Serial powering and high hit rate efficiency measurement for the Phase 2 Upgrade of the CMS Pixel Detector

Joint Annual Meeting of SPS and ÖPG 2019

Malte Backhaus, Vasilije Perovic, Branislav Ristic, Daniele Ruini, Rainer Wallny
Outline

1. Introduction – serial vs parallel powering
2. Shunt-LDO regulator for the RD53A chip
3. Serially powered chains of RD53A chips and modules
4. Readout efficiency in high hit rate environment
5. Summary
High Luminosity collider starting in 2026

Luminosity increase $\times 5$ means:

- up to 200 proton-proton collisions,
- $\mathcal{O}(10^4)$ particles produced,

every 25 ns.
High Luminosity collider starting in 2026

Luminosity increase $\times 5$ means:

- up to 200 proton-proton collisions,
- $O(10^4)$ particles produced,

every 25 ns.

Detector requirements driving power needs:

- High granularity to limit occupancy to $O(1\%\%) \rightarrow$ many pixels.
- Compliance with 12.5 $\mu$s trigger latency $\rightarrow$ lots of buffering.
  $\rightarrow$ smaller feature size technology (65 nm CMOS).
  $\rightarrow$ requires $\sim 8 \mu$A/pixel, $\sim 150k$ pixels/chip $\Rightarrow \sim 1.5 - 2$ A/chip.
RD53A: a large format IC prototype for Phase 2

- Designed by RD53 collaboration.
- Realized in 65 nm CMOS technology.
- Operates at 1 A and 1.4 V.
- Matrix of $192 \times 400$ pixels of $50 \times 50 \mu m^2$.
- Final chip will be about double in size.
Powering the pixel detector

A single chip works at $V_0$, $I_0$.

Parallel powering, power loss in cables:

$$P_{\text{cable}} = R_{\text{cable}} \cdot (nI_0)^2 \propto n^2.$$
Powering the pixel detector

A single chip works at $V_0, I_0$.

Parallel powering, power loss in cables:
$$P_{\text{cable}} = R_{\text{cable}} \cdot (n I_0)^2 \propto n^2.$$

Serial powering, power loss in cables:
$$P_{\text{cable}} = R_{\text{cable}} \cdot I_0^2.$$

The CMS Pixel Detector will consist of 13488 chips $\rightarrow$ serial powering.
Serial vs parallel powering

Advantages

• Huge improvement in power efficiency compared to parallel powering.
• Constant current consumption: avoids possibly lethal voltage transients.
• Only technically possible option to deliver power to the detector with an acceptable cable mass.

Disadvantages

• Increased fragility: failure of one element could compromise the entire chain.
• Increased chip complexity: an on-chip regulator is required to convert the input current into a constant input voltage.
Serial powering in CMS

- In CMS the cables will be \( \sim 100 \, \text{m} \) long.

- CMS plans to build serially powered chains of up to 11 modules.

- One module consists of two or four chips connected in parallel.

- In case of one (possibly two) chips failing, the remaining chips can shunt the extra current and keep working.
1 Introduction – serial vs parallel powering

2 Shunt-LDO regulator for the RD53A chip

3 Serially powered chains of RD53A chips and modules

4 Readout efficiency in high hit rate environment

5 Summary
Shunt-Low-Dropout (SLDO) regulator

- Provides constant voltage to the load (VDD).
- Shunts excess current in transistor M4.
- Configurable effective resistance and offset.
- Two SLDOs are integrated in RD53A for the analog and digital power domains.
Shunt-Low-Dropout (SLDO) regulator

- Provides constant voltage to the load (VDD).
- Shunts excess current in transistor M4.
- Configurable effective resistance and offset.
- Two SLDOs are integrated in RD53A for the analog and digital power domains.
Shunt-Low-Dropout (SLDO) regulator

- Provides constant voltage to the load (VDD).
- Shunts excess current in transistor M4.
- Configurable effective resistance and offset.
- Two SLDOs are integrated in RD53A for the analog and digital power domains.
Shunt-Low-Dropout (SLDO) regulator

• Provides constant voltage to the load (VDD).

• Shunts excess current in transistor M4.

• Configurable effective resistance and offset.

• Two SLDOs are integrated in RD53A for the analog and digital power domains.
Shunt-Low-Dropout (SLDO) regulator

- Provides constant voltage to the load (VDD).
- Shunts excess current in transistor M4.
- Configurable effective resistance and offset.
- Two SLDOs are integrated in RD53A for the analog and digital power domains.

\[ V_{\text{in}} = V_{\text{ofs}} + I_{\text{in}} R_3 \]

\[ V_{\text{out}} = 2V_{\text{ref}} \]
1 Introduction – serial vs parallel powering

2 Shunt-LDO regulator for the RD53A chip

3 Serially powered chains of RD53A chips and modules

4 Readout efficiency in high hit rate environment

5 Summary
Serially powered chains…

…of RD53A single chip cards:

…of RD53A modules:
Serially powered Shunt-LDOs

Single chip cards

- All chips and modules show the expected resistive behaviour.
- The total input voltage of the chain is the sum of the input voltages of the individual chips/modules.
Noise in a serially powered chain

Single chip cards

Chips on one module

- Distribution of pixel-by-pixel noise difference between operation of a single chip/module alone and in the chain.
- Operation in chain has no influence on the noise.
1. Introduction – serial vs parallel powering

2. Shunt-LDO regulator for the RD53A chip

3. Serially powered chains of RD53A chips and modules

4. Readout efficiency in high hit rate environment

5. Summary
Performance under high hit rate – X-ray setup at ETH
Performance under high hit rate – X-ray setup at ETH
Efficiency vs trigger latency – shunt mode

- Fast drop in efficiency caused by on-chip memory overflow.
- Memory buffers are optimized for clustered hits → isolated hits from photons are worst case.
Efficiency measurement vs simulation

• Simulation of random single hits (emulating X-ray photons) with 12.5 µs latency (thanks to Sara Marconi).
• Good agreement between measurement and simulation.
Efficiency vs trigger latency – powering modes

- Powering mode has no influence on the efficiency.
Summary and outlook

• Serial powering is the power distribution system for the Phase 2 Upgrade of the CMS pixel detector.
  – Affordable cable mass and power efficiency.
  – Tests show no impact on performance compared to single chip (parallel) operation.

• RD53A prototype chip is available and being tested.
  – 65 nm CMOS technology.
  – Integrates two shunt-LDOs for voltage and current regulation.
  – High hit rate efficiency measurements agree with simulation.

• Future:
  – Development of next chip version has started.
  – Further aspects (HV distribution, current sharing between parallel chips, . . . ) under study, showing promising results.
Backup
Analog front ends: three flavours for testing

• **Synchronous FE**: uses a baseline “auto-zeroing” instead of per-pixel threshold trimming.

• **Linear FE**: linear pulse amplification and comparison to a threshold voltage ← *results focused on this one.*

• **Differential FE**: differential gain stage in front of discriminator. Threshold is implemented by unbalancing the two branches.
Synchronous front end design

Charge Sensitive Amplifier with Krummenacher feedback

AC couplings to compensate mismatches between pixels

“Auto-zeroing” mechanism (no trim DAC): a baseline is acquired every 100 µs (during LHC abort gaps) and subtracted from signal.
Linear front end design

Signal compared to threshold voltage

Threshold tuning via 4-bit DAC

Charge Sensitive Amplifier with Krummenacher feedback
Differential front end design

Global threshold adjusted by \( V_{\text{thin}1} \) and \( V_{\text{thin}2} \), each trimmable by a 4-bit DAC

Classic continuous time comparator with output connected to pixel digital region

per pixel
global
Analog hit processing

- The signal from the sensor (bump pad) is amplified and compared to a threshold voltage.
- Signal is digitized counting the Time over Threshold (ToT).
Analog hit processing – sources of inefficiencies

- Two hits within few clock cycles are registered as one big hit with large Time over Threshold.
- Hit losses due to this effect scale linearly with hit rate.
Sources of inefficiencies – buffer architecture

- The pixel matrix is divided into *regions* of four pixels with shared timestamp buffer, optimized for clustered hits.
- Buffer overflow causes data loss in the pixel region.
Efficiency measurement procedure

1. Tune the chip to achieve a uniform threshold.

2. Place chip in X-ray beam.

3. Inject test signals in a few selected pixels and read out the full matrix (only linear FE for now).
   - Low trigger frequency (100 kHz) to avoid readout limitations.
   - Read out only the bunch crossing in which the signal was injected.

Results:

- Efficiency computed from number of signals recorded in selected pixel(s).
- X-ray hit rate computed from average occupancy of other pixels in the same pixel region (i.e. sharing the memory logic).
Efficiency simulation – realistic clusters

- Simulations of clustered hits are much closer to 99%.
- Further improvement possible by counting the ToT at 80 MHz (both clock edges).
SLDO X-ray irradiation to 6 MGy at CERN

- Performance of a 65 nm, 2 A SLDO prototype was tested for radiation hardness up to 6 MGy in April 2017.
- One chip irradiated at room temperature, one at \(-10^\circ\)C.
- Only effect: very small changes in $V_{in}$ and $V_{out}$.
SLDO irradiation to 6 MGy at CERN: results
SLDO irradiation to 6 MGy at CERN: results

- Cold: Vref = 600 mV, Vofs = 800 mV, Iin = 1.05 mA, Iload = 0 A
- Room: Vref = 600 mV, Vofs = 800 mV, Iin = 0.86 mA, Iload = 0.69 A

Output voltage:
- Cold: Vout ∼ 6 mV
- Room: Vout ∼ 8 mV
Requirements for the Phase 2 CMS tracker
as specified in the Technical Design Report

• Radiation hardness up to 10 MGy and $2.3 \times 10^{16} \text{n}_{\text{eq}}/\text{cm}^2$ for an integrated luminosity of 3000 fb$^{-1}$.
• Occupancy < 1‰ for 200 collisions per bunch crossing.
• Efficiency > 99% up to 3 GHz/cm$^2$ hit rate.
• Long trigger latency (12.5 µs), in order to use tracking information in the Level 1 trigger.
• Improved two-track separation.
• Low material budget.
• Tracking acceptance up to $|\eta| = 4$