Version 3.2, Dec. 13, 2015

RD53A Integrated Circuit Specifications

ABSTRACT: Specifications for the RD53 collaboration's first engineering wafer run of an integrated circuit (IC) for hybrid pixel detector readout, called RD53A. RD53A is intended to demonstrate in a large format IC the suitability of the technology (including radiation tolerance), the stable low threshold operation, and the high hit and trigger rate capabilities, required for HL-LHC upgrades of ATLAS and CMS. The wafer scale production will permit the experiments to prototype bump bonding assembly with realistic sensors in this new technology and to measure the performance of hybrid assemblies. RD53A is not intended to be a final production IC for use in an experiment, and will contain design variations for testing purposes, making the pixel matrix non-uniform.

Contents

1. Preface to Version 3.2

This version incorporates comments on Version 2.2, collected from the ATLAS and CMS collaborations between September and October, 2015. These were discussed at the October 14-16, 2015 RD53 general meeting, at which the scope of this document was finalized. Implementation of the RD53A design is now in progress. Two significant changes relative to Version 2.2 are: (1) The option of smaller size chip compatible with Multi Layer Mask (MLM) production has been dropped- see Sec. 5. (2) Technical detail relevant to implementation has been removed and will be collected in an Implementation "living document". The Implementation Document will grow along with the design and will eventually become a user guide for RD53A.

2. Introduction

The RD53 collaboration was established in 2013 to design a hybrid pixel readout chip for the high rate and radiation expected in the ATLAS and CMS phase 2 upgrades [1]. The goal was to deliver in a 3-year time frame the elements required for ATLAS and CMS to produce readout chips. The RD53A integrated circuit specified here embodies these deliverables. RD53A is intended to demonstrate in a large format IC the suitability of the chosen 65nm CMOS technology (including radiation tolerance), the stable low threshold operation, and the high hit and trigger rate capabilities, required for HL-LHC upgrades of ATLAS and CMS. RD53A is not intended to be a final production IC for use by the experiments, and will contain design variations for testing purposes, making the pixel matrix non-uniform. This document contains the specifications for the RD53A IC, not to be taken as the final specifications for either ATLAS or CMS production. RD53A must demonstrate that the ATLAS and CMS critical requirements can be met in a large format IC. The specifications to meet critical performance requirements are given in Sec. 3. Further specifications not central to demonstrating these critical items but nevertheless desirable are given in Sec. 4. In order to move forward with a design and achieve low technical and schedule risk, certain choices of parameters not central to demonstrating performance may deviate from the final wishes of the experiments, or may precede those wishes when the experiments had not yet reached a decision. In addition to performance, there will be functions or features that are important to demonstrate, have already been agreed, may influence the performance, or are necessary to make efficient use of the chip. These are specified in Sec. 8. Further features needed or interesting for the production chip, but not mandatory for RD53A are given in Sec. 9.

3. Critical Performance Specifications

While the experiments plan to use one readout chip throughout the entire pixel detector, the critical performance requirements are defined by the innermost layer, which will be at a radius of 3-4 cm from the HL-LHC interactions. Use in outer layers introduces additional requirements, such as large format for low cost assembly, aggregation of multi-chip module data, possibility to operate with reduced power and larger pixels, etc.; but the performance requirements are generally less demanding at higher radius. In simplest terms, the hybrid modules (readout chip plus sensor) must record the hit pixels and crossing times of 99% of incident charged particles, must hold this information until a trigger decision is received, and must read out the triggered information without loss, negligible fakes, and in a short enough time as needed for higher level triggering. These requirements give rise to the quantitative specifications in Table 1. While these may not exactly reflect the absolute final requirements for the ATLAS and CMS inner layers, we consider them representative enough for the RD53A prototype. It should be noted that it is not possible to cleanly separate chip requirements from sensor requirements. The chip has to work with the signal provided by the sensor and pull this signal away from the sensor capacitance. Thus the chip design must assume something about the sensor, and at the same time chip design considerations place constraints on sensor features. This is an iterative process. following discussions with ATLAS and CMS sensor developers, for RD53A we have assumed sensors have less than $100\,\mathrm{fF}$ per pixel and deliver a single pixel signal greater than $600\,\mathrm{e^-}$ in at least one pixel for 99% of incident particles,

Specification	Value	Comment or Test Conditions
Input polarity	Negative	
Interior pixel capacitance	<100 fF	this applies to most pixels
Edge pixel capacitance	<200 fF	see subsection on edge pixels
Interior pixel leakage current	<10 nA	
Edge* pixel leakage current	<20 nA	*see Sec. 5.2 for details
Min. stable threshold setting	600 e⁻	With 50 fF load, 4μA/pixel analog. For free running discriminated pixel. See Sec. 3.1
Min. charge above threshold resulting in <25 ns time walk	600 e⁻	With 50 fF load, 4μA/pixel analog. For free running discriminated pixel. See Sec. 3.1
Min. in-time threshold with free-running front end	1200 e⁻	With 50 fF load, 4μA/pixel analog. Simply the sum of the two above lines
Min. in-time threshold if using synchronous reset	750 e⁻	With 50 fF load, 4μA/pixel analog. See Sec. 3.1
Hit loss from in-pixel pileup	≤1%	at 75 kHz avg. hit rate. See Sec. 3.2
Recovery from saturation	<1 μs	See Sec. 3.2 for discussion
Trigger rate	1 MHz	
Trigger latency	12.5 μs	
Noise occupancy per pixel	$< 10^{-6}$	50 fF load; in a 25 ns interval
Single pixel noise (ENC)	design-dependent	See Sec. 3.1
Radiation dose	500 Mrad	delivered at -15 C. Room T annealing only
Temperature range	-40C to +40C	
Current consumption, analog	4 μA/pixel	periphery consumption not included
Current consumption, digital	<4 μA/pixel	periphery consumption not included
Current consumption, Total	<500 mA/cm^2	Note this is 1 W/cm^2 at 2 V input
SEU's affecting full chip	<0.05/hr/chip	in 1.5 GHz/cm^2 particle flux
SEU's affecting single pixel	<100/hr/chip	in 1.5 GHz/cm^2 particle flux
Charge scale shift	<2% / Mrad	change in mean with radiation
Threshold scale shift	<15 e⁻ / Mrad	change in mean with radiation
Threshold dispersion	<60 e⁻ / Mrad	added in quadrature
Charge meas. dispersion	<0.1 MIP/Mrad	added in quadrature

Table 1: Specifications related to critical performance requirements. See text for details. Note these specifications are to be met after 500 Mrad dose.

after radiation damage (worst case incidence angle). We are also assuming that the leakage current will remain at or below 10 nA per pixel at operating temperature after radiation damage. These characteristics seem achievable with both planar and 3D sensors of thickness in the range 100 to 150 μm and with pixel area of 2500 μm^2.

Not yet mentioned above is that the chip is required to perform throughout the life of the experiment and at the operating conditions. The $3\,\text{ab}^{-1}$ lifetime dose at the inner layer is estimated at 1 Grad, including a safety factor covering estimate uncertainty. However, both ATLAS and CMS are requiring their detector designs to be compatible with replacement of the inner layer (or layers), as the readout chip radiation tolerance is not the only consideration. Even without considering degradation of the readout chip, sensor degradation or component failures may require a replacement. Therefore, a chip radiation tolerance of 500 Mrad provides a viable technical solution. As our understanding of radiation effects beyond 500 Mrad is still evolving and will continue to be studied over the coming year, we place a specification of 500 Mrad on the RD53A design. This means that the chip must be simulated to meet requirements with corner models after such dose. The temperature and annealing conditions are also important for radiation damage. For RD53A we take advantage of the knowledge that the detector must operate cold, and should never see temperatures above 40C.

3.1 Threshold, Noise, and Noise Occupancy

Small pixels, thin sensors, and high radiation damage will result in significantly smaller signals than available today. However, signal charge cannot be considered in isolation and must be taken together with load capacitance. Charge distributions in small pixels after irradiation are being studied by groups working on sensor development and so final numbers are not available. RD53A will in fact enable such studies to be accurately carried out. Furthermore, we expect that the same readout chip may be used with a variety of sensors. But for efficient chip design we must specify something definite. A $600\,\text{e}^-$ threshold should be efficient for signals from $50\,\mu\text{m}$ MIP path length in silicon even with 50% charge loss after radiation damage (this implies a Landau peak of approximately $2000\,\text{e}^-$). By $600\,\text{e}^-$ we mean that a $600\,\text{e}^-$ signal will have a 50% probability of firing the discriminator, without any timing constraints, in a free-running discriminator design as used in present pixel detectors. At the HL-LHC with 25 ns bunch crossing period, time-walk is important, and we have therefore specified a $600\,\text{e}^-$ *overdrive*, which means that (for the $600\,\text{e}^-$ minimum threshold setting) a $1200\,\text{e}^-$ signal will have a 50% probability of firing the discriminator with a time delay within 25 ns of very large signals. This does not mean that signals smaller than this $1200\,\text{e}^-$ *in-time threshold* will be lost, because digital processing will be used for time-walk compensation, but such compensation may not be 100% efficient, so having an overdrive that is not very large is still important.

Note that we have specified the $600\,\text{e}^-$ threshold (and $1200\,\text{e}^-$ in-time threshold) for 50 fF capacitance per pixel, while the maximum pixel capacitance is specified as 100 fF. This reflects the need to have specifications that cover operation with a variety of sensor options. What this means in practice is that a sensor with capacitance close to the 100 fF limit must also provide more signal charge than one with 50 fF/pixel, so that one can operate with a slightly higher threshold to achieve the same required hit efficiency of 99%. However, designers need a specific point to benchmark performance using simulation results. The alternative of specifying the threshold as a function of capacitance would be too restrictive and cumbersome for benchmarking.

We have also specified a minimum in-time threshold of $750\,\text{e}^-$ for 50 fF capacitance in case of a front end design that is synchronously reset with each bunch crossing, and thus signals that do not fire the discriminator within 25 ns are permanently lost. In this case an in-time threshold of

$1200\,e^-$ would be less efficient than the free-running front end case, where digital processing can recover hits that fire the $600\,e^-$ threshold, but are delayed due to time-walk. However, demanding a $600\,e^-$ in-time threshold for a synchronous design seemed too aggressive, and $750\,e^-$ is proposed as a compromise.

The minimum stable threshold is a critical property of a pixel readout chip, which cannot be reliably simulated or demonstrated with a small test chip. It is limited by non-linear effects that cause positive feedback in noise occupancy and that depend on chip size as well as many other factors. Demonstrating stable low threshold operation is one of the key goals of RD53A. Note that we do not specify an input referred noise. This is because input referred noise is not a requirement. The requirement is stable threshold performance with a given input capacitance. Different designs may achieve this with higher or lower noise. We do specify a noise occupancy, because having a low contamination of noise hits in the data is important. We specify this as a maximum noise occupancy per pixel (a dark count probability in an arbitrary 25 ns interval). Note that 10^{-6} noise occupancy means 0.1 noise hits per bunch crossing in a 10^5 pixel chip, while the number of real hits from tracks in such a chip will be of order 100 (10) in the inner (outer) layer(s). Translating this into a dark count rate gives $1.6\,\text{MHz/cm}^2$.

It is reasonable to ask what equivalent input noise charge (ENC) is needed to achieve 10^{-6} noise occupancy at $600\,e^-$ threshold. Clearly one must have ENC$<126\,e^-$, since a 10^{-6} probability corresponds to a Gaussian tail beyond 4.75σ, but the this is not sufficient. It is the quadrature sum of the ENC, the static threshold dispersion, and the RMS threshold fluctuations vs. time that must be less than $126\,e^-$ equivalent input charge. Simply assuming that these three contributions are equal results in a $73\,e^-$ ENC estimate, which is a reasonable target, but it is not a strict specification, as the importance of the other two contributions can be traded off. Threshold dispersion of order $40\,e^-$ after tuning is achieved in current detectors, and therefore a similar value should be achieved by RD53A. Unfortunately, the dynamic threshold fluctuation is the "large chip effect" mentioned above that cannot be reliably simulated. RD53A must try to minimize threshold fluctuation through digital analog isolation, power and bias distribution, and digital and analog design, but a definite target value cannot be specified. Clearly, smaller ENC will generate more margin for this critical task.

3.2 Signal Shaping

The designers can use signal shaping in the analog front end to achieve different goals, so we do not specify shaping details. However, low *in-pixel pileup* is a critical performance requirement, as it is a source of inefficiency. In a free-running discriminator system, if a new hit arrives while the discriminator is still high from a prior hit, this second hit will be lost. Table 1 lists the condition for meeting the in-pixel pileup as 75 KHz average pixel hit rate. This is a proxy for the real test condition that should be a simulation of a train of signal pulses from a realistic inner layer sensor. This train of pulses will have a time and amplitude distribution produced by particles from pp collisions per ATLAS or CMS simulation. We expect such simulations to produce an average rate of signal pulses of approximately 75 KHz per pixel. A long enough time to evaluate steady state operation should be simulated (for example 0.5 seconds).

Note that we have allowed a 1% hit loss, which is about the total loss that the experiments would like to achieve, to come entirely from analog pileup. This is motivated by the expectation

that the digital processing should achieve much lower hit loss than the analog, so that roughly all loss should be in the analog section, where reducing loss costs significant power. Another way to say this is that allowing some analog loss is a much more effective power reduction tool than allowing digital loss. A final analog/digital hit loss budget will be optimized by the experiments after test data with RD53A are available. Note also that a synchronous reset design has negligible in-pixel pileup loss, but it will instead suffer losses due to time-walk, and the level of analog hit loss will be given by the minimum in-time threshold that can be achieved. We have specified $750\,e^-$ in-time threshold for a synchronous reset design, but it is not clear whether this will result in a loss of order 1% relative to a $600\,e^-$ threshold, as it will depend on the detailed shape of the single pixel charge distribution. This discussion should reinforce the fact that RD53A is a prototype, and the RD53A specifications should not be taken as the final specifications for the production chips.

A final point is the effect of very large signals. These can occur not only due to hits in operation, for example nuclear fragments, but also for other reasons in testing. It is hard to define a very large signal size- could be $100\,\mathrm{ke}^-$ or even $1\,\mathrm{Me}^-$. Such events will be very rare and therefore it is not necessary to consider them in estimating efficiency. The important thing is that a pixel should eventually recover from such an event. We have specified $<1\,\mu$s as a recovery time, but this is not a strict limit and even longer times would be acceptable. The important thing is to check that the design is compatible with such rare events.

4. Secondary Performance Specifications

Specification	Value	Comment or Test Conditions
Radiation dose	1000 Mrad	delivered at -15 C. Room T annealing only
Hit charge resolution	$600\,e^-$	analog measurement for every hit
Hit charge dynamic range	$\geq$4 bits	measurement for every hit
Slow analog resolution	$50\,e^-$	measurement for selected hits at low occupancy
Slow analog dynamic range	$\geq$8 bits	measurement for selected hits at low occupancy
Pixel leakage current resolution	1 nA	special measurement mode for every pixel
Larger pixel* capacitance	<300 fF	With $8\,\mu$A analog current. *See text
Larger pixel* leakage current	<20 nA	With $8\,\mu$A analog current. *See text
Larger pixel minimum stable threshold	$1000\,e^-$	With 150 fF load, 8μA/pixel analog. For free running discriminated pixel
Larger pixel in-time threshold	$2000\,e^-$	With 150 fF load, 8μA/pixel analog
Larger pixel in-pixel pileup	$\leq$1%	at 30 kHz/pixel avg. hit rate, 8μA/pixel analog
etc		

Table 2: Specifications related to secondary performance requirements.

This section includes specifications that may be important for final production, for sensor testing, or generally desirable, but not needed to show critical performance goals. RD53A will contain design variants in the pixel array and some of these may explicitly exclude specifications in this

section. Schedule and technical risk considerations may also favor not meeting all specifications in this section. Table 2 lists secondary specifications.

It will be interesting to explore analog designs that can serve either the 2500 μm^2 pixels specified in Sec.3, as well as larger pixels, with double to 4 times the area, and therefore significantly higher capacitance. Table 2 specifies capacitance and performance for larger pixels allowing double the operating current as Sec.3. The difference between these "larger pixels" and the edge pixels in Table 1 and Sec.5.2 is that the edge pixel design can be modified for working well *only* with higher load, whereas here the idea is that the same circuit can switch dynamically between two different modes: small pixel and large pixel mode. An analog front end that can handle a wide range of pixel sizes by adjusting the analog power is a greater design challenge, but it would be necessary to enable different granularity for inner and outer layers. To this end it should be possible to power down 1/2 or 1/4 of each analog quad. Accompanying reduction in digital consumption if possible.

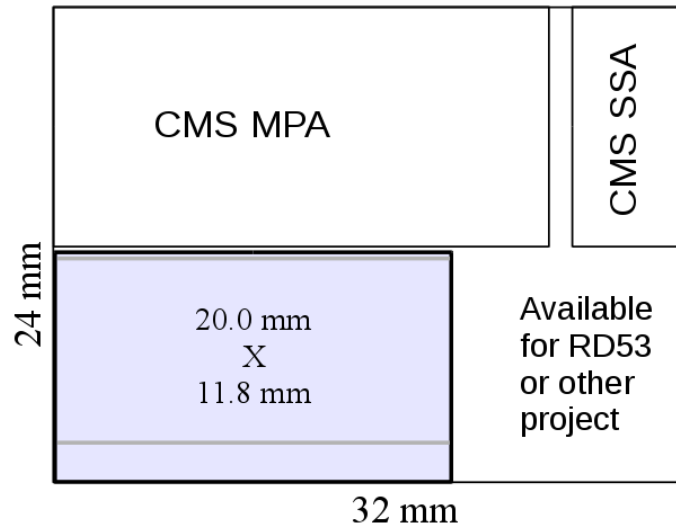## 5. Physical Dimensions, Bumps, and Pins

Figure 1: Chip outline shown inside shared process reticle. The bottom of chip region, indicated by a light solid line, is 2 mm tall.

A large format is important because many performance aspects scale with chip size, but the exact dimensions of RD53A are not fixed by physics requirements. They are instead driven by practical considerations. It is expected that a full reticle run will be produced, with the cost mitigated by other, non-RD53 circuits sharing the available silicon. At this time we anticipate a shared submission with the CMS MPA project, which requires 11.9 mm×32 mm, plus an 0.1 mm dicing street, out of the 24 mm×32 mm reticle. We further consider that many 20 mm and 40mm wide sensors are being prototyped with the FE-I4 chip, and there is fine pitch bump bonding experience with 20 mm wide chips. Combining these considerations, we propose 20 mm×11.8 mm as the RD53A circuit size (seal ring and crack stop are needed outside of that, so the physical dimension after dicing may be approximately 20.1 mm×11.9 mm). Allowing for a 2 mm bottom-of-chip, and

$200\,\mu$m space for test pads at the top of the chip, the active area will contain 400 pixels by 192 pixels. Note that while it would be possible to increase the 20 mm width dimension, this would risk unexpected issues with fine pitch bump bonding, and we therefore propose staying with 20 mm. The chip outline is shown in Fig. 1.

This expectation of sharing with MPA can not be guaranteed. If this sharing becomes impractical, other projects will be sought to share the silicon area, and the RD53A size will remain as indicated. The alternative of a smaller chip on an a Multi Layer Mask (MLM) submission is no longer considered, because the cost of an MLM run is 70% of a full run cost, and there are enough planned projects other than MPA that finding 30% cost sharing, if needed, is very likely. The additional space available on a full run will allow to include other interesting elements in the submission (for example a multiplexer prototype) that would not be possible with an MLM run.
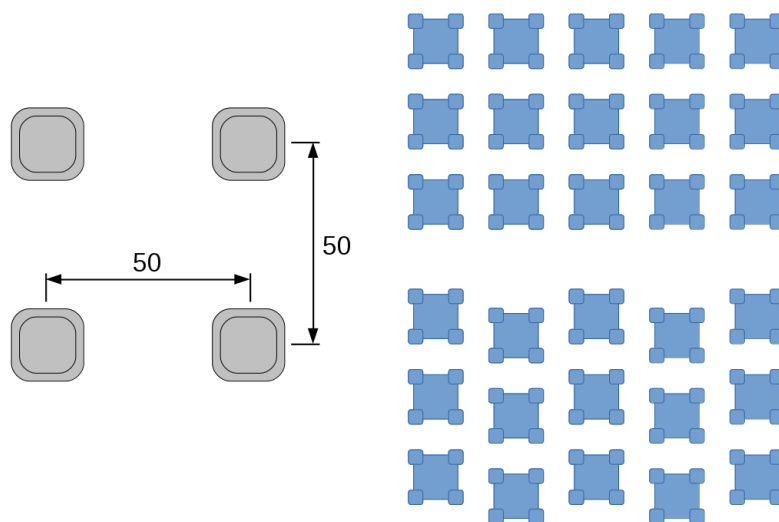
5.1 Pixel and Bump Pattern

Figure 2: Bump pattern basic cell known as analog quad (left), with regular (top right) and offset (bottom right) chip tiling options.

The agreed pixel area is $2500\,\mu\mathrm{m}^2$ and the minimum bump spacing must be $50\,\mu$m. Different bump placements are possible to respect these constraints, but as the bump pads on the chip use the same thick metal layer that is used for low resistance shield, ground or power distribution, the pattern of bumps on the chip will impact the performance. The "unit cell" shall therefore consist of 4 bump pads on the corners of a $50\,\mu$m by $50\,\mu$m square (Fig. 2 left). The area of this square can be devoted to analog circuitry and the top metal levels between the pads to analog power and bias routing. The actual "analog quad" circuit dimensions are not specified and to be optimized by the designers. For reference, in the FE65-P2 prototype test chip the analog quad dimensions are $67 \times 64\mu\mathrm{m}^2$. These analog quads will tile the entire chip surface as shown on the right side of Fig. 2, with space between them (blank in the figure) available for synthesized logic. The regular tiling pattern (top right of Fig. 2) has been implemented in the FE65-P2 prototype and works well for logic synthesis. The offset tiling pattern (bottom right of Fig. 2) may also work but has not

been investigated yet in practice. It is expected to lead to lower place and route density. Horizontal offsets are not possible as they would interfere with power and bias distribution.

The bump pad shape and dimensions are shown in fig. 3 and compared to the pad dimensions used in FE-I4. The layout rules of the 65 nm process do not allow the same pad geometry as used in FE-I4. The passivation opening is actually slightly bigger area than for FE-I4, due to the more square shape, while the metal is smaller, which reduces capacitance, crosstalk, and benefits power, ground, and shield routing.
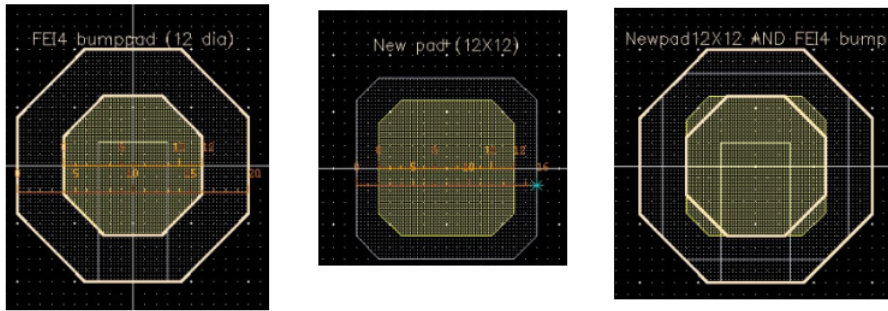

Figure 3: Bump pad for from FE-I4 chip (left), specified for RD53A (center), and both overlaid (right). All openings are $12\,\mu$m wide by $12\,\mu$m tall- only the shape of the corners varies.

5.2 Edge Pixels

Multi-chip modules require special treatment of the area of a sensor spanning the gap between two readout chips. Traditionally this has been done with *long* or *ganged* pixels. But as the pixel size shrinks both these approaches become problematic. In order cover inter-chip gaps that are larger than normal pixel size, it is necessary to "share the pain" among multiple pixels, rather than just the very last row or column. For RD53A we propose that the four rows or columns of pixels closest to an abuttable edge are biased differently than the rest of the pixels in order to be able to function with approximately double the load capacitance and leakage current, allowing for 50% higher threshold and in-time threshold. Top, left, and right side special biases must be independently controlled, so that the correct bias can be selected depending on the sensor used (1-chip, 2-chip horizontal, 2-chip vertical, quad...). Three rows or columns are probably enough geometrically (Fig. 4), but four is a better match to a circuit layout based on analog quads. The current consumption for such pixels is allowed to be more than double that of interior pixels, as needed, since this will have a negligible impact on the total chip power consumption. Pixels within 3 or 4 of an abuttable corner need no additional special treatment- they will simply have degraded performance. All such special pixels are not required to meet the in-pixel pileup specification. We expect it is acceptable to the experiments to have higher than 1% hit loss in such edge/corner pixels. Fig. 4 shows how sensor pixel boundaries and sensor single metal layer might look for interfacing to thee rows/columns of *double capacity* edge pixels.

5.3 Sensor Guard/Bias Ring Bumps

Special bumps for access to sensor guard rings or bias grids will be included at the bottom of the regular bump matrix. These will be connected by metal, without antenna diodes, to dedicated
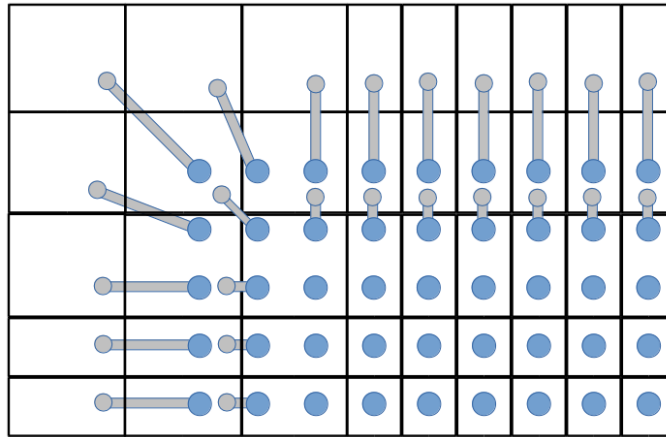
Figure 4: Diagram showing pixels in an abuttable corner of an example sensor, making use of the RD53A special edge pixels to cover inter-chip gaps. Sensor pixel boundaries and one-layer sensor metal are shown. The larger blue circles are on a $50 \times 50 \mu m^2$ grid matching the RD53A bump pad locations. 12 Normal pixels ($50 \times 50 \mu m^2$) can be seen at the bottom right of the figure. The remaining sensor pixels are larger, together creating $150\,\mu m$ of gap coverage, without any need for crossed metal routing.

wire bond pads. Note these are not rated to carry high voltage, but expected to be at near ground reference potentials. Four such special bumps at each end will be included as shown in Fig. 5. The four bumps at each end will be connected together to a common wire bond pad.
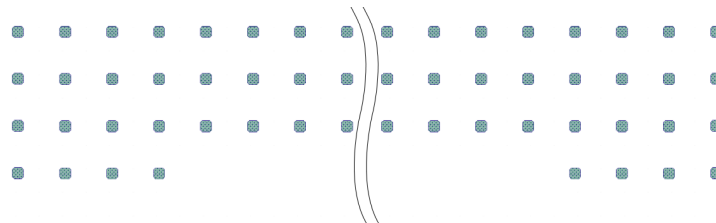

Figure 5: Location of guard ring / bias grid bumps. The placement is equivalent to an extra row, but covering the four end columns only (at both ends)

5.4 Alignment Marks

Alignment marks are needed for flip chip bump bonding. Two marks should be included at the extremes of bottom of chip periphery. Marks at the top of the chip are highly desirable, but may not be possible given the bump pad density of RD53A. The exact layout of the marks should be coordinated with bump vendors. The size of the marks used in the FE-I4 chip was $60 \times 60 \mu m^2$.

5.5 Wire Bond Pads

Wire bond pads will span the 20 mm width at the bottom of chip. Additionally, small test pads and test structures at the top of the chip are allowed for testing purposes (such top pads and structures

would not exist in a production chip). The chip shall be fully operational from the bottom pads alone, without any need for top connections. No pads are allowed on the sides of the pixel matrix. The bottom wire bond pads must be compatible with via last TSV processing. Additionally, via last test structures enabling daisy chains for TSV quality control should be included either at the bottom or top or both.

6. Power Supply and Regulation

The performance must be demonstrated under realistic power supply conditions. Additionally, both experiments have a baseline of serial (constant current) power distribution and this must be validated with the RD53A chip. RD53A must therefore operate from a single supply with voltage as high as 2.0 V. A very important consequence of this is that RD53a must contain voltage references that can operate from a 2 V input. The chip must include internal voltage regulators that generate the necessary internal rails (set relative to the above reference voltage) from one common external power rail. In order to avoid that all the current for a given internal rail must be concentrated in a single point, the internal regulators should be distributed and compatible with parallel operation. The internal regulators must also present a programmable, constant current load to the external power network, as needed for powering chips or modules in series (It should also be possible to disable the constant current consumption feature of the regulators- program it to zero). The possibility of programming via configuration, as opposed to an external resistor, should be evaluated, keeping in mind that it should not be possible for an SEU to include a serial chain power failure.

Constant current is not only a needed feature to serial power, but also expected to be helpful for achieving low stable threshold. Thus, the more constant the load on the internal power rails, the less work the regulators will have to do to draw constant current from the external source, and the better the low threshold performance is likely to be. This is a particularly important consideration for logic design, where common approaches that reduce power when digital activity is low, while highly desirable in other applications, would be counterproductive here.

While demonstrating the above power distribution method is an important goal, RD53A is also a prototype, and as such should have some flexibility for testing, studies, and contingency. It should therefore also be possible to power the chip by directly supplying the internal rail voltages, bypassing the regulation.

RD53A should include power-on reset circuitry to ensure reliable, safe startup in the entire temperature range. The correct startup of the voltage regulators (no oscillatory state) should be guaranteed in a wide range of ramp conditions.

7. Input and Output

This section covers digital communication as well as test analog I/O requirements. For digital communication, needed to operate the chip, two independent paths are required for RD53A. The first is a prototype for what will be needed in the experiment, and so we refer to it as "prototype I/O". The input and the output of the prototype I/O are described separately. The second will allow to more directly access functions within the chip, needed for debugging and testing, and is called "Test I/O". Analog I/O will allow further detailed debugging and testing.

7.1 Prototype Input

The production pixel chips must be operable with a minimal number of connections in order to enable the experiments to minimize services mass. The prototype input must contain all needed Trigger, Timing, and Control (TTC) information on a single differential pair (link). Furthermore, it should be possible for multiple chips to share a single TTC link. Each chip should have a 3-bit address that can be hard wired.

System level requirements for the RD53A prototype input are defined in Table 3. This is not a full specification of the protocol (that is, it will not possible to control the chip knowing only the information in this table). It is left to an Implementation Document to give the full details of the protocol, down to the valid bit sequences and the way the chip will interpret them. Note that an important, required feature is that configuration data can be sent any time. Thus the chip will not have a "configure mode" and a "run mode" as is currently typical. A priority sequence for encoding and decoding TTC information will be detailed in the Implementation Document.

Feature	Value	Comment or Test Conditions
Input data rate	160 Mbps	
Electrical protocol	DC-balanced, JEDEC SLVS	Note this is operable A/C coupled
Compatibility	LPGBT e-link	Must be able to use a GBT e-link
Clock	160 MHz	recovered from link transitions
Recovered Clock jitter	< 100 ps	RMS short-term jitter
Bunch crossing clock	40 MHz	denoted BX clock. 160 MHz clock/4
BX clock phase	recovered from link data	framing defines phase
BX phase offset	Programmable < 0.5 ns step, 50 ns range	Each chip can have its own offset
Trigger	Arbitrary, decoded from link	Supports any pattern of bunch crossings
Trig. encoding/ decoding latency	$<= 200$ ns	On top of transit time from DAQ to chip
Fault tolerance	single bit flip per 16 bits	if needed by DAQ
Configuration data effective bandwidth	> 10 Mbps	Can send any time (see text)

Table 3: System requirements for prototype input

7.2 Prototype Output

RD53A should have two outputs "ports", user-selectable between them. A "fast serial" output and a slower "parallel" output. The parallel output is mandatory, while the fast serial output is to be implemented on a best effort basis. Note that the parallel output total bandwidth (counting all bits sent in parallel) can match or exceed that of the fast serial output.

The fast serial output should be differential at ~ 5 Gbps maximum. The output bit rate should be programmable down to 1/16 of the maximum. It should be designed to drive cable models provided by the experiments with up to 6 m total length. The design should include a receiver

design, or identification of an available commercial receiver, as well as any necessary equalization in either driver, receiver, or both.

The parallel output should have 4 SLVS pairs, each capable of 2.56 Gbps maximum. The output bit rate should be programmable down to 1/8 of the maximum. The number of parallel outputs used should also be programmable, selecting all 4, 2, or just 1 (serial). At 1.28 Gbps or lower, these outputs should be compatible with driving LPGBT inputs. They should include necessary equalization (such as programmable pre-emphasis) to drive cable models of up to 6 m provided by the experiments.

The details of each of the prototype output ports, including the data encoding used, should be spelled out in the Implementation Document. Data should be DC balanced and therefore compatible with A/C coupling. The encoding used should allow for hit data as well as diagnostic and configuration readback data as needed. There should be a single output mode that supports all data types. Fault tolerance on output data is not required.

7.3 Test Digital I/O

Signal	Type	Function or comment
Slow Control	I/O	Multiple pads per protocol TBD (eg. JTAG, I2C, etc.)
BX clock	Input rising edge	bunch crossing clock
Trigger	Input level	latches on BX falling edge
Default	Input level	set default configuration
Reset-bar	Input level	resets all logic. Low is reset.
Hit_OR	SLVS output	hit signal from selected pixels
Monitor	Output	Select any internal dig. signal via register

Table 4: Minimum required signals on dedicated pads for Test Digital I/O interface. Additional pads can be defined in the Implementation Document as needed by designers.

As RD53A is a test chip, it is allowed and expected to have have "slow" I/O for dedicated tests of internal circuits, for example for scan chain control. This I/O can use CMOS levels and it should not be active during prototype I/O control in order to avoid noise injection. The slow control interface can use a standard protocol, for example I2C or JTAG, to be defined in the Implementation Document. Optional programmability of the internal registers via a slow interface is recommended. The Test Digital I/O should provide "back door access" to operate and read parts of the chip even if the prototype input fails. Thus the Test I/O must include at least the signals listed in Table 4 as dedicated pads. Use of Test I/O to override the standard input must be selectable either by dedicated mode pads or by a failsafe register that can always be programmed via Test Digital I/O interface.

7.4 Analog I/O

This is similar to Digital Test I/O, but for analog signals, such as direct spying in internal nodes of selected pixels, and analog biases. The following analog pads are required:

- One pad for each analog bias to monitor or override

- One or more multiplexed analog outputs to spy on single front end internal nodes

- A calibration injection voltage input (pulse or level to be determined by designers)

- An output-input pair of pads for each reference current, so that the current can be measured externally or an external reference can be supplied

- One pad for each reference voltage to monitor or override

8. Required Features

8.1 Default Configuration

Selectable by pin which forces all configuration bits to a hard-wired default. This should be very radiation hard.

8.2 Masking, Calibration Injection, Hit OR and Self-Trigger

The following functionality should be possible in RD53A

- Mask the output of each pixel independently.

- Select charge injection or not on each pixel independently. Should have the ability to inject up to $25 \, \mathrm{ke}^-$.

- ORed digital (hit/no-hit) output of each pixel. Can be ORed in columns, and then all columns, or several groups output on dedicated pads. It should also be possible to internally generate a trigger (with a programmable delay) from Hit OR signals.

- Separate mask (1 bit per pixel) for participation in above global OR.

- Support for externally supplied analog injection pulse.

- Internal calibration pulse generation, via chopper or internal pulse generator. this does not need to have a programmable delay, thanks to the clock phase adjustment of Sec. 8.4.

8.3 Threshold and Gain Equalization

Threshold trim bits for each pixel will be needed to achieve the required threshold dispersion. Depending on the ADC method, trim bits for fall time (which also affects gain) may also be needed. This up to the analog design. The threshold dispersion after equalization of *tuning* should be of order $40 \, e^-$, but the exact value is a design parameter as discussed in Sec. 3.1. The threshold and dispersion are allowed to vary with temperature, but such variation should be small such that small changes do not require a full retune. A dispersion variation of $< 5\%/^\circ\mathrm{C}$ and a threshold variation of $< 10 \, \mathrm{e}^-/^\circ\mathrm{C}$ can be taken as a guideline, but are not strict requirements.

8.4 Timing Dispersion and Clock Phase Adjustment

Timing dispersion (difference in time from pixel to pixel) can be caused by clock distribution within the chip and by variation in the front end rise time and discriminator delay.

Bunch crossing clock (BX) distribution to all the pixels should be approximately synchronous, but a skew of <4 ns is permitted as this may be desirable to avoid large power transients. The BX will be generated from the incoming serial command stream, after 160 MHz clock recovery. The reference phase (zero shift) is given by the sync frame as discussed in Sec. 7.1. The BX distributed to the pixels should have a programmable digital phase delay with 1 ns or smaller step size and a 50 ns or larger range. The high frequency data output clock generator can be used to derive the phase adjustment step.

Calibration injection distribution should be as synchronous as possible, ideally <1 ns difference to any pixel, in order to be able to characterize the timing performance. Timing dispersion caused by discriminator delay should be <2 ns RMS, either by design or after equalization, for example by discriminator current trimming.

8.5 Errors and Warnings

Error, warning, and status messages should be implemented in the prototype output.

8.6 Capacitor value calibration circuit

A standalone circuit should be included to provide a 1% precision or better absolute measurement of a reference capacitor, representative of injection and feedback capacitors.

9. Additional Features

A final production chip will have many features not covered in sections 3 and 8. However, RD53A is not intended to be a final production chip, and implementation of features that may be important in production, but not central to goals of RD53A, will increase both design time and technical risk. Therefore, the features listed in this section are to be considered on a best effort basis, if they do not add significant design time or technical risk. This section is intended to collect, for reference, all proposed features that may be needed or interesting in production. It is useful to have a comprehensive list, because it could well be that some feature that sounds exotic and low priority turns out to be effortless to implement already in RD53A.

9.1 Special Trigger / Readout Modes

Full readout of all hits at low rate without trigger.

9.2 Scan Chains and Design For Test

Use of scan flip flops mandatory in periphery to allow structural testing. In pixel array not required-up to designers how much testability can be included in the available space for RD53A.

9.3 Monitoring of Internal Voltages and Currents

Include appropriate ADC's and methods to read back via the data stream operating values of interest. These include supply input voltage, regulated internal rails, regulator currents, temperature, etc.

9.4 Data Compression

Minimizing the required output bandwidth will be critical at the high trigger and hit rates of the HL-LHC. Implementation of on-chip lossless compression methods is therefore desirable in RD53A. (Cluster pairing algorithms as used in FE-I4 are an example of lossless compression, but more sophisticated compression is expected to be used in the future).

9.5 Heartbeat

Upon power-up the data output should be immediately active producing a heartbeat signal, even without any serial command input. Signal should be usable for diagnostics- for example it would be different if there is no incoming command stream, corrupted incoming stream, or correctly locked and decoded incoming stream.

9.6 Self Test / Auto-Tuning

Sophisticated self-test features could be eventually included, such that only power and serial output are needed to have a fairly complete chip self-test, which would be very useful for wafer probing. Additionally, having the capability to internally generate test patterns could dramatically increase calibrations speed, which could be rather important in future experiments. In particular the feature of automatically performing threshold and charge measurement equalization (auto-tuning) may be very important to manage radiation-induced de-tuning at high luminosity (such auto-tuning could be implemented with programmable control patterns, as described here, or some other way). Threshold and other scans could be internally available in memory blocks, or more ambitions would be to have a processor and memory, so that long control patterns can be internally generated. None of this would be need to be SEU hard, as it would be used while no beam is present.

9.7 DC-DC Converter

An experimental version of a fully-on-chip DC-DC converter could be included in RD53A if available. This would not replace the regulators required in Sec. 6, but could be present in addition to them, for testing. To be competitive with serial powering of constant current devices, the DC-DC conversion ratio must be 4, although 3 may be acceptable. To allow for internal voltage rail voltage up to 1.2V, the input voltage rating must be 5 V or higher, while still meeting the radiation tolerance specifications.

9.8 Prompt Output Encoded in Output Data

A prompt output signal (for example a fast OR) could be encoded in the output data stream. This requires study as it must be compatible with the output protocol. Since clearly this can't be an asynchronous signal, it could be a special word with a payload encoding a wait time value from the asynchronous event to the actual time the special word was produced at the chip output.

9.9 Internal Processing and Histogramming

Programmable pattern search to count occurrence frequency in un-triggered data. For example histogramming of cluster length distribution. Histograms could be read out at will the same as configuration registers. Internal processing can also be used to feed internal self-trigger (Sec. 8.2).

9.10 Realistic Calibration Injection Patterns

Calibration injection in past pixel chips does not simulate real data patterns. The minimum time interval between successive injections is many bunch crossings, and the amount of charge injected is the same for all selected pixels in a given injection event. It would be useful to be able to execute more realistic injection patterns. For example, a large charge in pixel x and a small charge in pixel $x + 1$, separated in time by an arbitrary, programmable offset. This could be accomplished if there were two parallel injection circuits selectable pixel by pixel, or perhaps with the ability to have a digital injection pattern in parallel with conventional analog injection. Even if this is not implemented in RD53A, it will be useful to consider such eventual functionality in the Prototype Input protocol definition.

References

[1] RD53 Collaboration, "RD Collaboration Proposal: Development of pixel readout integrated circuits for extreme rate and radiation," CERN-LHCC-2013-008 ; LHCC-P-006 (2013).

[2] M. Karagouins *et. al,* "An Integrated Shunt-LDO Regulator for Serial Powered Systems," in Proc. of IEEE ESSCIRC '09, (2009).