D0 Note # 33nn
Draft 3.0

September 19, 1997

Study of Possible Run II Streaming Scenarios using a Run 1 Data Sample

Heidi Schellman, Iain Bertram
Northwestern University

1 Introduction

The new phenomena group [reference] have proposed an exclusive streaming model for Run II which achieves the convenience of streamed data without the substantial multiple copy overhead that plagued the run 1 streaming. The basic idea is to classify objects in an event into n types, for example: ' jets, electrons, missing et, hi pt muon, low pt muon, min-bias' and to classify events into $2^n - 1$ streams based on the objects in that event. For example, an event with 2 jets would be classified into the 'jet' stream, an event with 1 jet, an electron and missing ET would be classified into the 'jet + electron + missing ET' stream.

In this note we attempt to test this method on 100,000 high luminosity events taken with trigger list 9.0. The trigger bits for each event are decoded and the event is classified based on the trigger bits.

We then propose possible streaming schemes based on this classification.

2 Study Method

Each trigger is assigned a classification into one or more of 6 basic types. The classification is as follows:

Table I: Trigger types used in stream decisions
Triggers which fall into multiple classifications are given a type which is the OR of the multiple types.

Type	Type Code
Global Monitor	0
Jet	1
Electron	2
Missing ET	4
High pt mu	8
Low pt mu	16
Luminosity	32
Electron + Jet	3
High pt mu + Missing ET + electron + jet	15
etc ...	

The luminosty type includes Luminosity triggers, inspill, min-bias, zero-bias. Triggers which are classified as belonging to more than one basic type are assigned to a type which is the OR of the two (or more) basic types. An electron+jet trigger would be assigned type $(1+2) = 3$. At least in version 9.0, triggers had 1 basic type most of the time, 2 basic types occassionaly and 3 basic types (em2_ele_emf_ms has jet, electron and missing ET types $->$ type $= 7$) in only one case. (This coding method, where each type is assigned a bit will allow us to construct coded stream numbers which contain within them the full information about the content of the stream.)

The types corresponding to all triggers which fired in the event are then ORed to get the final event classification. An event in which a electron + missing ET trigger fired and a high PT muon fired would have two triggers, one assigned type (2+4) = 6 and the other assigned type 8. The event would be classified as type (6+8)= 14. This procedure assigns a unique type to each event which reflects the contents of that event. Someone wishing to study all events with electrons in them, would choose events in which the assigned stream type had bit 2 set in it (2,3,6,7,10,11,14,15 ...). Someone wishing events with jets and electrons would choose all events with stream type bits 1 and 2 set (3,7,11,15 ...).

Table II lists the classification for individual trigger bits in version 9.0. Note that triggers can be assigned more than one basic type.

Table II: Trigger Type Classification for Version 9.0

Trigger	Trigger bit	Type Assigned	Number of Times Fired
	0	0	0
jet_30	1	1	6674
mu_1_lo_jet	2	17	3040
mu_2_lo_y4	3	16	338
mu_2_low	4	16	270
mu_1_high	5	8	2956
mu_2_high	6	8	1012
em1_eistrkcc_ms	7	6	897
em2_eis_ms	8	4	161
mu_1_low	9	16	13214
ele_1_miss	10	6	369
jet_gap_nmi	11	1	133
ele_jet_high	12	3	1962
mu_ele	13	10	159
mu_jet_cal	14	9	1535
jet_gap_lhe	15	1	1124
jet_multi	16	1	2053
emjet_20	17	3	1586
jet_4_mon	18	1	2196
eletrk_1_mon	19	2	5167
missing_et	20	4	2183
jet_3_miss	21	5	179
jet_1_veto	22	5	2577
jet_far_end	23	1	7333
jet_50	24	1	21016
jet_85	25	1	2151
jet_max	26	1	520
jet_gap_low	27	1	2700
zero_bias	28	32	534
inspill	29	32	0
lumin_mon	30	32	197
heartbeat	31	32	0

Table II: Trigger Type Classification for Version 9.0 - Continued
Trigger names not shown

Trigger bit	Type Assigned	Number of Times Fired
32	0	546
33	17	43
34	8	296
35	8	532
36	8	1440
37	8	57
38	6	549
39	2	1651
40	2	1575
41	2	565
42	2	723
43	0	248
44	6	532
45	2	1407
46	2	1271
47	6	132
48	3	1442
49	2	221
50	2	1589
51	2	556
52	2	408
53	2	254
54	2	203
55	2	1034
56	7	751
57	6	129
58	1	2752
59	3	380
60	3	757
61	3	1795
62	3	601
63	10	42

Table II: Trigger Type Classification for Version 9.0 - Continued

Trigger bit	Type Assigned	Number of Times Fired
64	10	32
65	10	70
66	9	2017
67	9	1240
68	9	1497
69	1	1612
70	1	1843
71	1	297
72	1	208
73	1	556
74	1	3700
75	5	1968
76	1	214
77	3	783
78	1	1409
79	1	3331
80	1	2078
81	3	110
82	2	7973
83	4	682
84	5	452
85	5	60
86	5	422
87	5	600
88	0	2599
89	0	157
90	1	2504
91	32	409

3 Results for full exclusive streaming:

Table III shows the effects of event type classification. Each of the 100,000 events is assigned a unique type and 64 types (0-63) are possible. The last column shows how many events could have been assigned that type based on a single trigger, it is a measure of how often an event is reclassified because it had overlapping triggers which had different physics types. This is not a very informative number.

One finds that the basic event types, (jet, electron, missing et, high pt muon, low pt muon and luminosity), are reasonably well populated. There are also some other popular event types, (jet+Electron) and even 3 way combinations (Missing ET + Electron + Jet) which have significant populations. However, the 5-way type (31) has only 2 events and some types (30 - everything but a jet) have no population. This indicates that the 2^n event types could be mapped into a smaller number of physical streams without any real loss of functionality.

4 Possible Physical Stream designs

Given these data, we can see how many physical streams are really needed to cover the 64 logical streams proposed above.

First, the monitoring types should probably not be treated via the exclusive streaming method. They are a small fraction of the events, and should probably be written out as separate LUM, MIN, ZERO, INSPILL streams as they are vital for calibration. Events which also satisfy one of the main physics triggers (jet, electron) should also be written out to the relevant physics stream. These events show up in event types 33-63 and there are only 201 of them for a net overhead of 0.2%.

In addition, one could combine several logical streams which represent less than 1% of the total into a single physical stream. The merged streams must be small as anyone wanting any of the merged attributes will have to go through the whole stream to find those few events.

4.1 An example stream merging scheme

For the runs studied the muon triggers have little overlap with the other triggers aside from the jet trigger. The pure muon and muon+jets streams are quite large, the rest of the triggers can be sorted into two disjoint classes, muon + nojet + other and muon + jet + other where other stands for any type besides jets. These merged streams are quite small. This particular streaming scheme is shown in Table IV. The 64 physical

Table III: Event Type Classification for Version 9.0

Event type	Type Code	Number of Events assigned that Type	Triggers assigned that Type (not very useful)
Global Monitor?	0	0	3550
Jet	1	47460	66404
Electron	2	11465	24597
Electron + Jet	3	7778	9416
Missing ET	4	876	3026
Missing ET + Jet	5	5720	6258
Missing ET + Electron	6	780	2608
Missing ET + Electron +Jet	7	1707	751
High PT muon	8	4514	6293
etc. ..	9	1887	6289
	10	100	303
	11	57	0
	12	1	0
	13	56	0
	14	2	0
	15	21	0
Low PT muon	16	13218	13822
etc...	17	2715	3083
	18	0	0
	19	7	0
	20	0	0
	21	12	0
	22	0	0
	23	3	0
	24	166	0
	25	280	0
	26	0	0
	27	27	0
	28	0	0
	29	10	0
	30	0	0
	31	2	0

Table III: Event Type Classification for Version 9.0 - Continued
These are luminosity, min-bias etc...

Event type	Type Code	Number of Events	
Monitor	32	939	1140
etc ...	33	124	0
	34	14	0
	35	12	0
	36	2	0
	37	13	0
	38	0	0
	39	3	0
	40	0	0
	41	7	0
	42	0	0
	43	0	0
	44	0	0
	45	1	0
	46	0	0
	47	0	0
	48	15	0
	49	5	0
	50	0	0
	51	0	0
	52	0	0
	53	0	0
	54	0	0
	55	0	0
	56	0	0
	57	1	0
	58	0	0
	59	0	0
	60	0	0
	61	0	0
	62	0	0
	63	0	0

Table IV: Exclusive Stream design for Version 9.0
Logical Streams and Physical Streams
Numbers not corrected for 199 overlaps between Streams 1-31 and Min/Zero-bias

Stream Name	Type Codes for stream	Number of Events	Physical Stream Code
Jet	1	47460	1
Electron	2	11465	2
Electron + Jet	3	7778	3
Missing ET	4	876	4
Missing ET + Jet	5	5720	5
Missing ET + Electron	6	780	6
Missing ET + Electron +Jet	7	1707	7
High PT muon	8	4514	8
High PT muon + jet	9	1887	9
Merged High PT muon + nojet + others	10, 12, 14	103	14
Merged High PT muon + jet + others	11, 13 , 15	135	15
Low PT muon + nojet	16	13218	16
Low PT muon + jet	17	2715	17
Merged Low PT muon + nojet + others	18, 20, 22, 24, 26, 28, 30	166	30
Merged Low PT muon + jet + others	19, 21, 23, 25, 27, 29, 31	360	31
LUMINOSITY	32	197	32
ZEROBIAS	32	534	33?
MINBIAS	32	409	34?
INSPILL	32	0	35?

streams have been reduced to 19 of which only 5, the muon and monitoring streams, are merged.

5 Criticisms of the fully exclusive model

The fully exclusive model is very efficient, as it reduces the number of events written out to the absolute minimum but it does introduce some problems.

1. Quality cuts - we may wish to filter on quality cuts after RECO. For example, after RECO, we know much more about the event quality and most analyses reject a large fraction of events immediately. Events could be split into two further streams, high quality for things like jets with efficient triggers, electrons

with hits in a road ..., low quality for jets in inefficient regions, electrons with no evidence for a track ... The general user would concentrate on the high quality sample while those doing background and efficiency studies could still use the full sample. The different streaming models adapt themselves to quality streaming in different ways.

If the jet stream, for example, were inclusive and contained all jet events, one could require that all events in that stream possess a good jet at high efficiency to earn a good quality rating. But if jet events are divided between a jet stream and a combined electron-jet stream, one can no longer classify events with bad jets as low quality if they would pass the electron quality cuts. This forces different quality cuts for each stream, even though most analyses would include many stream. In principle one can flag the events that would have been called low quality and then make that cut at the analysis but the potential for user error in using a sample which has had a cut already applied in some cases but not others is large.

2. RECO tailoring - One cannot run tailored versions of RECO on different samples in a fully exclusive scheme. If one finds a fix (say muon tracking) which affects one analysis (muons) but not another (jets) one cannot reRECO the low PT muon + jet streams without either reprocessing the jet stream as well or keeping two copies of the muon+jet stream, one for jet analysis and one for muons.

3. Luminosity - it is quite likely that events with, for example, no electron trigger have a different luminosity dependence than events with an electron trigger. Thus the relation between number of events and luminosity in the 'jet' sample will not be the same as in the 'jet-electron' sample. This can introduce problems in luminosity book-keeping. While we know the luminosity for the 'jet' triggers, how it is divided among the exclusive subsamples which contain jets is less well defined.

4. Data integrity - this is related to Luminosity, in the old days, if we lost a raw or processed tape, randomly distributed events would be lost from within the larger sample. This only messed up the normalization. The integrated luminosity could easily be corrected for this loss. If one divides data up into exclusive subsamples, loss of a tape from one of the samples (worst case electron+high pt muon +missing ET + jet) will skew physics distributions as well as messing up the normalization. In addition, important samples may be concentrated on a small number of tapes. In such a model, the important small substreams will need duplicate copies, which defeats some of the purpose of exclusive streaming.

5. IO constraints - Multiple streams being written to a small number of tape drives means that some streams must be buffered while other streams are being written out. In the worst case, which is rare but possible, each stream would need close to a full tape worth of buffer space as one cannot guarantee that all streams buffers would not accidentally be 80% full.

6. Data access - If streams are merged to get around the IO constaints, then some inclusive measurements, such as electrons, may have to read many events not in that classification from merged streams. The proposed streaming method above was designed to prevent this case for jets, where the overhead is highest, but it cannot be prevented for all event types without inflating the number of streams.

6 The quasi-exclusive model

Another model, in which some inclusive streaming (for example of jets and monitor triggers) is done before streams are split up may be able to retain most of the advantages of the full exclusive streaming mode but provide less complex access to the data.

In this model, the data would first be streamed into 4 inclusive streams

- 1 - All jet triggers
- 14 - All triggers with any of electron, missing ET or high PT Muon
- 16 - All low Pt muon triggers
- 32 - All monitoring streams

Run 1 experience indicates that few people ever try to do analyses that cross these rough stream boundaries.

Any given event could go into any of the above streams that applied to it. Then these major streams would be split using the exclusive model described earlier. The size of the 4 primary streams is shown in Table V. The duplicate event overhead is 20% but may make analysis much simpler as it decouples samples which overlap incidentally.

Table V: Number of events assigned to 5 Exclusive Streams
In the second streaming model.
Events can occur in more than one stream

Stream	Type Code	Substreams included	Total Number
Jet	1	all odd	67908
Electron			
Missing ET			
High pt mu	14	any with bits 2,4,8	35524
Low pt mu	16	any with bit 16	16461
Luminosity	32	any with bit 32	1136
Total			121029

7 A Bad Idea - the old fashioned way

Finally, we looked at the number of events which would be written out in an inclusive streaming system in which each event was assigned one or more of the 5 types and events were written to as many streams as were applicable. The 5 streams are shown in Table VI. The result was 173,449 events written out instead of the 100,199 in the exclusive streaming model.

Table VI: A not overly bright streaming method
Number of events assigned each basic type
Events can occur in more than one type

Type	Type Code	Total Number
Jet	1	92201
Electron	2	37675
Missing ET	4	12643
High pt mu	8	12885
Low pt mu	16	16905
Luminosity	32	1140
Total		173449

8 Implementation

If we assume a reasonable Run II model:

- Data rate 40 Hz
- Event size 250 KB
- Tape drives which can write 10MHz/second
- Tapes which hold 100 GB
- No more than 100 files/tape
- Don't waste tape

The data rate is enough to keep 1 tape drive 100% busy. Any reasonable system must have 2 drives in case the event size or rates are higher or the drive speed is not the limiting factor. At least one more drive as a spare will be needed.

At these rates it will take around 3 hours to fill a tape. We can begin to estimate the impact of the 3 possible streaming modes.

Note that streams can exist at the file or tape level. A tape does not necessarily have to consist of 100 files, all from the same stream. Once a 1GB file contains a single stream, the overhead of checking each event is overcome and stream choice becomes a matter of accessing a known file on a tape. If streams which are likely to be read together are put on the same tape, tape mount overhead can also be reduced.

We consider the 3 streaming scenarios listed earlier.

8.1 N fully inclusive streams

One would just write the data out to N streams without further subdivision. Due to the overlap overhead, the output bandwidth needed would be around 18MHz, therefore at least 3 drives + spares are needed. If N>N_drives, some drives will need to alternate between streams. With three drives and 6 streams, 3 streams must always be buffered. As we wish to fill tapes, sufficient staging area would be needed to hold an entire tape full of the 3 streams not writing to tape at that moment + some cushion for mounts and file access. Staging needs are then 300-400 MB.

In order to access the data, only the tapes of the correct physics type for the analysis would need to be mounted.

8.2 4 exclusive streams with inclusive subdivisions

In this case the duplication overhead is 20% and 12MhZ output rates are needed. This is probably OK with 2 10MHz tape drives. Each inclusive stream would go to its own tape (thus needing alternation for the smaller streams) but within the major stream, substreams would be kept separate at the file level. With 4 streams and 2 drives, staging area for the two streams not being written + overhead will be needed.

Once again, the number of tape mounts for a given analysis is limited, although not as low as in the fully inclusive case.

8.3 Fully Exclusive streams

Here the total amount of data being transferred is lowest but, depending on the logic used, the amount of staging could be very high. If the 20 streams were each written to their own tapes, one could forsee a worst-case need for enough space for (20−N_drives)*100 GB worth of data waiting to be dumped. Logical combinations of streams with similar characteristics onto the same tape could reduce this down to the levels for the other options. For example, the jets stream would go to one tape, types 2-7 to another types 8-15 to another, types 15-31 to another, type 32 to another. This gives 5 tape streams.

Unfortunately, an analysis which required, for example, jets, would have to mount almost all tapes to get at the subsample of exclusive streams which contained jets.

9 Conclusions

We have used run I data to investigate streaming scenarios.

- A fully inclusive scenario with separate streams for individual physics analyses makes efficient use of analysis tape mounts and is robust against data loss but is prohibitively expensive in IO bandwidth and media costs.
- A scenario with fully exclusive streams minimizes data duplication but is vulnerable to data loss and removes flexibility in RECO quality cuts and reprocessing of special samples. Data access may be significantly slowed if streams are merged either at the tape or file level. But without such merging, buffering needs while writing the data to tape can become prohibitive.
- A mixed scheme with major inclusives streams at the tape level but exclusive substreams at the file level creates event duplication at the 20% level. Data loss and bookeeping are also a problem in this scheme. However, the buffering needs are similar to those for the fully inclusive case and much greater flexibility exists in optimizing RECO for each stream type.

Conclusion: We'd better think hard about this, these different scenarios have media/drive/disk costs which differ by large sums. In addition, the different schemes can have profound effects on data access and reconstruction patterns.

10 Thanks

Our thanks to Amber Boehnlein for suggesting the study and providing the version 9.0 trigger list. To Meenakshi Narain, Norm Amos and Joey Thompson for useful discussions on luminosity. To Kurt Ruthmansdorfer, Don Petravic, Lee Lueking and Vicky White for comments on the data handling aspect.