested modules, DCM23, exhibited no excessive disconnections even after 15 additional thermal cycles beyond the requirement, as can be seen in Figure D.20, DCM24 did. The occupancy maps before and after TC of DCM24 are shown in Figure 8.13. The columns with no occupancy in the SYNC of the right front-end chip in Figure 8.13(b) exactly correspond

to a core column of RD53A and are therefore probably caused by an electrical problem in the front-end chip and the rectangular area in the bottom left corner of both front-end chips was deliberately disabled, in order to deactivate a few extremely noisy pixels. However, the top left corner of the module is clearly disconnected after TC. On very close investigation, the bottom left corner of DCM14, shown in Figure D.19 also exhibits a cluster of disconnected pixels after TC. While this cluster is technically still small enough to meet the acceptance criteria, it is definitely concerning. The summary of the results of all three assessed modules for this vendor combination is presented in Table 8.5.

	Before TC			After TC		
Module	Poor	Disconnected	Unknown	Poor	Disconnected	Unknown
DCM14	16	168	103	23	192	53
DCM23	891	227	536	5	28	404
DCM24	2	7	141	3 442	7 272	657

Table 8.5: Summary of modules tested for vendor combination C3.

Vendor A already spotted irregularities with the hybridization process involving UBM supplied by Foundry 6 during their internal QC measurements and began an investigation, the results of which were kindly supplied for inclusion in this thesis. Figure 8.14 shows a cross section polish of a flip-chipped module of combination C3. Especially in comparison to the example in Figure 8.3, a potential problem with the UBM on the sensor side in combination with the employed solder bump-bonding process is observable. The sensor side UBM has almost completely dissolved during the flip-chip process, leaving only a weak direct interconnection between the solder bump and the sensor bump pad. This direct interconnection breaks much more easily than good bump-bond interconnections with sufficient UBM, when subjected to mechanical stress.

Conclusively, in the light of the presented results, vendor combination C3 cannot be recommended as supplier for ATLAS ITk modules. An improvement of the UBM layer, which, in this case is applied by Foundry 6, is necessary. For instance, an increased thickness of the UBM could help improve the bump-bond quality after thermally accelerated aging.

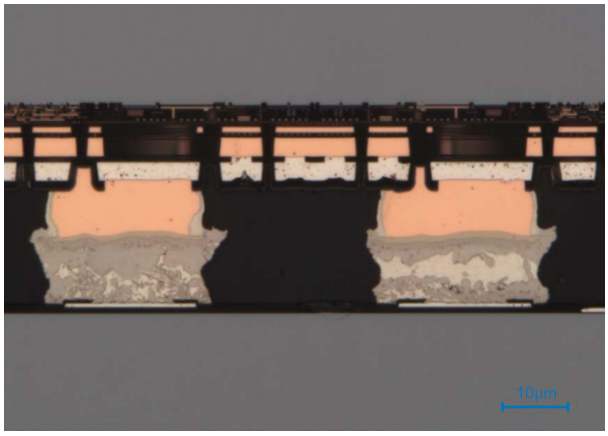


Figure 8.14: Cross section polish of a flip-chipped DCM of vendor combination C3. [76].

COMBINATIONS C4 AND C5

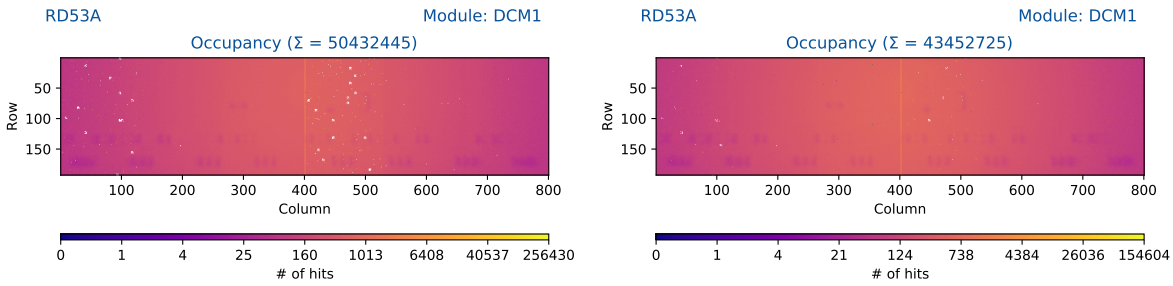
For vendor combinations C4 and C5, only one module of each variant was available for testing in Bonn, while the rest was delivered to other testing sites. Both available modules were assembled, but exhibited severe electrical problems of the front-end ASICs, preventing any detailed measurements. Unfortunately, these vendor combinations could therefore not be assessed for this thesis.

COMBINATION C6

Modules of vendor combination C6 were the first ones available in Bonn. Therefore, the largest number of samples of all combinations was assembled for this variant. However, during initial tests, severe differences in the sensor breakdown voltage before and after assembly of the first modules was observed. After an intensive investigation process, the culprit turned out to be an additional cleaning step of the modules during assembly that the sensors reacted badly to. This step is not foreseen in the predefined assembly procedure and was supplementarily implemented to improve wire-bonding quality. Beginning with DCM8, this assembly step was omitted, which solved the problem. However, in the course of the investigation of this effect, several modules were only partly assembled or deliberately disassembled again and can therefore no longer be assessed regarding the quality of their bump-bond interconnections. In the end, three modules of combination C6 could be fully evaluated: DCM1, DCM5, and DCM11. All three samples fulfill the acceptance criteria with very low numbers of poorly or entirely disconnected pixels. A qualitative summary of all results is shown in Table 8.6.

	Before TC			After TC		
Module	Poor	Disconnected	Unknown	Poor	Disconnected	Unknown
DCM1	3	3	521	1	9	105
DCM5	0	0	4 925	6	7	1 794
DCM11	8	108	413	6	10	65

Table 8.6: Summary of modules tested for vendor combination C6.



(a) DCM1 source scan occupancy map before TC. (b) DCM1 source scan occupancy map after TC.

Figure 8.15: Occupancy map of a Source Scan of DCM1 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.

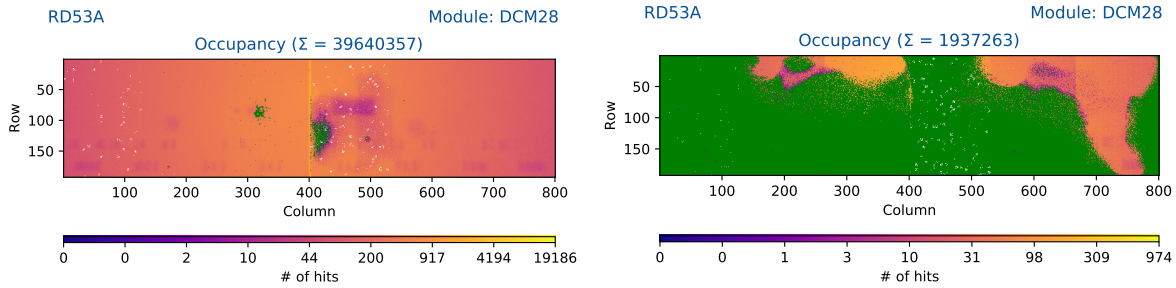
As an example, Figure 8.15 shows the occupancy maps before and after TC for DCM1. The

corresponding plots for DCM5 and DCM11 can be found in Figure D.21 and Figure D.22, respectively.

In summary, vendor combination C6 can be recommended without reservations as supplier for the production of ITk modules.

COMBINATION C7

All tested samples from vendor combination C7 exhibited insufficient bump-bond quality, already before TC. The documentation and shipping procedures for this vendor combination were inadequate, so that at least one module was destroyed before assembly due to inappropriate packaging. In the end, three modules could be assembled and fully tested. DCM21, DCM28, and DCM29 all exhibited massive disconnections even before TC. For DCM19, only the left readout chip could be assessed before TC due to reasons probably unrelated to the hybridization process. After TC however, both front-end chips responded and were evaluated.



(a) DCM28 source scan occupancy map before TC. (b) DCM28 source scan occupancy map after TC.

Figure 8.16: Occupancy map of a Source Scan of DCM28 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.

While all evaluated modules show a similar performance, Figure 8.16 shows the exemplary result of DCM28, the best of the tested samples. Large disconnected areas are clearly visible even before TC. The results for DCM19, DCM21, and DCM29 are shown in Figures D.23, D.24, and D.25, respectively.

	Before TC			After TC		
Module	Poor	Disconnected	Unknown	Poor	Disconnected	Unknown
DCM19	0	65 074	33 396	16 282	92 779	5 142
DCM21	15 139	49 495	17 076	7 563	102 687	6 921
DCM28	8 406	2 864	4 947	7 158	113 046	1 448
DCM29	20 145	37 024	334	8 972	135 276	4 015

Table 8.7: Summary of modules tested for vendor combination C7.

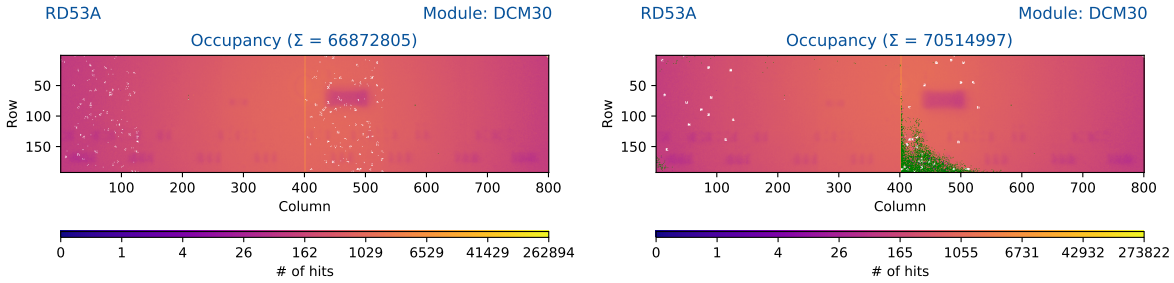
The quantitative results are summarized in Table 8.7 and support the visual impressions from the plots. All four tested samples for vendor combination C7 failed to fulfill the acceptance criteria by a large margin. Based on these results and experiences, vendor combination C7 cannot be recommended for the production of ITk modules at all.

COMBINATION C8

Last but not least, vendor combinations C8 and C9, both involving `Vendor M`, are the only two combinations using thermo-compressed In bumps instead of a solder-based process. For combination C8, two modules have been assembled and fully tested. Both `DCM30` and `DCM31` almost managed to stay below the required number of bump-bond failures after TC but ultimately failed to fulfill the acceptance criteria.

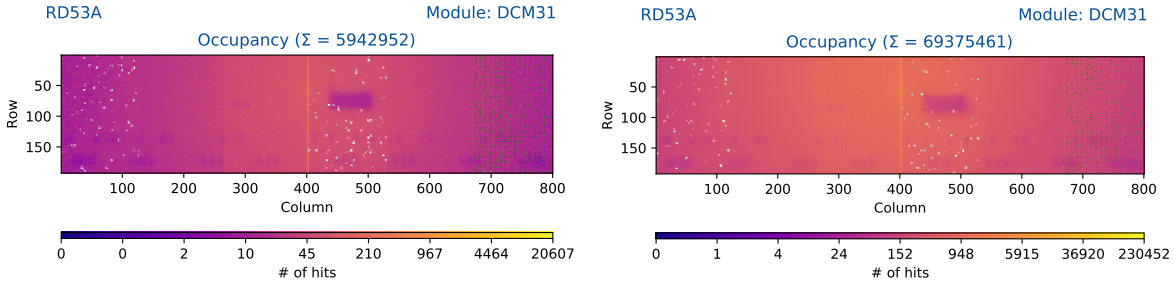
	Before TC			After TC		
Module	Poor	Disconnected	Unknown	Poor	Disconnected	Unknown
`DCM30`	37	6	700	39	2 762	462
`DCM31`	3 253	1 369	820	2 582	706	629

Table 8.8: Summary of modules tested for vendor combination C8.



(a) `DCM30` source scan occupancy map before TC. (b) `DCM30` source scan occupancy map after TC.

Figure 8.17: Occupancy map of a `Source Scan` of `DCM30` with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.

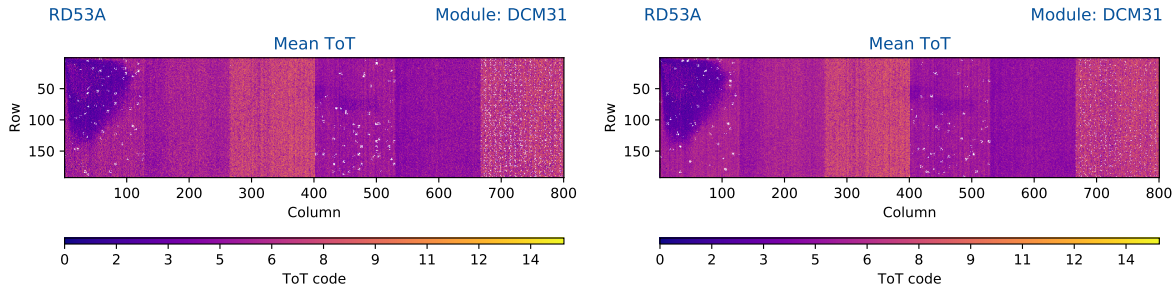


(a) `DCM31` source scan occupancy map before TC. (b) `DCM31` source scan occupancy map after TC.

Figure 8.18: Occupancy map of a `Source Scan` of `DCM31` with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.

The qualitative results are shown in Table 8.8 and the corresponding plots are shown in Figures 8.17 and 8.18. While `DCM30` is fine before TC, the lower left corner of the right front-end chip clearly lost connectivity during TC. `DCM31` shows no significant difference between the results before and after TC. The assessment of this module clearly suffers from many pixels with

low to no occupancy in the DIFF of the right front-end chip. The pattern resembles the bad pixels discussed in Section 5.5.1, leading to the conclusion, that the low occupancy here is probably caused by a feature of the AFE and cannot clearly be attributed to the hybridization. When ignoring these pixels, DCM31 would probably barely fulfill the acceptance criteria. However, upon closer investigation, another problematic area can be discovered in the top left corner of the module. The effect is much more pronounced in the ToT map of the same Source Scan, shown in Figure 8.19. Pixels in this area, and also a second cluster in the left third of the right readout ASIC, register significantly lower ToT values than surrounding pixels. This suggests a poor quality of the bump-bond interconnections in these areas. Therefore, the automated assessment of DCM31 can be confirmed.



(a) DCM31 source scan ToT map before TC. (b) DCM31 source scan ToT map after TC.

Figure 8.19: ToT map of a Source Scan of DCM31 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.

In summary, both tested samples seem to perform well at first sight, but ultimately fail to fulfill the acceptance criteria. However, the observed performance issues are comparably minor and Vendor M stated that they noticed irregularities during their internal QC process and are investigating the problem. Unfortunately, at the time of writing this thesis, no solution has been reported. Consequently, vendor combination C8 can only be recommended as supplier for the production of ITk modules, once the reported problems have been resolved.

COMBINATION C9

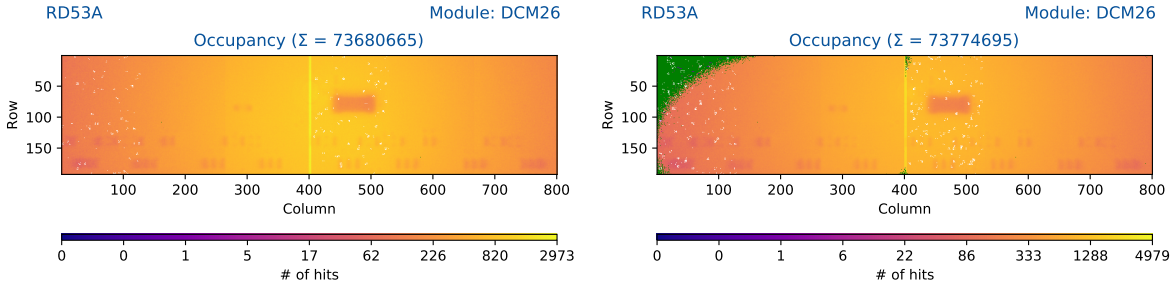
Finally, combination C9 has been assessed based on two assembled and fully tested samples, DCM25 and DCM26. Both samples behave similarly, as shown in Figure 8.20 for DCM26 and Figure D.26 for DCM25 and summarized in Table 8.9. The large number of pixels identified as poorly or entirely disconnected in the Source Scan of DCM25 can be attributed to the overall low occupancy, and therefore low statistical significance of this scan. An inspection of the occupancy map in Figure D.26 gives no reason for concern.

After TC, however, both samples show large clusters of disconnected bump-bonds in multiple corners, clearly suggesting an insufficient durability of the hybridization when exposed to thermal and mechanical stress. While the observed effects are similar to the results of vendor combination C8, they are more severe for combination C9.

In summary, similar to combination C8, combination C9 employing thermo-compression with In bumps cannot be recommended for the production of ITk modules, until the reported

	Before TC			After TC		
Module	Poor	Disconnected	Unknown	Poor	Disconnected	Unknown
DCM25	3 622	136	1 067	305	2 806	341
DCM26	2	0	524	33	6 006	1 039

Table 8.9: Summary of modules tested for vendor combination C9.



(a) DCM26 source scan occupancy map before TC. (b) DCM26 source scan occupancy map after TC.

Figure 8.20: Occupancy map of a Source Scan of DCM26 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.

problems can be resolved through an iterative feedback process with Vendor M.

8.5 SUMMARY AND CONCLUSIONS

Summing up the previous sections, results for 21 modules were successfully obtained and evaluated in the context of the ATLAS ITk hybridization MS as part of this work. Based on these results, recommendations about the suitability as supplier for the ATLAS ITk production can be expressed for seven of the nine vendor combinations. These recommendations are summarized in Table 8.10.

Two vendor combinations can be recommended without reservations based on the results presented in this chapter. Three more combinations require further investigation of the observed issues to be suitable for use during module mass production. Two vendor combinations could not be evaluated within the given constraints of this work due to the limited number of available samples. However, positive results of the other testing sites suggest that these combinations could be recommended as well. Finally, two vendor combinations cannot be recommended at all.

At the time of writing this thesis, the hybridization MS is still ongoing and no conclusion has been found. However, the results obtained by the other qualified testing sites so far are in good agreement with the results presented in this chapter.

Comb.	Recommendation
C1	Recommended without reservations.
C2	Recommended only for 100 µm thin sensors. Issues with 150 µm thick sensors require further investigation.
C3	Not recommended.
C4	Not evaluated.
C5	Not evaluated.
C6	Recommended without reservations.
C7	Not recommended.
C8	Not yet recommended. Problems need to be understood and resolved. Bump-bond interconnection quality needs to be improved.
C9	Not yet recommended. Problems need to be understood and resolved. Bump-bond interconnection quality needs to be improved.

Table 8.10: Summary of recommendations for vendor combinations for the hybridization MS.

CHAPTER 9

Conclusion and Outlook

The LHC will be upgraded to the HL-LHC after its third run period in the upcoming years. This upgrade will bring challenges in terms of readout speed, radiation hardness and other areas for the large-scale experiments at the LHC. The ATLAS experiment in particular will experience an almost complete rebuild to meet these new demands, the Phase-II upgrade.

Summing up the results presented in this thesis, several quality assurance and quality control measurement campaigns have been conducted for the ITk project of the ATLAS Phase-II upgrade. The presented results are, like most of the comprehensive pre-production efforts for the ITk project, based on RD53A, the first large-scale prototype of the new readout ASIC for the hybrid pixel detectors of the inner trackers of both, the upgraded ATLAS and CMS experiments. Even before the chip was available in silicon, a new readout and data acquisition system for the new generation of hybrid pixel particle detector readout ASICs, BDAQ53, was co-developed within the scope of this work. The system was initially developed based on RD53A chip simulations and, once the chip was available for testing, was evaluated in-depth and validated to give reproducible results suitable for characterizing the readout ASIC itself. Additionally, all the appropriate scan routines and algorithms were explained in-depth, that were co-developed and implemented within the scope of this work and served as a basis for the following measurements.

The measurements and results serving to characterize the bare chip performance of RD53A as well as its three different analog front-end flavors were presented and evaluated. In the scope of this work, several significant results were obtained that contribute to the community's understanding of the performance and limitations of the prototype chip. RD53A could be verified to work mostly within the expected parameters regarding the linearity of the global reference current IREF and several bias DACs. General functionality of the digital and analog functionality was confirmed by means of Digital- and Analog Scans, respectively.

A comprehensive performance comparison of the three analog front-end flavors within RD53A was conducted and presented in this thesis. The auto-zero feature of the synchronous analog front-end was shown to work as expected, reaching low threshold dispersion values without the need for a manual threshold tuning. The performance of the linear analog front-end was validated to satisfy all posed requirements. Despite the discovered issue in its implementation, the differential analog front-end was shown to perform far below the requirements,

reaching a minimum stable threshold of less than 450 e^-. These results served as a basis for Bonn's contribution as one of the community's major testing sites to the analog front-end reviews that were conducted by both, the RD53 and the ATLAS collaborations. All three flavors fulfill the posed requirements, but also considerable performance differences were observed and presented in this thesis. In the end, valuable feedback was provided for the development and design of the successor readout chip framework, RD53B.

Building on the QA measurements, the development and installation of a wafer probing setup and test routine for a large-scale QC campaign of RD53A were presented. Several thousand RD53A chips were successfully probed and analyzed using this setup within the scope of this work. The detailed analysis of the test results with an unprecedented sample size has revealed interesting and valuable insights into the peculiarities of RD53A. About 80 % of the tested chips were found to be suitable for further research and development purposes within the ITk pre-production program. Last but not least, a solid foundation and experience base could be established for the development of the readout chip QC procedure for the mass production phase of ATLAS ITk pixel detector modules.

Furthermore, the transition from the review of the bare readout chip's performance to a more comprehensive use case of complete hybrid particle detector modules was presented. RD53A could be verified to perform well in such a context and to be perfectly suited for advanced measurements and evaluation campaigns of other parts of the final pixel detector module design in the course of the ITk pre-production program. Multiple test beam campaigns have been conducted together with measurements of the analog front-ends' change in performance. Most importantly, despite the three substantially different analog front-end implementations, RD53A was shown to be a useful basis for the evaluation of sensors or hybridization techniques.

Finally, all the previous results and experiences have been utilized during an extensive measurement campaign for the official ITk hybridization market survey. More than twenty RD53A-based double-chip modules from seven different combinations of potential suppliers were successfully evaluated with respect to the quality of their hybridization. For many modules, adequate performance could be attested, even after an accelerated aging process, while also significant issues with some of the hybridization processes could be identified. In the end, well-founded recommendations were formulated for all evaluated processes that should serve as a solid basis for the conclusion of the hybridization market survey.

All in all, RD53A is going to play a substantial role during the pre-production of the ATLAS ITk for years to come. The data and categorization obtained from wafer probing in the scope of this work are going to be used for chip selection and matching throughout the pre-production program. A wafer probing setup for QC measurements of the next-generation readout chips can be easily derived from the setup that was presented in this thesis. While the characterization phase of RD53A is finished, some of the discovered issues and their solutions need to be followed up based on the second generation of readout chips. Last but not least, the hybridization measurements revealed potential weaknesses of some of the investigated processes that need to be followed up, if the corresponding vendors are selected for module production for the ITk. As availability of the successor chips increases and the limited inventory RD53A chips depletes over the course of the pre-production program, the community's focus will shift towards later chip generations. However, Some of the results and experiences obtained in the scope of this work should remain relevant throughout the production phase of the ATLAS ITk pixel detector.

APPENDIX A

SUPPLEMENTARY INFORMATION ON RD53A

This appendix contains additional information on RD53A, including a list of all global and in-pixel registers and a short explanation of their functions. Important signals and voltage- and current levels are explained together with an explanation of global pixel address encoding.

A.1 PIXEL ADDRESS ENCODING

Global pixel addressing in RD53A is implemented in a hierarchical fashion. A full overview of the process is shown in Figure A.1.

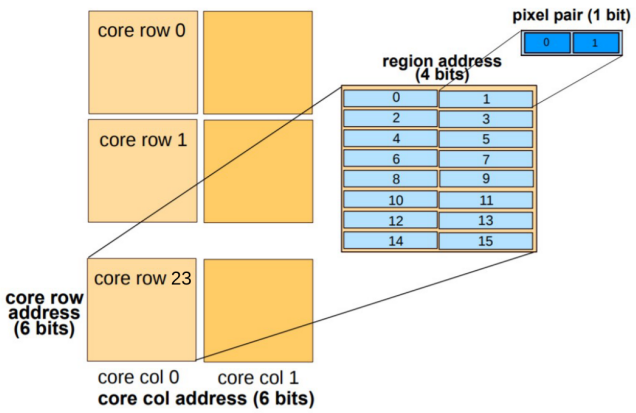


Figure A.1: RD53A pixel address encoding [33, p.73].

First, the pixel core is addressed globally, by specifying a core column number between [0,49] and a core row number from the range [0, 23]. A core consists of 8×8 pixels. These 64 pixels are structured in 16 regions, consisting of 4 adjacent pixels in a single row. The region in the already addressed core is addressed with 4 additional bit. One last bit is used to address the pixel pairs, each region is divided into. This pixel pair consists of two pixel, which share a single configuration register. So, when configuring one pixel, the second pixel in the pair needs to be configured simultaneously. The RD53A mask shifting algorithm of BDAQ53 makes use

of this by always enabling and disabling both pixels of a pixel pair at the same time (ref. to Section 4.3.1).

In total, pixel addressing in RD53A requires 17 bit, split into two registers: REGION_COL consists of the core column number, 1 bit to select the left or right region in this core column and another 1 bit to select the left or right pixel pair in the selected region, while CORE_ROW contains the core row address and 3 bit to select which of the 8 region rows to select.

Once these two address registers are written, the pixel pair configuration register can be accessed via PIX_PORTAL. [33, pp.72-73]

A.2 GLOBAL REGISTERS

Register Name	Description
PIX_PORTAL	Virtual register to access pixel matrix
REGION_COL	Region column address
REGION_ROW	Region row address
PIX_MODE	Mode bits: broadcast, autocol, autorow, broadcastmask
PIX_DEFAULT_CONFIG	Selects default configuration in pixels
IBIASP1_SYNC	Current of the main branch of the CSA
IBIASP2_SYNC	Current of the splitting branch of the CSA
IBIAS_SF_SYNC	Current of the preamplifier SF
IBIAS_KRUM_SYNC	Current of the Krummenacher feedback
IBIAS_DISC_SYNC	Current of the comparator differential amplifier
ICTRL_SYNCT_SYNC	Current of the oscillator delay line
VBL_SYNC	Baseline voltage for offset compensation
VTH_SYNC	Discriminator threshold voltage
VREF_KRUM_SYNC	Krummenacher voltage reference
PA_IN_BIAS_LIN	Preamplifier input branch current
FC_BIAS_LIN	Folded cascode branch current
KRUM_CURR_LIN	Krummenacher current
LDAC_LIN	Fine threshold
COMP_LIN	Comparator current
REF_KRUM_LIN	Krummenacher reference voltage
Vthreshold_LIN	Global threshold voltage
PRMP_DIFF	Preamplifier input stage current
FOL_DIFF	Preamplifier output follower current
PRECOMP_DIFF	Precomparator tail current
COMP_DIFF	Comparator total current
VFF_DIFF	Preamplifier feedback current (return-to-baseline)
VTH1_DIFF	Negative branch voltage offset
VTH2_DIFF	Positive branch voltage offset
LCC_DIFF	Leakage current compensation current
CONF_FE_DIFF	Configuration for DIFF
CONF_FE_SYNC	Configuration for SYNC

VOLTAGE_TRIM	Analog and digital voltage regulator trim
EN_CORE_COL_SYNC	Enable core (SYNC)
EN_CORE_COL_LIN_1-2	Enable core (LIN)
EN_CORE_COL_DIFF_1-2	Enable core (DIFF)
LATENCY_CONFIG	Latency configuration
WR_SYNC_DELAY_SYNC	Write synchronization delay (SYNC)
INJECTION_SELECT	Analog inj., digital inj., inj. fine delay
CLK_DATA_DELAY	Clock and data fine delay
VCAL_HIGH	VCAL_HIGH injection voltage
VCAL_MED	VCAL_MED injection voltage
CH_SYNC_CONF	Threshold and phase adjust for channel synchronizer
GLOBAL_PULSE_ROUTE	Global pulse routing select
MONITOR_FRAME_SKIP	How many data frames to skip before sending a monitor frame
EN_MACRO_COL_CAL_SYNC_1-4	Enable macrocolumn analog calibration for SYNC
EN_MACRO_COL_CAL_LIN_1-5	Enable macrocolumn analog calibration for LIN
EN_MACRO_COL_CAL_DIFF_1-5	Enable macrocolumn analog calibration for DIFF
DEBUG_CONFIG	Output channel and driver configuration
OUTPUT_CONFIG	Output channel and driver configuration
OUT_PAD_CONFIG	LVDS configuration
GP_LVDS_ROUTE	General purpose output routing configuration
CDR_CONFIG	CDR configuration
CDR_VCO_BUFF_BIAS	Bias current for VCO buffer of CDR
CDR_CP_IBIAS	Bias current for CP of CDR
CDR_VCO_IBIAS	Bias current for VCO of CDR
SER_SEL_OUT	20 bit serializer output select
CML_CONFIG	20 bit serializer output settings
CML_TAP0_BIAS	Bias current 0 for CML driver
CML_TAP1_BIAS	Bias current 1 for CML driver
CML_TAP2_BIAS	Bias current 2 for CML driver
AURORA_CC_CONFIG	Aurora configuration bits
AURORA_CB_CONFIG0	Aurora channel bonding configuration bits
AURORA_CB_CONFIG1	Aurora channel bonding configuration bits
AURORA_INIT_WAIT	Aurora init wait
MONITOR_SELECT	Current and voltage monitoring MUX selection
HITOR_0_MASK_SYNC	Mask bits for the HitOr_0 for SYNC
HITOR_1_MASK_SYNC	Mask bits for the HitOr_1 for SYNC
HITOR_2_MASK_SYNC	Mask bits for the HitOr_2 for SYNC
HITOR_3_MASK_SYNC	Mask bits for the HitOr_3 for SYNC
HITOR_0_MASK_LIN_0-1	Mask bits for the HitOr_0 for LIN
HITOR_1_MASK_LIN_0-1	Mask bits for the HitOr_1 for LIN
HITOR_2_MASK_LIN_0-1	Mask bits for the HitOr_2 for LIN
HITOR_3_MASK_LIN	Mask bits for the HitOr_3 for LIN
HITOR_0_MASK_DIFF_0-1	Mask bits for the HitOr_0 for DIFF
HITOR_1_MASK_DIFF_0-1	Mask bits for the HitOr_1 for DIFF

HITOR_2_MASK_DIFF_0-1	Mask bits for the HitOr_2 for DIFF
HITOR_3_MASK_DIFF_0-1	Mask bits for the HitOr_3 for DIFF
MONITOR_CONFIG	ADC band gap trim bits, ADC trim bits
SENSOR_CONFIG_0-1	Enable temp/rad sensors, dyn. el. matching bits
AutoRead0	Auto read register a for line 0
AutoRead1	Auto read register b for line 0
AutoRead2	Auto read register a for line 1
AutoRead3	Auto read register b for line 1
AutoRead4	Auto read register a for line 2
AutoRead5	Auto read register b for line 2
AutoRead6	Auto read register a for line 3
AutoRead7	Auto read register b for line 3
RING_OSC_ENABLE	Enable ring oscillators
RING_OSC_0-7	Counter value of ring oscillator #0 - #7
BCIDCnt	Bunch counter
TrigCnt	Counts all received triggers
LockLossCnt	Counts the number of times the channel sync lost lock state
BitFlipWngCnt	Counts the bit flip warning messages
BitFlipErrCnt	Counts the bit flip error messages
CmdErrCnt	Counts command decoder error messages
WngFifoFullCnt_0-3	Counters that hold the # of writes when FIFO was full
AI_REGION_COL	Allows to read the auto increment column value
AI_REGION_ROW	Allows to read the auto increment row value
HitOr_0_Cnt	HitOr_0 counter
HitOr_1_Cnt	HitOr_1 counter
HitOr_2_Cnt	HitOr_2 counter
HitOr_3_Cnt	HitOr_3 counter
SkippedTriggerCnt	Skipped trigger counter
ErrWngMask	Mask single error warning messages
MonitoringDataADC	Contains the value of the ADC to be read back

Table A.1: Gloabl register assignments of RD53A [33, pp.66-68].

A.3 PIXEL REGISTERS

Bit	Front End Flavor		
	DIFF	LIN	SYNC
0	Enable	Power and Enable	
1	Injection Enable		
2	HitOr Enable		
3	TDAC bit 0	TDAC bit 0	-
4	TDAC bit 1	TDAC bit 1	-
5	TDAC bit 2	TDAC bit 2	-
6	TDAC bit 3	TDAC bit 3	-
7	TDAC sign	Gain Select	-

Table A.2: Pixel register assignments of RD53A [33, p.72]. The SYNC only supports the pixel masks with the first three bits of its pixel registers, while DIFF and LIN additionally implement four TDAC bits. In addition, the DIFF features a TDAC sign bit, effectively doubling the amount of available local threshold settings, while the LIN has the option to select the gain of its amplifier for every pixel individually (see Section 3.1.2).

A.4 SIGNALS

Acronym	Name	Description
BX_CLK	Bunch Crossing Clock	40 MHz clock of the proton-proton collisions in the accelerator.
CMD_CLK	Command Clock	160 MHz base clock provided by the PLL locking to the input command stream.
SER_CLK	Serializer Clock	Clock signal for the output serializers. Can be chosen between 160 MHz and 1.28 GHz.
DEL_CLK	Delay Clock	Clock signal used to synchronize all other clock signals of the chip to the BX_CLK.
CMD	Command	RD53A command input with a nominal frequency of 160 MHz.
CAL_EDGE	Calibration injection signal	Signal issued to produce a calibration injection signal in form of a rising edge between VCAL_MED and VCAL_HIGH.
ECR	Event Counter Reset	Clears all pending hits and triggers inside the matrix and global logic of the chip.

Table A.3: Global signals of RD53A [33, pp.33, 47].

A.5 VOLTAGES AND CURRENTS

Acronym	Name	Description
VIN_A	Analog input voltage	Analog input voltage for the regulators.
VIN_D	Digital input voltage	Analog input voltage for the regulators.
VDDA	Analog supply voltage	Analog supply voltage for the chip.
VDDD	Digital supply voltage	Analog supply voltage for the chip.
VREF_A	Analog reference voltage	Reference voltage for the analog regulators.
VREF_D	Digital reference voltage	Reference voltage for the digital regulators.
VREF_ADC	ADC reference voltage	Reference voltage for the monitoring ADC as well as the calibration injection circuit.
IREF	Global reference current	The global reference current (nominal 4 µA) for almost all circuits and DACs.

Table A.4: Global voltages of RD53A [33].

A.6 POWERING MODES

Three different powering modes are available for RD53A, depending on the carrier board and wire-bond configuration.

- Direct Powering: In Direct Powering mode, the analog and digital supply voltages VDDA and VDDD are supplied directly to the respective rails. The on-chip voltage regulators are completely bypassed in this mode. The measured current consumption directly reflects the chips power consumption.
- LDO: In LDO mode, the chip's sLDO regulators are used as normal LDO voltage regulators. A nominal constant voltage of 1.5 V is supplied to the regulator inputs VIN_A and VIN_D. The two inputs may be shorted and supplied from a single power supply. The current consumption is proportional to the power consumption of the chip.
- Shunt: In Shunt mode, the chip is supplied with a constant current composed of the chip's actual power consumption and some additional headroom. The sLDO regulators supply the internal voltage rails with constant VDDA and VDDD and the necessary current. The surplus current is shunted to ground. The current measured at the power supply is constant. The regulators can also be supplied with a constant voltage in this mode, as long as the current limit is sufficient. This mode is the default operational mode for serially powered detector modules.

APPENDIX B

SUPPLEMENTARY INFORMATION ON THE RD53 AFE REVIEW

This appendix contains additional information on the RD53 AFE review conducted in December 2018 to evaluate the performance of the three Analog Front-End (AFE) flavors of RD53A: Synchronous Analog Front-End (SYNC), Linear Analog Front-End (LIN) and Differential Analog Front-End (DIFF).

B.1 CONFIGURATION GUIDELINES

This section lists the guidelines for configuration of the three AFEs as advised by the respective front-end designers.

Setting	Rec. Value	Description
VTH_SYNC	60-400	Global Threshold DAC (GDAC) for SYNC.
VBL_SYNC	350-450	Baseline voltage, defining the effective threshold together with VTH_SYNC.
VREF_KRUM_SYNC	350-500	Reference voltage for the Krummenacher circuit.
IBIASP1_SYNC	60-150	Defines the preamplifier bias current together with IBIAS_P2_SYNC. Higher values can decrease noise for the cost of more power consumed.
IBIASP2_SYNC	90-225	Should be 1.5 times the value of IBIASP1_SYNC.
IBIAS_SF_SYNC	50-200	Bias current of the preamplifier source follower.
IBIAS_KRUM_SYNC	10-160	Feedback current for the preamplifier. Defines the return-to-baseline and therefore is tuned to optimize the ToT response.
IBIAS_DISC_SYNC	100-400	Large influence on power consumption and slight influence on threshold.
ICTRL_SYNCT_SYNC	50-100	.

Table B.1: Configuration guidelines for the SYNC.

Setting	Rec. Value	Description
Vthreshold_LIN	408	Global threshold setting for LIN. Proportional to the resulting threshold. The effective threshold is the difference between the signal baseline defined by REF_KRUM_LIN and the voltage generated by Vthreshold_LIN.
KRUM_CURR_LIN	50	Feedback current for the preamplifier. Defines the return-to-baseline and therefore is tuned to optimize the ToT response.
LDAC_LIN	130	Defines the output range of the pixels' TDACs. Larger LDAC_LIN results in a larger step size, trading increased tuning range for granularity.
FC_BIAS_LIN	20	Current of the preamplifier folded cascode branch.
COMP_LIN	110	Reference current of the threshold discriminator input stage.
PA_IN_BIAS_LIN	300	Current in the preamplifier input branch. Main method of decreasing the AFEs current consumption for the cost of increased noise.
REF_KRUM_LIN	300	Preamplifier DC baseline.

Table B.2: Configuration guidelines for the LIN.

Setting	Rec. Value	Description
CONF_FE_DIFF	0b10	Choose feedback capacitance 0 and enable LCC.
VTH1_DIFF	14	Main GDAC for the DIFF. Defines the global threhsold together with VTH2_DIFF.
VTH2_DIFF	0	Secondary GDAC. Should be kept at 0 for small absolute thresholds.
PRMP_DIFF	533	Input stage current of the preamplifier. Trades increased power consumption for better performance.
FOL_DIFF	542	Preamplifier output follower current. Trades increased power consumption for better performance.
PRECOMP_DIFF	512	Precomparator bias. Used to trim the bit weight for the TDACs, trading granularity for range.
COMP_DIFF	528	Total current of the comparator. Trades increased power consumption for better performance.
VFF_DIFF	140	Feedback current for the preamplifier. Defines the return-to-baseline and therefore is tuned to optimize the ToT response.
LCC_DIFF	20	Leakage current compensation. Used to decrease noise with present leakage current from a sensor.

Table B.3: Configuration guidelines for the DIFF.

APPENDIX C

SUPPLEMENTARY INFORMATION ON THE HYBRIDIZATION MARKET SURVEY

This appendix contains additional information on the ATLAS ITk hybridization Market Survey (MS), conducted to evaluate the performance of different vendor combinations applying for the production of the final ATLAS ITk pixel detector modules.

C.1 ACCEPTANCE CRITERIA

All acceptance criteria for the produced, RD53A-based prototype DCMs for the hybridization MS are listed below [70].

- For planar sensors:
  - Width of the sensor tile: (41.100 ± 0.050) mm.
  - Height of the sensor tile: $(10.900 - 11.200) \pm 0.015$ mm.
  - Total thickness of the bare module: (0.325 ± 0.015) mm for 150 µm thick sensors and (0.275 ± 0.015) mm for 100 µm thick sensors.
- For 3D single-chip sensors:
  - Width of the sensor tile: (20.300 ± 0.050) mm.
  - Height of the sensor tile: (19.500 ± 0.050) mm.
  - Total thickness of the bare module: (0.425 ± 0.015) mm.
- Ratio of leakage current at operating voltage with respect to sensor acceptance test < 10 V.
- Leakage current $< 1.5\,\mu\text{A}\,\text{cm}^{-2}$ at $V = V_{\text{depl}} + 50$ V , at 20 °C and at a relative humidity of less than 50 %.
  - Leakage current stability better than 20 %, when measured over 48 h.
- Breakdown voltage $> V_{\text{depl}} + 80$ V.

- No ASICs failing after assembly.

- Number of total individual bump failures per SCM < 600.

- Number of total individual bump failures per DCM $< 1\,200$.

- Number of individual bump failures per ASIC, which are grouped in a cluster (cluster size) < 50.

C.2 TUNING PARAMETERS

This section describes the tuning parameters for the `RD53A` asics in the tested DCMs for the `Source Scan` or `Bump Connectivity Scan`, if applicable.

C.2.1 SOURCE SCAN

Running a `Source Scan` with a ^{241}Am X-ray source does not require a particularly low threshold, especially for unirradiated senors, since the amount of deposited charge is typically larger than $10\,\mathrm{ke}^{-}$. However, if the result of the scan should be used for the analysis of the module's bump-bond connectivity, as is the case for the measurements presented here, a homogeneous threshold distribution , especially over the AFE borders is desirable.

To achieve this, the `Global-` and `Local Threshold Tuning` routines, described in Section 4.6 are executed, followed by a `Noise Occupancy Scan` and a `Stuck Pixel Scan`, to disable noisy and stuck pixels, respectively. This series of routines typically results in a set of register settings, that is depicted in Table C.1.

Setting	Value	Description
IBIASP1_SYNC	90	Defines the preamplifier bias current together with IBIAS_P2_SYNC. Higher values can decrease noise for the cost of more power consumed.
IBIASP2_SYNC	140	Should be 1.5 times the value of IBIASP1_SYNC.
IBIAS_SF_SYNC	80	Bias current of the preamplifier source follower.
IBIAS_KRUM_SYNC	55	Feedback current for preamplifier. Defines the return-to-baseline and is tuned to optimize the ToT response.
IBIAS_DISC_SYNC	300	Large influence on power consumption and slight influence on threshold.
ICTRL_SYNCT_SYNC	100	
VBL_SYNC	380	Baseline voltage, defining the effective threshold together with VTH_SYNC.
VTH_SYNC	250 - 300	Global Threshold DAC (GDAC) for SYNC.
VREF_KRUM_SYNC	450	Reference voltage for the Krummenacher circuit.
CONF_FE_SYNC	0x0002	
PA_IN_BIAS_LIN	350	Current in the preamplifier input branch. Decreases the AFEs current consumption for the cost of more noise.
FC_BIAS_LIN	20	Current of the preamplifier folded cascode branch.
KRUM_CURR_LIN	32	Feedback current for the preamplifier. Defines the return-to-baseline and is tuned to optimize the ToT response.
LDAC_LIN	130	Defines the output range of the pixels' TDACs. Larger LDAC_LIN results in a larger step size, trading increased tuning range for granularity.
COMP_LIN	110	Ref. current of the threshold discriminator input stage.
REF_KRUM_LIN	300	Preamplifier DC baseline.
Vthreshold_LIN	325 - 425	Global threshold setting for LIN. Proportional to the resulting threshold. The effective threshold is the difference between the signal baseline defined by REF_KRUM_LIN and the voltage generated by Vthreshold_LIN.
PRMP_DIFF	511	Input stage current of the preamplifier. Trades increased power consumption for better performance.
FOL_DIFF	542	Preamplifier output follower current. Trades increased power consumption for better performance.
PRECOMP_DIFF	512	Precomparator bias. Used to trim the bit weight for the TDACs, trading granularity for range.
COMP_DIFF	1023	Total current of the comparator. Trades increased power consumption for better performance. Used to equalize the performance of the pixel matrix.
VFF_DIFF	40	Feedback current for the preamplifier. Defines the return-to-baseline and is tuned to tune the ToT response.
VTH1_DIFF	300 - 600	Main GDAC for the DIFF. Defines the global threshold together with VTH2_DIFF.
VTH2_DIFF	50	Secondary GDAC. Should be kept at 0 for low thresholds.
LCC_DIFF	20	Leakage current compensation. Used to decrease noise with present leakage current from a sensor.
CONF_FE_DIFF	0b10	Choose feedback capacitance 0 and enable LCC.

Table C.1: AFE conf. parameters for Source Scan measurements during the hybridization MS.

C.2.2 Bump Connectivity Scan

Tuning for a Bump Connectivity Scan is counter-intuitive, because the AFE's have to be optimized for the highest possible susceptibility to leakage current, where usually the opposite is desirable. For the results presented in this thesis, some registers were set manually to the values depicted in Table C.2 and the threshold was tuned to about 600 ΔVCAL for the SYNC and 1 000 ΔVCAL for the LIN and DIFF. Table C.2 shows only parameters that differ from Table C.1.

Setting	Value	Description
IBIASP1_SYNC	38	Defines the preamplifier bias current together with IBIAS_P2_SYNC. Higher values can decrease noise for the cost of more power consumed.
IBIASP2_SYNC	10	Should be 1.5 times the value of IBIASP1_SYNC.
IBIAS_SF_SYNC	100	Bias current of the preamplifier source follower.
VTH_SYNC	370 - 390	Global Threshold DAC (GDAC) for SYNC.
PA_IN_BIAS_LIN	30	Current in the preamplifier input branch. Main method of decreasing the AFEs current consumption for the cost of increased noise.
Vthreshold_LIN	400 - 500	Global threshold setting for LIN. Proportional to the resulting threshold. The effective threshold is the difference between the signal baseline defined by REF_KRUM_LIN and the voltage generated by Vthreshold_LIN.
PRMP_DIFF	10	Input stage current of the preamplifier. Trades increased power consumption for better performance.
VTH1_DIFF	750 - 1023	Main GDAC for the DIFF. Defines the global threhsold together with VTH2_DIFF.
VTH2_DIFF	50	Secondary GDAC. Should be kept at 0 for small absolute thresholds.
LCC_DIFF	20	Leakage current compensation. Used to decrease noise with present leakage current from a sensor.
CONF_FE_DIFF	0b0	Choose feedback capacitance 0 and disable LCC.

Table C.2: AFE configuration parameters for Bump Connectivity Scan measurements during the hybridization MS.

APPENDIX D

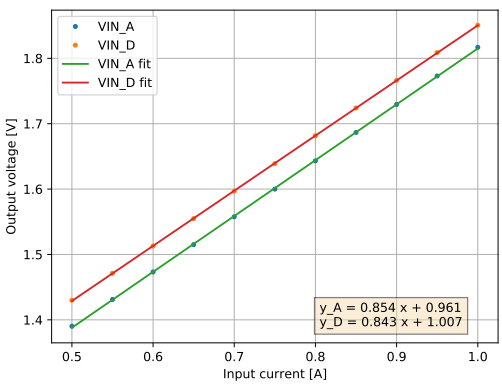
ADDITIONAL RESULTS

This appendix contains additional results and plots that are not included in the main text body.

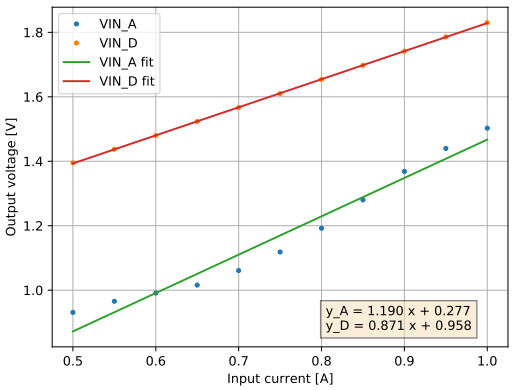
D.1 WAFER-LEVEL TESTS OF RD53A

In this section, additional results from wafer level testing of RD53A, as described in Chapter 6, are presented and shortly summarized.

D.1.1 REGULATOR IV CURVES



(a) Example for the IV curves of working sLDO regulators.



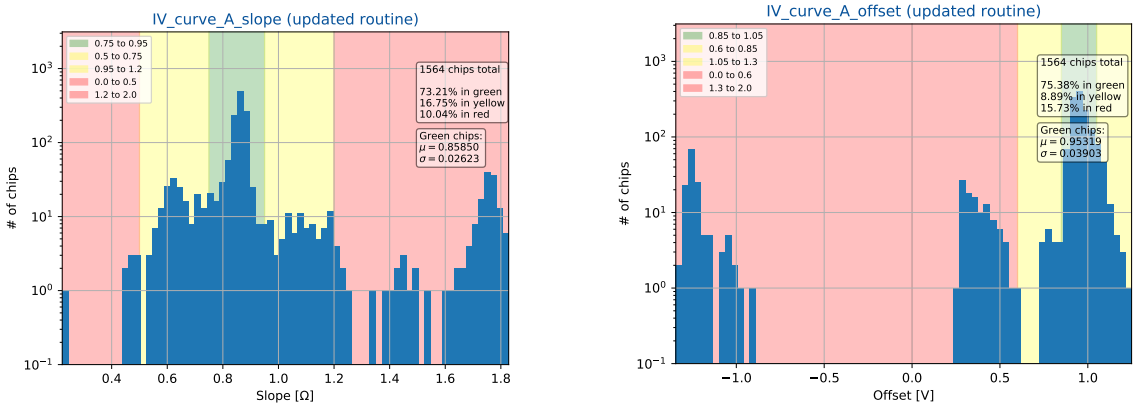
(b) Example for the IV curves of a working digital and a malfunctioning analog sLDO regulator.

Figure D.1: Two examples of sLDO regulator IV curves..

Figure D.1 shows two examples of Shunt Low-Dropout Regulator (sLDO) input IV curves as measured during wafer probing. In Figure D.1(a), both, the digital and analog regulator exhibit a linear behavior and the slope and offset values extracted from a straight line fit to the measured points nicely fall into the ranges required for a *working* chip. In Figure D.1(b) on the

other hand, the analog sLDO exhibits a non-linear behavior and the extracted slope and offset will be classified as *malfunctioning* and *not working*, respectively.

This measurement is performed on every tested chip, though for chips tested with the original routine, the IV curve was measured for increasing input currents instead of starting with the highest value and decreasing the constant input current. This method led to an artificially lowered yield, since many regulators had problems starting up with the slowly increasing input current, while they would be ramped up directly to their optimal working point much faster during actual operation. For this reason, only IV curves of chips tested with the updated routine are included into the evaluation of this test stage.



(a) Histogram of the slope values extracted from the analog regulator IV curves of chips tested with the updated routine. The background colors depict regions used to classify the chips. The yield for this test is shown in the top right corner.

(b) Histogram of the offset values extracted from the analog regulator IV curves of chips tested with the updated routine. The background colors depict the regions used to classify the chips. The yield for this test is shown in the top right corner.

Figure D.2: Results of the IV curve test for the analog regulator.

Figure D.2 shows the results of the two IV curve test stages for the analog sLDO regulator. About 73.2 % of all chips that were tested with the updated routine exhibit an extracted slope value that is categorized as *working* and 16.8 % are classified as *malfunctioning*. For the extracted offset, 75.4 % of chips are *working* and 8.9 % are *malfunctioning*.

D.1.2 VREF TRIMMING

Figure D.3 shows the results of the trimming test for VREF_ADC. 92.2 % of all tested chips can be trimmed to have a mean VREF_ADC of (0.894 ± 0.020) V, very close to the required 900 mV.

The distributions of optimal trim bit settings for VREF_D, VREF_A, and VREF_ADC are shown in Figure D.4. While the allocation of values is shifted towards the higher end of the range for both, VREF_D and VREF_A, it is almost centered for VREF_ADC. In all three cases, the overflow bin is not overly used, suggesting that all three references can still be trimmed to the optimal value in most cases.

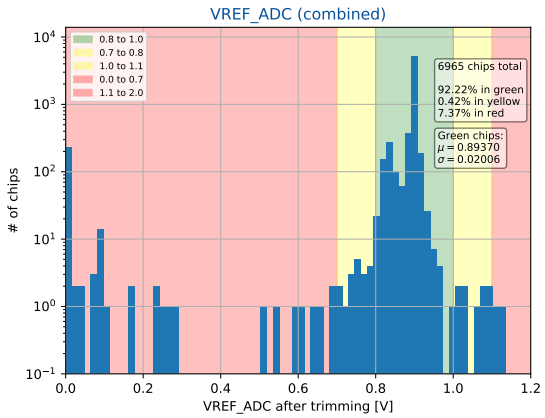
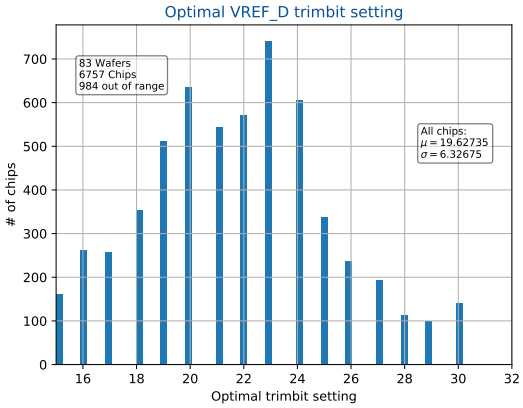
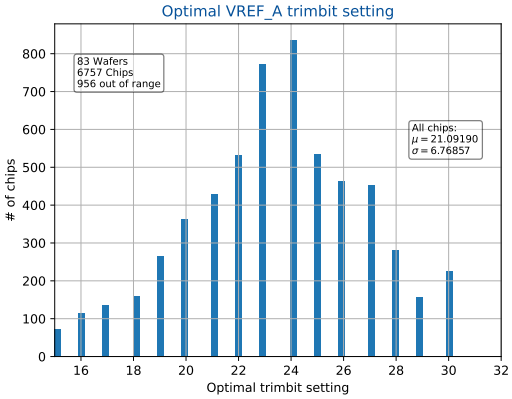


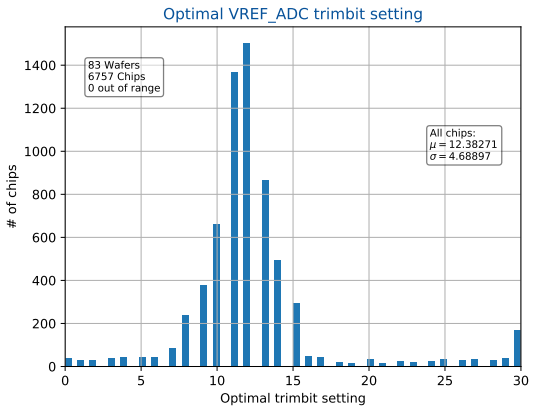
Figure D.3: Results of the VREF_ADC trimming test. The background colors depict the regions used to classify the chips. The individual yield for each test is shown in the top right corner.



(a) Histogram of the optimal VREF_D trim bit values.



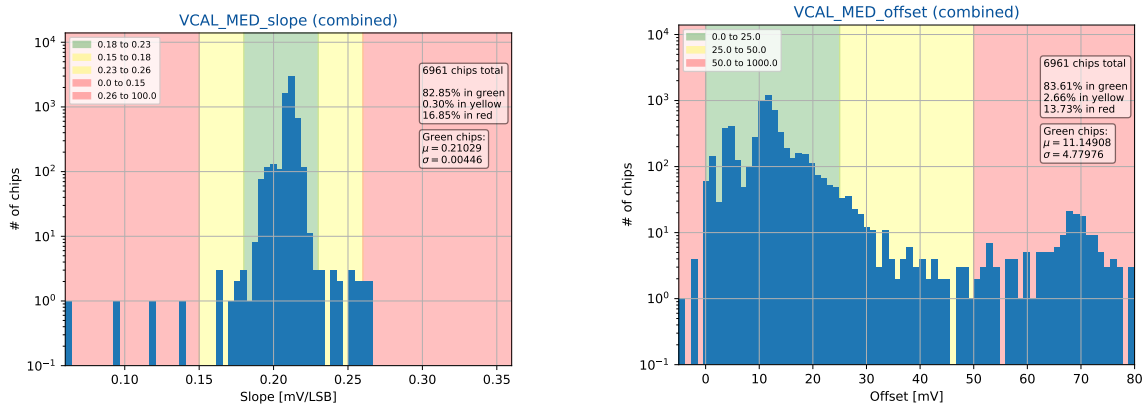
(b) Histogram of the optimal VREF_A trim bit values.



(c) Histogram of the optimal VREF_ADC trim bit values.

Figure D.4: Distribution of the optimal trim bit values for the different reference voltages of all tested chips.

D.1.3 INJECTION DAC CALIBRATION



(a) Histogram of the extracted VCAL_MED slope values of all tested chips.

(b) Histogram of the extracted VCAL_MED offset values of all tested chips.

Figure D.5: Results of the calibration measurement for VCAL_MED. The background colors depict the regions used to classify the chips. The yield for this test is shown in the top right corner.

Figure D.5 shows the results of the calibration measurements of the second calibration injection DAC of RD53A, VCAL_MED. The distributions are comparable to the results for VCAL_HIGH, about 85 % of all tested chips exhibit an extracted slope and offset value within the expected range and are categorized as *working*.

D.1.4 TOTAL POWER CONSUMPTION

The total power consumption is assessed during wafer probing by measuring the sum of the current consumption of the digital and analog domains at constant voltage in LDO mode at the power supply after configuring the chip to a known state. The result for this test stage is shown in Figure D.6. 99.6 % of all tested chips show a power consumption within the expected range and are classified as *working*.

D.1.5 REGISTER TEST

Figure D.7 shows the result distribution of the Register Test test stage. 100 % of all chips tested with the updated routine show no bad registers at all. This suggests that broken global registers are not a common failure mode for RD53A.

D.1.6 NOISE OCCUPANCY SCAN

The results of the Noise Occupancy Scan test are shown in Figure D.8. 81.6 % of all tested chips exhibit less than 1 % of noisy pixels and are categorized as *working*. 9.3 % of chips are classified as *malfunctioning*, with 1 to 5 % noisy pixels.

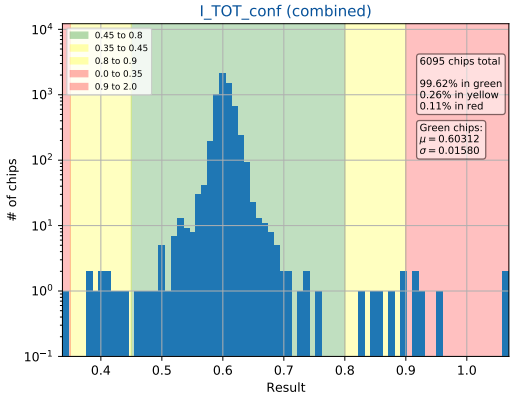


Figure D.6: Results of the total power consumption test of all tested chips. The background color depicts the regions used to classify the chips. The yield for this test is shown in the top right corner.

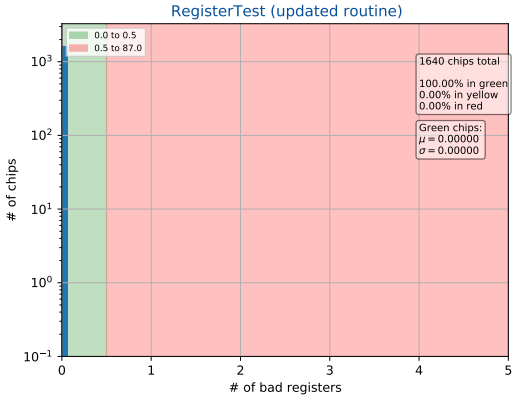


Figure D.7: Result of the `Register Test` of all chips tested with the updated probing routine. The x-axis shows the absolute amount of working global registers. The background colors depict the regions used to classify the chips. The yield for this test is shown in the top right corner.

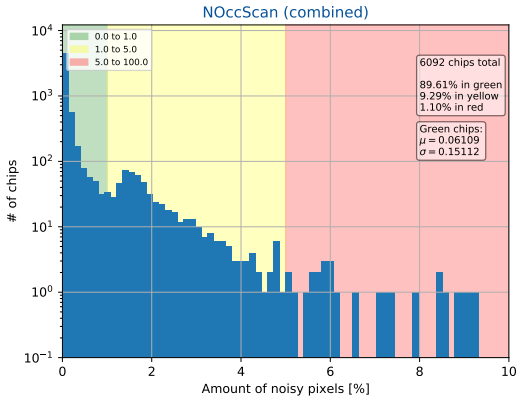


Figure D.8: Result of the `Noise Occupancy Scan` of all tested chips. The x-axis shows the relative amount of noisy pixels. The background colors depict the regions used to classify the chips. The yield for this test is shown in the top right corner.

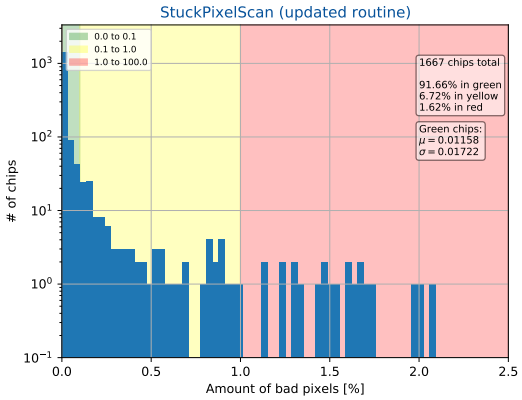
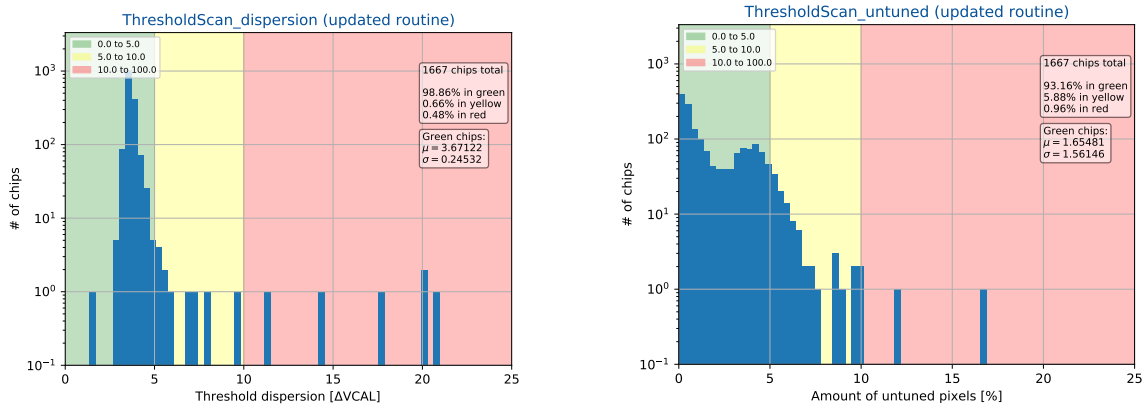


Figure D.9: Result of the `Stuck Pixel Scan` of all chips tested with the updated probing routine. The x-axis shows the relative amount of stuck pixels. The background colors depict the regions used to classify the chips. The yield for this test is shown in the top right corner.

D.1.7 STUCK PIXEL SCAN

Figure D.9 shows the result distribution of the `Stuck Pixel Scan`. 91.7 % of all chips tested with the updated routine are classified as *working*, exhibiting less than 0.1 % of stuck pixels. Another 6.7 % of chips exhibit 0.1 to 1 % of stuck pixels and are therefore classified as *malfunctioning*.

D.1.8 THRESHOLD SCAN



(a) Histogram of the threshold dispersion of all chips tested with the updated probing routine.

(b) Histogram of the amount of untuned pixels of all chips tested with the updated probing routine.

Figure D.10: Results of the threshold dispersion and amount of untuned pixels from a `Threshold Scan` of all chips tested with the updated probing routine. The background colors depict the regions used to classify the chips. The yield for this test is shown in the top right corner.

The results of the mean dispersion test and the untuned pixel test stages based on the `Threshold Scan` after `Threshold Tuning`, that are evaluated on all chips tested with the updated probing routine are shown in Figure D.10. 98.9 % of tested chips have a mean threshold dispersion below 5.0 % and are classified as *working*. Considering the amount of untuned pixels, 93.2 % of tested chips are classified as *working*, due to exhibiting less than 5.0 % of untuned pixels. Almost all other chips are categorized as *malfunctioning* and less than 1 % of all tested chips is classified as *not working*.

Finally, Figure D.11 shows the result of the mean threshold test stage for all chips that were tested with the original probing routine. Since in this routine no `Threshold Tuning` was performed, the distribution is much wider than for chips tested with the updated routine. Still, 80.6 % of all tested chips are ranked as *working* and another 15.8 % are classified as *malfunctioning*.

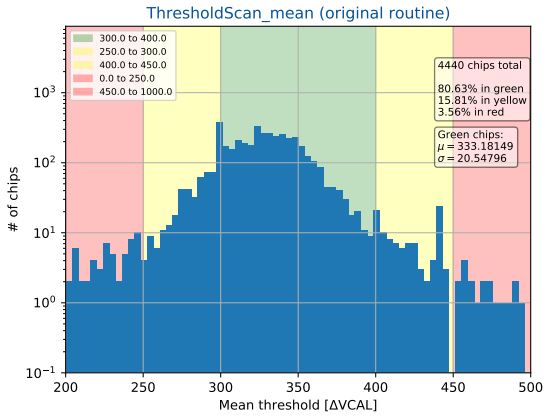


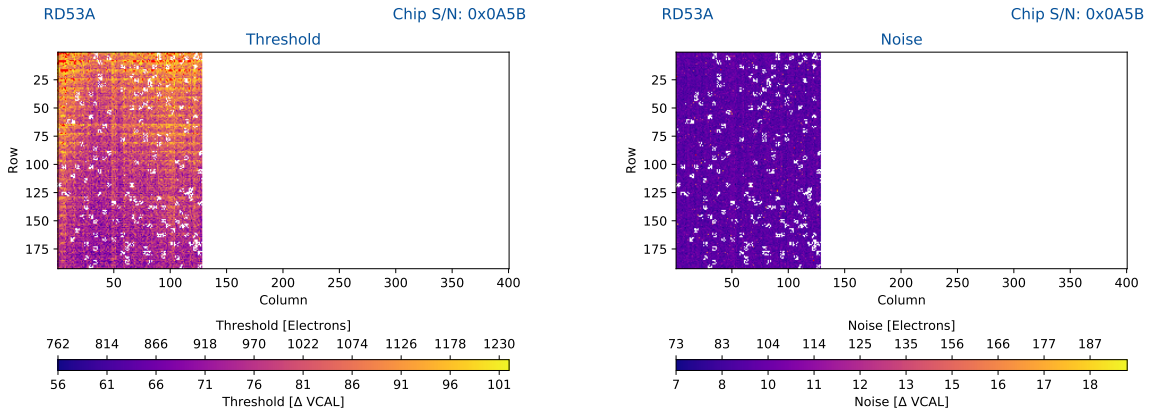
Figure D.11: Result of the `Threshold Scan` of all chips tested with the original probing routine. The x-axis shows the relative amount of stuck pixels. The background colors depict the regions used to classify the chips. The yield for this test is shown in the top right corner.

D.2 PIXEL DETECTOR MODULES BASED ON RD53A

This section contains additional results of RD53A SCMs, presented in Chapter 7.

D.2.1 ANALOG FRONT-END PERFORMANCE WITH SENSOR

The AFE performance measurements on an RD53A SCM were performed with a reverse bias voltage of 80 V and a resulting leakage current of about 100 nA.



(a) Threshold map of the SYNC in an SCM after tuning.

(b) Noise map of the SYNC in an SCM after tuning.

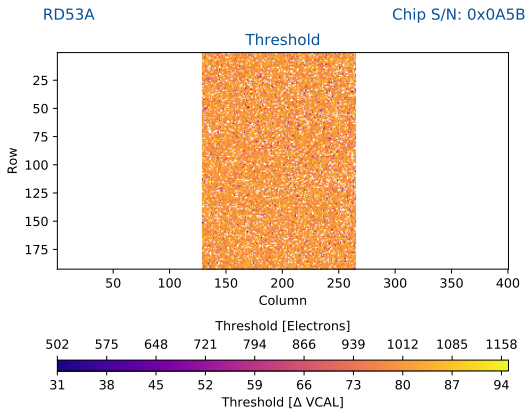
Figure D.12: Additional threshold and noise results of the SYNC in an SCM after tuning to a threshold of about 1 000 e^-.

The threshold map of the SYNC, shown in Figure D.12(a), exhibits the same gradient of the threshold from higher values at the top of the matrix to lower values in the bottom part. As this feature is already observed on a bare chip and discussed in Section 5.3.2, it is apparently not a feature of the SYNC in a module, but rather a general feature.

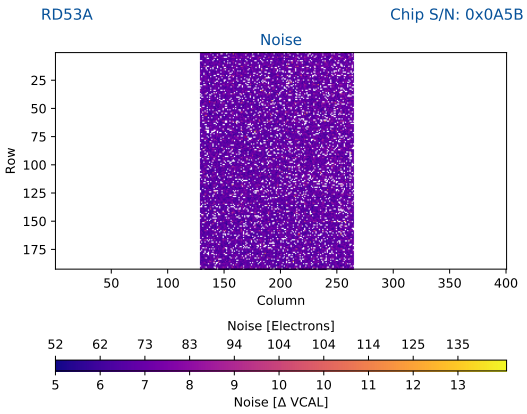
Figure D.12(b) shows the noise map of the SYNC in an SCM, which exhibits no gradients or irregularities.

Figure D.13 shows the threshold and noise map associated with the distributions discussed in Section 7.3. No gradients, hotspots or other irregularities can be observed.

The threshold and noise map of the DIFF in an SCM after tuning the good pixels to a threshold of about 1 000 e^- is shown in Figure D.14. Like for a bare chip, no remarkable features can be observed in these maps.

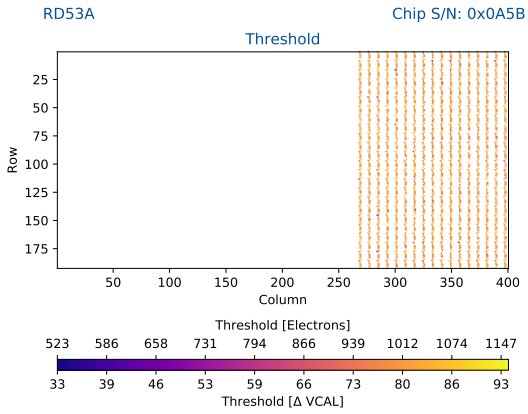


(a) Threshold map of the LIN in an SCM after tuning.

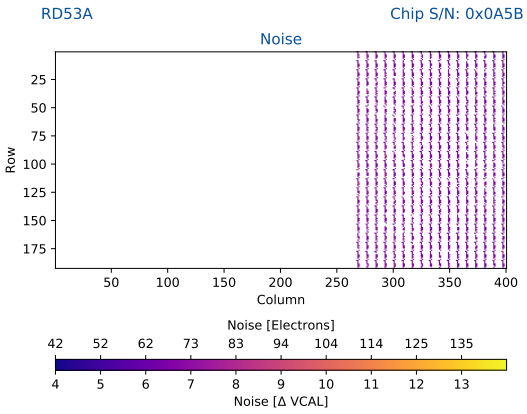


(b) Noise map of the LIN in an SCM after tuning.

Figure D.13: Additional threshold and noise results of the LIN in an SCM after tuning to a threshold of about 1 000 e^-.



(a) Threshold map of the DIFF in an SCM after tuning.



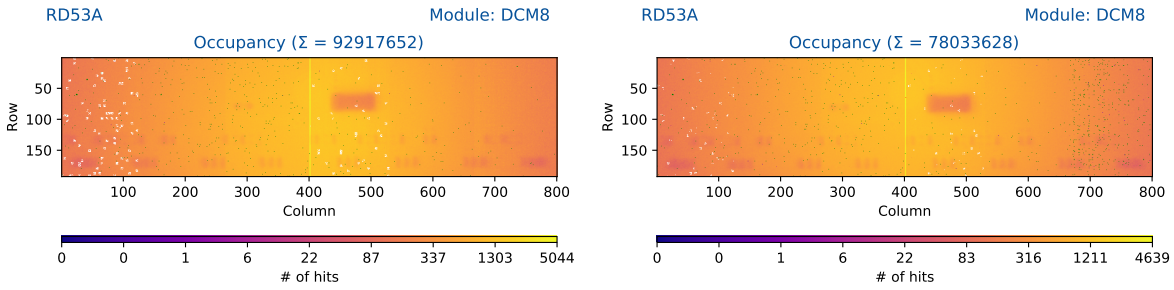
(b) Noise map of the DIFF in an SCM after tuning.

Figure D.14: Additional threshold and noise results of the DIFF in an SCM after tuning to a threshold of about 1 000 e^-.

D.3 QUALIFICATION OF THE HYBRIDIZATION PROCESS

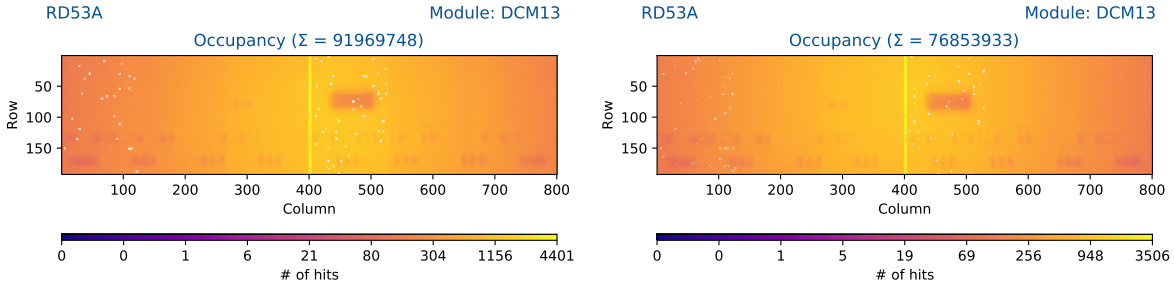
This section contains additional results of RD53A DCMs for the hybridization MS, that is presented in detail in Chapter 8.

D.3.1 VENDOR COMBINATION C1



(a) DCM8 source scan occupancy map before TC. (b) DCM8 source scan occupancy map after TC.

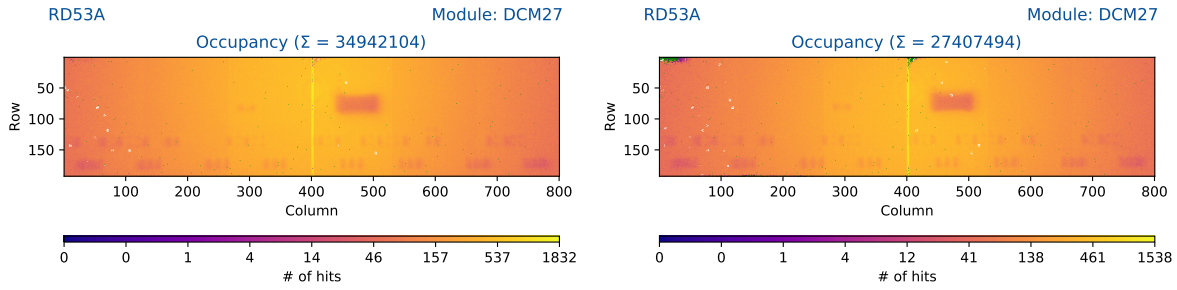
Figure D.15: Occupancy map of a Source Scan of DCM8 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.



(a) DCM13 source scan occupancy map before TC. (b) DCM13 source scan occupancy map after TC.

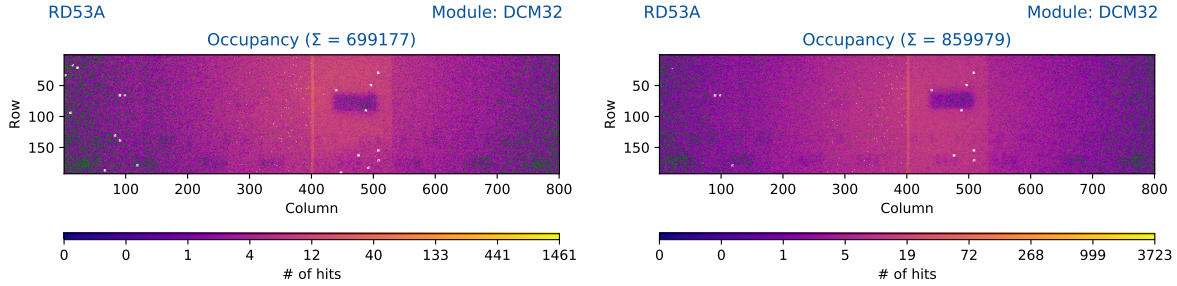
Figure D.16: Occupancy map of a Source Scan of DCM13 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.

D.3.2 VENDOR COMBINATION C2



(a) DCM27 source scan occupancy map before TC. (b) DCM27 source scan occupancy map after TC.

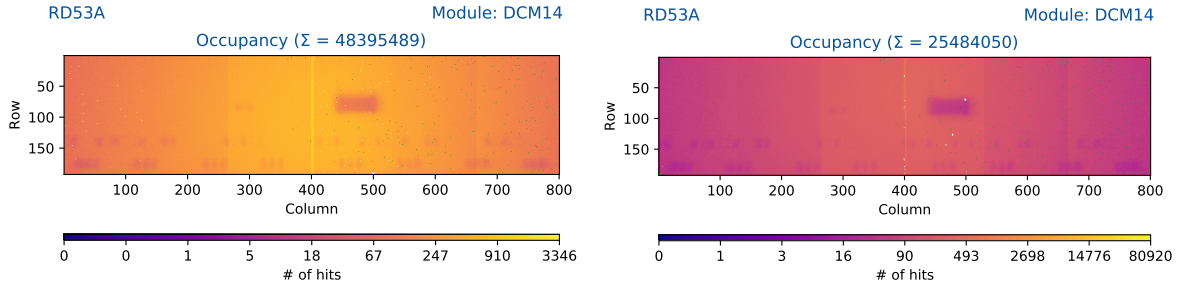
Figure D.17: Occupancy map of a Source Scan of DCM27 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.



(a) DCM32 source scan occupancy map before TC. (b) DCM32 source scan occupancy map after TC.

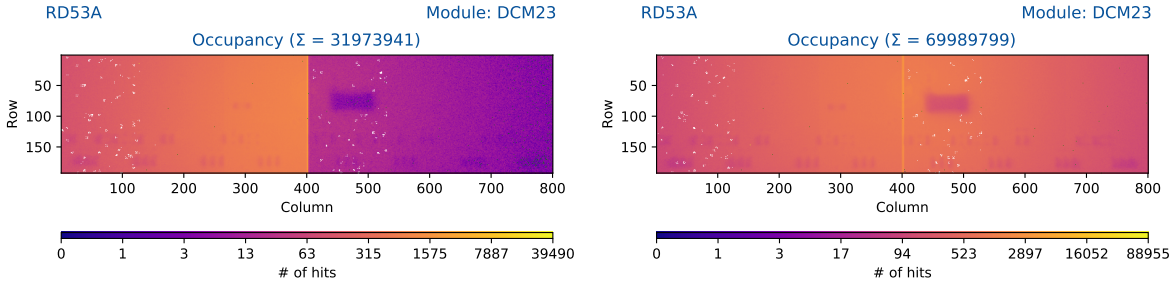
Figure D.18: Occupancy map of a Source Scan of DCM32 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.

D.3.3 VENDOR COMBINATION C3



(a) DCM14 source scan occupancy map before TC. (b) DCM14 source scan occupancy map after TC.

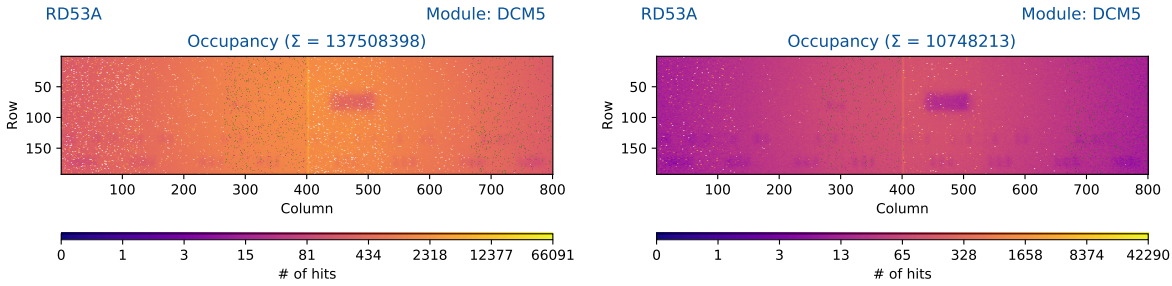
Figure D.19: Occupancy map of a Source Scan of DCM14 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.



(a) DCM23 source scan occupancy map before TC. (b) DCM23 source scan occupancy map after TC.

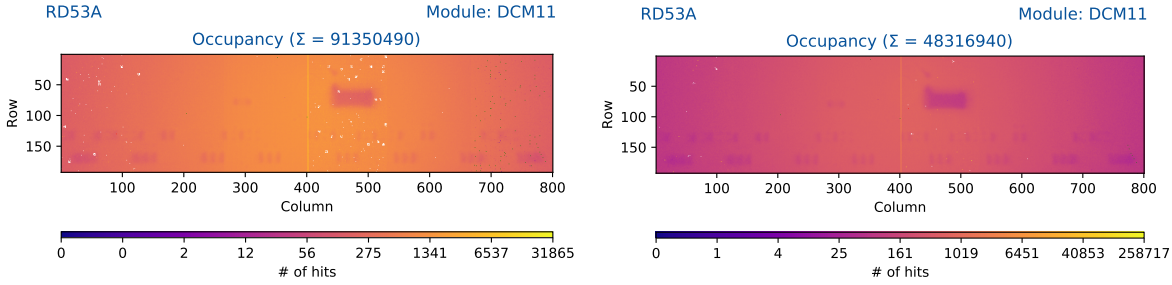
Figure D.20: Occupancy map of a Source Scan of DCM23 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.

D.3.4 VENDOR COMBINATION C6



(a) DCM5 source scan occupancy map before TC. (b) DCM5 source scan occupancy map after TC.

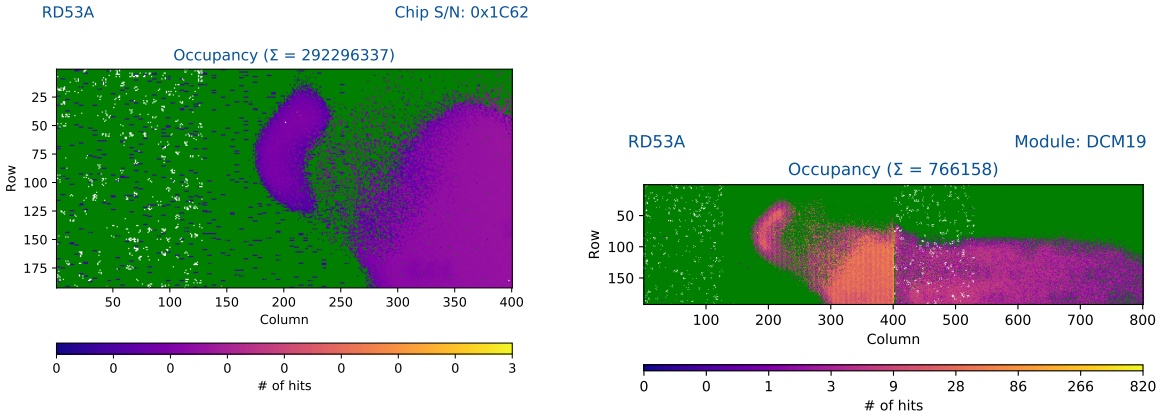
Figure D.21: Occupancy map of a Source Scan of DCM5 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.



(a) DCM11 source scan occupancy map before TC. (b) DCM11 source scan occupancy map after TC.

Figure D.22: Occupancy map of a Source Scan of DCM11 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.

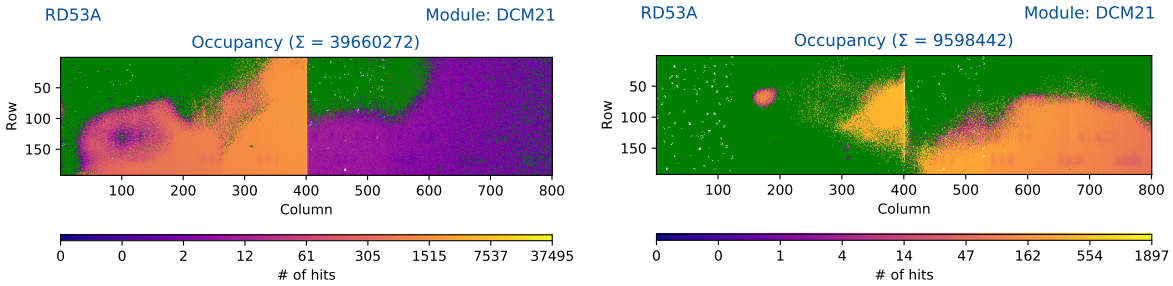
D.3.5 VENDOR COMBINATION C7



(a) DCM19 source scan occupancy map before TC. Only the left front-end chip is shown.

(b) DCM19 source scan occupancy map after TC.

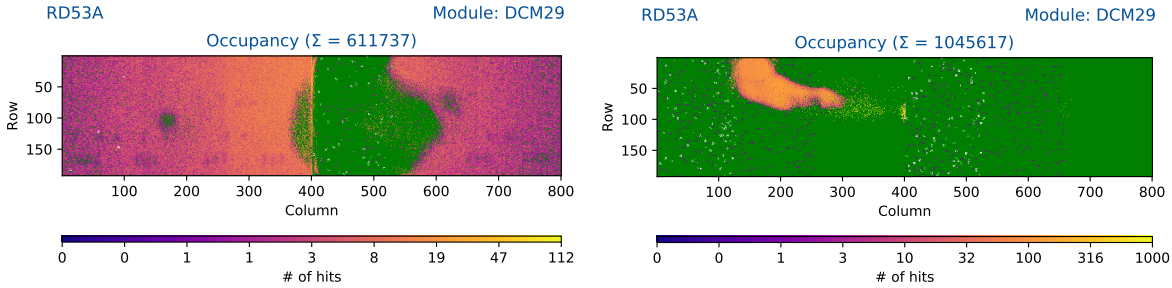
Figure D.23: Occupancy map of a Source Scan of DCM19 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.



(a) DCM21 source scan occupancy map before TC.

(b) DCM21 source scan occupancy map after TC.

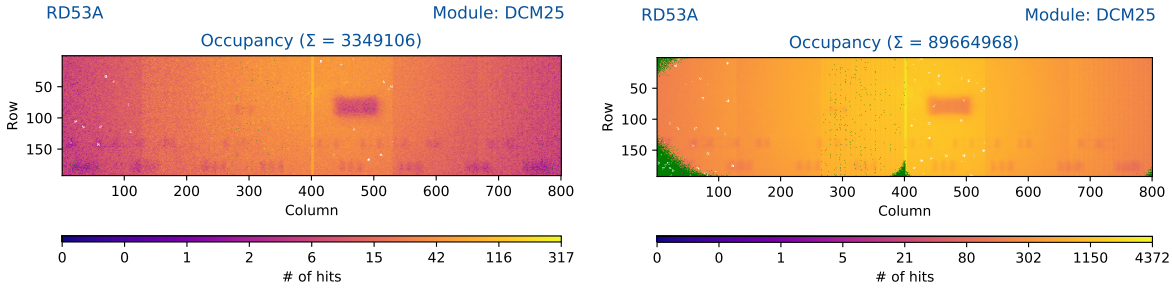
Figure D.24: Occupancy map of a Source Scan of DCM21 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.



(a) DCM29 source scan occupancy map before TC. (b) DCM29 source scan occupancy map after TC.

Figure D.25: Occupancy map of a Source Scan of DCM29 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.

D.3.6 VENDOR COMBINATION C9



(a) DCM25 source scan occupancy map before TC. (b) DCM25 source scan occupancy map after TC.

Figure D.26: Occupancy map of a Source Scan of DCM25 with a ^{241}Am X-ray source before and after TC. White pixels are disabled. Green pixels recorded exactly 0 hits.

BIBLIOGRAPHY

[1] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3**.08 (2008) S08001.

[2] P. Mouche, *Overall view of the LHC*, 2014-06-12, URL: https://cds.cern.ch/record/1708847.

[3] CERN, *HL-LHC Upgrade Schedule*, 2020, URL: https://project-hl-lhc-industry.web.cern.ch/content/project-schedule.

[4] P.A. Zyla et al. (Particle Data Group), *Review of Particle Physics*, Progress of Theoretical and Experimental Physics **2020**.8 (2020).

[5] CERN Yellow Reports: Monographs, *CERN Yellow Reports: Monographs, Vol. 10 (2020): High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*, tech. rep., CERN, 2020, URL: https://e-publishing.cern.ch/index.php/CYRM/issue/view/127.

[6] J. Pequenao, *Computer generated image of the whole ATLAS detector*, 2008-03-27, URL: https://cds.cern.ch/record/1095924.

[7] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 1*, tech. rep., 1997, URL: https://cds.cern.ch/record/331063.

[8] J. Pequenao, *Computer generated image of the ATLAS inner detector*, 2008-03-27, URL: https://cds.cern.ch/record/1095926.

[9] ATLAS Collaboration, *ATLAS Pixel Detector: Technical Design Report*, tech. rep. CERN-LHCC-98-13, 1998.

[10] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, tech. rep. CERN-LHCC-2010-013. ATLAS-TDR-19, 2010-09, URL: http://cds.cern.ch/record/1291633.

[11] ATLAS Collaboration, *TRT performance results from 13 TeV collision data (2015/2016)*, 2016, URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/TRT-2016-001/.

[12] ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report*, tech. rep., CERN, 1996, URL: https://cds.cern.ch/record/331061.

[13] ATLAS Collaboration, *ATLAS muon spectrometer: Technical Design Report*, tech. rep., CERN, 1997, URL: https://cds.cern.ch/record/331068.

[14] ATLAS Collaboration, *ATLAS 13 TeV event displays - 2017 Summer Conferences*, 2017-07-12, URL: https://cds.cern.ch/record/2274095.

[15] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report, 1*, tech. rep., CERN, 1999, URL: http://cds.cern.ch/record/391176.

[16] ATLAS Outreach, *ATLAS Fact Sheet : To raise awareness of the ATLAS detector and collaboration on the LHC*, 2010, URL: https://cds.cern.ch/record/1457044.

[17] C. Gemme, *The ATLAS upgrade program*, tech. rep. ATL-UPGRADE-PROC-2014-004, CERN, 2014-09, URL: http://cds.cern.ch/record/1756237.

[18] Atlas Collaboration, *ITk Pixel Layout Updates*, 2020, URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/ITK-2020-002/.

[19] K. Hoepfner, *CMS upgrade and future plans*, EPJ Web of Conferences **95** (2015) 03016.

[20] N. N. Greenwood and A. Earnshaw, *Chemistry of the Elements*, Butterworth-Heinemann, 2012, ISBN: 9780080501093.

[21] H. Kolanoski and N. Wermes, *Teilchendetektoren: Grundlagen und Anwendungen*, 1st ed., Springer-Verlag GmbH, 2016-03-04, ISBN: 978-3-662-45349-0, URL: https://www.ebook.de/de/product/25501763/hermann_kolanoski_norbert_wermes_teilchendetektoren.html.

[22] L. Rossi, P. Fischer, T. Rohe, and N. Wermes, *Pixel Detectors: From Fundamentals to Applications*, Particle Acceleration and Detection, Springer Berlin Heidelberg, 2006, ISBN: 9783540283331.

[23] ATLAS collaboration, *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*, tech. rep. CERN-LHCC-2017-021. ATLAS-TDR-030, CERN, 2017, URL: https://cds.cern.ch/record/2285585.

[24] M. S. Lazo, D. M. Woodall, and P. J. McDaniel, *Silicon and silicon dioxide neutron damage functions*, Proc. Fast Burt React. Workshop **1** (1986).

[25] N. Wermes, *Pixel detectors ... where do we stand?*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **924** (2019) 44, ISSN: 0168-9002, URL: http://www.sciencedirect.com/science/article/pii/S016890021830826X.

[26] D.-L. Pohl, *3D-Silicon and Passive CMOS Sensors for Pixel Detectors in High Radiation Environments*, PhD thesis: Universität Bonn, 2019.

[27] J. Chistiansen and M. Garcia-Sciveres, *RD Collaboration Proposal: Development of pixel readout integrated circuits for extreme rate and radiation*, tech. rep. CERN-LHCC-2013-008. LHCC-P-006, CERN, 2013, URL: https://cds.cern.ch/record/1553467.

[28] M. Garcia-sciveres et al., *Results of FE65-P2 Pixel Readout Test Chip for High Luminosity LHC Upgrades*, Proceedings of the 38th International Conference on High Energy Physics, PoS (ICHEP2016) (2017).

[29] L. Pacher et al., *Results from CHIPIX-FE0, a small-scale prototype of a new generation pixel readout ASIC in 65 nm CMOS for HL-LHC*, Proceedings of Topical Workshop on Electronics for Particle Physics - PoS (TWEPP2017) (2018) 024.

[30] J. Große-Knetter, "Internal communication: Analog FE Decision (Access restricted to ATLAS collaboration)", 2019-03-08, URL: https://indico.cern.ch/event/804139/.

[31] S. Mersi et al., "Report on the IT ASIC Front-End evaluation", *CMS Tracker Week*, 2019-06-05, URL: https://indico.cern.ch/event/823864/contributions/3445661/attachments/1857001/3050512/mersi_20190605_AnalogueFrontEndEvaluationSummary.pdf.

[32] F. Arteche Gonzalez et al., *Extension of RD53*, tech. rep. CERN-LHCC-2018-028. LHCC-SR-008, CERN, 2018, URL: http://cds.cern.ch/record/2637453.

[33] M. Garcia-Sciveres, *The RD53A Integrated Circuit*, tech. rep. CERN-RD53-PUB-17-001, CERN, 2017, URL: https://cds.cern.ch/record/2287593.

[34] S. Marconi, E. Conti, P. Placidi, J. Christiansen, and T. Hemperek, *The RD53 Collaboration's SystemVerilog-UVM Simulation Framework and its General Applicability to Design of Advanced Pixel Readout Chips*, Journal of Instrumentation **9** (2014-10-07).

[35] L. Gaioni et al., *Design of analog front-ends for the RD53 demonstrator chip*, Proceedings of The 25th International workshop on vertex detectors PoS (Vertex 2016) (2017).

[36] F. Krummenacher, *Pixel detectors with local intelligence: an IC designer point of view*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **305**.3 (1991) 527.

[37] L. Pacher, E. Monteil, A. Rivetti, N. Demaria, and M. Da Rocha Rolo, *A low-power low-noise synchronous pixel front-end chain in 65 nm CMOS technology with local fast ToT encoding and autozeroing for extreme rate and radiation at HL-LHC*, 2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC) (2015) 1.

[38] L. Gaioni et al., *65 nm CMOS analog front-end for pixel detectors at the HL-LHC*, Journal of Instrumentation **11**.02 (2016).

[39] P. Rymaszewski, *Design and characterization of pixel IC electronics and sensors for a new pixel detector generations (In preparation)*, PhD thesis: Universität Bonn, 2021.

[40] M. Karagounis et al., *An integrated Shunt-LDO regulator for serial powered systems*, Proceedings of the 35th European Solid-State Circuits Conference - PoS ESSCIRC2009 (2009).

[41] S. Orfanelli et al., *Serial Powering Optimization for CMS and ATLAS Pixel Detectors within RD53 Collaboration for HL-LHC: System Level Simulations and Testing*, Proceedings of Topical Workshop on Electronics for Particle Physics - PoS (TWEPP2017) (2018).

[42] A. Dimitrievska and A. Stiller, *RD53A: A large-scale prototype chip for the phase-II upgrade in the serially powered HL-LHC pixel detectors*,
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **958** (2020) 162091.

[43] F. Hinterkeuser, *Development and Evaluation of Serial Powering Prototypes for the ATLAS ITk Pixel Detector (In preparation)*, PhD thesis: Universität Bonn, 2021.

[44] M. Standke, *Characterization of the Joined ATLAS and CMS RD53A Pixel Chip*,
MA thesis: Universität Bonn, 2019.

[45] M. Garcia-Sciveres et al., *The FE-I4 pixel readout integrated circuit*,
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **636**.1, Supplement (2011) S155,
URL: http://dx.doi.org/10.1016/j.nima.2010.04.101.

[46] G. Giakoustidis, *Charge Calibration and Crosstalk Measurements with RD53A Assemblies*,
MA thesis: Universität Bonn, 2019, URL: https://web.physik.uni-bonn.de/group/view.php?&group=2&lang=en&c=t&id=649.

[47] T. Wang et al.,
A high speed transmitter circuit for the ATLAS/CMS HL-LHC pixel readout chip, Proceedings of Topical Workshop on Electronics for Particle Physics - PoS (TWEPP2018) **17** (2019) 21.

[48] Xilinx Inc., *Aurora 64B/66B Protocol Specication*, version SP011 (v1.3), 2014-10-01,
URL: https://www.xilinx.com/support/documentation/ip_documentation/aurora_64b66b_protocol_spec_sp011.pdf.

[49] Video Electronics Standards Organization (VESA), *VESA DisplayPort Standard*,
version 1, Revision 1a, 2008-01-11,
URL: http://file.yizimg.com/383992/2014090921252964.pdf.

[50] M. Daas et al., *BDAQ53, a versatile pixel detector readout and test system for the ATLAS and CMS HL-LHC upgrades*, Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors, and Associated Equipment **986** (2021) 164721,
ISSN: 0168-9002.

[51] M. Backhaus et al.,
Development of a versatile and modular test system for ATLAS hybrid pixel detectors,
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **650**.1 (2011-11) 37.

[52] J. J. D.-L. Pohl, *pyBAR - Bonn ATLAS Readout in Python*, 2018,
URL: https://github.com/SiLab-Bonn/pyBAR.

[53] G. Van Rossum, *The Python language reference*,
Python Software Foundation, SoHo Books, 2010, ISBN: 1441412697.

[54] *IEEE Standard for Verilog Hardware Description Language*, 2006.

[55] Xilinx Inc., *7 Series FPGAs Datasheet*, 2014-10-01, URL: https://www.xilinx.com/support/documentation/data_sheets/ds180_7Series_Overview.pdf.

[56] Enclustra, *Enclustra Mercury+ KX2*, 2020, URL: https://www.enclustra.com/en/products/fpga-modules/mercury-kx2/.

[57] SiLab, Physikalisches Institut, Universität Bonn, *Basil Documentation*, URL: http://basil.readthedocs.io.

[58] M. Frohne, *Developments for a multi-module test system for ATLAS ITk pixel detector modules*, MA thesis: Universität Bonn, 2020, URL: https://web.physik.uni-bonn.de/group/view.php?&group=2&lang=en&c=t&id=647.

[59] J. Janssen, *Test beam results of ATLAS DBM pCVD diamond detectors using a novel threshold tuning method*, Journal of Instrumentation **12**.03 (2017) C03072.

[60] H. Jansen et al., *Performance of the EUDET-type beam telescopes*, EPJ Techniques and Instrumentation **3**.1 (2016).

[61] D. Cussans, *Description of the JRA1 Trigger Logic Unit (TLU), v0.2c*, tech. rep., 2009, URL: https://www.eudet.org/e26/e28/e42441/e57298/EUDET-MEMO-2009-04.pdf (visited on 2020-09-26).

[62] M. Backhaus, *High bandwidth pixel detector modules for the ATLAS Insertable B-Layer*, PhD thesis: Universität Bonn, 2014-01.

[63] A. Krieger, D. Gnani, and A. Mekkaoui, "Internal communication: RD53A Differential Analog Front-End Design and Simulation (Access restricted to RD53 collaboration)", 2018-10-04, URL: https://indico.cern.ch/event/769894/.

[64] RD53 Collaboration, *Internal communication: RD53 Front-End Review Conclusion Letter (Access restricted to RD53, ATLAS and CMS collaborations)*, tech. rep., CERN, 2019-01-21.

[65] M. Vogt, *Development of the Readout System BDAQ53 and Verification and Characterization of the Hybrid Pixel Readout Chip RD53A (In preparation)*, PhD thesis: Universität Bonn, 2021.

[66] FormFactor, *Cascade PA300 300mm Semi-automatic Probe System Datasheet*, 2019-04-01, URL: https://www.formfactor.com/download/pa300-data-sheet/ (visited on 2020-10-17).

[67] J.-C. Beyer, *Optimisation of Pixel Modules for the ATLAS Inner Tracker at the High-Luminosity LHC*, PhD thesis: LMU München, 2019, URL: http://hdl.handle.net/21.11116/0000-0005-D6DD-1.

[68] P. Ahlburg et al., *EUDAQ - a data acquisition software framework for common beam telescopes*, Journal of Instrumentation **15**.01 (2020) P01038.

[69] H. Jansen et al., *Performance of the EUDET-type beam telescopes*, EPJ Techniques and Instrumentation **3**.1 (2016).

[70] F. Hügging and P. Riedler, *Internal communication: Technical Specification: Hybridization of Hybrid Pixel Detector Modules for the ATLAS ITk (Access restricted to ATLAS collaboration)*, (2019).

[71] J. John, L. Zimmermann, P. D. Moor, and C. V. Hoof,
High-density hybrid interconnect methodologies,
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **531**.1-2 (2004) 202.

[72] T. Fritzsch et al., *Flip chip assembly of thinned chips for hybrid pixel detector applications*,
Journal of Instrumentation **9**.05 (2014) C05039.

[73] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*,
Journal of Instrumentation **13**.05 (2018) T05008.

[74] G. Aad et al., *ATLAS pixel detector electronics and sensors*,
Journal of Instrumentation **3**.07 (2008) P07007.

[75] P. S. Teo, Y.-W. Huang, C. H. Tung, M. Marks, and T. B. Lim,
Investigation of under bump metallization systems for flip-chip assemblies, 2000 Proceedings. 50th Electronic Components and Technology Conference (Cat. No.00CH37070) (2000).

[76] M. Rothermund, *Private Communication: Comparison of Micron and IZM UBM*,
IZM Berlin (2020-12-14).

[77] T. Fritzsch et al., *Investigation of low temperature bonding process using indium bumps*,
Journal of Instrumentation **13**.11 (2018) C11007.

[78] Mitutoyo Corporation, *Measuring Microscopes MF/MF-U Series*, (2013),
URL: https://www.mitutoyo.com/wp-content/uploads/2013/07/E14003_MF_MF-U.pdf.

[79] K. Mauer, *Characterization of Hybrid Pixel-DetectorModules for the ATLAS Phase-II Upgrade*,
Universität Bonn, 2020-10.

[80] Weiss Umwelttechnik GmbH, *Tischgeräte für Temperaturprüfungen LabEvent*, 2018,
URL: https://www.weiss-technik.com/fileadmin/Redakteur/Mediathek/Broschueren/WeissTechnik/Umweltsimulation/Weiss-Technik-LabEvent-Tisch-DE.pdf.

GLOSSARY

Notation	Description
0b	Binary numbers are denoted with a leading 0b, e.g. $8_{10} = 1000_2$ = 0b1000.
0x	Hexadecimal numbers are denoted with a leading 0x, e.g. $255_{10} = FF_{16}$ = 0xFF.
ACB	Analog Chip Bottom. Refers to the analog part of the global bottom logic block of a readout chip.
ADC	Analog-to-Digital Converter / Conversion. A circuit that generates discrete digital values from a continuous analog input signal.
AFE	Analog Front-End. First stage of analog circuitry in a particle detector. Amplifies, shapes and digitizes the signal of the respective charge collecting element.
ASIC	Application-Specific Integrated Circuit.
AZ	Auto-Zero. Feature of the SYNC of `RD53A` to automatically minimize threshold dispersion without manually tuning.
Band Gap	A voltage reference circuit commonly used in IC design. Based on the physical band gap of silicon, this circuit produces a constant voltage that is relatively independent of external influences like circuit load, power supply variations or temperature.
BCID	Bunch Crossing ID. Unit of time used in LHC experiments. Corresponds to 25 ns.
Binary Search	Binary search is a search algorithm commonly used in programming to find a specific element in an ordered set. It works by testing the mid element of a subset against the desired value, eliminating one half of the tested subset. The other half is used as subset for the next iteration. This leads to a logarithmic run time in the worst case.
CBA	Central Buffer Architecture.
CDR	Clock/Data Recovery is a type of circuit that is typically used together with a PLL to generate a phase-related clock signal from a serialized data stream.

Notation	Description
CERN	European Organization for Nuclear Research, from French: **C**onseil **E**uropéen pour la **R**echerche **N**ucléaire. With its 23 member states, CERN is the largest international research organization for fundamental particle physics in the world. It was established in 1954 by twelve original member states, including Germany, which today is the largest total contributor to CERN's budget.
CI	Continuous Integration. A practice in programming with the goal to avoid fragmentation and divergence of code during development. Used in conjunction with Test-Driven Development (TDD) to automatically test small parts of the code base to detect potential bugs, based on automatic unit tests.
CML	Current-Mode Logic. A digital logic family typically employed in highspeed data transmission.
CMOS	Complementary Metal-Oxide-Semiconductor. Semiconductor fabrication processes employing both n- and p-type logic.
CRC	Cyclic Redundancy Check. A hash function used to detect errors in data transmission.
CSA	A Charge Sensitive Amplifier is a current-integrating amplifier circuit with capacitive feedback, typically used as preamplifier for silicon particle detectors.
CSC	A Cathode Strip Chamber is a type of gaseous particle tracking detector. It uses a multiwire proportional chamber where the readout cathode is subdivided into strips to provide one-dimensional spacial resolution.
CTE	The coefficient of thermal expansion is a measure of how much the physical size of an object changes with temperature. Composites of layered materials of different CTE tend to bend with temperature due to the different linear expansion and contraction of the different materials.
DAC	Digital-to-Analog Converter / Conversion. A circuit that outputs an analog voltage depending on a digital register setting.
DAQ	Data Acquisition. A readout system used to interface with a readout ASIC, process and visualize its data.
DBA	Distributed Buffer Architecture.
DCB	Digital Chip Bottom. Refers to the digital part of the global bottom logic block of a readout chip.
DCM	Double Chip Module. A silicon pixel detector module or assembly consisting of two readout ASICs bump-bonded to a single passive sensor.
DCS	Detector Control System. The entirety of systems and power supplies necessary to operate a detector in the final experiment.
DDR	Double Data Rate. A data transfer strategy where both the rising and falling edge of the clock are used to execute a given function.
DIFF	Differential Analog Front-End of `RD53A`.

Notation	Description
DMAPS	Depleted Monolithic Active Pixel Sensor.
DNL	Differential Non-Linearity.
DP	DisplayPort.
DUT	Device Under Test. In the context of DAQ systems, the DUT is the ASIC or module that is being investigated.
ENC	Equivalent Noise Charge. Typical measure for the noise of a charge-integrating amplifier system. Defined as the charge at the input that would produce a signal at the output equivalent to the measured noise amplitude.
FE	Front-End. See AFE.
FIFO	First In First Out. A type of data buffer where elements are processed in the order they arrived in.
flex	Flexible PCB. A thin and flexible type of PCB typically during pixel detector module assembly to provide an interface between the readout ASICs and the detector services. Can have stiffeners to provide rigidity in some areas.
flip-chip	Flip-chip is the final step of a hybridization process of aligning and connecting the two prepared chips by melting the previously applied solder bumps.
FPGA	Field-Programmable Gate Array. An IC whose logic gates can be programmed by the user using an Hardware Description Language (HDL).
GDAC	Global Threshold DAC. Global DAC defining the mean charge detection threshold of the pixel matrix.
HDL	Hardware Description Language. A programming language used to create configurations for FPGAs.
HitOr	A single output signal of a readout ASIC comprising the logical OR of the analog hit signal of all connected pixels.
HL-LHC	High Luminosity LHC.
HV	In the context of silicon pixel detector modules, high voltage describes the reverse bias voltage applied to the sensor in order to deplete its volume.
I^2C	A serial communication bus used for low-speed communication between peripheral ICs.
I/O	Input/Output.
IBL	The ATLAS Insertable B-layer is the innermost pixel detector layer of the current ATLAS Inner Detector and was added as an upgrade in 2013.
IC	Integrated Circuit.

Notation	Description
ID	The Inner Detector is the tracking detector of the current ATLAS detector. It consists of three subdetectors: A four-layer hybrid silicon Pixel Detector, four double layers of silicon strip detectors (SCT) and a straw-tube TRT.
INL	Integrated Non-Linearity.
ITk	Inner Tracker. The tracking detector of the ATLAS Phase-II upgrade.
LCC	Leakage Current Compensation.
LDO	Low-Dropout Regulator.
LHC	Large Hadron Collider.
LIN	Linear Analog Front-End.
LSB	Least Significant Bit. The lowest value bit in a binary number. Commonly used as unit for binary register values.
MAPS	Monolithic Active Pixel Sensor.
MDT	A Monitored Drift Tube is a type of gaseous particle tracking detector. It consists of tubes filled with a pressurized active gas mixture that is ionized by traversing particles. An electrical field is applied between the tube wall and a wire in the center and a signal is created by the ions drifting in this electric field.
MGT	Multi-Gigabit Transceiver. A common serializer / deserializer circuit with data rates above 1 Gbit/s.
MIP	A Minimum Ionizing Particle is a charged particle with a relativistic velocity $\beta\gamma \approx 3$. Its mean energy loss per path length in an absorber is minimal as shown by the Bethe-Bloch Formula (refer to Section 2.1.1).
MS	A Market Survey is a special market research project aimed at identifying and assessing possible vendors or service providers for a specific product or service.
MUX	A multiplexer is an active electronic component that allows to select between several analog or digital signals to be forwarded to a single output line.
NIEL	Non-Ionizing Energy Loss. Energy loss of particles in matter by effects other than ionization of the absorber material, mostly collisions with nucleons. The term is mostly used in the context of radiation damage of semiconductor materials.
NTC	Negative Temperature Coefficient thermistor. A passive component used for temperature measurements.
PCB	Printed Circuit Board.
PLL	Phase-Locked Loop. A control loop generating an output signal with a phase related to a given input signal.

Notation	Description
POR	Power-on Reset. Reset function of an ASIC that is executed automatically on start up.
QA	Quality Assurance. Measurements to qualify a general design. In the scope of this work, QA refers to the characterization measurements of different components, assuring their general capability to fulfill the requirements.
QC	Quality Control. Measurements performed on every manufactured entity or a random sample of all entities to control the consistency of production quality. In the scope of this work, QC refers to measurement routines performed on every readout chip during wafer probing as well as test procedures for fully assembled modules in preparation for production.
QCM	Quad-Chip Module. A silicon pixel detector module or assembly consisting of four readout ASICs bump-bonded to a single quadratic passive sensor.
ROI	A Region of Interest is a part of a particle detector that potentially contains interesting information. Defining and only reading out a ROI instead of the whole detector can be a means of reducing data rates.
RPC	A Resistive Plate Chamber is a fast gaseous detector with digital readout. In a trong electric field between two parallel resistive plates, a primary ionization leads to an avalanche that produces amplified pulses.
SAR	Successive Approximation Register. Refers to a type of ADC that uses binary search to find the digital output setting best matching the analog input.
SCC	Single-Chip Card. Test PCB that ASICs are wire-bonded to that provide connections for laboratory equipment.
SCM	Single-Chip Module. A silicon pixel detector module or assembly consisting of one readout ASIC bump-bonded to a passive sensor of the same size.
SCT	Semi-Conductor Tracker. The middle subdetector of the current ATLAS Inner Detector. Based on silicon microstrips.
SEE	Single-Event Effect. General class of radiation effects in electronic circuits. Comprises Single-Event Upsets (SEUs), Single-Event Latch-Up (SELs) and other effects.
sLDO	Shunt Low-Dropout Regulator.
SM	The Standard Model of Particle Physics is the current baseline model for describing the fundamental particles of the universe and their interactions using three of the four known fundamental forces.
SMU	A Source Measure Unit is a type of test equipment that is capable of both sourcing and measuring voltages and currents at the same time. In detector testing, SMUsare typically used as high-voltage power supply to measure sensor IV curves.

Notation	Description
SPS	Super Proton Synchrotron. The second-largest accelerator at CERN with a circumference of 6.9 km and a center-of-mass energy of currently up to 450 GeV. It is used as injector for the LHC (and for LEP before that) and supplies several other experiments and test beam areas with 400 GeV proton beams.
SYNC	Synchronous Analog Front-End of RD53A.
TC	Thermal cycling.
TCP	Transmission Control Protocol. Common data transport protocol with error checking.
TDAC	Local Threshold DAC. In-pixel DAC used as a local offset to the reference voltage generated by the GDAC. Commonly used to fine-tune the pixel threshold to minimize dispersion over the matrix.
TDC	Time-to-Digital Converter / Conversion. A circuit that digitizes a given time interval.
TGC	A Thin Gap Chamber is a gaseous particle tracking detector similar to a multiwire proportional chamber. It is operated in saturation for an optimized timing resolution.
TID	Total Ionizing Dose. Measure of radiation damage inflicted by ionizing radiation.
TLU	Trigger Logic Unit. Circuitry designed to generate trigger pulses from different inputs, e.g. scintillators or an EUDET beam telescope [60]. [61]
ToT	Time-over-Threshold. Method of charge measurement in pixel detectors.
Triplet	A special silicon pixel detector module or assembly consisting of three SCMs fixed to a common support structure.
TRT	Transition Radiation Tracker. Outermost subdetector of the current ATLAS Inner Detector.
UBM	Under-bump metallization describes the additional metal layer added to an IC's interconnection areas to help with bump-bonding.
UDP	User Datagram Protocol. Common data transport protocol with low overhead and latency but without error checking.
yield	Relative amount of working integrated circuits after manufacturing.

List of Figures

List of Tables