Development of a Stitched Monolithic Pixel Sensor prototype (MOSS chip)

towards the ITS3 upgrade of the ALICE Inner Tracking System

TWEPP'22, Bergen, Norway. September 21th 2022

Pedro Leitao <Pedro.Leitao@cern.ch>
On behalf of all the designers and contributors to the MOSS chip

Overview

Motivation

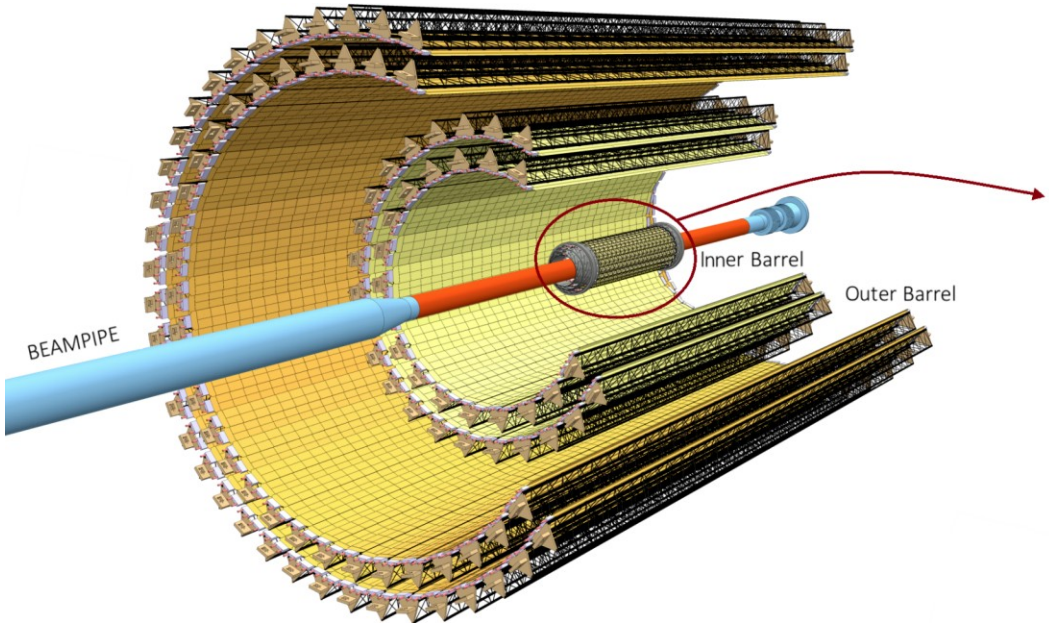
MOSS Prototype

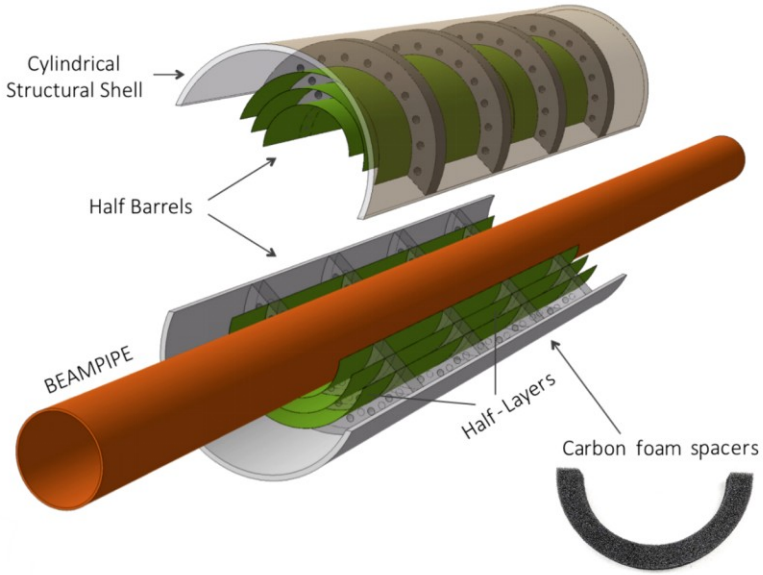
MOSS challenges

- Yield
- Power distribution
- Data transmission

Summary

ALICE ITS3 upgrade





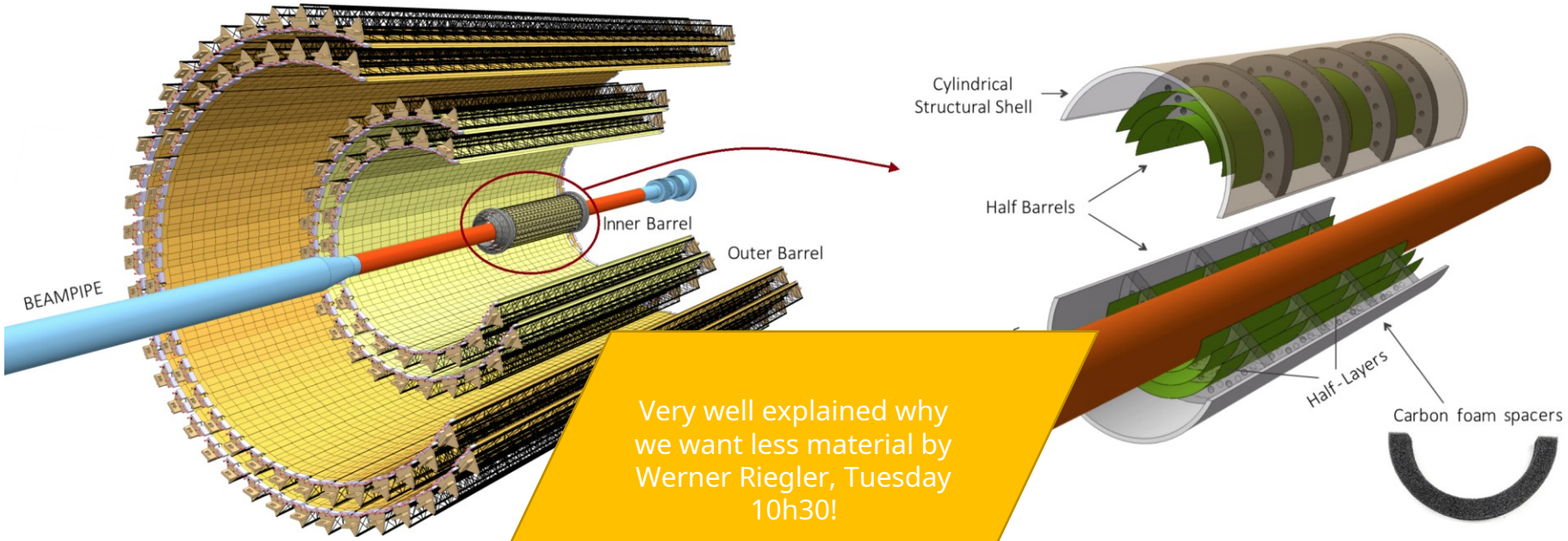
ALICE ITS2 to ITS3 upgrade

- Less material equals better physics
- Less power consumption allows for air cooling
- Onchip data transmission allows for no flex
- Bended Si wafers allow a stable mechanical structure
- ITS3 Letter of Intent proposes the use of **wafer scale stitched particle detectors**

Dummy silicon model

https://indico.cern.ch/event/1071914, *ALICE ITS3 – a next generation vertex detector based on bent, wafer-scale CMOS* sensors, Magnus Mager (CERN)
https://cds.cern.ch/record/2703140/files/LHCC-I-034.pdf - *Letter of Intent for an ALICE ITS Upgrade in LS3*

ALICE ITS3 upgrade



Very well explained why we want less material by Werner Riegler, Tuesday 10h30!

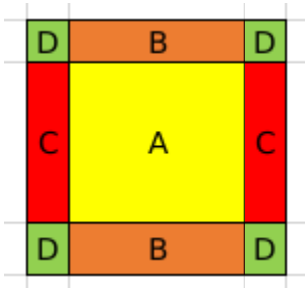
ALICE ITS2 to ITS3 upgrade

- Less material equals better physics
- Less power consumption allows for air cooling
- Onchip data transmission allows for no flex
- Bended Si wafers allow a stable mechanical structure
- ITS3 Letter of Intent proposes the use of **wafer scale stitched particle detectors**

Dummy silicon model

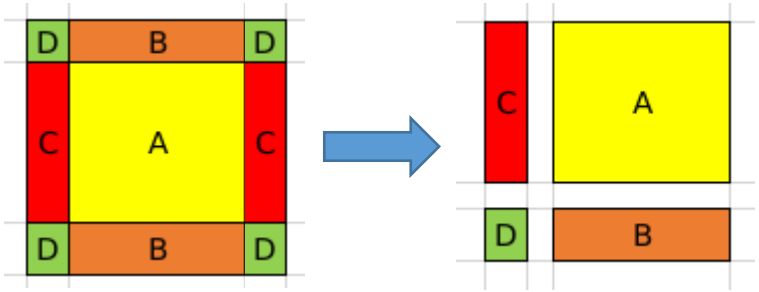
https://indico.cern.ch/event/1071914, *ALICE ITS3 – a next generation vertex detector based on bent, wafer-scale CMOS* sensors, Magnus Mager (CERN)
https://cds.cern.ch/record/2703140/files/LHCC-I-034.pdf - *Letter of Intent for an ALICE ITS Upgrade in LS3*

Wafer scale stitched sensors



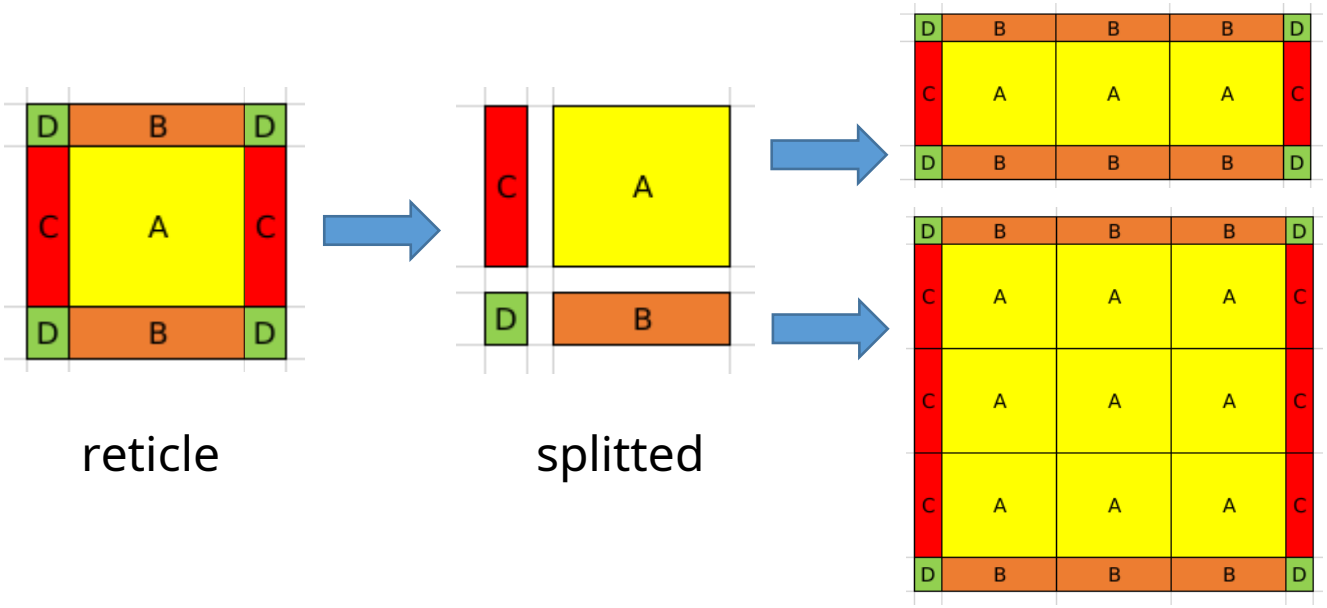
reticle

Wafer scale stitched sensors



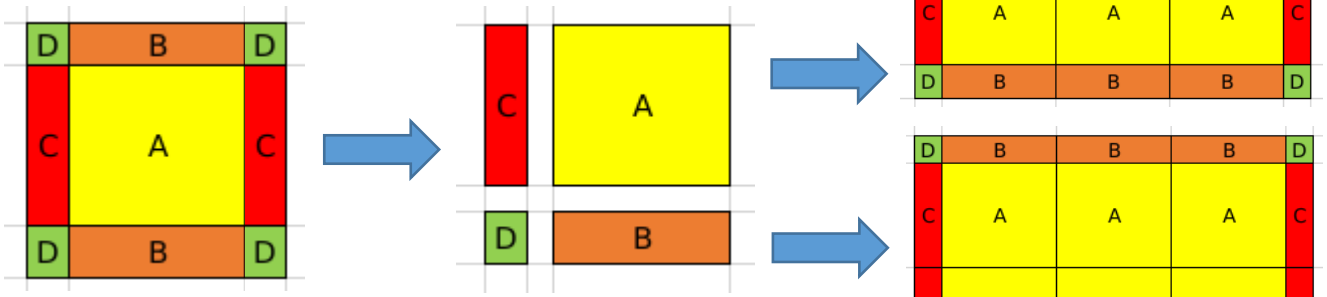
reticle splitted

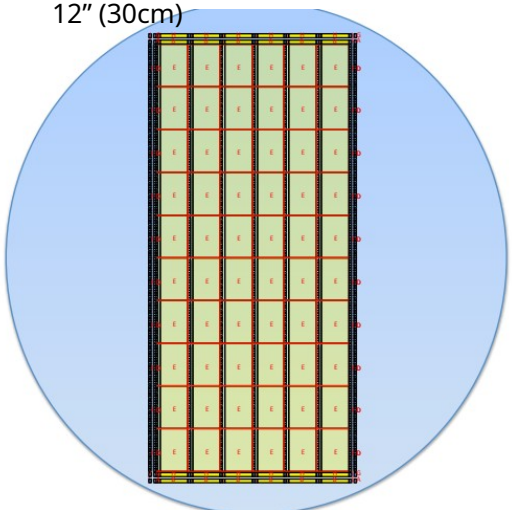
Wafer scale stitched sensors



reticle splitted stitched objects

Wafer scale stitched sensors





thinned wafers to < 50 um
slice wafers to size depending on radii

6 wafers make the entire ITS3 production

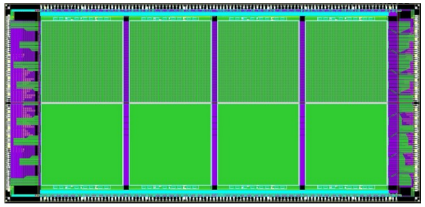
Dummy silicon model

MOnolithic **S**titched **S**ensor (MOSS)

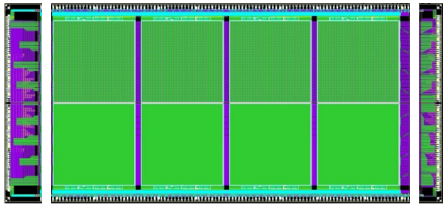
We want to understand if we can design a stitched monolithic particle detector with satisfactory yield. The MOSS prototype is a proof-of-concept.

Thus, the primary goals are:

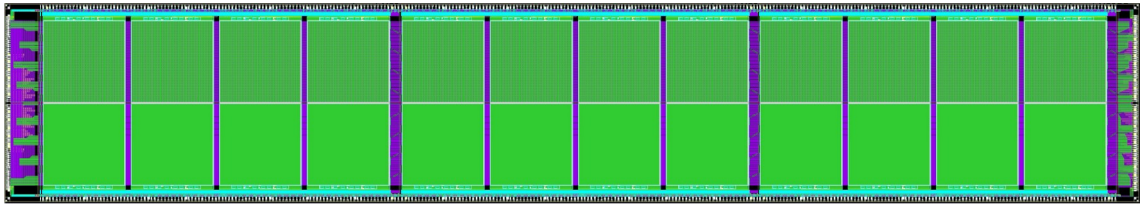
- Learn stitching techniques
- Interconnects
- Learn about yield and design-for-manufacturing (DFM)
- Study power schemes, leakage, spread, noise and speed
- Develop inhouse stitching methodology



reticle

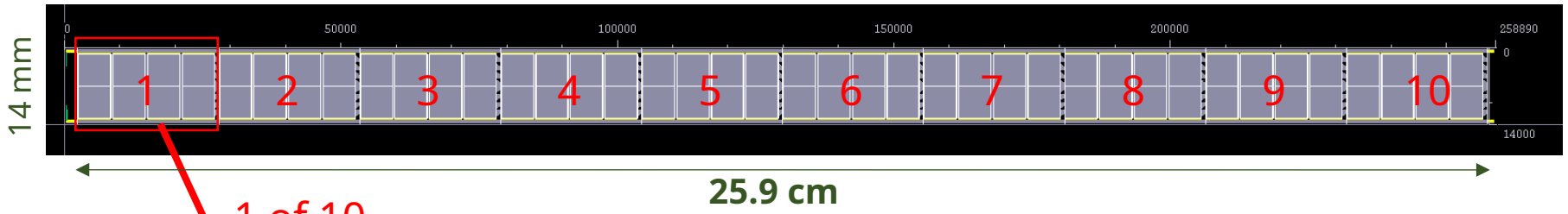


splitted

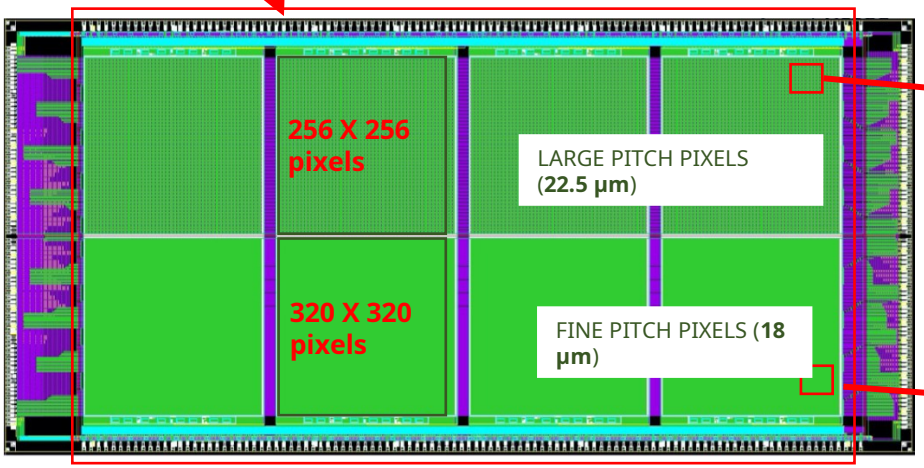


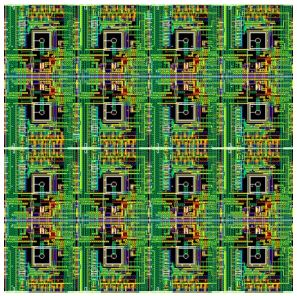
stitched object

MOSS Prototype summary

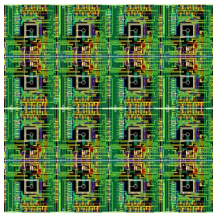


1 of 10





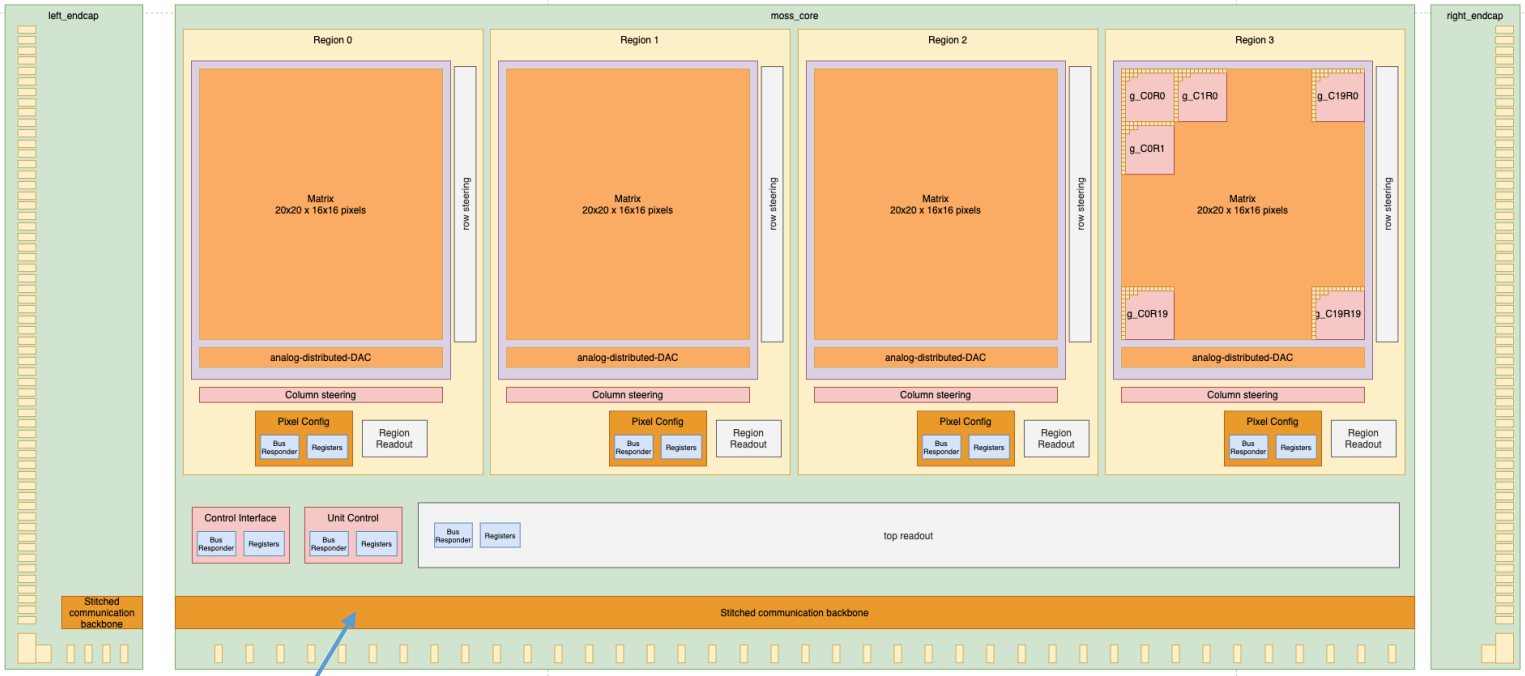
Pitch 22.5 µm
- Conservative layout
- 7 mW/cm2 (analog FE)
- 1us peaking time



Pitch 18 µm
- Compact layout
- 11 mW/cm2 (analog FE)
- 1us peaking time

- 1.4 x 26 cm monolithic stitched sensor
- Binary readout with parameterizable strobe duration
- In-pixel latch with fast OR for column and row signals
- Analog and digital pulse testing per pixel
- Periphery designed with custom DFM std cell library
- Matrix designed with DFM rules in mind
- 736.3 Million transistors
- 1.67 Million pixels

MOSS half unit architecture

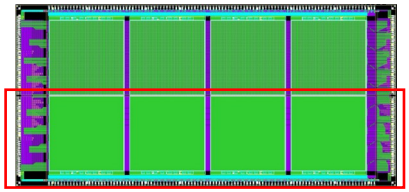


Left endcap (LEC)

Repeated sensor unit (RSU)

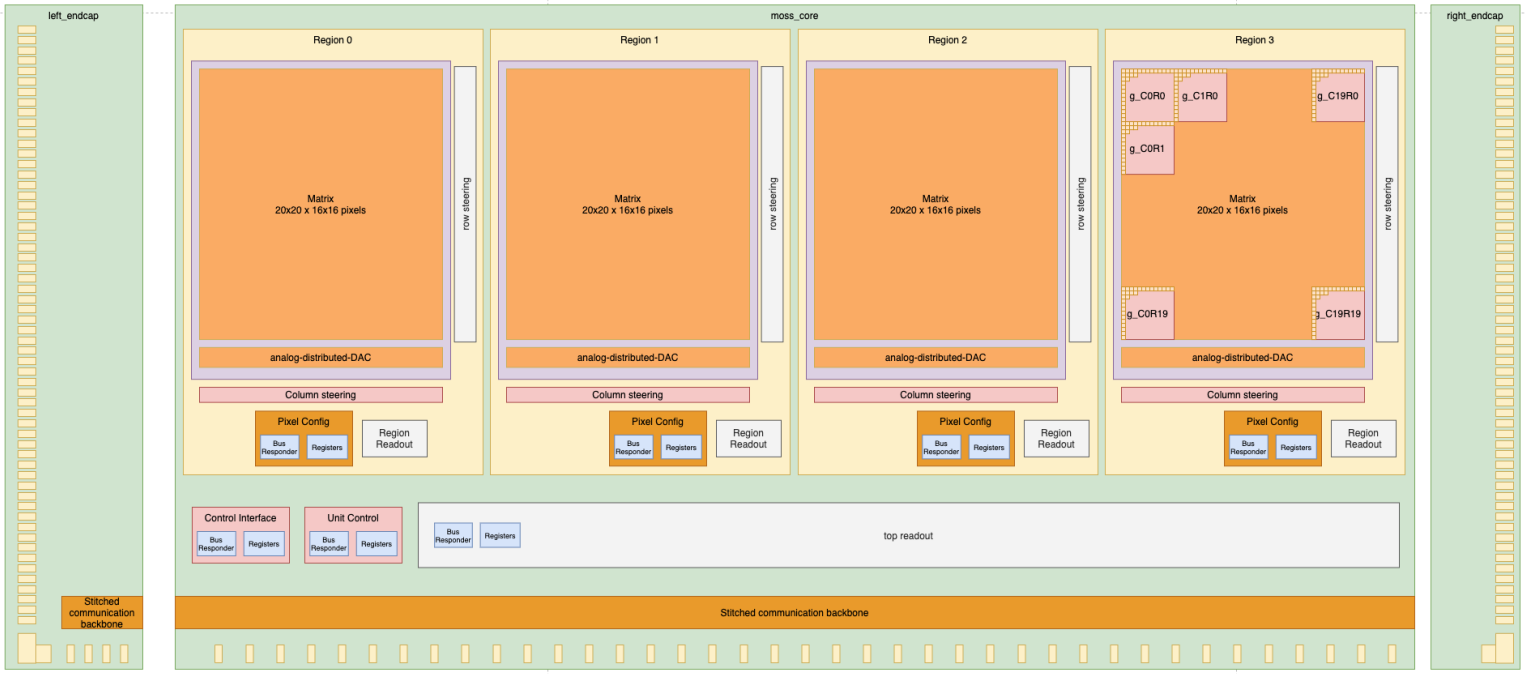
Right endcap (REC)

stitched backbone



more verbose architecture description in the backup slides

MOSS half unit (long-edge control)



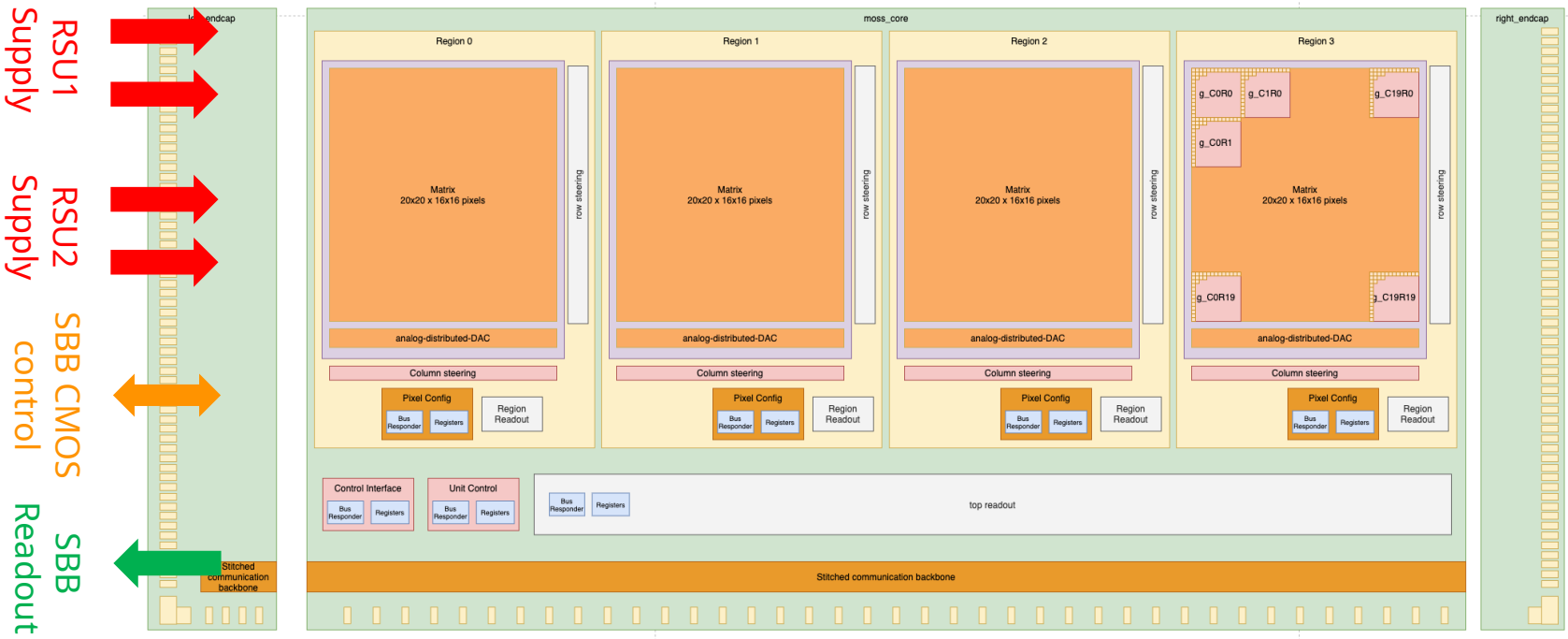
RSU Supply

RSU CMOS control

RSU Readout

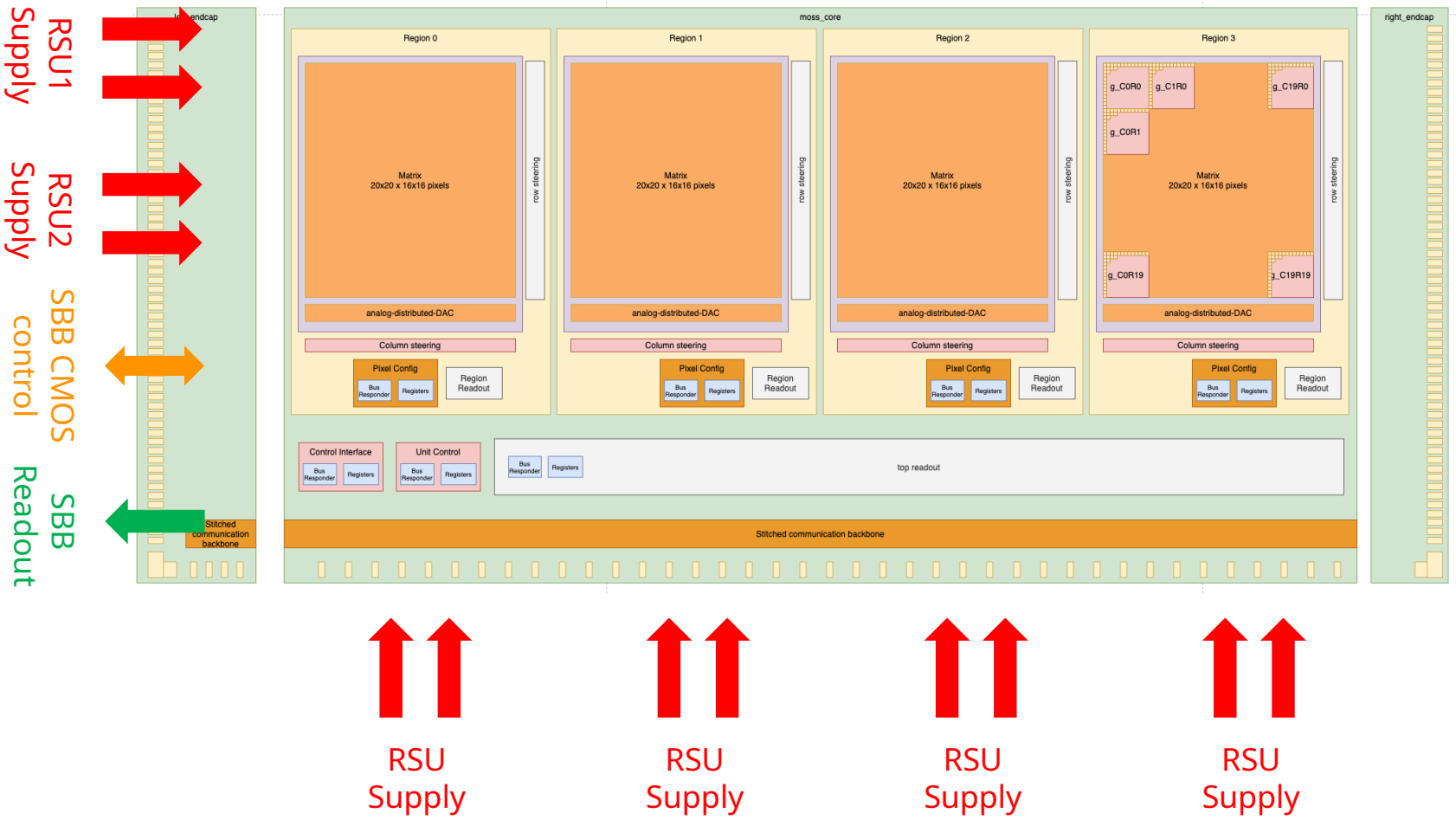
RSU Analog control override

MOSS half unit (short-edge control)



- Stitched backbone (SBB) allows for inter stitching communication
- Power can be applied/monitored over the short-edge

MOSS half unit (hybrid control)



- Stitched backbone (SBB) allows for inter stitching communication
- Power can be applied/monitored over the short-edge
- When controlling through the short-edge, one can use long-edge power pads to decrease IRdrop
- Possibility to mux short/long control and readout

Challenges of stitched design

What are the biggest challenges of a wafer stitched particle detector, which has 1.4 cm (short-edge) x 30 cm (long-edge)?

- Yield
  - Manufacturing defects can have high impact depending on the chosen architecture
- Power distribution
  - How to distribute power only using the “short-edge”?
- Data transmission
  - How to transmit data over the “short-edge”?
    - readout can be in the Gbps range while slow control in the Mbps range

MOSS yield improvements

To improve yield we have followed:

- All width/spacing were maximized whenever possible
- Multi-cut vias are of the utmost importance (min double-via)
- A custom designed-for-manufacturing (DFM) standard cell library was implemented with multi-cut contact vias
- Architecture plays a major part. In order to prevent a short rendering the entire chip unusable:
  - Having granularity on the power distribution will prevent a power short to compromise the chip
  - Having fault tolerant architecture which allows to switch off section of the logic

MOSS Power distribution

There are:

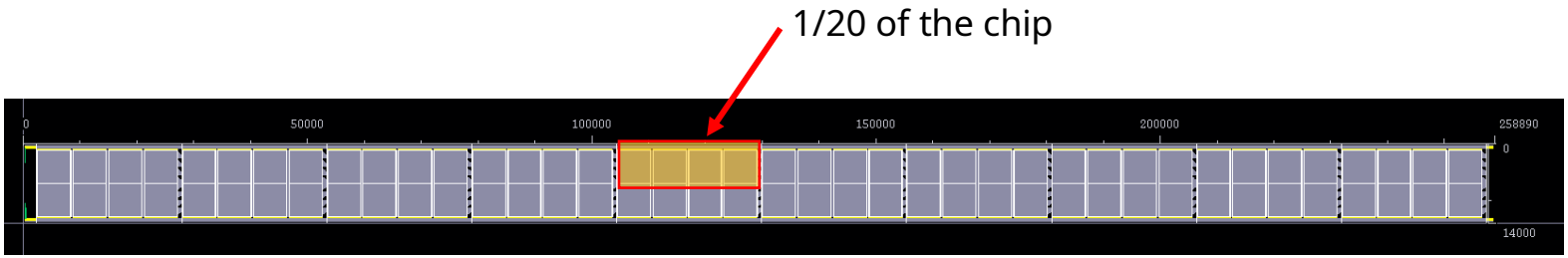
- 44 power/ground lines for pixel, periphery and stitched backbone (1.2V / 0V)
- 22 IO domain (1.8V) for CMOS pads
- 1 PSUB which can be reversed bias (0V / -0.6V)
- this means up to 67 power domains!!

This requires:

- well isolation between power supplies, in pad ring and circuitry
- level shifters
- at least one pin per power/ground line

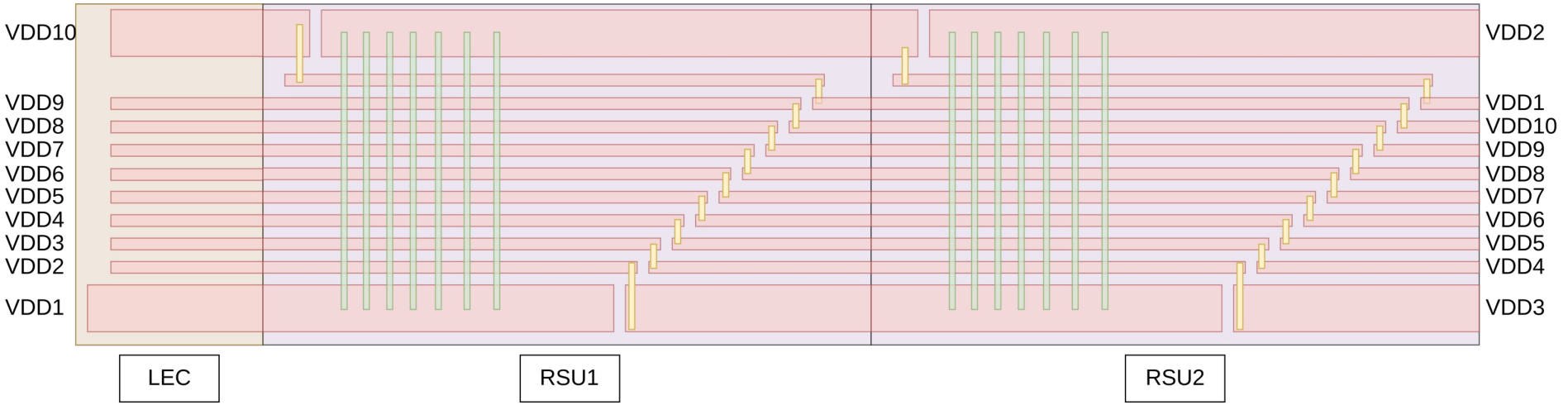
But allows for:

- in case of shorts, we can simply switch-off 1/20 of the chip



MOSS Power distribution

Powering & monitoring from the short-edge (example for VDD)



With the same physical RSU, it is necessary to forward power from the short edges to the most far RSUs

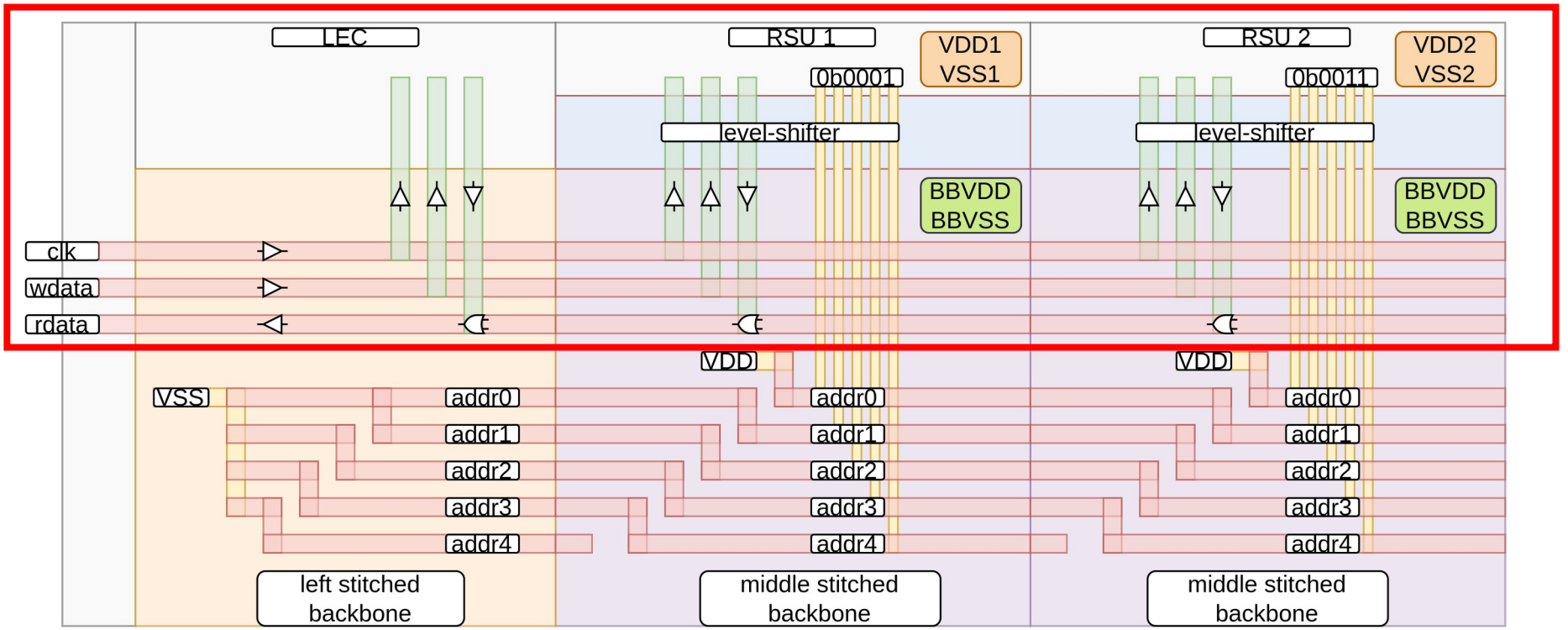
We achieve this by performing *line-hopping*

For the MOSS prototype, the current metal stack only allows to power the most left (RSU1) or most right RSU (RSU10). All the other power pads are only used for *monitoring* in order to verify the stitching. More on the backup slides.

MOSS SBB slow control

Stitched backbone

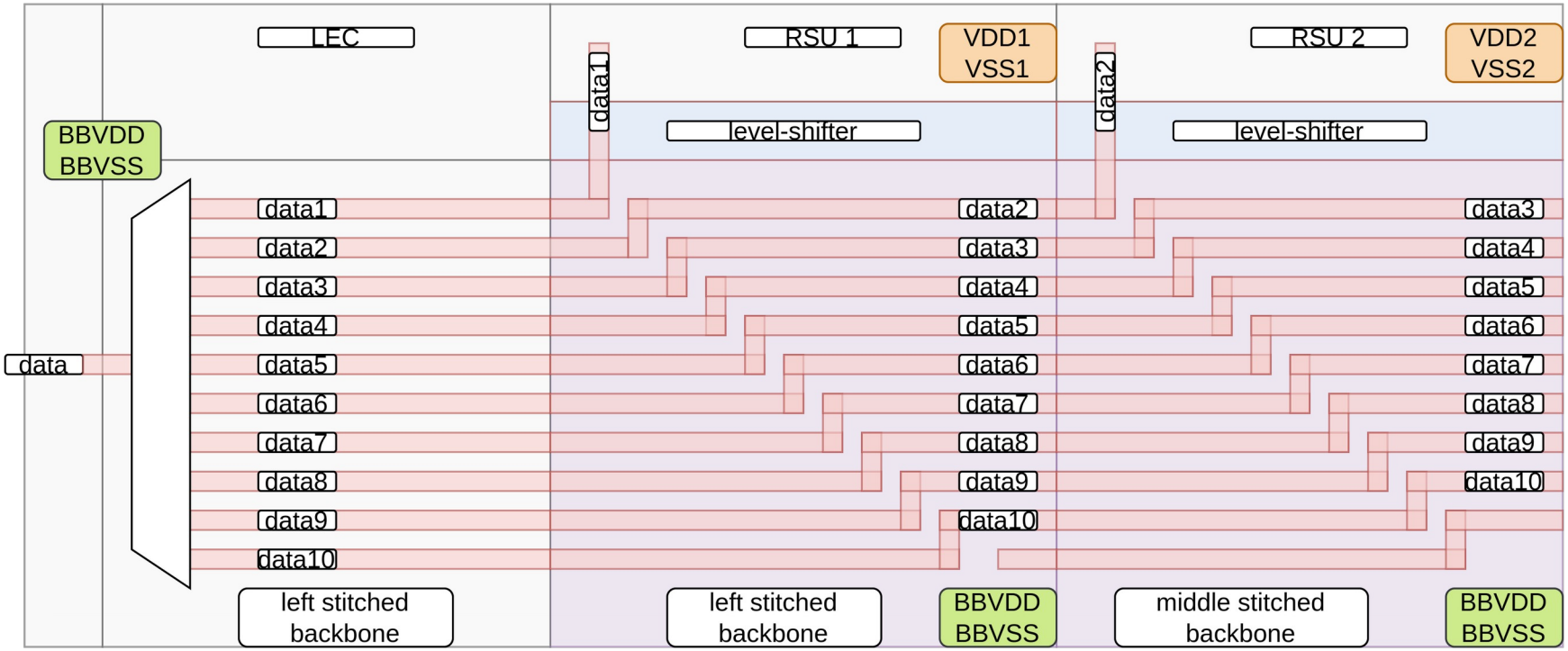
- The MOSS prototype uses serial slow control
- Write transactions are simple point-to-point
- Read transactions are OR'ed together
- No retiming between RSUs greatly simplifies timing
- Necessary a level shifter between power-domains
- RSU adressing is also handled by *line-hopping*. More on backup slides.



MOSS SBB data readout

Stitched backbone readout

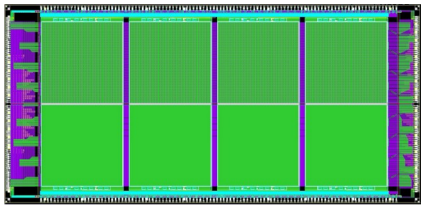
- Readout requests are sent over the serial slow control
- We can read one RSU at a time (1x 15 Mbps)
- Line-hopping allows to connect all 10 RSUs to a mux in the LEC
- The mux is controlled by means of the slow control



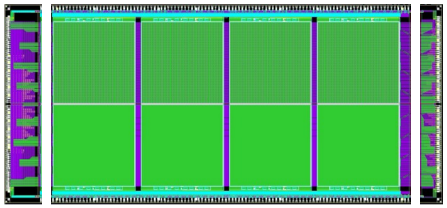
Special readout mode: all RSUs at once (10x 3.75 Mbps)

Inhouse stitching methodology

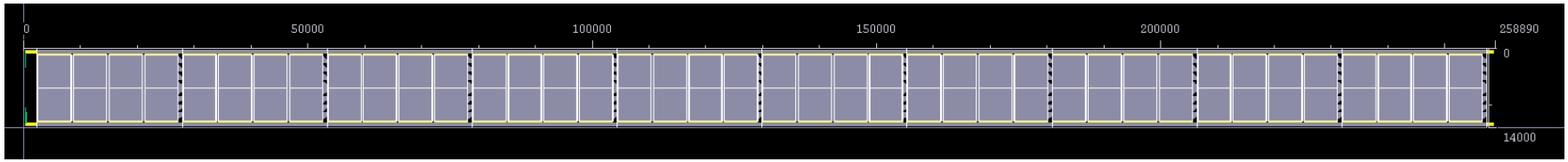
Digital on top implementation



split in left, middle, right



stitched 10x version



To ensure Quality-of-Results (QoR) DRC and LVS should be ran at every step and for every view

Executive summary

The MOSS prototype tries to answer if a stitched monolithic particle detector is possible

- Yield improvements were presented
- Power and data inter stitching connections were presented
- Methodology was presented

The MOSS Prototype will be submitted in Q4 2022 with ER1 in the WP1.2 framework

- ER1 wafer being currently assembled at CERN
- MOSS is passing DRC/LVS signoff checks
- MOSS is passing UVM signoff regression
- MOSS power integrity being revised using Voltus

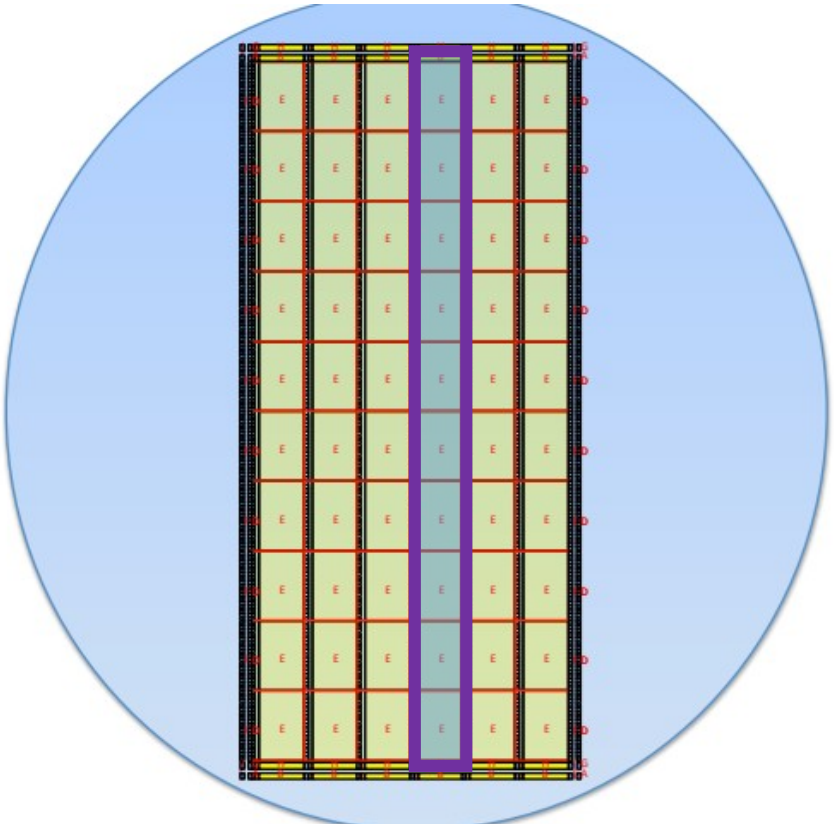
After the MOSS submission there will be an extensive testing campaign

A MOSS successor is being planned to start after submission where it will have to meet more requirements towards ALICE ITS3

- We plan to take all the lessons from MOSS and ER1 and improve on
  - yield, power distribution, stitching, data rate, power consumption, methodology, …

Reference material

ER1 submission



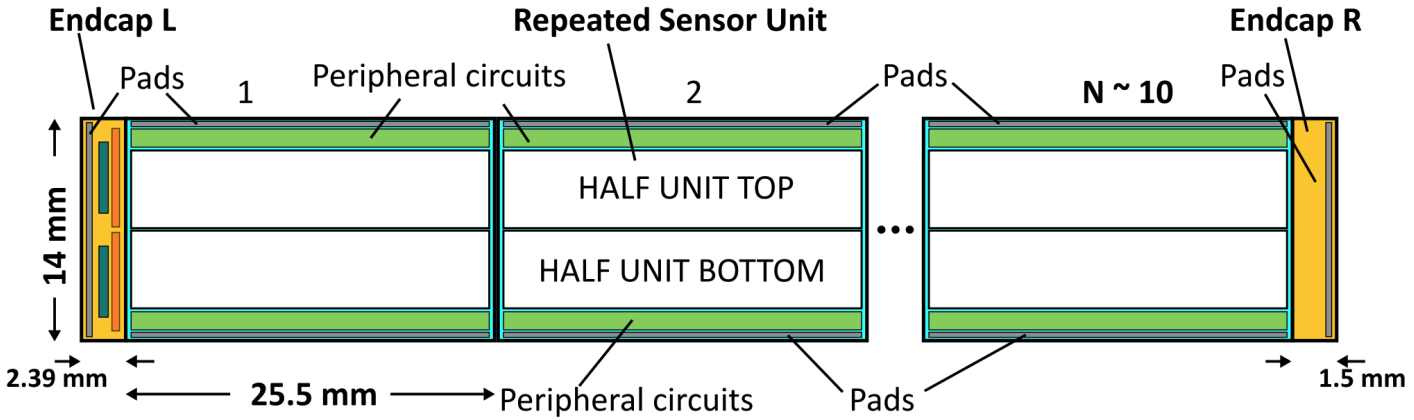
Submission planned by Q4 2022

6 x 10 central fields: **MOSS** and MOST chips
17 test sites 1.5 mm x 1.5 mm (60 x per wafer)

6 top and 6 bottom fields
Endcaps MOSS and MOST chips

10 left and 10 right fields
~ 40 test sites 1.5 mm x 1.5 mm (20 x per wafer)

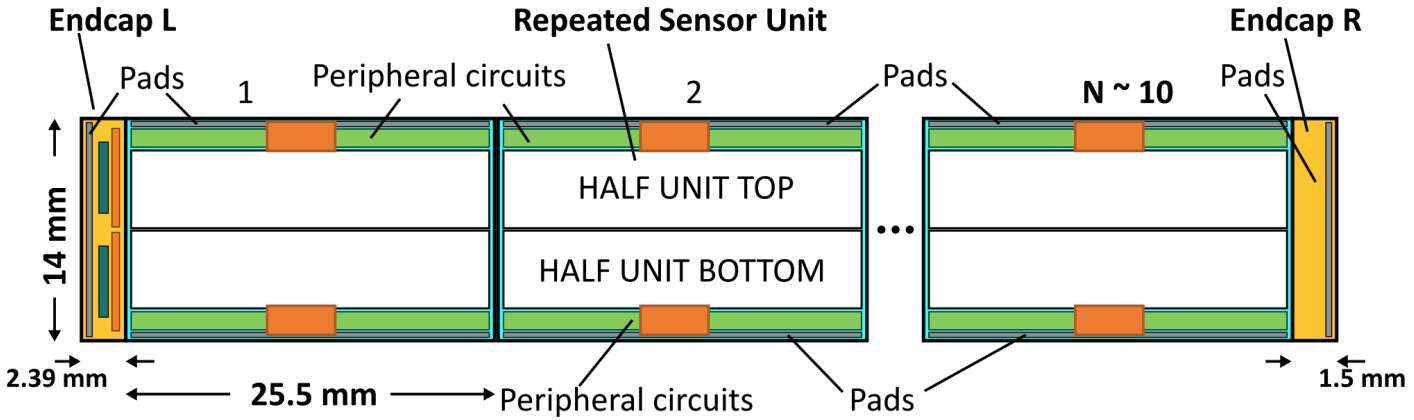
MOSS Prototype architecture



MOSS (1.4 cm x 26 cm) is composed of “3 units”:

- Left endcap (“left short-edge”)
  - Contains power and data pads with minimal digital logic
- Repeated sensor unit (RSU), replicated 10 times
  - Each RSU is subdivided in “half unit top” and “half unit bottom”
  - “half unit top” features 22.5 um pixel pitch
  - “half unit bottom” features 18.0 um pixel pitch
  - “Top” and “bottom” are physically independent and even have their own independent analog & digital power supplies
  - Can be controlled by the “long-edge”
- Right endcap (“right short-edge”)
  - Contains only power pads and no digital logic

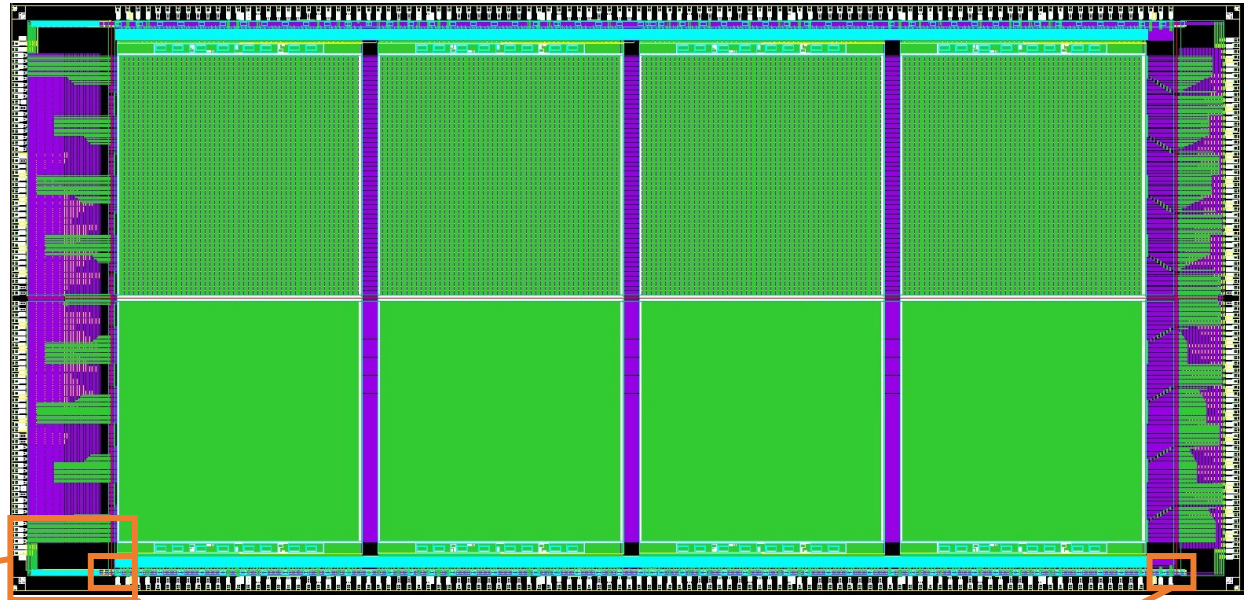
MOSS Prototype architecture

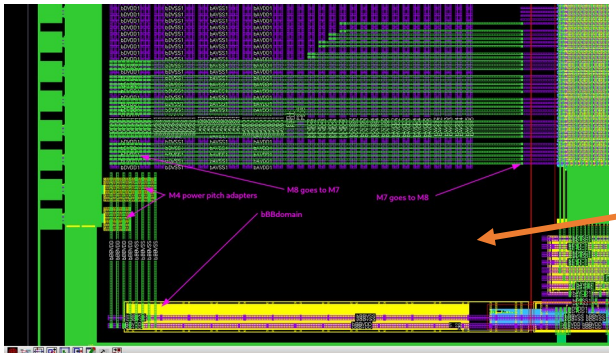


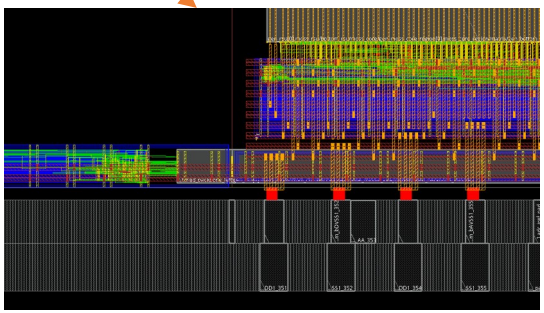
The stitched backbone (SBB) is a set of transmissions lines:

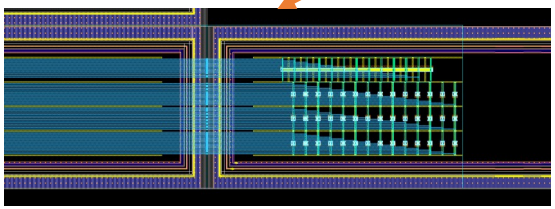
- The stitched backbone has its own independent VDD/VSS
- Every “half unit” can be disconnected electrically from the stitched backbone
- This prototypes
  - data communication across power domains
  - data communication across stitching boundaries (inter chip)
  - long range communication (26 cm)

MOSS Implementation details

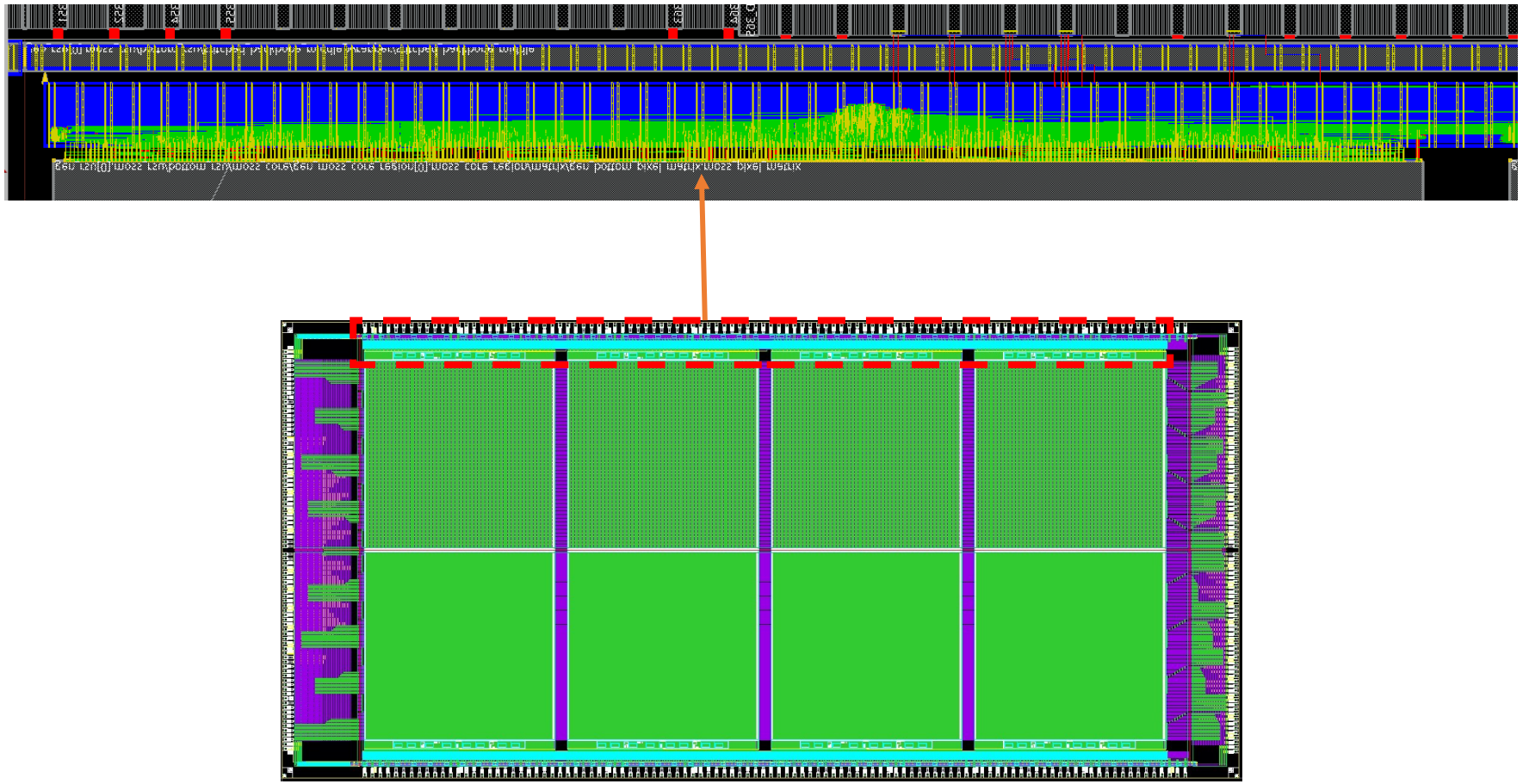




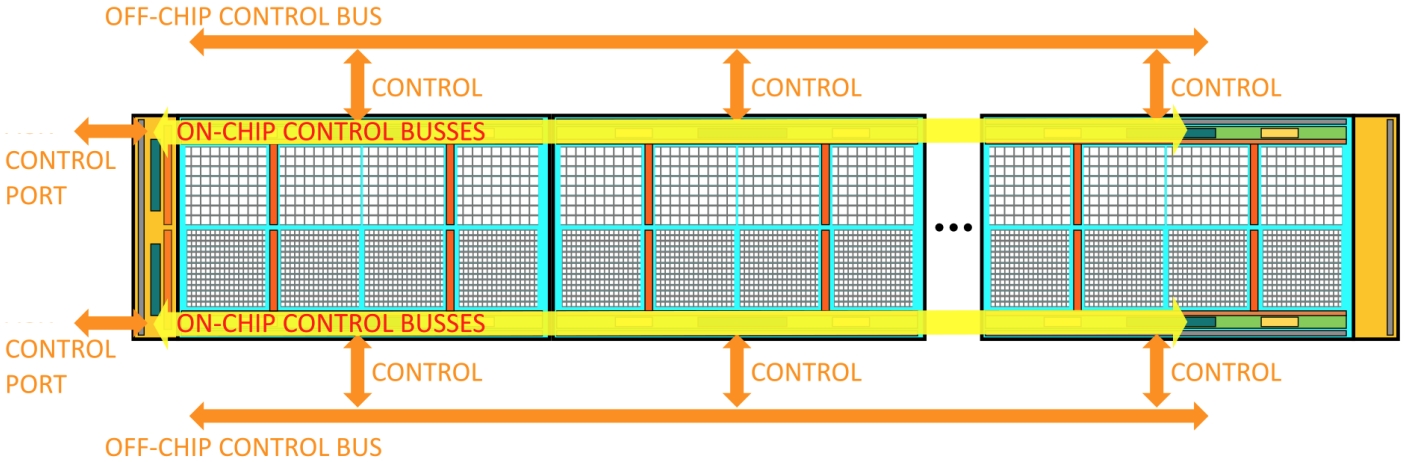




MOSS Implementation details



MOSS Architecture - slow control



The “stitched backbone” is responsible to forward the on-chip control bus and readout data

- CMOS signaling with regeneration across 26 cm
- Only to be used from the left endcap, capable of addressing the 10 sensors

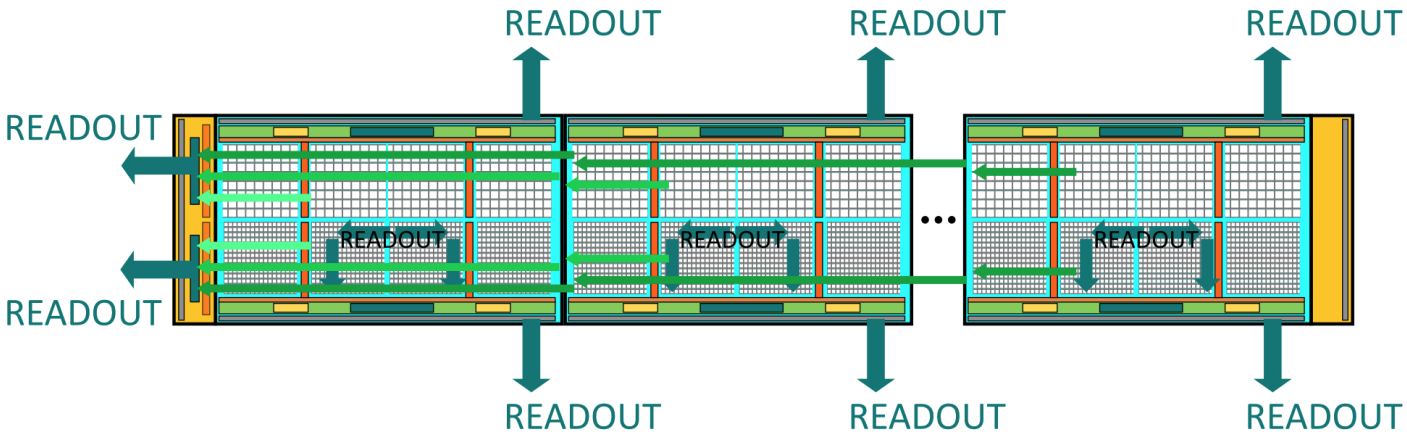
Detaching the sensor will allow full control from the peripheral pads

Off-chip communication is serial using 8b10b encoding at 40 MHz/40 Mbps with custom protocol

- The special 8b10b words are used for synchronization and fast commands (eg, STROBE, RESET, REG_READ, REG_WRITE, PULSE, etc)

On-chip communication is a simplified wishbone

MOSS Architecture - data readout



Data readout can have several modes

From the peripheral pads:

- 8b @ 40 Mbps

From the left endcap, by means of the stitched-backbone (only has 4 readout lines per sensor):

- 4b @ 40 Mbps / sensor, reading one sensor at a time (10 readout pads on left endcap)
- 2b @ 40 Mbps / sensor, reading one sensor at a time (10 readout pads on left endcap)
- 1b @ 40 Mbps while reading all sensors at once (10 readout pads on left endcap)

MOSS Architecture - data readout frame format

	Length	Binary code
IDLE*	8 bits	**1111_1111**
UNIT_FRAME_HEADER	8 bits	**1101_**<unit_id[3:0]>
UNIT_FRAME_TRAILER	8 bits	**1110_0000**
REGION_HEADER	8 bits	**1100_00_**<region_id[1:0]>
DATA	8 bits	**00_**<hit_row_position[8:3]>
DATA	8 bits	**01_**<hit_row_position[2:0]>_<hit_cln_position<8:6>
DATA	8 bits	**10_**<hit_cln_position<5:0>

A readout packet is comprised of header, data and trailer

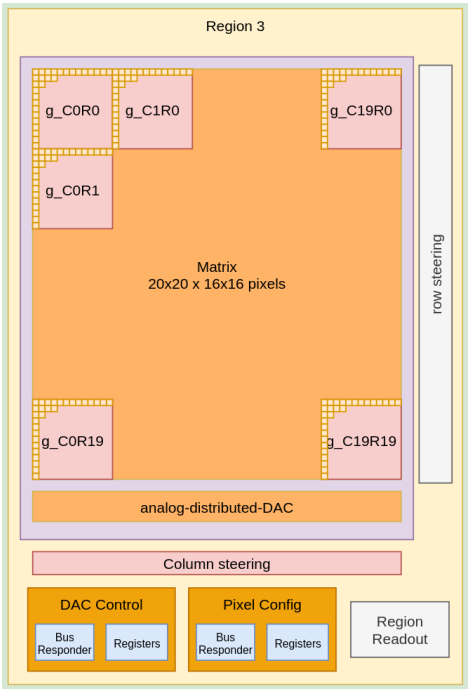
Each word is 8 bits

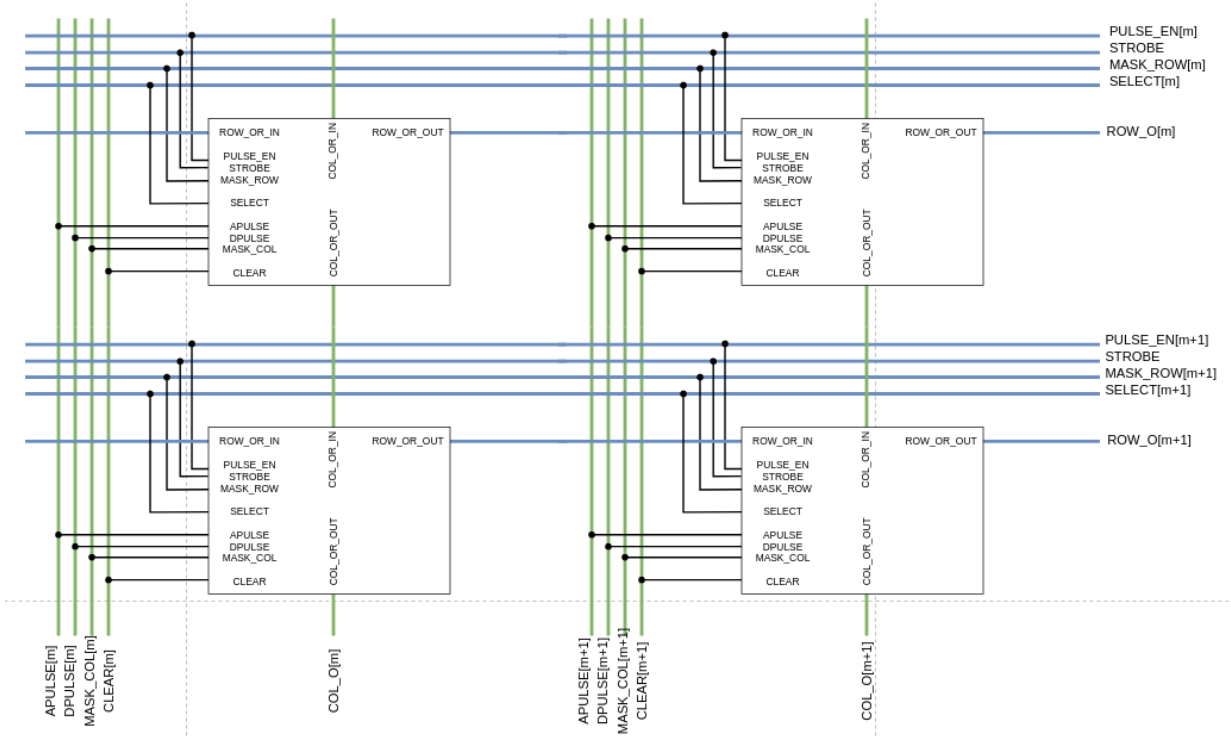
The peripheral pads have access to the whole 8 bit word

The stitched backbone, due to routing and power resources, only has access to 4 bits per sensor

Redundancy is added in the eventuality of one of the readout lines in the stitched backbone is broken

MOSS bottom half region and pixel logic



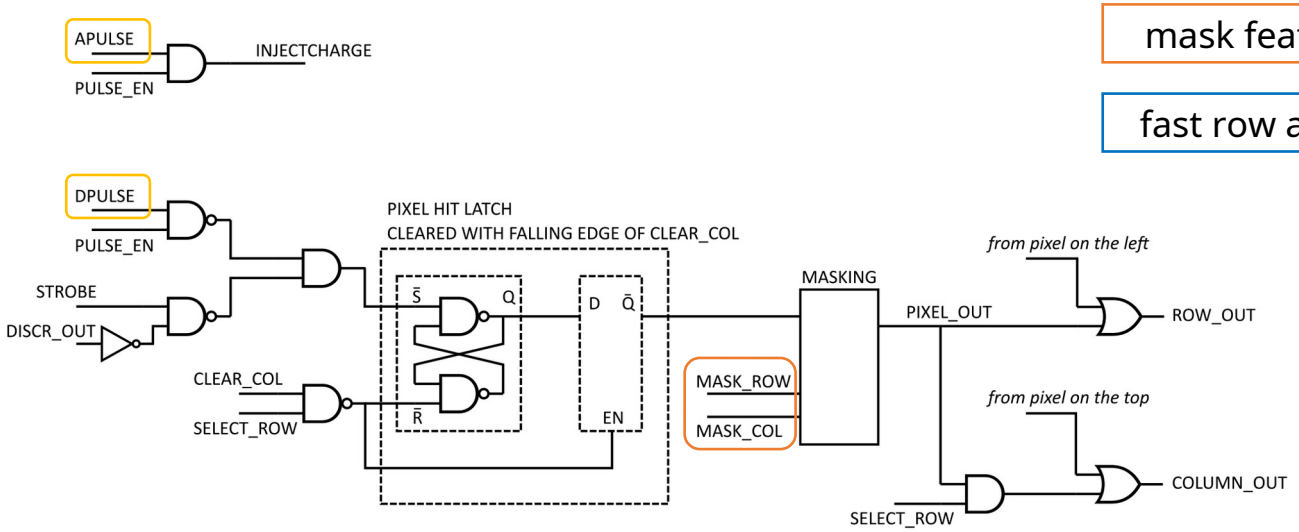


MOSS bottom half region and pixel logic

test features

mask features

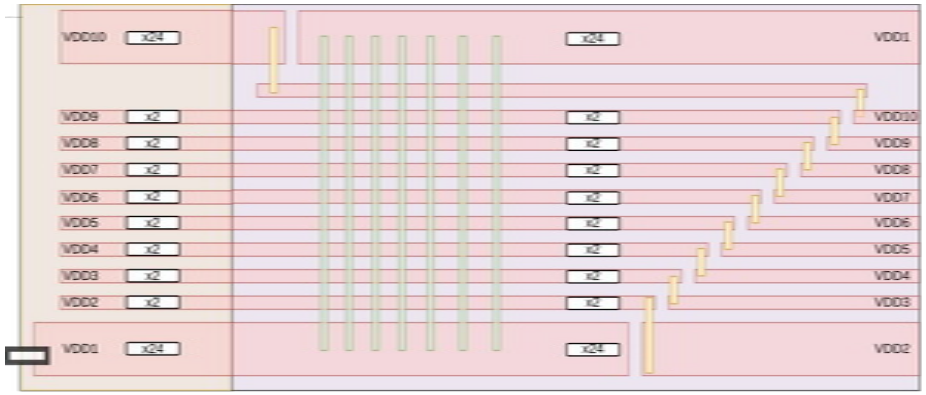
fast row and column OR



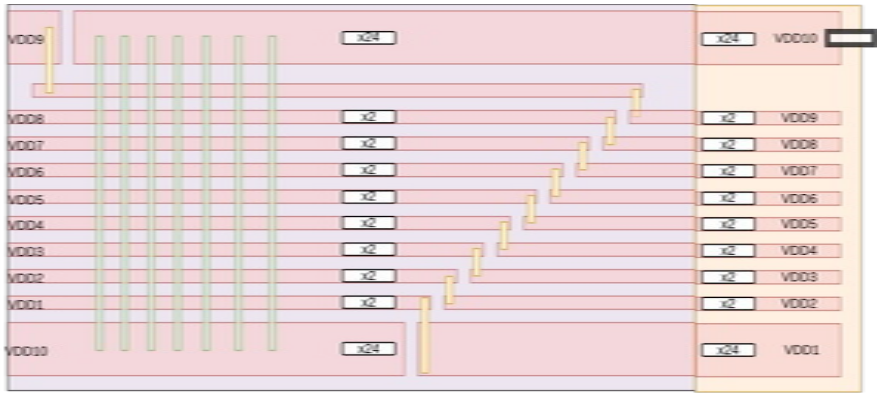
MOSS Power distribution

For the MOSS prototype, the current metal stack only allows to power the most left (RSU1) or most right RSU (RSU10)

All the other power pads are only used for *monitoring* in order to verify the stitching



LEC RSU1



RSU10 REC

Additional power pads are placed on the long edge to allow to test the other RSUs

The same power scheme is used for the stitched backbone

MOSS SBB RSU addressing

Stitched backbone RSU addressing

- Each RSU is address is set by means of line-hopping in the stitched backbone
- A local tie-high cell ensures the address increment

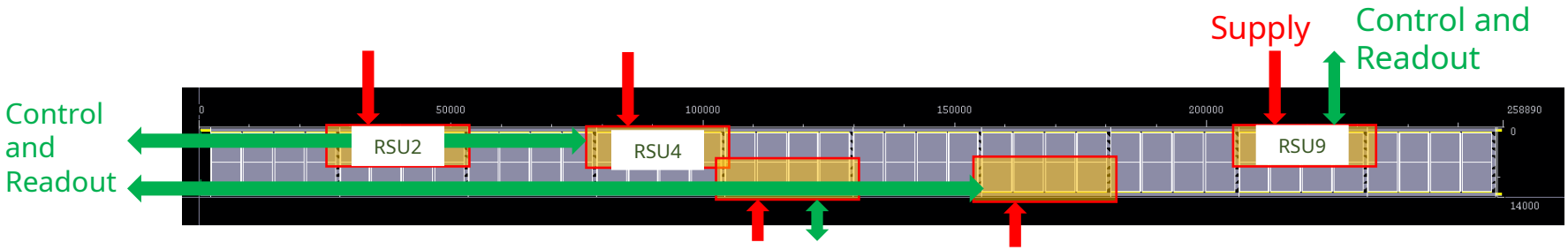
MOSS Testing Scenarios - Examples

Possible to test half units independently

- Top (example)
  - RSU2, RSU4 are connected to the stitched backbone
  - RSU9 is electrically disconnected from the stitched backbone. It is controlled by means of the long edge

This allows to

- Test half units independently
- Check functional yield at block, column/row/pixel granularity
- Study noise, threshold, pixel performance
- Study stitching interconnects (power and data communication)



Contributors

Contributing institutes (in no particular order)

CERN
IPHC
INFN, Bari
INFN, Cagliari
Nikhef
STFC (RAL)
Yonsei University
Central China Normal University