

Screaming Streaming

Greg Landsberg

Experiment Integration Meeting

June 20, 2001

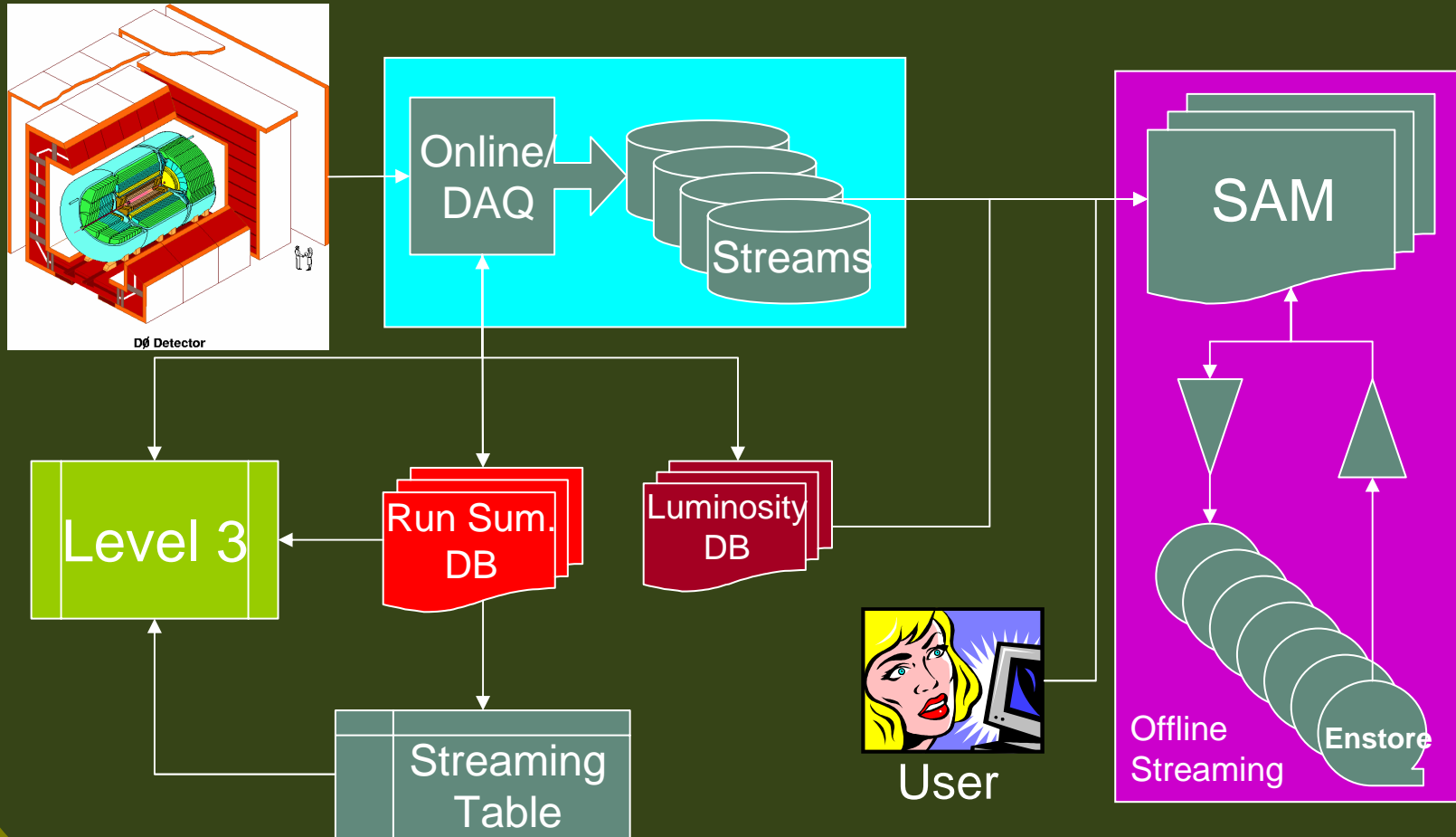
DØ Streaming Model Overview

- Run 2A @ nominal luminosity yields 6×10^8 events/year, so it's very desirable to split the data in several smaller streams to allow for faster reprocessing
- Limited operational budget and the tape robot capacity does not allow for multiple copies of one and the same event
- Hence, exclusive streaming model: every physics event goes into one and only one stream
- If implemented correctly, streaming should be completely transparent to the end-user, while significantly reducing data processing time (cf. optimized option in a good compiler)
- Correct implementation, however, requires intimate understanding of the limitations coming from physics and the online and offline software and hardware

Streaming Constraints

- Physics constraints:
 - Ability to **extract data by** a **trigger** name
 - Precise integrated **luminosity tracking**
 - Very **low data loss** rate
 - Ability to add **small** and efficient **streams for high-priority analyses**
 - Ability to **modify stream definitions** in a simple and transparent way
- Infrastructure-driven constraints:
 - **Compatibility with** the existing **L3, Online, and SAM** infrastructures
 - Automatic database-powered **bookkeeping** of all changes
 - Ability to **merge small** streamed data **files** and put **multiple streams on a single tape**

Simplified Streaming Diagram



Streaming at Level 3

- Streaming at Level 3 consists of the following **four distinctive mappings**:
 - Mapping from **physics objects to primal streams** (e.g., L3FEle → EM stream)
 - Up to **32 primal streams** are currently supported
 - The result of the mapping is a **primal stream bitmap**, which is a collection of bits for all the primal streams (bit is set if the event passes the trigger requirement for a particular primal stream, or 0 otherwise)
 - Mapping from the **primal stream bitmap to a physical stream**, as defined by the **Stream Definition Table**
 - Mapping every **physical stream to a stream ID**, provided by **COOR**
 - Mapping **trigger bits to physical streams** for each **Trigger Menu** and **Stream Definition Table** for further use in the offline streaming and physics analyses
- In order for system to be robust, **number of physical streams** should be manageable (**~20 or less**)
- **Hierarchical mapping** from bitmap to physical stream **is preferred**

Exclusive Streaming Basics

- Exclusive streaming is based on mapping of 2^N-1 possible combinations of N Level 3 primal streams into n physical streams, where n is small
- The association can be done on the bit mask basis:
 - Define n physical streams by creating n unique N -bit masks OR of which is an N -bit word with all the bits set (*complete*), and AND of any of the two bitmask is equal to zero (*exclusive*)
 - An event with certain L3 primal stream bitmap is assigned to the only stream for which the result of logical AND between the primal stream bitmap and the corresponding bitmask is not equal to zero
- Alternatively, it can be done hierarchically:
 - Define $n-1$ unique bitmasks and order them hierarchically
 - Test an event against each of them and assign it to the stream, associated with the first bitmask it passes via AND
 - If the event fails all the bitmasks, it is put in the n^{th} , “Everything Else” stream
- Either of the two approaches is reversible, as long as the mapping between the Level 3 triggers and the primal stream bitmap is based solely on the AND operation, which is currently enforced, and must remain enforced in the future

Streaming and Online

- COOR provides **mapping between the stream ID and the file name** for the corresponding stream
 - Datalogger directs all the events with one and the same stream ID to one file
- COOR also provides **mapping between the stream ID and the file family** (physical tape)
- Based on current resources, the following **constraints** are dictated by the online system:
 - Total number of **stream ID's: 13-16 or less**
 - Total number of **file families: 8 or less**
- COOR and Datalogger ensure that **one luminosity block is not spanned across the file boundaries** by keeping both the old and the new files open for some time, driven by the DAQ/L3 time-out
- Nevertheless, **there could be gaps in the luminosity blocks** in the files that correspond to rare streams
 - There are ways to **recover these gaps** using the **luminosity database**

Luminosity Tracking

- Each file always starts and ends on a boundary of a luminosity block (60-second long)
- All the luminosity-related information per trigger, per luminosity block is stored in the luminosity database
- In order for events not to get lost from low-rate streams due to time-outs and to avoid empty luminosity blocks, there should be no streams with size <1% of the entire data set (~10 events/block)
- Events in a given file might not be ordered by the luminosity block – a certain care is required
- Information about the luminosity blocks in a particular raw data file must be stored in the database or in the file itself
- Luminosity information for any processed file (i.e. filtered, merged, etc.) file can be traced through the inheritance tree to the original raw data file(s) and thus does not have to be stored for every file
- Special care is required for the cases when one trigger is split over several physical streams:
 - A loss of a single partition in any of these streams would result in the lack of means to recover the exact luminosity information for the same luminosity blocks in other streams that the trigger spans
 - Therefore, these luminosity blocks in all other streams corresponding to a given trigger must be marked as “dead”
 - Data from “dead” luminosity blocks must not be used if the exact luminosity knowledge is required

Physics Considerations

- Physics analyses naturally **fall into one of the three categories**:
 - **Luminosity-sensitive analyses**, which care about the exact luminosity information
 - **W/Z cross section measurement; top cross section measurement**
 - **Statistics-sensitive analyses**, which care about processing as high fraction of the collected data as possible, ideally all 100%
 - **Search for new physics; W-mass measurement**
 - **Systematics-limited analyses**, which do care neither about exact knowledge of integrated-luminosity, nor about full statistics
 - **$\sin(2\beta)$ measurement; jet color flow measurement**
- The first two categories **use orthogonal approaches**; the third one can use either one of the two:
 - **Cross section measurements** do not mind to **sacrifice small fraction of statistics** in order to **ensure that the luminosity is known** precisely
 - **Searches** do not mind a **slight increase in the luminosity uncertainty** in order to **ensure that as much data as possible** is processed
- Hence, the **streaming should provide solutions for both** types of analyses:
 - **Remove data from “dead” luminosity blocks** when delivering a data set **for a luminosity-sensitive analysis**
 - **Keep all available data** in the data sets **for statistics-sensitive analyses**, and **estimate** the fraction of **lost luminosity** as precise as possible based on the luminosity blocks, adjacent to the lost ones

Stream Definition Table

- **Stream Definition Table (SDT)** is a **database algorithm** that encodes the **mapping between the primal streams** and the **physical streams**
- **SDT** contains the **entire streaming logic** for a given trigger menu
- All **physics-driven decisions** on the number, priority, and definition of the physics streams **are reflected** in this table
- The goal is to have **SDT change less frequently than the Trigger Menu**
- **Several SDT's** for different instantaneous luminosities can be **provided per Trigger Menu** in order to reflect changing trigger prescales
- **Mapping** of the Level 3 **primal streams** in the SDT **must be reversible**, i.e. it must be possible to determine for every Level 3 trigger which streams it can go into
 - A **utility** that builds the reverse association **exists** (**Zhong Yu**)
- **Trigger database** is the **natural place for SDT's** to be kept

Run Summary Database

- **Run Summary Database** contains the **record of what actually happened** in the run, as opposed of what was expected to happen (e.g., actual prescales vs. desired prescales)
- Apart from this crucial information, it **will contain other useful information for streaming**, such as:
 - The **first and the last luminosity blocks** for any run
 - **Trigger Menu** version and **SDT version** used in each run
 - Actual **mapping between the physical streams and the file families**
 - Possibly, the **luminosity information** for each trigger, integrated over **each run**
 - **Mapping between the Level 3 triggers and physical streams**
 - Additional **redundant information**, which is **useful for** fast look-up of the parameters needed for the offline **streaming**

Strawman Streaming Proposal

- Basic considerations:

1. Keep the most luminosity-sensitive **W/Z cross section data** in as **few streams** as possible
2. Provide **separate streams** for main types of objects identified by the DØ detector
3. Keep the **number of streams small** for possibility to store them on separate tapes
4. Keep the **size of the streams more or less equal** for the most efficient streaming

- Reality check:

- ~**10% of Run I data** taken with **trigger version 10.2** spanning runs 87804-89517, corresponding to **5,671,746 events**
- Although **trigger terms will be different**, **bandwidth allocation** to Run 2 physics goals **is expected to be similar**

Run I Trigger Bandwidth Studies

Trigger name	Brief Description	Number of Events	Prescale	Effective fraction
EM_1_MON	$P_T(e) > 16 \text{ GeV}$	28,154	100	33%
MU_1_MAX	$P_T(\mu) > 15 \text{ GeV}$	136,792	1	2.4%
MISSING_ET	$ME_T > 40 \text{ GeV}$	115,410	1	2.0%
MU_JET_CAL	$P_T(\mu) > 10 \text{ GeV}$ $E_T(j) > 15 \text{ GeV}$	83,375	1	1.5%
MU_ELE	$P_T(\mu) > 8 \text{ GeV}$ $E_T(e) > 7 \text{ GeV}$	300,548	1	5.3%
JET_MULTI	$E_T(j) > 10 \text{ GeV}$ $N(j) \geq 5$ $H_T > 115 \text{ GeV}$	141,598	1	2.5%
JET_4_MON	$E_T(j) > 10 \text{ GeV}$ $N(j) \geq 4$	21,187	25	8.5%

Strawman Streaming Table

Priority	Stream Name	Stream Definition	Comment
1	ELE_HIGH	$P_T(e) > 10-15 \text{ GeV}$	W/Z cross section in electron channel; New physics; Top/Higgs
2	MU_HIGH	$P_T(\mu) > 5-10 \text{ GeV}$	W/Z cross section in muon channel; New physics; B-physics; Top/Higgs
3	PHOTON	$P_T(\gamma) > 10-15 \text{ GeV}$	New physics, QCD, Electroweak
4	TAU_B	$P_T(e) > 3-5 \text{ GeV}$ or $P_T(\mu) > 3-5 \text{ GeV}$ or τ -jet or Isolated track with $P_T > 5-7 \text{ GeV}$ and low occupancy	New physics; W/Z cross section in τ -channel; B-physics; Top/Higgs
5	MISSING_ET	$ME_T > 25-40 \text{ GeV}$	New physics; Top/Higgs
6	JET_MANY (or some other jetty stream, e.g. HT)	$P_T(j) > 15-20 \text{ GeV};$ $N(j) > 3-4$	New physics; Top/Higgs; QCD
7	QCD	Everything else	QCD; Top/Higgs

Physical Stream Grouping

- Online system allows only for a **small number of file families** (currently **eight**)
- **Stream Definition Table** should also **specify preferred stream grouping** in the case of limited resources (e.g., temporary resource allocation by calibration or special runs)
- **Actual grouping** used in a particular run **is stored in the Run Summary Database**

# of file families	Grouping						
7	MU_HIGH	ELE_HIGH	PHOTON	TAU	MISSING_ET	JET_MANY	QCD
6	MU_HIGH + ELE_HIGH		PHOTON	TAU	MISSING_ET	JET_MANY	QCD
5	MU_HIGH + ELE_HIGH		PHOTON + TAU		MISSING_ET	JET_MANY	QCD
4	MU_HIGH + ELE_HIGH		PHOTON + TAU		MISSING_ET + JET_MANY		QCD
3	MU_HIGH + ELE_HIGH		PHOTON + TAU		MISSING_ET + JET_MANY + QCD		
2	MU_HIGH + ELE_HIGH + PHOTON + TAU				MISSING_ET + JET_MANY + QCD		
1	MU_HIGH + ELE_HIGH + PHOTON + TAU + MISSING_ET + JET_MANY + QCD						

People Currently Involved

- Online:
 - Jerry Guglielmo
 - Stu Fuess
- Level 3:
 - Amber Boehnlein
 - Jon Hays
 - Heidi Schellman
- Online databases
 - Elizabeth Gallas
- Offline databases:
 - Lee Lueking
 - Vicky White
- Luminosity:
 - Michael Begel
- Stream definitions, physics and technical issues:
 - John Hobbs
 - GL
 - Harry Melanson
- Most of the key players are oversubscribed - help is needed with the implementation!

Schedule and Plans

- Streaming Web page:
<http://hep.brown.edu/users/Greg/streaming/index.htm>
- **Weekly meetings** (time/place to be finalized)
- Achieved **basic understanding of the communication** between the online and offline systems **and constraints**
- **Streaming TDR** is being written
- **Will implement strawman trigger model for initial physics data-taking** this summer
- **Will use the experience of initial running to fine-tune the model** and add new streams