Data volume estimations

Irakli Mandjavidze

Irfu, CEA Saclay
Gif-sur-Yvette, 91191
France

MPGD-DSC
7 / May / 2026

Outlook

- MPGD data volumes need to be reevaluated per sub-system
  - → Important for FEB design
    - Number of active Salsa TX links and required lpGBT devices to be verified
  - → This is important to determine the number of RDO and/or DAM units
- The background needs to taken into account
  - → On-going within the MPGD “physics” groups
- In the current baseline FEB, Salsa sends data to lpGBT over a single 1.24 Gbit/s link
- Here the limits of channel hit rates and data volumes are estimated allowing stable operation
  - → Within the constraint of a single Salsa-lpGBT link
  - → If the physics simulations will not fit within the limits, the FEB design shall be revisited
- RDO and DAM data throughput is also estimated

FEB with 4 Salsas, an lpGBT an a VTRX+

- 4 64-channel Salsas per FE
  - → 256 channels
- 4 lpGBT eLink groups
  - → For data, clock, synchro
  - → An eLink group / Salsa
  - → 1.28 Gb/s data link / Salsa
    - 64 channels per data link
- I2C chain for configuration
- lpGBT GPIO and analog peripherals
  - → Control / monitor IO pins of Salsa
    - Reset, PLL lock, error, etc.
  - → Control / monitor on board resources
    - ADC, DAC

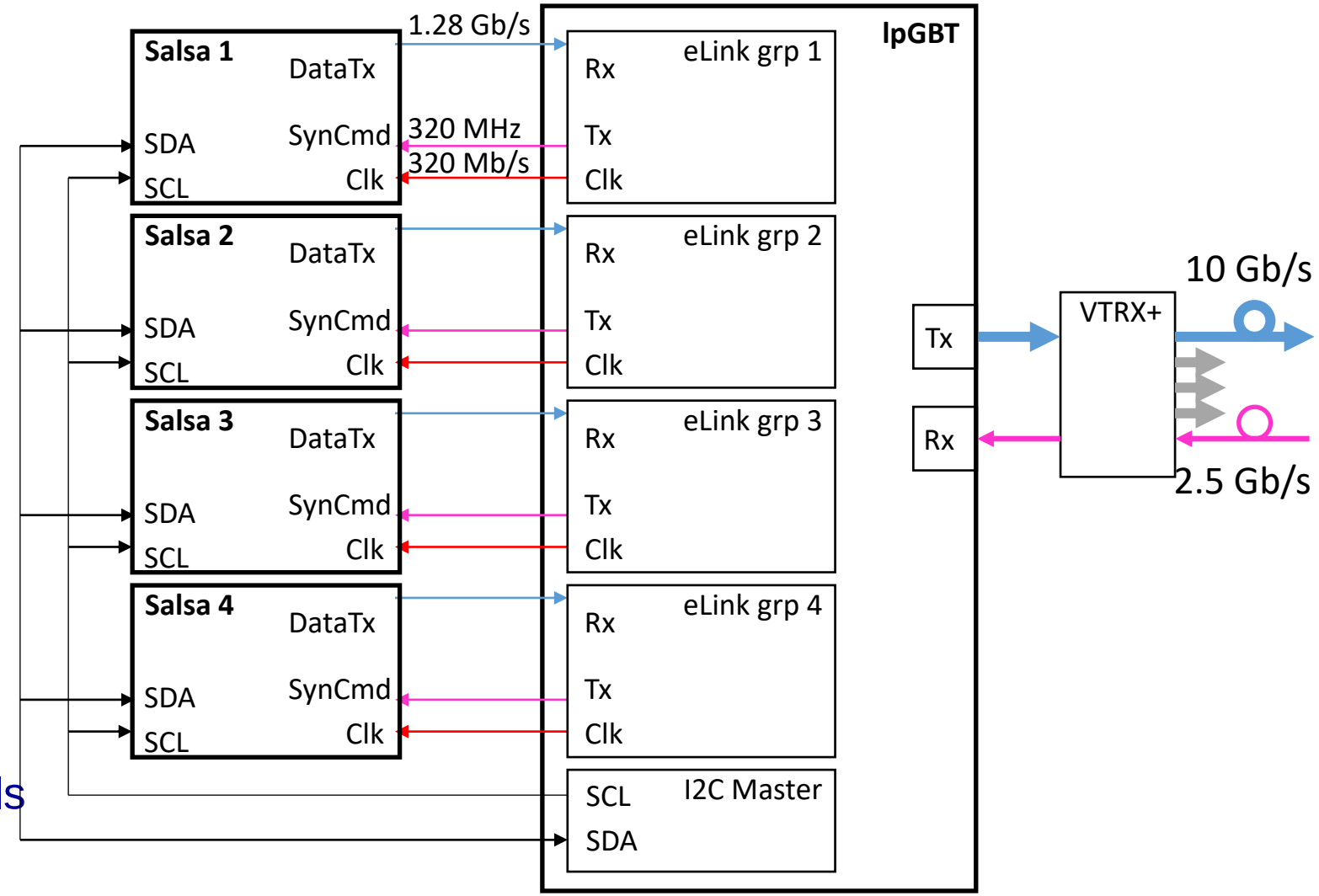
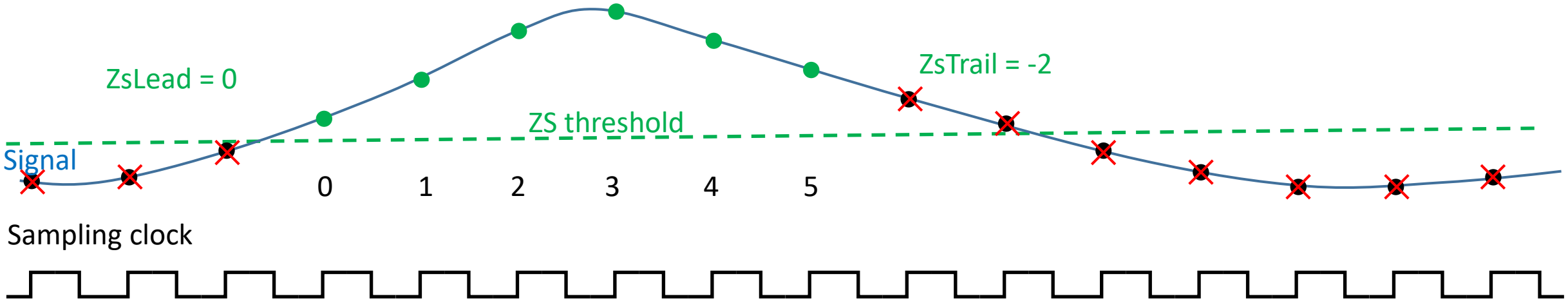


Illustration of a channel data packet : signal shape readout



Bits	31							24	23						16	15				11			8	7							1	0
Header	P	1	1	Type of packet 3b			Chip address 3b		Channel number 6b						Time frame number 4b					Sample 0 timestamp 13b (within time frame)													
Sample word	P	0	0	Flags 5b					Value of sample 1 12b											Value of sample 0 12b													
Sample word	P	0	0	Flags 5b					Value of sample 3 12b											Value of sample 2 12b													
Sample word	P	0	0	Flags 5b					Value of sample 5 12b											Value of sample 4 12b													
Trailer 1	P	1	0	Size of packet in number of words 6b						Infos 3b			Checksum of packet 8b							Link ID 3b			Packet number in link 9b									
Trailer 2	!P	1	0	Vertical parity word 29b																													

- Channels are “independent” : for a hit cluster of N channels N packets will be produced and sent

Inputs to estimate Salsa link bandwidth

- Number of channels per Salsa : 64
- Sampling rate : 40 MHz
- Peaking time : 200 ns
  → Signal shape 500 ns
- Signal shape readout
  → Number of retained samples can be set to
    - Either full signal shape : time over threshold + few samples below threshold
      - 20 signal samples in average + 4 under threshold samples
    - Or a fixed number of samples
      - 8, 10, 16, …
- Two 12-bit samples are packed in a 32-bit sample word
- A packet is formed of sample words + 3 32-bit overhead words
  → Time stamp header, link and parity trailers
- A packet is sent over 1.28 Gbit/s link with no encoding overhead
- At least 2 32-bit idle words are sent after every packet to maintain synchronization

Salsa to lpGBT 1.28 Gbit/s data link occupancy

- Determine channel hit rate resulting to 50% and 70% occupancies of the link

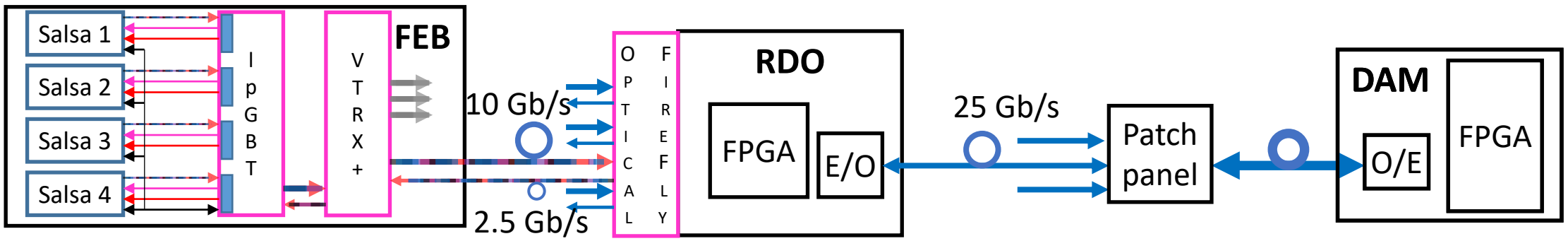
Channel rate limits as a number of retained samples per hit and link occupancy

Load %	8-samples kHz	10-samples kHz	16-samples kHz	ToT (20+4)-samples kHz
50	35	31	24	19
70	49	44	34	26

- Link occupancies above 60-70% may result to unstable operation due of the Poissonian traffic
  → On-line calibration may add large chunks of non-ZS data on top of the “normal” ZS data
  → At high link occupancies absorbing of perturbations may not be possible

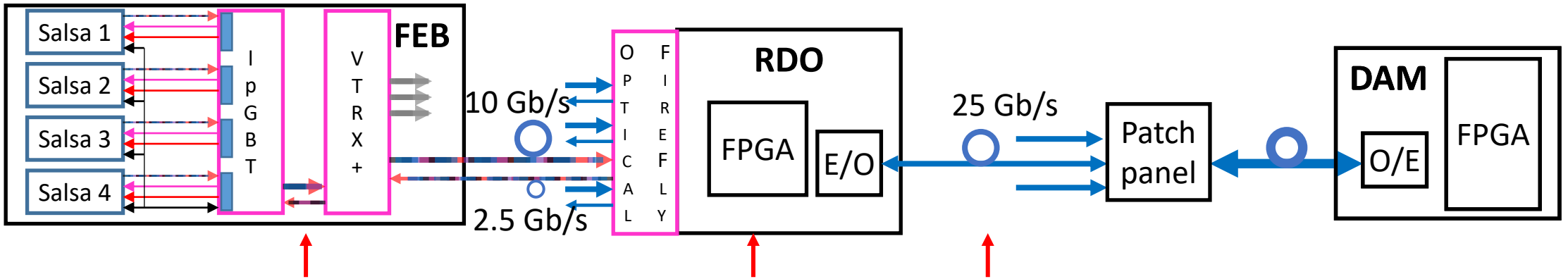
- Once more : we need to check that expected data volumes fit the limits
  → With a reasonable safety factor

RDO data throughout estimation



- Assume 10 samples are retained per channel belonging to a hit cluster
  → 250 ns of signal shape
- Assume RDO aggregates data from 12 FEBs
- Assume RDO is a “transparent” aggregator
  → Low or no data reduction
  - Unless there will be an on-line data reduction e.g. cluster finding
  - And/or some (lossless) data compression implemented in firmware with significant data reduction
    - But I am not aware of this effort – certainly an interesting domain to explore (with some AI algorithms)
- Assume 25 Gbit/s link bandwidth between RDO and DAM

RDO data throughout estimation : time / amplitude readout

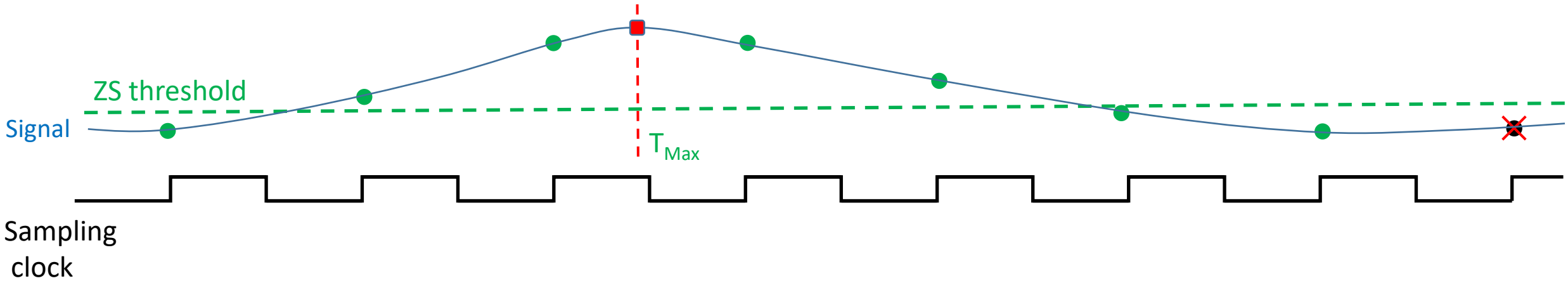


- Assume RDO aggregates data from 12 FEBs
- Assume RDO is a “transparent” aggregator

Channel rate kHz	Salsa-lpGBT link load %	RDO input BW Gb/s	RDO-DAM link load %
10	16	7.8	31
20	32	15.7	63
30	48	23.6	94

- If this assumptions stand :
  - → The RDO must have significant data throughput capability
  - → The RDO to DAM link limit may be hit pretty quickly
  - → Possibility to use less than 12 aggregation factor increasing RDO count and probably DAM count too

Illustration of a channel data packet : time / amplitude readout

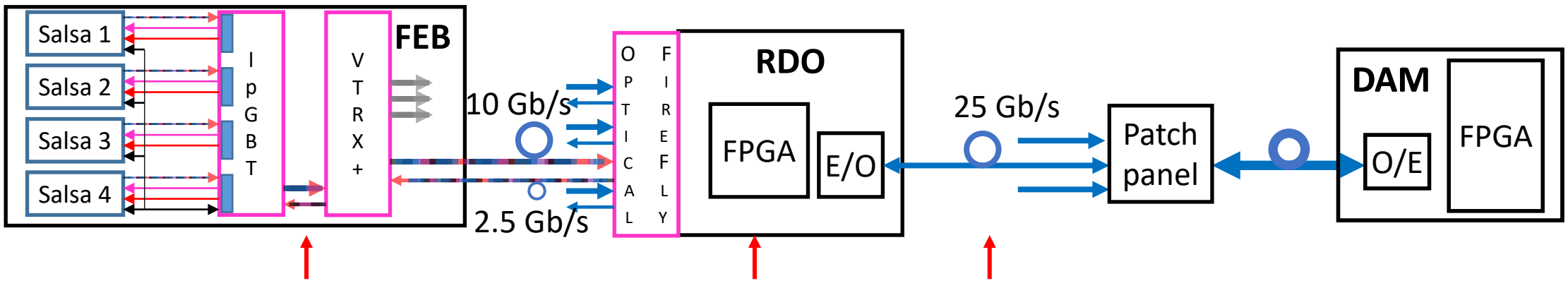


Preliminary, under definition

Bits	31							24	23						16	15				11			8	7						1	0
Header	P	1	1	Type of packet 3b			Chip address 3b			Channel number 6b						Time frame number 4b				Sampling timestamp of the 1st sample in convolution product 13b												
Data word	P	0	0	Signal features including amplitude, time of max or time of arrival, time over threshold																												
Trailer 1	P	1	0	Size of packet in number of words 6b						Infos 3b			Checksum of packet 8b							Link ID 3b			Packet number in link 9b								
Trailer 2	!P	1	0	Vertical parity word 29b																												

- Channels are “independent” : for a hit cluster of N channels N packets will be produced and sent

RDO data throughout estimation : time / amplitude readout



- Assume RDO aggregates data from 12 FEBs
- Assume RDO is a “transparent” aggregator

Channel rate kHz	Salsa-lpGBT link load %	RDO input BW Gb/s	RDO-DAM link load %
10	9	3.4	13
20	18	6.9	27
30	26	10	41

- Comfortable

  → But time / amplitude extraction algorithm in Salsa yet to be validated for all 3 MPGDs

DAM data throughout estimation

- FEB-RDO-DAM : 36K channels per DAM
  → Single DAM for entire CyMBaL
    - No partitioning, in case of DAM failure the whole CyMBaL is out

- FEB-DAM : 9K channels per DAM
  → 4 DAMs per CyMBaL

	Signal shape readout		Time / amplitude readout	
Channel rate kHz	FEB-RDO-DAM Gbit/s	FEB-DAM Gbit/s	FEB-RDO-DAM Gbit/s	FEB-DAM Gbit/s
10	94	24	41	10
20	189	47	82	21
30	283	70	123	31

  - Channel occupancies are not uniform – presented cases have to be considered as high end

- RDO-DAM architecture requires significant data processing capabilities from DAM
  → Either prove the ability to handle the data
  → Or distribute processing adding DAMs
  → Or delegate some data processing to RDOs

Summary

- Need firm figures on expected particle hit rates and cluster sizes
  → Including physics and background

- Validate that single 1.28 Gbit/s link between Salsa and lpGBT is adequate
  → Otherwise rethink FEB design

- Study data reduction algorithms and their on-line implementation in firmware
  → Time / amplitude extraction from signal shape
    - To be used in Salsa or in RDO or in DAM
  → Cluster finding and its feature extraction
    - To be used either in RDO or in DAM
  → Data compression algorithms between RDO and DAM if data throughput exceed a reasonable link load

- Validate factor 12 aggregation feasibility in RDO
- Decide if a single DAM per 36K-channel sub-detector is righteous approach and validate its feasibility

- DAM quantities for MPGDs should allow
  → Independent operations (in home institutes) during the development phase : 3 systems
  → A functional MPGD test bench during experience life-time