

Event Builder / Level-3 Upgrade for CDF Run IIb

Christoph Paus, MIT

Lehman Review
September, 2002

Overview of CDF DAQ

Event Builder IIa

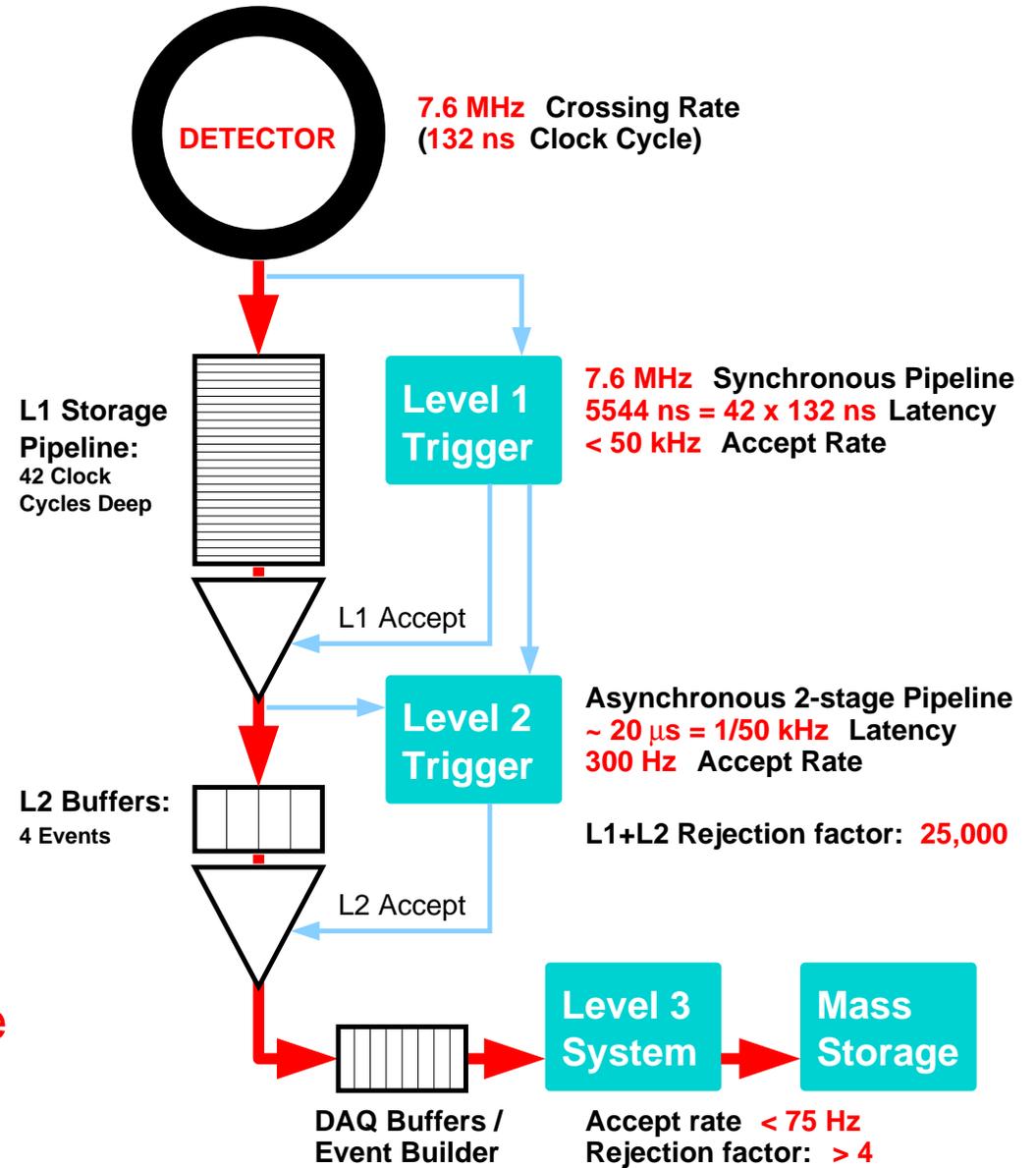
ATM switch	32 ports
input rate [Hz]	300
event size [kB]	150
total flow [MB/s]	44

Level-3 Processing IIa

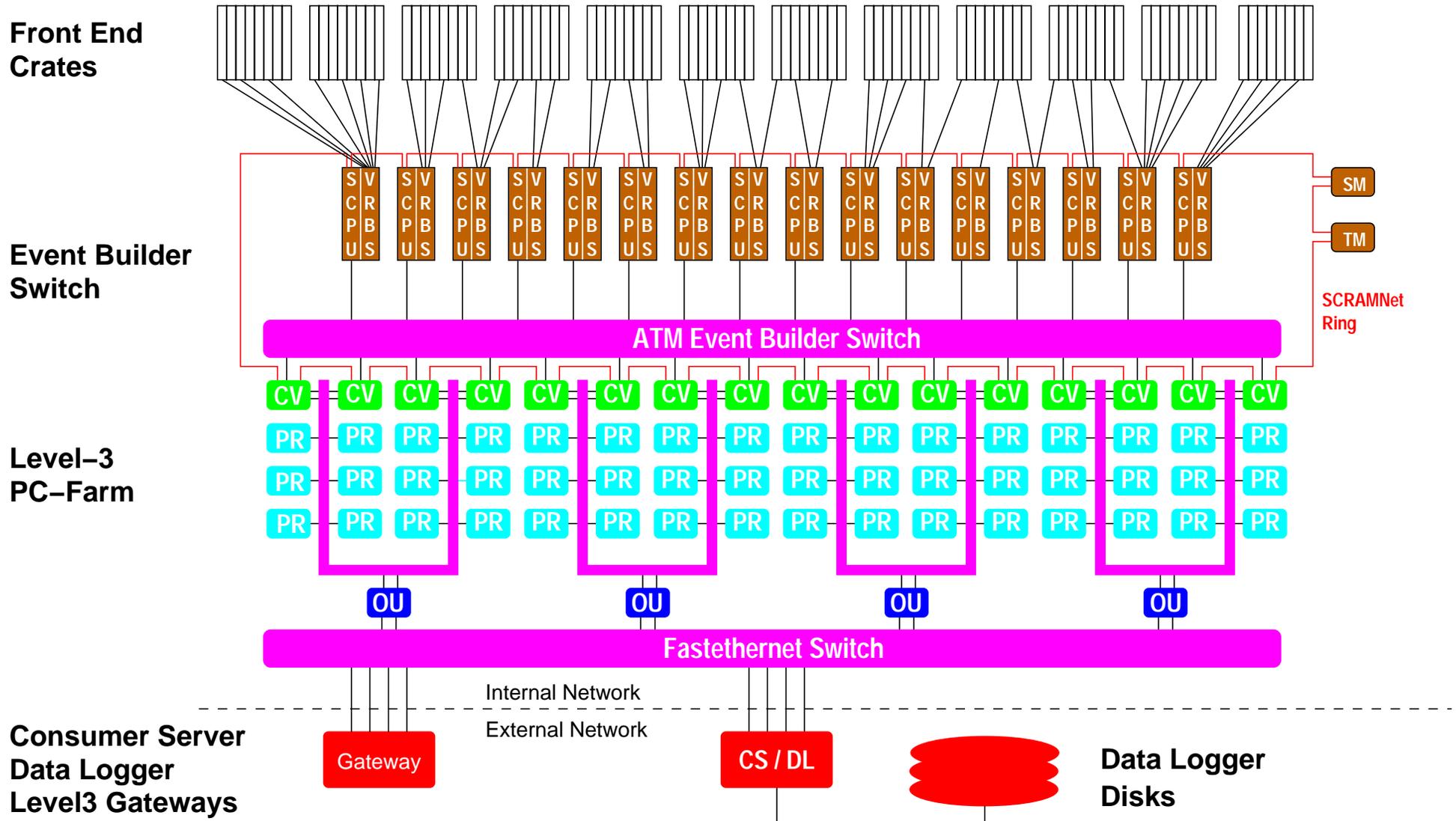
PC workers (dual)	144
input rate [Hz]	300
output rate [Hz]	30
rejection rate	≈ 10
logging flow [MB/s]	4.4

Status

- ☞ completed in time on budget
- ☞ 1 year operation under battle condition
- ☞ no limitations to data taking



Event Builder and Level-3 PC Farm



Performance Specifications

Event Building, Upgrade

	Ila Spec	Ila now	Ilb spec
L2 accept rate [Hz]	300	300(500)	750
event size [kB]	150	250	500
total flow [MB/s]	44	70-120	375

Level-3 Processing, Upgrade

	Ila Spec	Ila now	Ilb spec
L3 input rate [Hz]	300	300(500)	750
output rate [Hz]	30-75	75	85
workers [duals]	144	270	270

Event Builder Performance - Rate

Theoretical Limit

$$\frac{\text{ATM link bandwidth} * N_{\text{SCPU}}}{\text{Average EventSize}} = \frac{16 \text{ MB/s} * 15}{250 \text{ kB}} \approx 1000 \text{ Hz}$$

Practical Limitations

- ➡ non uniform data distribution
- ➡ data volume fluctuations
- ➡ limitations in other parts of DAQ (ex. L2)
- ➡ protocol overheads, latencies in control

Performance so far

- ➡ reached 300 (500) Hz sustained rate

Event Builder Performance - Data Volume

Theoretical Limit

$$\text{ATM link bandwidth} * N_{\text{SCPU}} = 16 \text{ MB/s} * 15 \approx 240 \text{ MB/s}$$

Practical Limitations

- ➡ non uniform data distribution
- ➡ data volume fluctuations
- ➡ limitations in other parts of DAQ (ex. L2)
- ➡ protocol overheads, latencies in control

Performance Run IIb estimate

- ➡ event size: 500 kB
- ➡ event rate: 750 Hz (1.1 kHz peaks)
- ➡ data volume: 375 MB/s (550 MB/s peaks)

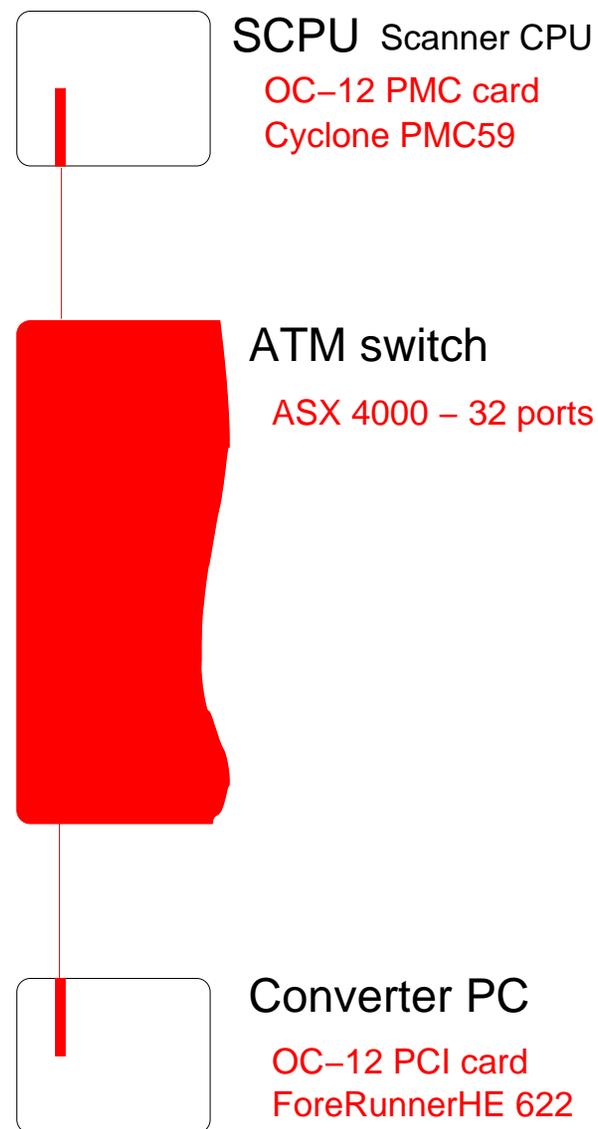
Event Builder Upgrade

Upgrade Strategy

- ☞ needs to fit into present system
- ☞ required rate increase: factor of 2
- ☞ minimal change to the software
- ☞ upgrade OC-3 to OC-12 connections
- ☞ new ATM switch ASX 4000

Quote from December 2001

ASX 4000 32 ports	1	\$215k
OC-12 PMC card	15	\$30k
OC-12 PCI card	16	\$60k
System Cost		\$305k
Spares ASX 4000	1/1	\$91k
Spare PMC card	3	\$12k
Spare PCI card	3	\$6k
System Cost with spares		\$414k
Total Cost (incl. 30% cont)		\$538k



Event Builder: Schedule / Resources

Software development

- ➡ **new Linux/PCI, VxWorks PMC drivers**
- ➡ adjust SCRAMNet control software
- ➡ adjust SCPU and L3 receiver code
- ➡ start end '03
- ➡ almost 1 year
- ➡ 2 students 50%, 1 researcher 20% (MIT)

Hardware upgrade

- ➡ construct prototype: Feb '04 (70 days)
- ➡ construct full system: May '04 (261 days)
- ➡ system ready May '05
- ➡ 2 students 50%, 1 researcher 20% (MIT)

Commissioning

- ➡ commissioning: May '05 (101 days)
- ➡ need DAQ/detector available
- ➡ need real data for final tuning
- ➡ finished until Oct '05 finished
- ➡ 2 students 50%, 1 researcher 20% (MIT)

Contingency: 1 computer expert

Summary

Event Builder Upgrade

- ➡ need factor of 2 upgrade in data throughput
- ➡ choose economic scheme: as similar as possible
- ➡ ATM switch with OC-12 ports: bandwidth factor 3-4
- ➡ \$414k including spares
- ➡ \$538k including spares and contingency (30%)
- ➡ human resources 1.5 + 1 FTE (2 years)
- ➡ additional labor available in emergency

Level-3 PC Farm / DAQ

Run Ila Purchase Scheme

- 👉 buy PCs as late as possible, staged
- 👉 1998: 20 nodes (200 MHz)
- 👉 1999: +30 nodes (300 MHz)
- 👉 2000: +100 nodes (450 MHz)
- 👉 2001: +120 nodes (850 MHz)

Level-3/DAQ Maintenance

- 👉 after 3 years PCs are obsolete
- 👉 performance growth by factor of ≈ 3
- 👉 potential hardware failures
- 👉 move to rack mount for L3

Level-3 Upgrade

- 👉 increased occupancy \rightarrow increased processing

Scheme for Maintenance / Upgrade

- 👉 buy new PCs (210 L3 / 45 DAQ) total
- 👉 price per PC is about \$1500
- 👉 total cost: \$383k (over three years)

2 students 50%, 1 researcher 20% (MIT)

Level-3 Monitor

Level 3 Display _ □ ×

File Edit Help

Converters

c01	c02	c03	c04	c05	c06	c07	c08	c09	c10	c11	c12	c13	c14	c15	c16
0	42941	22482	22463	22400	22460	24437	21450	19662	27320	24268	22555	22517	24199	25287	23485
0	42947	22446	22473	22388	22441	24427	21412	19663	27324	24272	22542	22483	24220	25295	23475
0.0 Hz	40.7 Hz	18.8 Hz	19.5 Hz	19.1 Hz	20.0 Hz	20.8 Hz	18.0 Hz	16.8 Hz	23.6 Hz	21.0 Hz	19.7 Hz	19.4 Hz	20.1 Hz	21.6 Hz	19.9 Hz

Processors

001	019	035	051	067	083	099	115	131	145	159	173	187	201	215	229
002	020	036	052	068	084	100	116	132	146	160	174	188	202	216	230
003	021	037	053	069	085	101	117	133	147	161	175	189	203	217	231
004	022	038	054	070	086	102	118	134	148	162	176	190	204	218	232
005	023	039	055	071	087	103	119	135	149	163	177	191	205	219	233
006	024	040	056	072	088	104	120	136	150	164	178	192	206	220	234
007	025	041	057	073	089	105	121	137	151	165	179	193	207	221	235
008	026	042	058	074	090	106	122	138	152	166	180	194	208	222	236
009	027	043	059	075	091	107	123	139	153	167	181	195	209	223	237
010	028	044	060	076	092	108	124	140	154	168	182	196	210	224	238
011	029	045	061	077	093	109	125	141	155	169	183	197	211	225	239
012	030	046	062	078	094	110	126	142	156	170	184	198	212	226	240
013	031	047	063	079	095	111	127	143	157	171	185	199	213	227	241
014	032	048	064	080	096	112	128	144	158	172	186	200	214	228	242
015	033	049	065	081	097	113	129								243
016	034	050	066	082	098	114	130								

Output nodes

u01	85921	0	u02	90079	0	u03	89481	0	u04	91663	0	u05	93922	0	u06	93711	0	u07	93500	0	u08	97599	0
40.3 Hz	2.8 MB/s		39.7 Hz	2.7 MB/s		39.7 Hz	2.7 MB/s		38.8 Hz	2.7 MB/s		38.4 Hz	2.6 MB/s		40.1 Hz	2.7 MB/s		40.9 Hz	2.8 MB/s		43.0 Hz	2.9 MB/s	

	State/Transition	Phase #/out of #	Time spent
Partition 0:	Active	In state: 1/1	00:38:40
Partition 1:	Not defined	In state: 1/1	02:05:29
Partition 2:	Not defined	In state: 1/1	02:05:29
Partition 3:	Not defined	In state: 1/1	02:05:29
Partition 4:	Not defined	In state: 1/1	02:05:29
Partition 5:	Not defined	In state: 1/1	02:05:29
Partition 6:	Not defined	In state: 1/1	02:05:29
Partition 7:	Not defined	In state: 1/1	02:05:29

0 1 2 3 4 5 6 7

Level3 Summary

	Total	Inst. rate
Input	722844	317 Hz
Reformatter rejected	0 (0.00%)	0.00 %
Filters rejected	0 (0.00%)	-----
Output	722844	320 Hz

Last heartbeat: Mon Apr 15, 12:43:12 2002
 Current time: Mon Apr 15, 12:43:16 2002

Ch. Paus, Lehman Review, September, 2002 - 10

Summary Maintenance/Upgrade

Level-3/DAQ PC Maintenance

- ➡ PCs are obsolete after 3 years
- ➡ purchase new hardware over several years
- ➡ cost of about \$400k until 2005

Level-3 PC Upgrade

- ➡ higher occupancy means higher processing time
- ➡ amount is difficult to estimate
- ➡ probably smaller than 2

About Recommendations

Alternative Technologies

At present only the ATM technology has been carefully evaluated at CDF and implemented for the Run IIa system. The upgrade in that directions seems straight forward although it will involve significant amount of work and is probably more expensive than some of the alternative solutions.

Other technologies like Gigabit ethernet or MyriNet provide on paper the necessary performance but have not yet been carefully studied at CDF. Before the system is bought alternative solutions, in particular Gigabit ethernet are going to be carefully evaluated. It is not unlikely that the Run IIb Event-Builder will in the end use one of the alternative technologies.

About Recommendations

Linux and VxWorks ATM Drivers / Upgrades

Since ATM is no commodity component drivers for the network cards, in particular the VxWorks driver, are an issue. For the existing system significant expertise with very similar cards has been acquired; one MIT researcher and one MIT student. The drivers have been developed in a time period of roughly one year.

At present the CPUs driving the switch are no limitation to its performance and there is no compelling reason for regular upgrades of the operating system. Upgrades of the drivers to a new version of the Linux operating systems have been performed twice and have in both cases not taken longer than two weeks.

About Recommendations

Risk Analysis: Linux and VxWorks ATM Drivers / Upgrades / Personnel

The expertise in the MIT group is concentrated in one researcher but has to some extent been propagated to another postdoc. It is at present unlikely that the researcher leaves the group and it is expected that he will propagate his expertise to two new students and another postdoc.

In general the group is computer savy and eventual departures could be accommodated by other personnel. For the unlikely case that there is a shortage of expertize we plan for one additional external computer expert, which would be able to help solve a potential problem. We see this as a contingency and believe the risk for the timely and proper completion of the project is low.

The presented plan is basically identical to the implementation of the Run Ila system which was completed on budget and very timely.