Electronics Architecture of the LHCb Upgrade

LHCb Technical Note

Issue:
Revision: 2.2

Reference: LHCb-PUB-2011-011
Created: July 2012
Last modified:

Prepared By: Ken Wyllie, Federico Alessio, Richard Jacobsson, Niko Neufeld

Abstract

The electronics architecture of the upgraded LHCb experiment is defined. This covers all data processing from the detectors up to the input of the data acquisition system. Building blocks for implementation are also described.

Document Status Sheet

Table 1 Document Status Sheet

1. Document Title: Electronics architecture of the LHCb upgrade			
2. Document Reference Number: EDMS			
3. Issue	**4. Revision**	**5. Date**	**6. Reason for change**
Draft			
	1.0	March 2011	TFC requirements added (RJ & FA) New GBT format added
	2.0	June 2012	GBT formats updated
	2.1,2.2	July 2012	Addition of new specifications

Table of Contents

1. Introduction

This note describes the electronics architecture adopted for the upgrade of the LHCb experiment. The upgrade is described in a Letter-of-Intent [1] and is based on a highly efficient trigger algorithm run on a farm of CPUs and using data from all bunch-crossings. No hardware trigger is applied to the on-detector electronics, so sub-detectors must transfer the data from each bunch crossing at 40MHz. It is expected that only seventy-five percent of bunch-crossings will contain interactions leading to interesting physics, so a 'low level trigger' (LLT) based on a simple algorithm will be applied to the data at the level of the off-detector electronics in the counting room. This will reduce the event rate to about 30MHz.

The following sections will describe the architecture chosen to meet these general requirements.

2. General Architecture

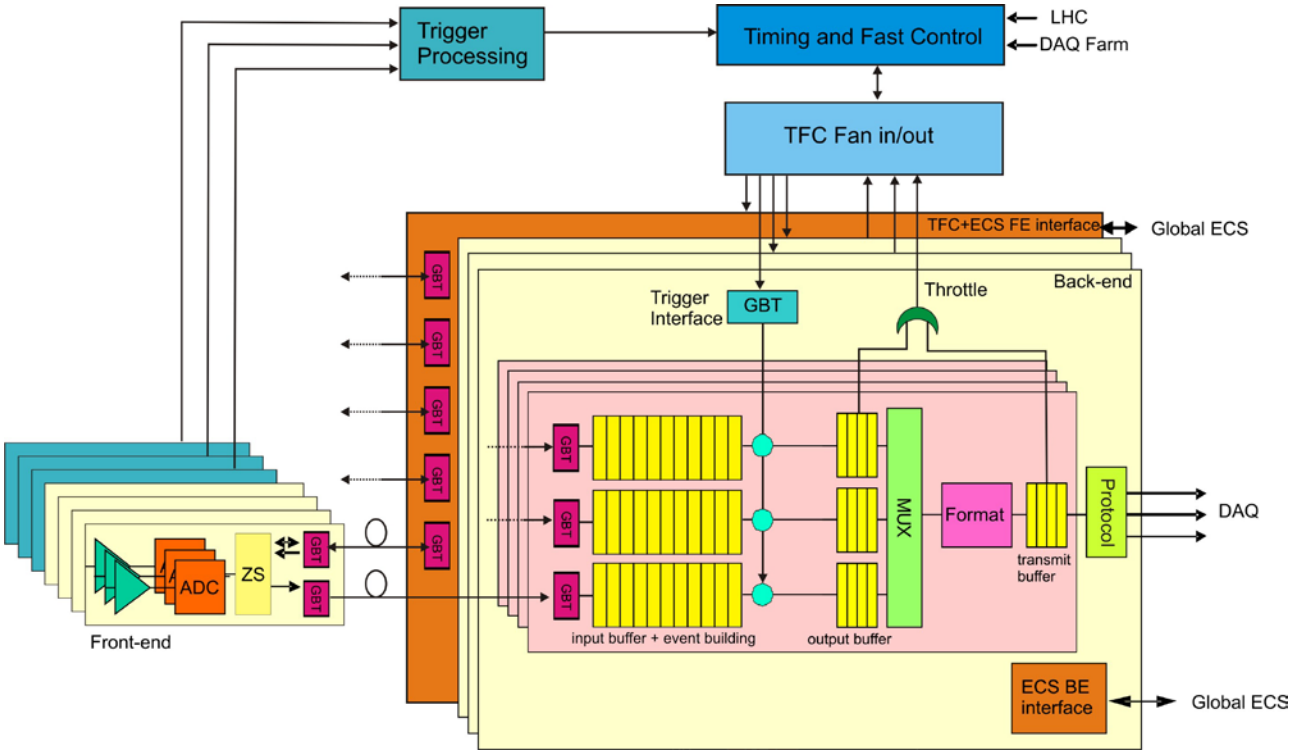

Figure 1: General architecture

The general architecture is shown in Figure 1. The front-end (FE) amplifies and shapes the signals generated within the particle detectors. These signals are digitised, zero-suppressed and then transmitted down a generic high-speed optical link running at a serial rate of 4.8 Gbit/s. All components of the front-end electronics are located on or close to the detector, and are therefore exposed to some level of radiation.

The back-end electronics (BE) sit in the counting room and receive the data from the optical links. After buffering and filtering by the LLT, data are formatted for transmission to the data acquisition system. The counting-room is a radiation-free environment and commercial components can be used for the implementation of the electronics.

Data from certain sub-detectors are extracted via an independent transmission system to the trigger processors to generate the LLT. This is then distributed to the back-end electronics by a Timing and Fast Control (TFC) system. Configuration and monitoring of the BE and FE electronics are through an interface to the Experiment Control System (ECS).

The LHC bunch structure will consist of 3564 bunches spaced by 24.95ns. Regular resets of the system will be issued to guarantee synchronisation whilst also minimising data loss.

[Discussion

Point from Federico: 119-BX gap is not guaranteed, so the front-ends should recover from a reset as quickly as possible.

Each of the components in the global architecture is described in the following sections.

3. Front-end

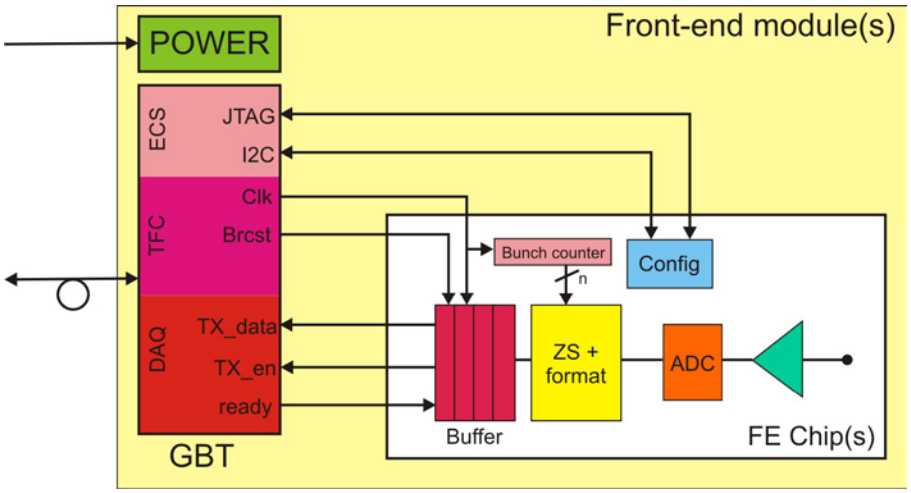

Figure 2: A typical front-end module

A typical FE module is shown in Figure 2. It consists of an FE chip(s) connected to a high-speed bi-directional optical link, implemented using the Gigabit Bidirectional Trigger and Data link (GBT)[2]. The FE chips should be tolerant to the radiation environment of the detector. High radiation areas often require a custom design but commercial devices can be used as long as they are proven tolerant to the expected radiation levels.

The FE chip(s) should amplify, shape and digitise the detector signals according to the requirements of the detector. The rise time of the shaped pulse must be sufficiently short to allow the correct identification of a hit with its bunch crossing while the recovery time must meet the detector specifications on, for example, double-pulse resolution or pile-up. Each sub-detector must choose its digitisation scheme as an optimisation of performance, complexity and cost. For example, the simplest case is a one-bit digitisation implemented using a discriminator. Digitisation at this stage is mandatory to allow high speed digital data transmission.

Digital data are synchronised with the 40MHz clock provided by the GBT. This clock should be phase adjusted to find the optimum efficiency with respect to the bunch collisions within the 25ns clock period. Data are then zero-suppressed, buffered and formatted. The buffer will absorb statistical fluctuations in the zero-suppressed event size and allow an optimal use of the data bandwidth provided by the GBT. However, this implies that data from different FE modules will arrive asynchronously at the BE modules. Additional information is therefore added to the data frames to allow reconstruction of the complete events. The minimum formatting required is the attachment of a header containing all or some bits of the bunch-counter (BXID). This BXID must be generated within the front-end chip synchronously with the clock. These bits are used to track the data later in the system. The data format is discussed in more detail below.

After formatting, data are transferred to the DAQ interface of the GBT chip. Data are then transmitted off the detector through the optical link.

3.1. TFC/ECS interface to FE

Fast, synchronous controls are provided by the GBT link transmitting from the counting room, known as the master-GBT. There is one frame broadcast every 25ns clock period. These commands are broadcasts so every FE device will receive the same data. The broadcast frame is decoded by the FE chip to carry out a number of tasks, the most important of which are the following:

1. generation of a FE reset
2. generation of a calibration pulse
3. generation of the BXID reset

The encoding of the TFC commands is described in Section 6. To allow the local correction of latencies introduced by different link lengths, each FE device must have the ability to delay the TFC commands by up to 16 clock cycles. This will allow the correction of any latency differences between FE devices connected to the same TFC command bus.

This link also provides the local 40MHz clock used by the FE devices. The FE module must have the capability of phase-shifting this clock to find the optimum timing with bunch collisions. One possibility is to use one of the eight 40MHz clock signals from the GBT system whose phases can be adjusted independently. Note that the phase of the TFC command is not adjustable.

Configuration of the FE chip(s) is done through the ECS interface provided by the GBT chipset. The FE chip(s) should support one of the standard protocols provided by the GBT system. To allow the read-back of configuration data, each ECS link must be constructed as a point-to-point link between the GBT system and the FE device.

The bandwidth available in the GBT for transmitting ECS/TFC information is large. So in many cases, it is efficient to use one duplex GBT link to provide ECS/TFC information to many front-end components.

Discussion:

Are 16 clock cycles enough to guarantee complete latency correction? Sub-detectors should answer.

3.2. Zero-suppression

Data are zero-suppressed before transmission off the FE modules. The amount of data and the time to zero-suppress will have some statistical fluctuation based on the detector occupancy in each bunch crossing. A buffer should therefore be implemented to allow for these fluctuations and the FE module should carefully monitor the behaviour of this buffer. If the buffer occupancy reaches its limit, the FE module must truncate the data and allow the buffer to recover. The module continues to send header information but with no data until the buffer is sufficiently empty. The activation of data truncation must be indicated in the header.

The FE module should be capable of masking noisy or faulty channels that could limit the efficiency of a zero suppression algorithm.

The implementation of zero-suppression in the FE electronics implies that the parameters must be well understood before the hardware is designed and constructed. For example, the efficiency of the algorithms and the amount of buffering must be tested carefully using data from realistic physics simulations. Depending on the detector radiation environment, it may be possible to use programmable devices for the zero-suppression, which will allow some flexibility in the implementation. The use of such devices is addressed in Section 11. The important parameters that should be quantified are:

1. Inefficiency = number of truncated bunch-crossing packets/number of bunch-crossings.

2. Maximum size of the data packets. This has a large impact on the buffering implemented in the BE modules.

3. Maximum readout latency. Depending on the algorithms, there could be a wide spread in the latency between the bunch-crossing and the corresponding data packet arriving at the BE module. This also has a large impact on the buffering in the BE modules and on the number of bits of BXID needed to prevent ambiguities in the time-stamping of packets.

3.2.1. Transmitting non-zero-suppressed data

In some regions of some sub-detectors, the hit occupancy can be high enough that there is no benefit in zero-suppressing the data. Transmitting non-zero-suppressed data is allowed in these cases as long as the header protocol is consistent with zero-suppressed data.

Discussion:

Do sub-detectors require NZS data?

Should it be ‘on-the-fly’? ie an NZS event is transmitted amongst ZS events during normal data taking? (this would limit wasting beam time)

For a single BX, do they need both the ZS and NZS data? Do they need to correlate this event with NZS-data from the same event but from a different sub-detector?

OLD TEXT: All sub-detectors must include a special running mode where the front-end module transmits non-zero-suppressed data at a limited rate. This can be used during calibration and commissioning. Such a running mode implies a reconfiguration of the readout system, including the BE and TFC, and hence adequate partitioning to allow a particular sub-detector to run alone in this mode.

3.3. Signal Synchronisation using bunch counter

Overall synchronisation of the system should be done using the BXID information included in the data. Some possible sources of de-synchronisation are the following:

1. In some FE modules, signals from different parts of the detector may arrive at different times after the interaction

2. Counter reset pulses (BXID reset) will have different arrival times from module to module due to cable delays.

To allow synchronisation, a local offset (preset value) of the BXID counter should be implemented. This offset is loaded into the counter on the arrival of a BXID reset and its value should be programmable via the ECS interface.

3.4. Data framing

Data are transmitted in a format containing a header of fixed length followed by the bunch crossing data of varying length. This is illustrated by the example in Figure 3, which shows the transmission of a number of data packets of varying sizes and the use of the front-end buffer.

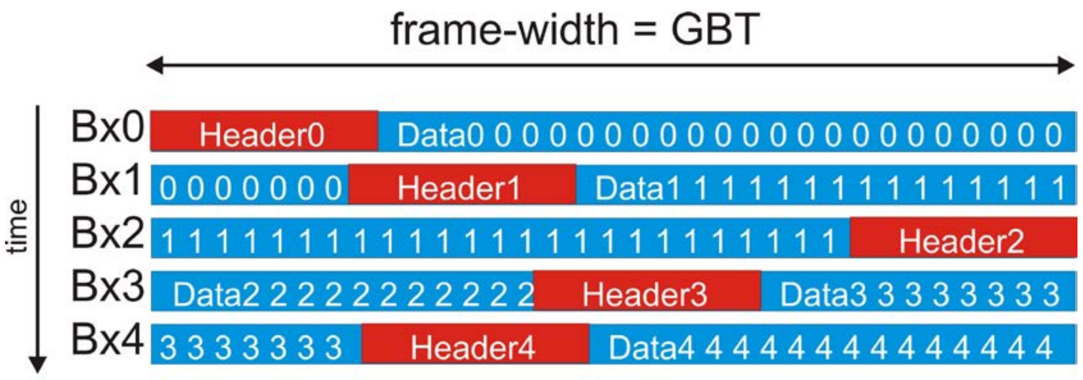

Figure 3: Example of data frame structure

In this example, the width of the frame is limited by the bandwidth of the GBT. The front-end buffer allows the transmission of data packets longer than the GBT frame. However, buffer overflow will occur if a number of consecutive bunch crossings exceed this limit. At this point, headers continue to be sent but transmission of data frames should stop until the buffer occupancy falls below the limit.

3.4.1. Header definition

The header contains a fixed number of bits and these are defined in Table 2.

# Bits	Definition
12	BXID
1	Truncation Status: 0 = normal, 1 = data truncated
n	Length of data packet

Table 2: Header definition

The parameter ***n*** is defined by the sub-detector according to the maximum length of data it expects to transmit per bunch crossing. This field can also be used to indicate if the event data is non-zero-suppressed.

If truncation has occurred then the truncation bit is asserted.

In certain cases, a reduced number of BXID bits can be transmitted if the full 12 bits use a significant fraction of the data bandwidth. However, in this case it should be guaranteed by the sub-detector that this will not lead to ambiguities in the identification of the data with the absolute bunch-crossing. The remaining bits of the BXID must then be added back into the data header by the BE module before the LLT is applied.

The back-end electronics will use the contents of the data-length field to extract the data and will also monitor the increment of the BXID field as a cross-check of the synchronisation.

3.5. Resets of the FE

The FE electronics must react to a set of reset signals in a well-defined fashion to ensure that correct error recovery and synchronization can be obtained at start-up or after a detected malfunction. Synchronous resets are sent as part of the TFC frame via the GBT interface as defined in Section 6. All of these reset signals are asserted for one clock cycle. None of the reset signals defined here should affect the configuration of the electronics.

3.5.1. BXID reset

The BXID reset is used to maintain a correct synchronization of the entire LHCb readout system with the bunch collisions. When asserted, the FE modules should reset (or preset) their bunch counters. No other electronics are affected by this action. The BXID reset will be automatically asserted once every machine orbit to maintain overall synchronization of the system. Note that the BXID counter should not be affected by any other reset signal.

3.5.2. FE reset

The FE reset is used to reset the components on the FE modules through which data flows. This includes buffer address pointers and other digital logic related to the data flow, including operational counters and status bits. However, the ECS configuration and control registers should be left intact, including time alignment delay settings of all types. Note that the local BXID counter should not be reset with this signal.

The FE reset will be preceded by an extended period of the Header Only signal set to empty the FE output buffers and the BE input buffers. The duration of the reset time is programmable in the TFC system and should not be linked to a specific BXID.

3.6. Control commands to the FE

The FE electronics must react to a set of commands signals in a well-defined fashion so that correct readout of data is ensured. Synchronous commands are sent as part of the TFC frame via the GBT interface as defined in Section 6. All of these commands are asserted for one clock cycle.

3.6.1. Header Only command

If the Header Only signal is set for a particular BXID in the TFC word, the FE electronics should only transmit header information to the BE for that particular event.

Note that the Header Only signal is used for operational purposes only and that it is not used to control the rate nor is it correlated with the filling scheme.

3.6.2. Bunch Crossing Veto (BX Veto)

If the BX Veto bit is set, the FE electronics should only transmit header information for the BXID for which the bit was received as given by the BXID in the TFC command. The signal is used to reject the LHC crossings with no bunch crossings already at the level of the FE electronics. The bit is exclusively based on the LHC Bunch Filling Schemes, which are loaded into a state-machine in the TFC system.

3.6.3. Calibration Type command

The Calibration Type command is composed of four bits allowing four different calibration types. The signals should be used to fire calibration systems either within the sub-detectors or within the FE electronics. The association between a calibration system of a sub-detector and a bit in the Calibration Type command should be configurable in the FE electronics via ECS.

3.6.4. Non-Zero Suppressed mode (NZS)

If the NZS Mode bit is set for a particular BXID in the TFC word, the FE electronics should transmit data non-zero suppressed and without compression of data. As in most cases (if not all) this will require sending the event data across several GBT data frames, the sub-sequent BXIDs will have the Header Only bit set which implies that the FE data formatting should pack the headers together in one GBT frame to catch up. All the FE electronics must strictly support this mechanism whether they have zero-suppression/compression or not.

3.7. Special running modes

The FE can be put into special running modes by configuration through the ECS interface. These modes are the following.

3.7.1. TFC-alignment mode

The FE module continues to transmit data packets with the standard headers, but with the detector data replaced by the current value on the TFC command bus. This will allow the calculation of latencies in each system and synchronisation checks.

3.7.2. Status-snapshot TFC command

When this command is issued through the TFC interface, the state of all status and monitoring registers in the FE device should be sampled for that BXID. These sampled values should be readable via the ECS interface.

Discussion: Should we include a ‘synch’ mode, where the FE will send a fixed pattern that covers a complete GBT frame and allows the BE to synchronise (eg it tells the BE when the next header will arrive). This could be triggered by a fast command from the TFC. A FE reset could do the same, but the FE reset may also have other effects on the system.

3.8. Mandatory specifications for FE implementation

The following features are mandatory for every implantation of the FE electronics:

The FE module generates the required phase-adjusted clocks for all modes of operation based on the clocks provided by the GBT.

A bunch-counter of 12 bits is implemented and counts from 0 to 3563. When the count reaches 3563, on the next clock it will wrap around to 0. When the BXID reset bit of the TFC command is asserted, the counter is set to a preset value. This preset value is loaded via ECS. The preset action is triggered only by the BXID reset and no other signal. This BXID reset is asserted for one clock cycle (25ns).

A data packet is transmitted for every BXID. The minimum-sized packet contains a header-field.

The header field contains all 12 bits of the BXID, or a subset of bits if correct operation is guaranteed by the sub-detector.

The data field of the packet is truncated if the occupancy of the local buffer goes beyond a pre-defined limit. A bit in the header is set if the packet is truncated.

The FE module has an interface compatible with the GBT data interface.

The FE module has a TFC interface compatible with the master-GBT link.

All TFC commands used by the FE module can be delayed by up to 16 clock cycles. The choice of delay is made via ECS.

When the FE reset bit of the TFC command is asserted, the digital logic is reset. This FE reset is asserted for one clock cycle (25ns).

The FE module supports the Status-Snapshot TFC command.

The FE module can operate in TFC-alignment mode. This mode is set via ECS.

The FE module has an ECS interface based on a master-GBT link (bidirectional).

All ECS registers are writeable/readable via this link.

4. Back-end

The back-end (BE) electronics are situated in the counting room and act as an interface between the FE modules, DAQ, TFC and ECS systems. The counting room is a radiation-free environment and commercial components can be used without the need for radiation qualification. Electronics modules should be implemented in industrial standard formats to allow the use of standardised mechanics, cooling and power supplies.

The right hand side of Figure 1 shows the blocks of the BE, and these are described below.

4.1. Data Receiver and synchronisation

This is based on the GBT system and uses the GBT functionality implemented in an FPGA. Data are received on the optical links arriving from the FE.

[Discussion:

The latency for data to arrive at the BE will vary (see 3.2). A maximum will be defined. We have to decide where this is enforced: FE or BE. Doing it on the BE is more flexible, but means that the FE will waste bandwidth in transmitting packets that will be immediately suppressed by the BE....]

4.2. Input buffer and low level trigger

The LLT is applied to filter out non-interesting bunch-crossings. When running at full rate, its frequency is about 30MHz. It will also be used to reduce the data rate to the DAQ if, for example, the system cannot run at full capacity or buffers are close to overflowing. It is transmitted from the TFC system by means of a GBT interface. Trigger decisions will be saved in a buffer as they arrive from the TFC.

Data arriving from the FE are stored in the input buffer to allow some latency for making the LLT decision. The occupancy of this buffer should be continuously monitored. If it reaches its limit, the buffer controller must truncate the data and allow the buffer to recover. The buffer continues to send header information but with no data until the buffer is sufficiently empty. The activation of data truncation must be indicated in the header.

With zero-suppression being performed in the FE modules, there is no longer a fixed time between the bunch-crossing and the arrival of its data at the BE input buffer. This requires that the LLT decisions include the BXID of that bunch crossing. Accepts and rejects are done by comparing the BXID in the data header with that of the LLTs stored in the trigger buffer. At this point, data accepted by the trigger are tagged with an event number.

Another consequence of zero-suppression on the FE is that there may be a wide spread in the latencies of data from the same bunch crossing but from different channels arriving at the input buffer. This spread should be incorporated into the size of the input buffer and the trigger buffer.

Data accepted by the LLT are given an Event-ID extracted from a local Event-ID counter.

[Discussion points:

trigger latency – we should define a maximum

Event-ID – how many bits?

Do we want to build complete events on TELL40 and then transmit to DAQ? ie What is the allowed time spread on transmitting events from the different TELL40 units to the DAQ?] → where is the buffering done?

4.3. DAQ formatting and transmission

Data are merged from different sources according to the bandwidth of the link to the DAQ. To optimise the use of the bandwidth, it is advantageous to construct multi-event packets (MEPs) before transmission. This formatting strongly depends on the event size and the occupancy of the output buffers at this stage will be monitored. If they are close to being full, then an alarm is back-propagated to initiate throttling.

Packets of data are then transmitted to the DAQ system on one or more links whose total bandwidth can sustain the data throughput. The destination address to where the BE module should transmit is broadcast to the modules by the TFC system.

The technology of choice for this interface is based on 10 Gbit Ethernet. The interface is described in more detail in Section 8.

4.4. Trigger throttling

Each of the BE modules should monitor the states of all buffers which are involved in the data processing and which have a variable behaviour. Buffer occupancy limits should be defined for all buffers which are located after the event rejection in the BE in association with an overflow warning signal that is used to compile a trigger throttle signal. This signal should be transmitted to the TFC+ECSInterface. Buffers before the event rejection must be deterministic.

Sufficient buffering beyond the high-level water mark must exist in the buffers after the event rejection logic to allow for latency between asserting the throttle bit and the actual throttling of the LLT. The configurable buffer limits should be determined in simulation. This latency has been

estimated to be approximately 30 bunch clock cycles. The throttle signal should be transmitted together with the BXID for which the throttle was raised.

The TFC+ECSInterface boards compute at 40 MHz a throttle word by fanning-in and aligning the throttle bits from the connected BE modules. This word is transmitted to the TFC system via the TFC links. As soon as this detects a throttle bit from any of the BE boards, the sub-sequent triggers are rejected until the bit is removed to allow the BE to evacuate events and free the buffers. In this regard, the throttling mechanism is based on a logical OR of the BE boards in a partition.

4.5. TFC interface to BE

Fast, synchronous controls are provided by links from the central TFC system, one frame broadcast every 25ns clock period. These commands are broadcasts so every BE device will receive the same data. The broadcast command is decoded by the BE module to carry out a number of tasks, such as resets, MEP destination assignment and the LLT. The encoding of the frame is described in Section 6. The BE modules receive the bunch clock from the TFC/ECS interface module, as described in Section 6.

4.6. Resets of the BE

The BE electronics must react to a set of reset signals in a well-defined fashion to ensure that correct error recovery and synchronization can be obtained at start-up or after a detected malfunction. Synchronous resets are sent as part of the TFC frame via a GBT interface as defined in Section 6. All of these reset signals are asserted for one clock cycle. None of the reset signals defined here should affect the configuration of the electronics.

4.6.1. BXID reset

The BXID reset is used to maintain a correct synchronization of the entire LHCb readout system with the bunch collisions. When asserted, the BE modules should reset (or preset) their bunch counters. No other electronics are affected by this action. The BXID reset will be automatically asserted once every machine orbit to maintain overall synchronization of the system. Note that the BXID counter should not be affected by any other reset signal.

4.6.2. Event-ID reset

The event ID appended to data accepted by the LLT is based on a counter that increments with each positive trigger response. This counter is reset by the Event-ID reset. No other electronics are

affected by this action. Note that the Event-ID counter should not be affected by any other reset signal.

4.6.3. FE reset

The FE reset is defined to reset the components on the FE modules through which data flows. The data logic of the BE should not be affected by this reset. The only exception is an emulator of the FE module that has been implemented in a BE module.

4.6.4. BE reset

The BE reset is defined to reset the components on the BE modules through which data flows. This includes buffer address pointers and other digital logic related to the data flow, such as operational counters and status bits. However, the ECS configuration and control registers should be left intact, including time alignment delay settings of all types.

Note that the local Bunch and Event Identifier Counters should not be reset with this signal. In order for the system to resume data taking cleanly, a FE Reset is always performed in parallel with a BE Reset. In addition the Header Only command is sent continuously during the entire duration of the reset of the BE module.

4.7. Control commands to the BE

4.7.1. Trigger bit

The global trigger decision is a compilation of the LLT decision, the TFC internal triggers, forced triggers, and the various throttle sources. This global trigger decision determines whether the BE should immediately discard an event or process an event and transmit it to the Event Filter Farm.

The decision is distributed with the shortest possible latency to the BE such that the trigger decisions arrive at the BE at a constant latency with respect to the actual bunch crossing and always before the data arrives from the FE electronics.

4.7.2. Trigger type

Each positive trigger decision is accompanied by a trigger type that defines the principal type of an accepted event. It is a 4-bit word encoding up to 15 different types (Trigger Type = 0 is reserved). The type may be used to control the processing in the BE boards. The principal trigger type is

defined in a priority scheme which takes care of the cases where several different sources of triggers fire simultaneously. If two triggers are fired in the same BXID, the one with the highest priority in the priority scheme is taken into account.

Examples of different trigger types include:

- physics trigger
- beam-gas trigger
- luminosity trigger,
- calibration trigger,
- random trigger,
- timing alignment trigger,
- NZS trigger
- auxiliary trigger

4.7.3. Event Destination command

The TFC broadcasts to the BE boards the destination address to where the events should be sent in the Event Filter Farm.

Upon reception of the Event Destination command the BE module should close the current Multi-Event Packet, add the transport header for the DAQ network using the destination address, and transmit it to the Event Filter Farm.

4.7.4. Status and Counter Snapshot command

The Snapshot command should trigger all status and monitoring counters in the BE boards to be sampled in ECS registers for a particular BXID. The information may then be read via ECS.

4.8. Special running modes

4.8.1. TFC-alignment mode

The TFC decoder/encoder block can be configured via ECS to send back the TFC data on the same link in place of the throttle word.

4.9. Implementation of back-end

[To be expanded by TELL40 people]

The GBT link offers bi-directionality and is conceived to provide DAQ, TFC and ECS in one link. However, the large imbalance between the required bandwidths from and to the detector together with the relatively coarse modularity of the LHCb sub-detectors (in contrast to the trackers of ATLAS and CMS) means that it is more efficient to use dedicated simplex GBT links for DAQ and dedicated duplex links for ECS/TFC. This implies the implementation of a BE module dedicated and optimised for receiving DAQ data and another module dedicated to the ECS/TFC interface of the front-ends.

5. Low Level Trigger

The low level trigger (LLT) is included in the system to provide the following functions:

1. Allow staging of the DAQ system by controlling the rate.

2. Allow a throttling mechanism if buffers overflow or the CPU farm has insufficient resources to process the data

The LLT will make decisions based on detector data and is based around the existing LHCb L0 trigger [3]. It will be formed with data from the Calorimeter and the Muon detectors. The trigger will be broadcast from the TFC system with a fixed latency.

Figure 4 illustrates the principle. The ECAL and HCAL detectors are readout by FE boards equipped to handle the new architecture but with a separate data path for the trigger. These transmit data to a Validation Card where initial selection is carried out. This is the same module used in the existing experiment. Data from the Validation Card are then transmitted off-detector to two BE cards (TRIG40) equipped with the appropriate input links running at 1.6 Gbit/s with 8bit/10bit encoded data. Decision-making algorithms are carried out on these TRIG40s which then transmit the results directly to the TFC system. The Muon trigger is based almost entirely on the existing L0-muon system, where data from the off-detector electronics (ODE) are transmitted to the counting room at 40MHz. The existing L0-muon system then processes this data and transmits decisions to a third TRIG40. Final muon trigger decisions are made in this TRIG40 which interfaces directly to the TFC. Note that the TRIG40s play the role of the L0 decision unit in the existing LHCb. The TRIG40 modules are also equipped with links to transmit data to the DAQ in a manner identical to the standard BE modules.

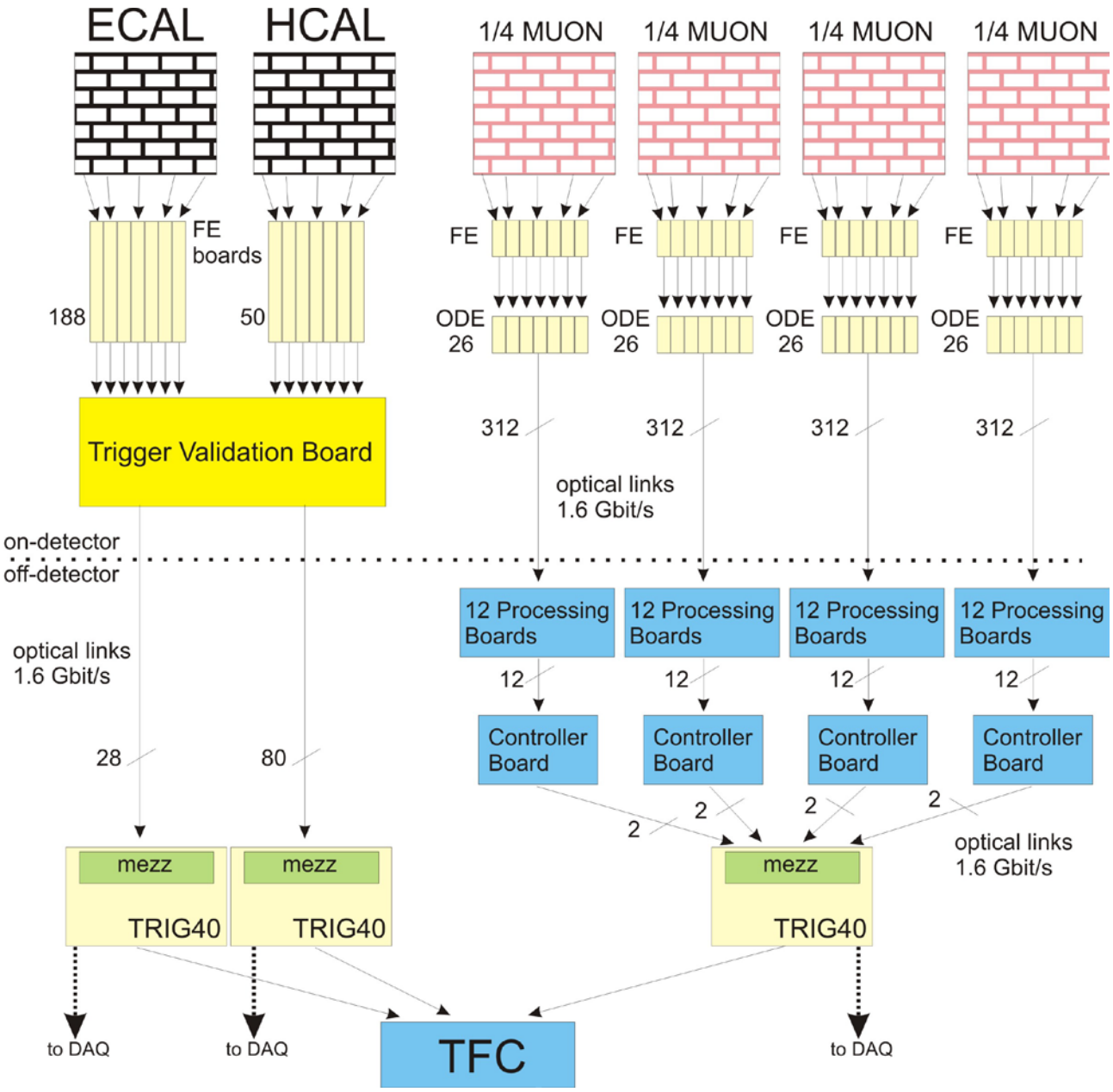

Figure 4: Low level trigger system

6. Timing and Fast Control (TFC)

The distribution and control of the experiment timing, and the control of the entire event flow from the Front-End electronics to the Event Filter Farm should be assured by the Timing and Fast Control system (TFC) [4]. The control is performed by a TFC Master which has a direct link to the LHC systems for radio-frequency and machine timing. This distributes the LHC beam-synchronous clocks to the FE and BE electronics together with synchronous resets, fast controls commands and the LLT.

The clock reception should provide means of aligning the global timing of the experiment. The TFC distribution network should allow transmitting a clock to the readout electronics with a known and stable phase at the level of ~50ps and very low jitter (<10ps). It must also allow stable control of the latency of the distributed information. Local alignment at the FE and the BE of the individual TFC links and the synchronous reset commands together with BXID and Event-ID checks are required to assure synchronicity of the experiment.

The TFC communication network should be bi-directional to allow monitoring of the input and output buffer status of the BE electronics in order to protect against overflows. The buffer information is used in order to control the rate at which events are accepted in the BE modules.

The TFC architecture must allow partitioning, that is the possibility of running autonomously one or any ensemble of sub-detectors in a special running mode independently of all the others. In practice this means that the system should contain a set of independent TFC Masters, each of which may be invoked for local sub-detector activities or used to run the whole of LHCb in global data taking, and a configurable internal switch fabric.

The TFC Master should have three interfaces to receive the trigger information from the ECAL, HCAL and MUON sub-triggers, as described above. Trigger buffering is required in order to absorb the different sub-trigger latencies, align the trigger information to the maximum latency, and perform the final trigger decision. In addition to the dynamic rate control based on the buffer status and the LLT, the TFC Master will transmit a bunch crossing veto to the FE which is based on the LHC filling scheme and which allows the FE to only send event headers for crossings with no collisions in order to optimize the readout bandwidth. Exceptions to this veto are generated by the TFC Master for calibration and monitoring purposes.

The system should provide means to synchronously distribute to the BE electronics the destination of data within the DAQ farm for each event. This function should also include a request mechanism by which the farm nodes declare themselves as ready to receive the next set of events for processing. The event transfer from the BE electronics is thus a push scheme with a passive pull mechanism. The scheme avoids the risk of sending events to non-functional links or nodes, and produces a level of load balancing as well as an additional rate control in the intermediate upgrade phase with a staged farm. Ultimately this would be the only emergency control of the rate when the system has been fully upgraded to 40 MHz readout.

The TFC Master should have several mechanisms to generate forced trigger accepts internally for calibration and detector monitoring purposes. This should also include transmission of synchronous commands to generate calibration pulses, such as special data patterns, LED or laser pulses, etc. These triggers override the collision scheme veto and the LLT and may also be transmitted to a special destination in the farm.

In addition, since the proposed FE modules will perform zero-suppression in most cases, a scheme must be envisaged which allows occasional non-zero suppressed readout for special purposes. Although the bandwidth does not allow this at 40 MHz, there is no requirement for high-rate so the TFC Master should synchronize a sequence in which the readout of a non-zero suppressed event spans over several consecutive bunch crossings. The mechanism effectively suppresses the data of the other crossings by forcing pure header transfer via the same mechanism as the collision scheme veto.

A data bank containing information about the identity of an event (Run Number, Orbit Number, Event Number, Universal Time) and trigger source information should be transmitted by the TFC Master to the Event Filter Farm for each event as part of the event data. In order to save on the bandwidth out of the TFC Master, there may be a reduced bank for local sub-detector runs.

In order to replace the current readout electronics and commission the new electronics in steps, and make use of the L0 trigger system which is already operating at 40MHz, the new TFC system must support the old TTC system, at least for a period of time during the upgrade phase.

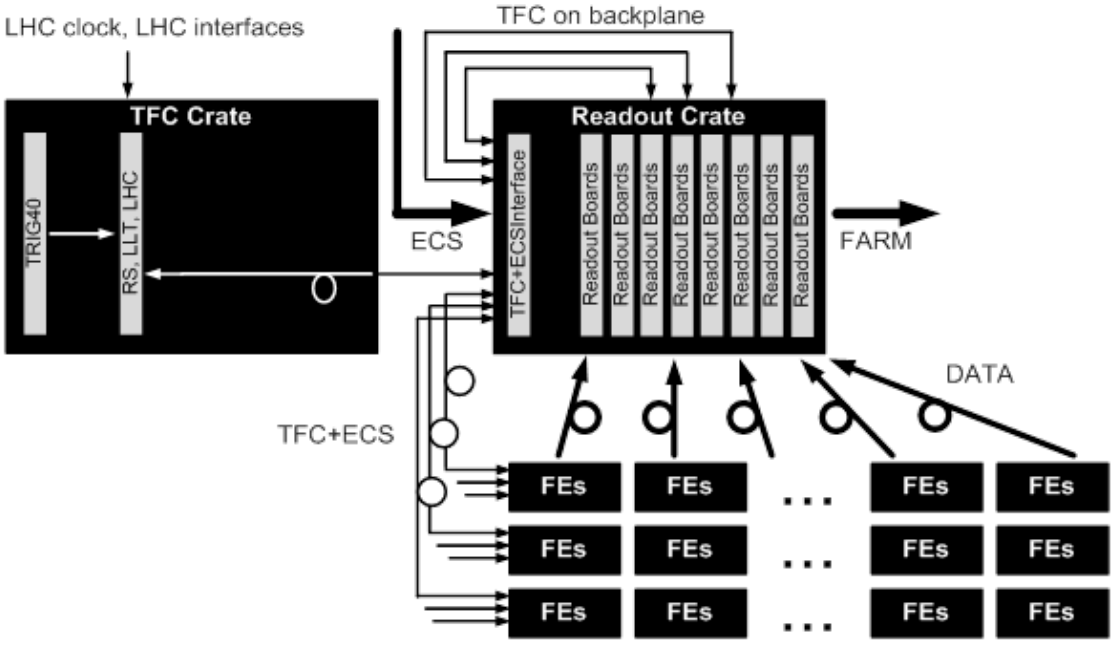

Figure 5: Architecture of the upgraded S-TFC system for the upgraded LHCb readout architecture. A single readout slice composed of a Readout Crate and its FE electronics is represented in the figure.

Figure 5 shows the proposed TFC architecture. In the upgraded architecture, a pool of TFC Masters is instantiated in one single Super Readout Supervisor ("Super-ODIN") based on a single large FPGA for all TFC functions. In order to have a manageable set of transceivers on the Super-ODIN, each BE crate contains a fan-out/fan-in module, called TFC+ECSInterface. Thus, physically, each transceiver on the Super-ODIN is connected via a bidirectional optical link to a TFC+ECSInterface board which functions as a fan-out/fan-in for the BE modules in the crate and the FE modules connected to that readout slice.

The TFC commands and their corresponding codes are listed in Table 3 and Table 4. As described in [4], a total of 44 bits are considered enough in order to transmit the TFC information to. Including the encoding of the protocol and the error correction, a total of 60-bits are transmitted at 40 MHz. An instantiated TFC decoding/encoding block in the TFC+ECSInterface is responsible for relaying a subset of the synchronous control commands to the Front-End electronics via the TFC+ECSInterface, and separately to the BE modules. In order to check the synchronicity, all TFC words will carry the bunch crossing identifier of the event to which they belong.

43 .. 32	31 .. 16	15 .. 12	11..8
BXID(11..0)	**MEP Dest(15..0)**	**Trigger Type(3..0)**	**Calibration Type(3..0)**

7	6	5	4	3	2	1	0
Trigger	**BX Veto**	**NZS Mode**	**Header Only**	**BE Reset**	**FE Reset**	**EID Reset**	**BXID Reset**

Table 3: TFC information encoding commands and resets to be transmitted to the BE boards via the TFC+ECSInterface boards.

It should be noted that the TFC word per bunch crossing will be transmitted at a constant phase with respect to the actual crossing in such a way that the information is available at the proper time in the FE electronics. There are two consequences of this. Firstly, the information which is subsequently needed in the BE electronics must be buffered to account for the variable data processing time in the FE and the data transfer. Secondly, unavoidably the LLT will be transmitted together with the TFC information of a later bunch crossing at a constant offset corresponding to the maximum latency of the LLT. The decoding and correction for this offset is handled by the TFC+ECSInterface by assuring that all TFC information contained in one TFC word will correspond to the BXID included in the TFC word.

A TFC word of 24 bits incorporated in the GBT protocol has been considered sufficient to encode the synchronous commands for the FE together with the bunch crossing identifier. Table 4 shows the preliminary encoding of the FE TFC word. The transmission of these 24 bits, as described in Section 10, requires the use of 48 bits of the GBT frame. Since the GBT protocol consists of 80 user bits per frame, this leaves an ECS field of 32 bits (with 2 more bits dedicated to the GBT slow control (SC) port) available for the FE. The details of the various commands and reset sent to the FE and BE are extensively described in [5].

23 .. 12	11 .. 10	9	8 .. 5	4	3	2	1	0
BXID(11..0)	**Reserve**	**Snapshot**	**Calibration Type(3..0)**	**BX Veto**	**NZS Mode**	**Header Only**	**FE Reset**	**BXID Reset**

Table 4: TFC commands encoded in the standard GBT protocol employed between the TFC+ECSInterface boards and the FE electronics.

The throttle and trigger protocol basically consist of transmitting a local throttle bit together with the bunch crossing identifier of the last event. Since the system is asynchronous at this level, the bunch crossing identifier serves mainly for monitoring and additional checks. The protocol between the BE modules and the TFC+ECSInterface may thus be reduced to words of less than 20 user bits.

It is important to note that the TFC commands corresponding to a certain BXID will be sent well before the actual bunch crossing takes place.

7. Experiment Control System (ECS) Interface

This interface will allow the configuration of the electronics components in the system, the reading of the detector status and the monitoring of local parameters. All configuration registers implemented in the electronics should be both writeable and readable. The sections below describe the different types of interface.

7.1. FE ECS interface

The ECS interface to the FE components will be implemented using the GBT system, as described in Section 10. This bi-directional link allows the writing and reading of configuration and monitoring data. The GBT-SCA chip provides an interface between the GBT and standard protocols such as I2C and JTAG and can be mounted on an FE module.

Transmission of ECS and TFC to and from the FE electronics will be done on the same physical GBT-based link. This will be driven by a common module sitting in the counting room, known as TELL-TFC/ECS. The full GBT protocol, including forward-error correction as described in Section 10, will be used to guarantee the link robustness. Thus 80 data bits will be available, 48 of which are reserved for TFC information. The remaining bits can be used by the sub-detectors and can connect to, for example, more GBT-SCA chips, as described in section 10.1.

7.2. BE ECS interface

The ECS interface to the BE components will be implemented using a credit-card PC (CCPC) mounted on a mezzanine daughter-card. This will provide standard control interfaces for hardware configuration and monitoring. The communication protocol between the CCPC and the central ECS system will be TCP/IP.

7.3. FE and BE control and monitoring resource requirements

7.3.1. Control registers

All configurable ECS resources should support both read and write actions. The resources should be readable at all times without interference with the operation of the readout logic. For efficiency reasons, several functional parameters may be located together in the same hardware register.

7.3.2. Status registers and counters

The FE and BE modules should have ECS monitoring counters for all the received TFC commands, and all readout related operational statistics. Moreover, all the TFC+ECS master GBT internal registers should be made available for monitoring.

With the exception of the BXID and the Event ID counters, the range of all counters should allow at least an hour operation at maximum counting rate.

It should be possible to obtain a set of consistent values for all monitoring counters in all devices. This requires that an update action or *snapshot* is implemented which is triggered either by an ECS write/read action to a particular register or by a TFC snapshot command. The action should trigger the registering of the values of all counters in the same bunch clock into separate ECS registers which may then be read out sequentially through the ECS interface. Triggering the snapshot via the TFC allows obtaining consistent information across the experiment for a particular BXID. The monitoring counters should continue undisturbed during the sampling operation.

Status registers reflecting the states of operational logic should be available on the ECS bus in three versions. The live value should be readable at anytime. In addition, there should be a value which is registered via the snapshot action described above in order to store values consistent with the counters. Finally, there should be a registered value available on the ECS bus which is latched if the status bit changes to an abnormal value during a single bunch clock cycle in order to be able to monitor instantaneous and rare erroneous state changes. The latter keeps its value until reset via the ECS.

8. Data Acquisition (DAQ) Interface

This section describes the requirements imposed on the BE by the upgraded DAQ system.

8.1. Protocol

The full IP stack as described in RFC 791 must be implemented, as well as ARP RPC 826 and ICMP and ICMP (ping) as defined in RFC 1122. We will embed MEP in UDP so the UDP protocol (RFC 768) is also needed. However, only 4 UDP ports need to be supported.

8.2. Buffering

In order to cope with buffer limitations in the network it should be possible to buffer each 10 Gbit stream for at least 4 seconds. This means at least 4 Gigabyte of memory for each 10 Gbit link. This memory must be capable of supporting the bandwidth of the interface of 10 Gbit/s full-duplex.

8.3. ECS

It must be possible take a snapshot of a number of MEPs by latching them into a memory which can be accessed from the ECS. The ECS must be accessible via a Gbit Ethernet link which is strictly independent from the data links. It must be possible to reset the ECS interface out-of-band and to do a consistent latch of all counters pertaining to the data-flow in a coherent way so that they can be read by the ECS. The ‘master’ counters must continue counting during the time the shadow counters are latched. Write-only registers are not permitted.

8.4. Debugging

A data-generator must be provided which works at the interface rate for all output links and is completely independent of the input data from the FE.

9. Testing and Commissioning

All FE and BE modules should contain functionality for in-situ testing, calibration and commissioning. In this architecture, a crucial task will be the correct synchronisation of the bunch counters across the experiment and mechanisms should be included in the electronics to allow this. Examples of these are:

1. Test pulse injection. A pulse is injected into the front-end amplifiers, triggered by a synchronous command issued by the FE TFC interface. The phase and amplitude of this pulse must be adjustable. This technique can be used for the bunch-counter synchronisation as long as path differences in the distribution of the TFC command across the experiment are well known. The transmission of this TFC command must be done at a known time after transmission of the BXID reset, described in Section 6.

2. Triggered light source. If the detector is sensitive to light, then a single pulsed light source illuminating many channels can be used to test and time align these channels relative to each other. The trigger of the light source is generated by a synchronous command from the GBT-TFC interface and its phase is adjustable. The transmission of this TFC command must be done at a known time after transmission of the BXID reset.

3. Cosmic rays. These can be used for an absolute time alignment of the bunch counters, but care has to be taken because of the asynchronous arrival time of the particles.

4. Digital pattern. The ability to transmit a fixed or predictable pattern from the FE module can be useful for the verification of the links and the receivers of the BE modules. The generation of this pattern can be implemented in the FE digital logic and should run independently of the actual signals from the detector elements.

   [Discussion:

   One case of this is TFC-alignment mode.

   Another possibility is for the sub-detectors to test the system bottlenecks by controlling the occupancy of these artificial events.]

9.1. Hybrid operation

The upgraded TFC system will support the old TTC-based distribution system. This is necessary since the muon LLT sub-trigger will be based on the current L0 electronics which is already operating at 40 MHz. An important bi-product of this requirement is that it allows operating the whole readout system in a hybrid mode with both the current and the upgraded electronics operating simultaneously and even together at 1 MHz. This is done by interfacing the current TFC system

with the upgraded TFC system, where the current TFC system will control the old electronics and the upgraded TFC system will control the upgraded electronics.

10. Use of the GBT and Versatile Link

The GigaBit Transceiver chip-set and Versatile Link have been developed as generic building blocks for data transmission, TFC and slow-control systems. The GBT chip-set consists of radiation-tolerant components for mounting in the FE modules and compatible firmware for commercial FPGAs in the BE modules. The GBT is designed to be operated in duplex or simplex mode. The Versatile Link project offers radiation qualified electro-optical components to implement a complete optical transmission system.

Details of the GBT can be found in [2], but the most relevant points for LHCb are summarised here, together with suggestions for implementation. The components of the chip-set and the versatile link are shown in Figure 6. The trans-impedance amplifier (TIA), PIN diode (PD), laser driver (LD) and laser will be mounted together in a bi-directional Small-Form-Pluggable (SFP) Package. Another option will be a dual-transmitter SFP where the receiver components are replaced by a second transmitter channel.

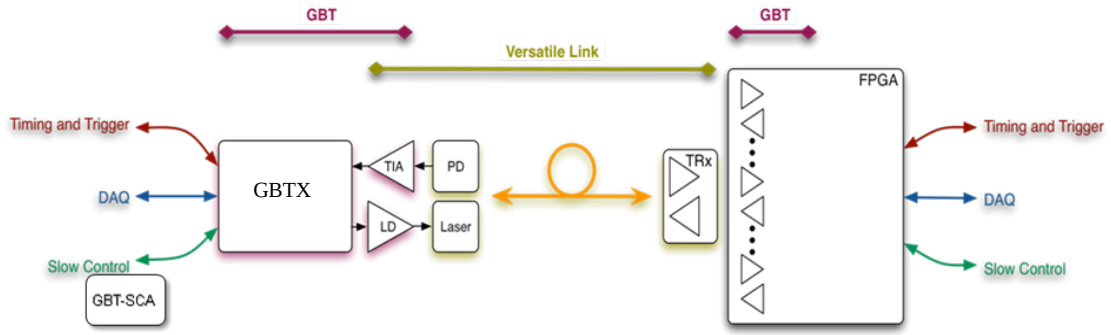

Figure 6: GBT chipset

The GBTX is the serialiser/deserialiser chip operating at a serial rate of 4.8 Gbit/s, and transmits 120-bit words of data arriving every 25ns. The allocation of the 120 bits is shown in Figure 7.

1. The header (H) and forward-error-correction (FEC) fields are not available to the user.

2. The data (D) field is fully available to the user for data transmission. The GBT accepts user data in a parallel or serial mode using an interface based on ‘E-ports’ or ‘Elinks’[2]. Parallel mode uses a 40-bit bus running at 80 Mbit/s. Serial mode has different configurations, from forty E-ports running at 80 Mbit/s down to eight E-ports running at 320 Mbit/s.

3. The slow-control (SC) field is divided into 2 bits reserved for internal configuration of the GBT and 2 bits available to the user. These 2 user bits will be transmitted/received by an E-port running at 80 Mbit/s. This E-port can be interfaced to the GBT-SCA (slow control adapter) which translates into standard protocols such as I2C and JTAG.

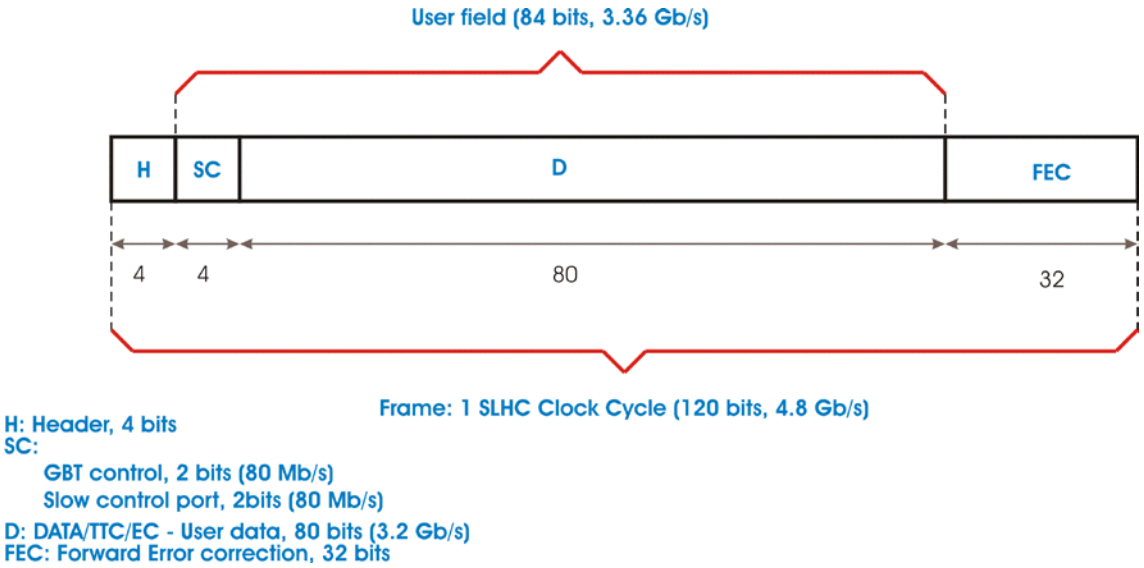

Figure 7: GBT transmission format with FEC

Two alternative frame formats can be used and are shown in Figure 8 and Figure 9. In Figure 8 (8b10b mode), the 120 bits are divided into 12 segments of 10 bits. Each segment is an 8b/10b encoded data word. The first segment contains a fixed comma character and this is used as a frame delimiter. The remaining 11 segments are available to the user and provide 88 bits for data transfer. Note that no error correction is possible using this format so it is more susceptible to bit errors than the format of Figure 7. The interface between the GBT and FE is similar to the format of Figure 7 but a group of 4 output E-ports are re-configured as inputs to receive the additional 8 bits of data. Note that the SC fields are not included in this mode.

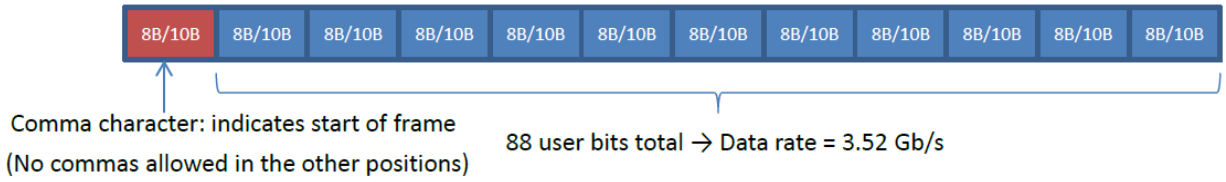

Figure 8: GBT transmission format with 8b/10b segments

In Figure 9 (Widebus mode), the FEC is not used and these 32 bits of the frame are available for data transmission. Note that no error correction is included in this format so it is more susceptible to bit errors than the format of Figure 7. In this mode, 16 output Eports are re-configured as inputs to receive the additional 32 bits of data.

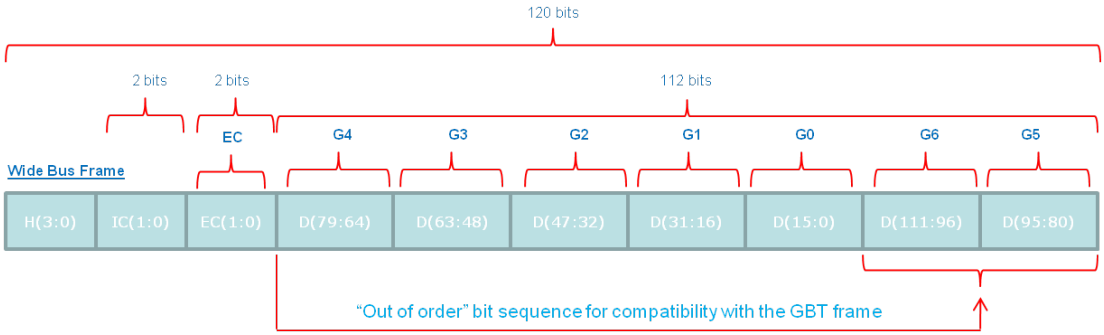

Figure 9: GBT transmission in Widebus mode

The GBTX can be used in one of three different transceiver modes. In all cases, an external clock source (for example, a crystal oscillator or another GBTX) is needed to help in the initial locking procedure. Note that unused blocks can be powered off in cases 2 and 3.

1. Transceiver mode: The system is a duplex transceiver. Configuration (write/read) of the GBTX is done by the receiver circuitry. The clock output of the GBTX is recovered from the receiver data.

2. Receiver mode: The system is a simplex receiver. Configuration (write/read) of the GBTX is done by interfacing the chip to an auxiliary system that has bi-directional capability. The clock output of the GBTX is recovered from the receiver data.

3. Transmitter mode: The system is a simplex transmitter. Configuration (write/read) of the GBTX is done by interfacing the chip to an auxiliary system that has bi-directional capability. An external clock source must be used, for example another GBTX in one of the other two modes above.

10.1. Implementation in LHCb

The data bandwidth from the LHCb detector to the counting room will be many times larger than the TFC and ECS traffic in the opposite direction. So it is likely that the GBTX will be widely used in simplex-transmitter mode. However, duplex transceivers will be required for writing and reading ECS information and the receiver mode for TFC operations (clock, calibration, resets).

An example of a possible implementation is shown in Figure 10 for a system with a FE module containing many channels and hence a high data bandwidth. The FE chips drive multiple GBTXs configured in simplex-transmitter mode. These then drive a number of dual-transmitter optical packages connected to fibres that fan into a 12-way fibre ribbon. This arrives at a 12-way receiver on the BE board. To maximize data bandwidth at the expense of link integrity, the user could use the GBT format of Figure 9.

A single GBTX (master) on this FE module is configured in duplex-transceiver mode to accept TFC and ECS data from a TELL-TFC/ECS module. This distributes TFC information to the GBTXs and other FE components, and provides the phase-adjustable 40MHz clocks. It also provides the local ECS interface through the GBT-SCA. For example, the GBT-SCA has a number of I2C busses that can be used to configure and monitor the front-end components. To guarantee the integrity of this link and minimise bit errors that may corrupt the configuration of the FE, the GBT transmission should be done using the format of Figure 7.

The implementation of the links has to be optimized according to the bandwidth and integration requirements of each sub-detector.

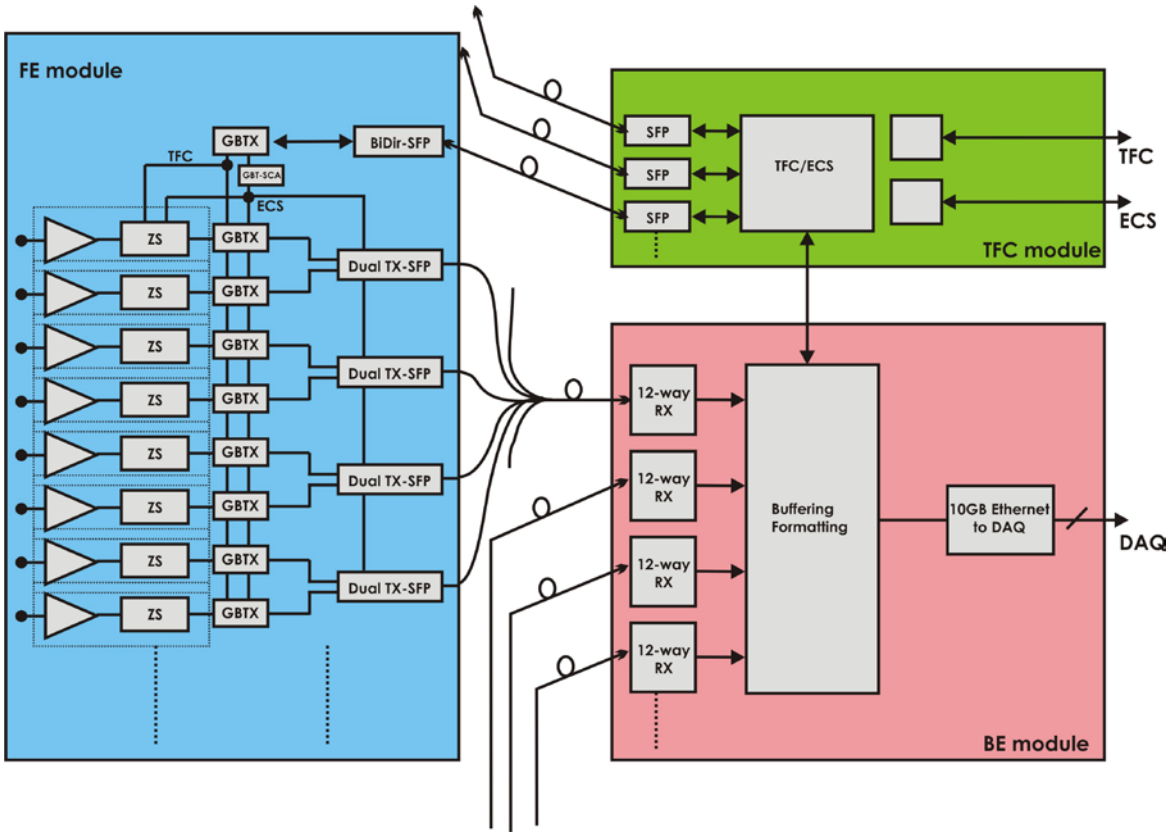

Figure 10: A possible link implementation in the LHCb upgrade

10.2. TFC communication to the FE electronics

The reception at the FE electronics of the timing and the TFC commands, and also the ECS communication, should be performed by a dedicated TFC+ECS master GBTX chip which is associated to a set of FE electronics as shown in Figure 10. The location of this master GBTX chip and the architecture below may differ (control card in crate, repeater board, service box, etc.) but the connection and fan-out to the actual FE logic must respect the jitter and delay requirements.

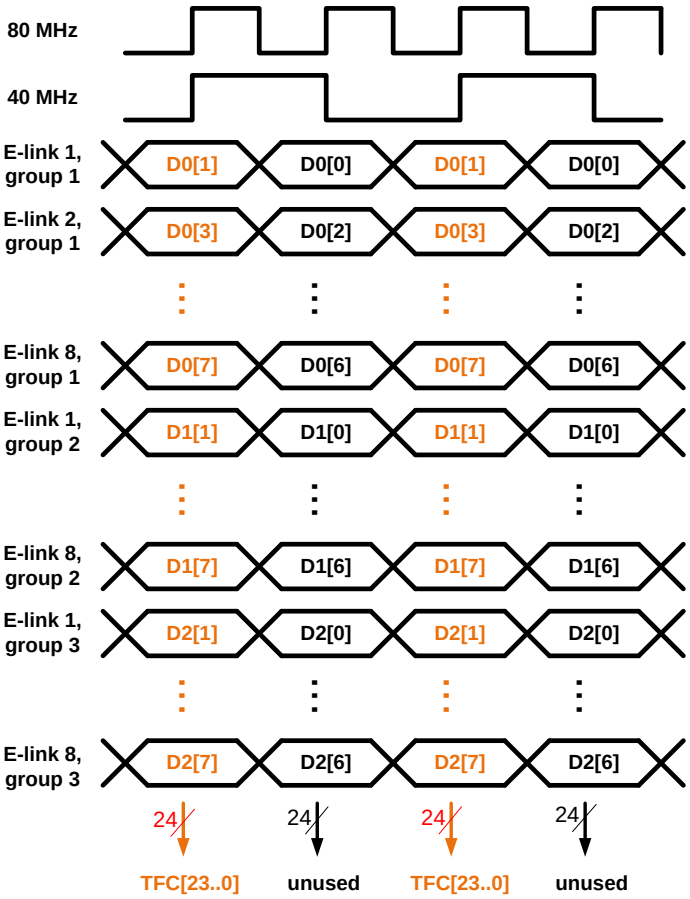

Figure 11: Illustration of the decoding of the TFC commands in the FE by using the GBT e-links.

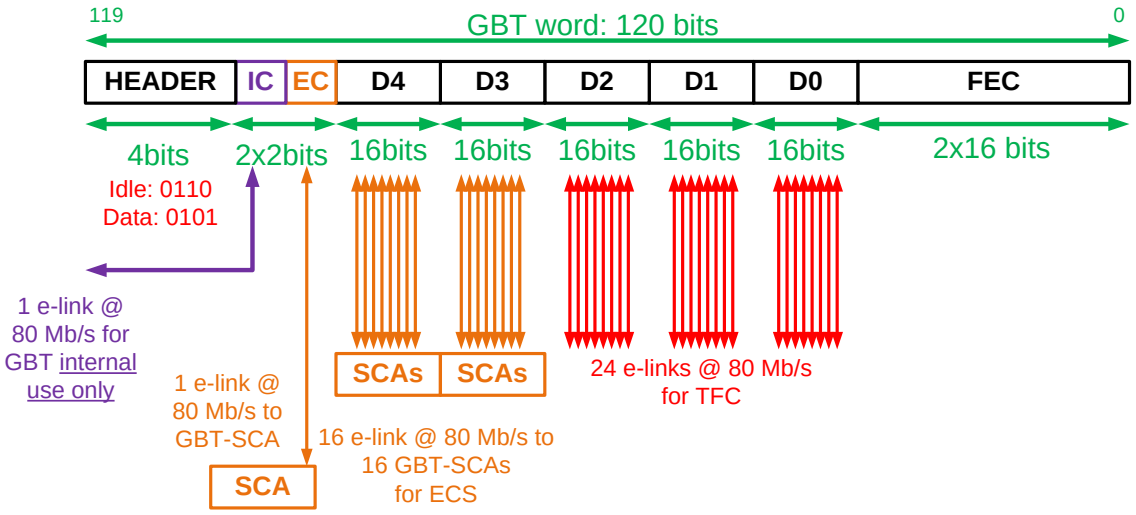

Figure 12: Description of the GBT frame as seen at the output of a GBT chip at the FE.

The definition of the TFC readout control protocol encapsulated in the standard 120-bit GBT frame has direct consequences on the mode in which the master GBTX can be operated and how it must be implemented in the FE electronics:

- a total of 3 groups of 8 E-links organized in an E-link lane, i.e. a parallel bus, of 24 bits are reserved for the TFC commands.
- each E-link dedicated to the decoding of the TFC commands should be configured in 2x data mode, i.e. at 80 Mbit/s. In practice this means that the TFC word will use up to 48 bits of the GBT frame, transmitted as two words of 24 bits each at 80 MHz. The TFC E-links are used in simplex receiver mode in the sense that the FE is not supposed to transmit back anything to the TFC+ECSInterface boards in the TFC data field of the GBT frame.
- The TFC commands will be compiled in a way that all the TFC commands will come out on the same rising edge of the 80 MHz clock (i.e. TFC commands are all msb or for easy reading they correspond to the odd positions of bits, Figure 11). Therefore, the 24 bits of TFC commands will be available in parallel without need of pipelining. This also means that the lsb (even bits) corresponding to the TFC commands are unused for the time being or the TFC commands could be repeated.
- Programmable coarse delays (bunch clock pipelining) must be envisaged for each of the TFC commands encoded in the protocol to adjust to the local operational FE logic. In addition, calibration commands may require fine adjustment delays to trigger a pulsing system at the proper time.
- The rest of the GBT user data frame, 34 bits at a total of 1.36 Gb/s of bandwidth per GBT, is dedicated to ECS for control and monitoring purposes. The 34 bits allow driving up to 17 bi-directional e-link ports which depending on the use may be routed directly to a FE resource or used together with an SCA chip to drive several other bus types (I2C, JTAG, SPI, PCI). A clock is also available for each E-link port to drive the SCA chip or an ECS E-link slave implementation if necessary.

The details of the GBT protocol as it should be used for TFC and ECS communication to the FE is illustrated in Figure 12

11. Use of programmable commercial devices in FE modules

Zero-suppression and data packing must be carried out in the FE electronics. In many cases, this requires sophisticated algorithms that depend heavily on the expected occupancy of the detector. It is advantageous to have the possibility of changing or tuning these algorithms and using programmable devices allows such an optimization.

Such a device has to be proven to resist the radiation environment of the detector according to the following requirements:

1. Any increase in power consumption due to total-ionising-dose must be estimated and the power infrastructure should be capable of supporting this.

2. The device must be tolerant to single-event-upsets (SEUs) to a level acceptable for the efficiency of the detector. For example, it may be acceptable to have occasional bit errors in the data, but critical circuits, such as buffer pointers, must be much more tolerant to SEUs. Mitigation techniques, such as triple-redundancy, should be used.

3. The configuration of the device must resist the radiation environment.

4. The device must resist destructive single-event-effects.

Devices employing flash programming from ACTEL/MICROSEMI have been shown to resist upsets in their configuration. The latest versions of these devices are currently under study for their radiation resistance.

12. References

[1] Letter of Intent for the LHCb Upgrade, CERN/LHCC/2011-001.

[2] https://espace.cern.ch/GBT-Project/default.aspx

[3] LHCb Trigger System, Technical Design Report, CERN-LHCC-2003-031, September 2003

[4] F. Alessio et al., "System-level specifications of the Timing and Fast Control System for the LHCb Upgrade", CERN-LHCb-2012-001, January 2012.

[5] F. Alessio et al., "Readout Control Specifications for the Front-End and Back-End of the LHCb Upgrade", LHCb-INT-2012-018, July 2012.