A PRIMER ON DETECTORS IN HIGH LUMINOSITY ENVIRONMENT

R. Huson, L. M. Lederman and R. Schwitters
Fermi National Accelerator Laboratory*
Batavia, Illinois 60510

I. History

The following remarks are relevant to the problem of balancing luminosity versus energy in new HEP construction.

In a 1973 Isabelle Summer study,[1] it was stated that the only experiment that would succeed at a luminosity of $10^{33}cm^{-2}sec^{-1}$ was one in which the apparatus was shielded from the collision region by massive quantity of steel. In 1981, this opinion was confirmed by an authority no less than S.C.C. Ting.[2] It may be instructive to review the progress of collider detectors over the past decade. In 1973, the time resolution or, better, the integrating time of tracking detectors was ~100 ns. In 1982, this time has remained the same since PWC's are still the fastest tracking devices available. The fundamental limit is the saturated drift velocity of electrons in gases. Better resolution and three dimensional properties have led to the choice of drift chambers and TPC's which have considerably longer integration times. A new characteristic of 1982 detectors is the increasing pervasiveness of calorimeters which have become indispensable devices for measurement of electromagnetic and hadronic energy, especially at momenta where magnetic measurements become imprecise. Calorimeters, because of their innate geometric dimensions set by the nuclear mean free path and their distance from the interaction point have integration times of ~200-1000 ns. Of course this is the present state of the art which depends on the properties of BBQ, gas chambers, liquid argon, lead glass, etc.

The conclusion is that things have only gotten worse since 1973.

II. Integration Time - Tracking

What are the implications of long integration times? We are facing collision energies so high that the charged and neutral multiplicities, $\overline{M}$ average about 60 particles near 1 TeV. These typical multiplicities have surprisingly large fluctuations, such that Gaussian or Poisson statistics do not apply.[3] For example, the probability of having 2 $\overline{M}$ particles is one quarter that of having $\overline{M}$ particles. A track detector that integrates over, say, N events (with its integrating time of $\gtrsim$100ns) must add N times the average multiplicity to the number of particles in the triggering event. If this is a typical hard collision it may well have a track multiplicity many times higher than the average multiplicity.[3] At $10^{33}cm^{-2}sec^{-1}$, ±100ns integrates over an average of 10 events. If each event generates an average of <u>30</u> charged[3] particles (and ~30 neutral particles) one must add <u>an average of</u> 300 particles to the trigger induced event. Not all of these will conveniently stay in the beam pipe. (See typical events attached.) According to UA1[3] an average of 50 particles enter the central calorimeter at $\sqrt{s}$ = 540 GeV in minimum bias events. Many others will strike flanges, supports, pole pieces, etc. and shower with very high multiplicities, the end products of which give rise to noise or albedo, i.e., single hits in detectors or random tracks. This has severe implications for tracking efficiency; there is in fact a fair likelihood that these high multiplicities will render any of the tracking devices, as we now understand them, inoperable. PWC's have operated at ambient singles rates of 10 Mcps with fairly simple track configurations. However, experience with 20-30 tracks, e.g., at the ISR's Split Field Magnet or at various multiparticle spectrometers suggest a CDC 7600 CPU analysis time <u>per event</u> of hundreds of milliseconds up to <u>~5 sec!</u> To contemplate the functioning of a track chamber with several hundreds of tracks, many of low and "curling" energies (even given scintillation tagging) clearly requires a major advance. As a dramatic example, look at Fig. 1 and imagine superposing 2, 3 or 5 such events in a single trigger.

We should note that before one can reject tracks for pointing incorrectly one must be able to do the pattern recognition. A more quantitative tabulation of the influence of finite integrating time is presented in Tables I and II.

III. Calorimetry

To this tale of woe we must add the problem of the calorimeters. Now we have ~30 charged and 30 neutral particles incident upon the calorimeter which has an optimistic integrating time of ±200ns. This is at ~1 TeV. Multiplicities will about double at 10 TeV. It is true that a typical event may add negligibly to a (say) 100 GeV/c transverse momentum trigger. Some fraction of good events would be confused by the integration, but it is also clear that a large enough number of random accumulations of 10 or 20 minimum bias events can generate fake physics. These may provide a background for a large fraction of the anticipated physics signatures. During the interval between real 100 GeV/c jets say (at the rate of 10 per day) there would be $\sim 5\times10^{11}$ accumulations of twenty random events! If each charged particle generates a transverse energy of 500 Mev[3] and each photon 250 Mev, a minimum bias event produces an average of ~20 GeV of E_t. Twenty events yields 400 GeV!! Gating may reduce this to ~200 GeV. A patient Monte Carloist can decide how often these will fluctuate and cluster so as to fake a PT = 100 GeV/c event. However, this intrepid soul must be sure he is using the correct distribution function for fluctuations around the "typical" minimum bias trigger. This does assume either a breakthrough in tracking or, more likely, ability to see jets without tracks.

IV. Current State of the Art

There is ample data from 1982 experiments that support this pessimism. Charm was discovered in 1975. In spite of eight years and three generations of experiments at Fermilab, ISR, SPS and AGS the total number of clear charm events observed in hadron collisions is about one hundred! Nevertheless, literally millions of charmed particles were produced in the targets of the dozens of experiments looking for charm. It is obviously even worse for bottom mesons. Why? The primary problem is that the hadronic production cross section is less than 0.1% of the total cross section. Then, high (5-10 tracks) multiplicities, combinatorials, backgrounds, i.e., the

*Operated by Universities Research Association, Inc. under contract with the U.S. Department of Energy.

cruel and unforgiving real world enters. Surely the difficulty is not the lack of luminosity.

Recent data from a BNL-CERN et al. group at the ISR[4] on jet production notes that "...remaining background (predominantly from pile-up) is less than 9% (21%) for E_T = 6(14)GeV. This background would tend to mask any existing structure of the real events." The average luminosity was 1.4×10^{31} and the mean multiplicity is 10 particles. The lesson is that using reactions (like charm production) to evaluate the luminosity requirements of a future accelerator is misleading unless realistic estimates of backgrounds are included.

It is useful to survey fixed target experiments. Those that subtend very large CM solid angles have generally been limited to $\sim10^6$ interactions per second, i.e., a luminosity of $\sim3\times10^{31}\text{cm}^{-2}\text{sec}^{-1}$ (σ_T=30 mb). Experiments that subtend smaller solid angles can of course use higher luminosity. The primary proton experiment that discovered the upsilon resonances at Fermilab was able to withstand 7×10^{11} interactions per second at the target with a 0.7% solid angle and with a shield of twenty feet of Beryllium to absorb all hadrons. Again we find a limitation at about 10^{31} (two factors of 10^{-3}). In all cases, attempts to increase the data rates lead to pile-up, chamber inefficiencies, domination of random rates and unreliable data.

V. CW vs Bunched Beams

In discussing the impact and counter measures for the influence of finite integration times, we must discuss bunched beams and CW beams separately. We start with a few theorems.

A. In bunched beams, one divides the luminosity between the number of collisions per crossing and the number of crossings per sec. The latter quantity is just the number of bunches per machine period. If the number of collisions per crossing is too high, Poisson fluctuations will give large numbers of collisions coming essentially simultaneously and these will either kill the trigger or, worse, generate fake physics. Suppose X is the probability per crossing of a collision with cross section σ which is the optimum balance between rate and safety. Suppose the bunch spacing Δ is adjusted to correspond to the integrating time of the detector. Then the luminosity is just

$$L = \frac{X}{\Delta\sigma} \qquad (1)$$

In CW beams, if Δ is the integrating time and we dare take no more than X collisions within that time, the luminosity is again

$$L = \frac{X}{\Delta\sigma} \qquad (2)$$

i.e., the same luminosity. It is not clear whether X dare be higher in CW beams vs. bunched beams. Scintillation counters can detect multiple events with a time resolution ≃10% of the integrating time in CW beams and long interaction length makes for multiple vertex detection in bunched beams. It may be a wash.

B. Suppose we have a "sure fire" way of knowing when we have multiple collisions within the bunch (say by detecting multiple vertices) and we desire to reject these events in order to have a clean sample of triggered events. If this rejection is 100% effective, then optimized luminosity for bunched beams is given by the following argument.

$$X = L\sigma\Delta \qquad (3)$$

i.e., the average number of collisions per bunch crossing is the number of collisions per second times the bunch spacing.

$$\text{No. of collisions/sec} = L\sigma \qquad (4)$$

$$\begin{array}{l}\text{No. of events/sec}\\ \text{with only 1 collision}\\ \text{in a crossing}\end{array} = \frac{P(1)}{\Delta} \qquad (5)$$

where

$$P(N) = \frac{X^N e^{-X}}{N!} \qquad (6)$$

$$= Xe^{-X} \quad \text{for } N = 1$$

$$\begin{array}{l}\text{No. of single}\\ \text{collision}\\ \text{events/sec}\end{array} = \frac{Xe^{-X}}{\Delta} \qquad (7)$$

$$\equiv L_{eff}\sigma \qquad (8)$$

$$L_{eff} = \frac{X}{\sigma\Delta}\, e^{-X}$$

$$L_{eff} = Le^{-L\sigma\Delta} \qquad (9)$$

This has a maximum of L/e for $L_{opt} = 1/\sigma\Delta$ i.e. X = 1. Note: For an optimistic bunch spacing of 200ns

$$L_{opt} = \frac{10^{-9}}{5\times10^{-26}\times200} = 10^{32} \text{ (machine luminosity)} \qquad (10)$$

and the $L_{effective}$ (what is recorded) = 4×10^{31}. This implies for example, 100 bunches in TeV I -- very optimistic.

For CW beams Δ is the integrating time of the detector. For CW beams, we need each collision to be preceded and followed by intervals of Δ seconds with no collisions:

$$\begin{array}{l}\text{Number of such}\\ \text{events/sec}\end{array} = \frac{X}{\Delta}\,[P(0)]^2$$

$$= \frac{X}{\Delta}\, e^{-2x}$$

$$\equiv \sigma\, L_{eff}$$

where b is the effective cross section for the non-luminosity-associated background. The higher order terms, L^p, are neglected here. Statistical rules then give us the relation

$$\frac{\Delta S}{S} = \frac{1}{\sqrt{S}} \sqrt{1 + 2B/S} \qquad (12)$$

where

$$S = \sigma_S \, LT \qquad (13)$$

$$B = \sigma_B \, LT$$

and σ_S is the cross section which gives rise to S events in a time T.

We can now define an effective luminosity for comparison with background-free data from Eq. (12).

$$L_{eff} = \frac{L}{1+2B/S} \qquad (14)$$

Useful events (equivalent to background-free events) are

$$S_{eff} = \sigma_S \, L_{eff} \, T = \sigma_S \, T \, \frac{L}{1+2B/S} \qquad (15)$$

1. It is the denominator that enables the much lower luminosity e^+e^- machines to dominate the charmed physics field over hadron machines as discussed previously.

Substituting from (13) and (11):

$$S_{eff} = \frac{\sigma_S TL}{1+2\sigma_B/\sigma_S}$$

In the high luminosity limit (b<<aL) and $\sigma_B \gg \sigma_S$:

$$S_{eff} = \frac{\sigma_S{}^2 T}{2a} \qquad (16)$$

2. It is likely that backgrounds scale as the total cross section and therefore rise slowly with energy whereas interesting signals increase rapidly with energy. As an example, the Z^0 cross section increases by a factor of 10 from 0.8 TeV to 2 TeV. Other ratios are given in the following table:[5]

Table III

Increase of Cross-Sections with Energy

Production	Energy Increase	Cross Section Ratio
100 GeV jets	1-3 TeV	~15
100 GeV $\phi\bar{\phi}$	1-3	40
100 GeV Z^0	1-3	8
240 GeV η_T	3-10	30
1 TeV Z^0	3-10	500

Equation (16) indicates that the effective luminosity in a high background regime will then increase as the square of the cross section. In the cited case of Z^0 production, the effective sensitivity goes up by a factor of 100. Equation (16) indicates that this cannot be compensated by increasing luminosity. We would approach this condition if the residual pile up is larger than twice the signal. Table III indicates that the gain in sensitivity for a factor of three increase in energy can be very substantial.

3. We can use an example of how background is generated by multiple events as we described in Tables I and II. UA2 recently reported that a large fraction of high E_T events have jets. Thus we may have many triggers with simulated two jet events. Suppose we have a signal cross section σ_s, a cross section, σ_2, for two or more single events that make background that occur within the resolution time and a cross section σ_1, for single event background. Then the background-to-signal ratio can be calculated as

$$\frac{B}{S} = \frac{\sigma_1 L\Delta + \varepsilon(\sigma_2 L\Delta)^2/2}{\sigma_s L\Delta} \qquad (17)$$

where ε is the probability that the multiple collision event is misidentified as a signal event. Δ is the integrating time or the time between bunches. We can write this as:

$$\frac{B}{S} = r_1 + \frac{L}{L_o} \quad \text{where } L_o^{-1} = \frac{1}{2}\,\varepsilon\,\frac{\sigma_2{}^2\Delta}{\sigma_s}$$

The modification of luminosity due to background (Eq. 14) is now

$$L_{eff} = \frac{L}{1+2r_1+2L/L_o} \leq \frac{L_o}{2}$$

Suppose $\varepsilon = 0.01$, $\sigma_s = 1$ nb, $\sigma_2 = 1$ mb and $\sigma_1 \sim 0$

$$L_o = 10^{30}\text{cm}^{-2}\text{sec}^{-1}$$

This is a frighteningly low luminosity for plausible cases. Clearly refined rejection criteria will improve the situation. To appreciate the problem we must face the difficulty of estimating the probability that large accumulations of high multiplicity events will simulate the physics signal. Indeed, if we simply use the measured jet cross sections from UA2, and ignore the possibility of fake jets, the optimum luminosity goes above 10^{32}. Again we emphasize that higher energy accelerators (e.g., 20 TeV) can afford drastic E_T cuts of, say, 500 GeV, with small physics penalty.

VIII. Conclusion

All of this is not to say that these problems are incapable of being solved. Great progress has indeed been made in the cost per channel of microelectronics. Track processors which will rescue the large laboratory computer are "in the wind." There are clever tricks in chamber arrangements which promise factors of 2 improvement here and there. However, those of us[6] who have struggled mightily for many

$$L_{eff} = Le^{-2L\sigma T}$$

$$L_{eff}(\max) = \frac{1}{2e\sigma\Delta} = 1.8\times10^{31}$$

$$L_{opt} = \frac{1}{2\sigma\Delta} = 5\times10^{31}$$

Here we took the optimistic integrating time Δ = 200ns.

The apparent advantage (compare Eq. 10) of a factor of 2 of bunched over CW beams for this clean event scenario is somewhat decreased because, for realistic energy distributions in calorimeters, only half the energy of a second event (on the average) is added to the trigger.

In general CW beams can reduce the effect of multiple events for moderate luminosities ($<10^{32}$) by using scintillators to veto multiple events. Bunched beams can use vertex detection over the ≥1 meter crossing region. Both techniques are imperfect and tend to fail badly (tracking will fail first) as the luminosity approaches 10^{32} (see Table I).

VI. Pile-up Effects at High Luminosity

So far, the results are well defined. Now we discuss the consequences of incomplete elimination of multiple events. To set the scale, let us calculate the frequency of accumulations of events for track chambers assumed to be PWC's at ±100ns integrating time and calorimeters with ±200ns integrating time. We believe these times to be optimistic for presently foreseeable detectors (Eq. 6):

Table I

Probability of an accumulation of more than N events within the integrating time of tracking (±100ns)

	10^{31}	10^{32}	10^{33}
N≥2	.0047	.26	1.0
N≥4	3×10^{-5}	.019	.99
N≥10	~0	2×10^{-5}	.54
N≥20	~0	~0	.0035
Total sample (in 10^7 sec)	5×10^{12}	5×10^{13}	5×10^{14}

In a run of 10^7 sec where 100 good physics events are a minimum detectable signal, we will have 5×10^{14} collisions at $10^{33}cm^{-2}sec^{-1}$ which will come in clumps of 2, 3, ... within the resolution time of the tracking, i.e., 54% will have more than 10 collisions, each with an average of 30 tracks. Although 10^{32} looks better, remember that triggers (e.g., UA1 experience) use analogue calorimetry information which will clearly favor high multiplicity and make things worse. As it is, in the standard 10^{32} run, 10^{12} events will contain four or more collisions. We may detect 99% of these as multiple collisions (a great feat!) leaving 10^{10} to fluctuate towards the desired signature. At 10^{31}, the required rejection is only 10^{3} in order to have a background equal to signal. Table II gives the calorimetry problem:

Table II

Probability of accumulating more than N events within the integrating time of calorimetry (±200ns)

	10^{31}	10^{32}	10^{33}	$<E_T>$
N≥2	.018	.59	1.0	20 GeV
N≥4	7×10^{-5}	.14	1.0	40 GeV
N≥8	~0	.011	1.0	80 GeV
N≥10	~0	3×10^{-5}	1.0	100 GeV
N≥20	~0	~0	.53	200 GeV
N≥40	0	0	1×10^{-4}	400 GeV
	5×10^{12}	5×10^{13}	5×10^{14}	

Here we have used UA1's estimate of $E_T\simeq20$ GeV per collision. We have divided this by two because gating will protect the calorimeter by this factor. At 10^{33}, the rate of events with E_T > 400 GeV would be 5000 per second! For reference UA1 triggered at $\sim E_T$ = 30 GeV. At 10^{32}, an acceptable triggering rate can only be achieved with ~100 GeV threshold. Furthermore, 5×10^{11} situations will arise with eight or more collisions integrated into the resolving time of the calorimeter. How many of these will, by fluctuations, give clusters simulating jets of very significant energy? Clearly rejection mechanisms will have to be developed; the burden will fall on the calorimetry but it is not at all clear that this can be reduced to the levels sought by the high luminosity design. Some kind of clustering requirements will have to accompany the E_T trigger threshold. However, it seems obvious that any 100 event or even 10,000 event signal will have to deal with significant backgrounds that arise from fluctuations.

How well can we reject supermultiplicity events? It is not clear that one can easily find criteria that do not bias the physics. Cuts, on E_T for example, will reduce the effective backgrounds. These cuts are essential for triggering the subsequent logic circuits. However, the severity of the physics bias will depend on the energy of the E_T cut relative to the energy of the accelerator. Further discussion of the influence of these pile-up effects requires analysis of specific reactions and eons of Monte Carlo computer time. However, we have ample justification for assuming the existence of a luminosity-induced background due to these pile-up effects. Tables I & II indicate that a substantial fraction of the calculable and anticipated reactions will be compromised at 10^{32}. Electron identification will be compromised by the failure of tracking or by random energy adding to the "E" in the E/P rejection process. Muon identification will be burdened by the manyfold increase of pions and subsequent pi mu decay.

One final comment about Table II. We expect the E_T's in the last column to go up logarithmically with energy of the accelerator. Whereas a 400 GeV threshhold would be devastating at $\sqrt{s}$ = 1 TeV, a 1 TeV threshhold for E_T could be quite acceptable at 20 TeV, i.e., the higher the primary energy, the more luminosity the calorimetry can take.

VII. Effective Luminosity; Backgrounds

From the foregoing it is clear that luminosity brings with it inefficiency, dead time, but also background events that will be added to any true physics signal. We assume that part of the background is proportional to the luminosity:

$$\sigma_B = b + aL \tag{11}$$

years to handle high rates and get physics out have a very deep respect for these problems.

A pp collider offers the unique feature of factors of 10-100 in potential collision rate over $\bar{p}p$ at the very significant added cost of an additional ring. This luminosity is strongly motivated by anticipated "new" physics cross-sections. Yet the confidence that these rates are usable is very far from being demonstrated. Before we invest heavily in luminosity, we need a great deal of confidence that the detectors can be dramatically improved. One solution is to use the extra ring money in order to go to higher energy. This tends to raise cross sections for these processes, e.g., Table III, in two different ways and therefore also the signal to noise. Higher energy results in bigger cross sections for masses approaching or exceeding ~10% of $\sqrt{s}$. Also, since most of the data at very high energy machines are at low x, the QCD effects tend to raise parton flux and therefore effectively again raise cross sections. As we have seen, i.e., Eq. 16, with backgrounds present, we gain with a power of cross section which is larger than one. Since backgrounds increase with multiplicity which scales logarithmically with energy, cuts applied to reduce background are much less likely to injure the physics at higher energy. Of course the strongest drive for high energy is the totally unpredictable phenomena we may see. We should recall that every accelerator that has opened a new energy region in the past thirty years has yielded unanticipated results.[7] It is also at high energy (>10 TeV) where there is some possibility that the 10^{32} luminosities can be profitably utilized. Of course, new physics may very well be nicely explored with modest luminosity. We must go there to see.

The prognosis for instrumental breakthrough is mixed. Serious studies of high luminosity colliders started in 1972. We can look at this as a 15 year program of which 10 years have already been spent. Nevertheless, (and this is the principal motivation of this paper), work must continue on decreasing the integrating time of tracking detectors, preferably without breaking the bank by infinite readout channels. Calorimetry is fundamentally ugly; a cure here would be to improve resolution, decrease integrating time and find a cheap substitute for steel.

References

1. L.M. Lederman, Muon Pairs Revisited, Crisp 72-43 and BNL 17 522 (1972) p.406.
2. S.C.C. Ting et al. ISABELLE Summer Study, 1981.
3. UA5 Collaboration - CERN/EP 81-15 3.
   UA1 Collaboration - CERN EP/81-155; also presentations at the Paris (Rochester) Conference.
4. T. Akesson et al (Axial Field Spectrometer Collaboration) CERN EP/82 - 139 (1982).
5. F. Paige, BNL Report. G. Kane, Snowmass graphs.
6. One of the authors (LML) has been doing experiments almost exclusively with primary protons since ~1964, including several generations of ISR experiments in the period 1971-1978.
7. This was stressed by T.D. Lee in private communication.

Figure Caption

Typical UA1 events taken at $\sqrt{s}$ = 540 GeV and very low luminosity.

Fig. 1(a), (b), and (c). Typical UA_1 events.