Status of SimBES: A New Simulation Program for BES-I/II

Jiangchuan Chen, Baosen Cheng, Canguo Li, Lin Zhang
Monte Carlo Group, Division 1, IHEP
http://hpws3/~ licg/mc.html

June 20,1997

1 Why SimBES?

For many years, the simulation program for the BES detector has been SOBER, a homemade program based on the MARK3 framework and EGS4 electro-magnetic shower simulation package. It is worrisome that the BES physics analysis is based on such an outdated, rather poorly understood and ill-maintained simulation program.

Nowadays, most high energy experiments base their detector simulation programs on GEANT3 which is a universal geometry (detector description) and tracking package (tracking is the kernal of a simulation code). GEANT3 has been used by, hence tested by, the experiments worldwide for many years, and is constantly maintained by a group of experts at CERN.

Considering the relatively low cost on the software and the high reward on the quality of the BES physics analysis, it is imperative that the BES be simulated by a GEANT-based program. However, it is not feasible to GEANTize the SOBER code, because GEANT and SOBER are fundamentally dissimilar, and SOBER contains many dubious parts. A new simulation program must be written in the GEANT3 framework. The design of the new program was reported on the last annual meeting. SimBES is a Fortran program written in the framework of GEANT3.21, the latest version of GEANT3. The digitization part of SimBES is written with SOBER digitization algorithms as reference.

At the time of SimBES's development, Fortran begins to give way to C++ in the HEP world. GEANT3 will soon be superseded by GEANT4 which is written in C++. While SimBES is written in Fortran and based on GEANT3.21 to be on the safe side, it is well-structured to make the transition to GEANT4 and C++ smooth.

2 Design of SimBES

Strict hierarchy and weak coupling are guiding principles in the design of SimBES. Main features of SimBES are:

(1) standalone programs: GenBES, SimBES, DRUNK(the existing reconstruction program which itself definitely needs a reconstruction). Advantages: isolation of problems in each program, reduction of runtime memory size.

(2) decoupled phases: geometry, tracking, digitization. Advantages: effects of changes are confined to a phase, phases can be executed individually or combined, result of each phase can be written to and read from the disk.

(3) tree-like code reference structure: multi-path to the same function (most subprograms in SimBES are FUNCTION) is avoided except for the service and mathematical functions(which has no static data, hence no history). Similarly, GOTO is avoided within a function except for the error exit. With the structure looking more static, the logic of the code is much more understandable.

(4) hierarchical and decentralized: self-contained initialization and termination in all program modules(implemented by multi-ENTRY). Strict hierarchy demands that details(e.g. initialization) in lower levels should be hidden from the higher levels. This greatly facilitates the code development and maintenance work.

(5) modular: subdetectors can be turned on/off at runtime via data cards. This proves to be extremely useful for the code diagnosis and detector study.

(6) detector parameters are read in by (free format) cards: this makes changing of parameters much easier and quicker. It seems unnecessary to change the detector constants at runtime. However, in order to understand the detector and its sysmatic errors, the simulation code must allow variation of constants, although strangely few simulation codes do so. This feature also makes SimBES code good for both BES-I and BES-II, by using different sets of data cards. Currently, SimBES has about 100 data cards grouped into 8 card files. SOBER's constant files are used in case of large amount of data such as MDC wire survey data.

(7) explicit data flow: no COMMON block is defined for SimBES internal usage (communication between SimBES and GEANT3 still relies on GEANT3 COMMON blocks). This anti-COMMON feature may look very uncommon in the HEP Fortran world, but is normal in the C/C++ world. Access to static data should always be constrained, to make the debugging easier. In addition, the inconvenience of data exchange between subprograms without COMMON block forces one to design more carefully the code reference structure. Readability of the code is improved as a result.

(8) hidden data format: no indexing by index. While Fortran relies on array, C/C++ uses *structure*. SimBES mimics *structure* by EQUIVALENCE (the previous anti-COMMON feature also depends on the extensive use of EQUIVALENCE). This makes changing of data format a lot simpler and safer.

(9) at all stages from design, coding to compile and link, measures are taken to prevent, expose, and trap the errors. We always maximize the effect of bugs. e.g., automatic data is used whenever static data is not necessary. We do not bypass any error, e.g., floating point underflow. These make the program more alert to bugs. The built-in debug codes can be turned on/off either at precompile stage or at runtime by cards. As a result of these measures, debugging has never been a significant activity in the SimBES project.

3 Status of Programming

SimBES project consists of two parts: engineering and detector study.

Engineering includes design and coding of geometry(detector description for tracking purpose), primary event interface(primary tracks), hits storage/retrieval, interface to digitization code, packing of raw data, book keeeping of Monte Carlo truths and environmental data, i/o service and error handling. This task is suitable for one person and has been undertaken by C. G. Li. It is basically finished except for the book keeping of Monte Carlo truths and environmental data still in progress. In addition, GenBES (driving program package for SOBER event generators) is finished by C. G. Li.

Detector study includes coding and tuning of digitization code. The challenge in this part is not so much in coding but in understanding the detector. Each subdetector usually demands long term attention by a devoted expert. This section covers the status of coding only(the physics tuning is reported in Section 5).

Digitization of ESC is undertaken by J. C. Chen. The coding is finished.

B. S. Cheng is responsible for both VC and BSC digitization. The coding is finished.

Digitization of BTOF and ETOF is undertaken by L. Zhang. The coding is in progress.

As for digitization of CDC, MDC and MUO, we still wait for experts. It is worthwhile to mention that it took an expert only a couple of weeks to write the digitization code in case of BSC and ESC.

GEANT3 has a graphical and interactive version GXINT. SimBES thus has a sister XSimBES as a free byproduct, since no extra effort is needed to make the interactive version once the batch version is coded. Since one is unlikely to use XSimBES to produce digitized data, XSimBES is essentially finished.

XSimBES plays two roles: it serves as a debugging tool for the batch version SimBES; it can also be used as a visual learning tool of particle physics. Perhaps the second role is more interesting to most users.

A very important missing piece is the manual for SimBES.

4 Run Performance

Performance tests were conducted on IHEP's HP workstations (hpws3,4,5).

Crash test passed: 3 types of events, 10,000 events each type, generated by *radee, rhopi, p2mumu* are processed by SimBES successfully.

cpu time: cpu time is mainly spent on the shower simulation in the dense materials. cpu time is a function of the cutoff energies for electrons and photon. On our overloaded machines, cpu time varies also with occupation status of the machine(in weekdays cpu time may increase by 50% than in weekends). Speed of SimBES is measured at 3 sets of cutoff energies. Speed of SOBER is tested at the default cutoff energies.

When comparing the speeds of SimBES and SOBER, one should remember that (1) SOBER does not simulate the hadronic processes, (2) SOBER simulates electro-magnetic shower in aluminum walls of MDC, in TOF and SC only, not in COIL and MUO. (3) SOBER's cutoff energy for electron and photon is higher(BSC: $E_e = 2.0MeV, E_\gamma = 1.0MeV$; ESC: $E_e = 3.6MeV, E_\gamma = 2.0MeV$).

SimBES's speed is OK down to cutoff energy at $E_e = 1.0MeV, E_\gamma = E_e + 1.0MeV$. Runtime memory size: OK.

5 Status of Physics Tuning

As simulation consisits of geometry, tracking and digitization, tuning contains the same three stages. Hits(position, energy loss, t.o.f.) are the results of geometry and tracking. Pulse height and time are the results of digitization.

Geometry parameters do not need radical adjustment. Tracking parameters (including those of physics processes and those controlling tracking precision) are mostly handled by GEANT3.21 without need of user intervention(what a release!). So far, SimBES only needs to set the shower cutoff energies. This is a great advantage of the GEANT3-based program over SOBER.

Now that we have few things to worry for the tracking phase, we can concentrate on the digitization process. Algorithms and parameters of the SOBER digitization code are referenced in SimBES's digitization code. Note that SimBES does not interface to SOBER. Interfacing means that the other side is treated as a black box(e.g., GEANT3 tracking is a black box to SimBES). Digitization is an essential part of SimBES that must be fully understood. Therefore, digitization codes of all subdetectors have been, or will be, written from scratch. SOBER algorithms have been, or will be, critically reviewed. In some cases, new algorithms have been developed for SimBES. Our goal is not merely producing a running program, or a program that can reproduce the SOBER data. SimBES aims to reproduce the real data, and to understand how/why this can/cannot be achieved. This might be too lofty a goal for SOBER, but mimimal for any decent simulation and honest physics analysis.

5.1 BSC (by B. S. Cheng)

BSC is made of 24 gas layers interleaved with 23 layers of aluminum-lead-aluminum plates(thickness 1/2 radiation length). Each gas layer is segmented into 560 ϕ cells by aluminum I-beams, and into 4 zones along z direction by aluminum ribs for mechanical strengthening. The 24 layers of tubes are combined into 6 readout layers(1-2, 3-4, 5-6, 7-9, 10-12, 13-24), while the 560 ϕ cells are read out individually from both ends.

The pulse height is determined by the *single-particle spectrum*, the number of charged tracks passing the tubes(visible part of the shower development), and the charge division relation. Sum of the pulse heights correlates to the energy deposited in the lead layers.

Table 1: SimBES run performance: cpu time per event

E_e cutoff	1.0 MeV	2.0 MeV	4.0 MeV	(SOBER cpu)
Bhabha	2.0	1.0	1.1	(0.58)
di-muon	0.25	0.25	0.21	(0.13)
$\rho\pi$	1.1	0.91	0.76	(0.26)

The underlying *single-particle spectrum* can be extracted from the real di-muon data. In SimBES, this spectrum is approximated by a Gaussian function, instead of the double Gaussian function used by SOBER. Figure 5 compares the observed and simulated single-channel(both ends added) charge spectra in each readout layer. The apparent multi-peak feature of the charge spectrum in the figure is due to the combined effects of the radial layer ganging and the very fine ϕ segmentation. When the pulse heights in a readout layer are summed, much of the simple Gaussian feature is recovered as shown in Figure 6.

In SimBES, the tracks(showers) are generated by GEANT3. Since the BES shower counter has very fine segmention in the plane transverse to the incident particles, details of the lateral development of the electromagnetic showers can be observed. To reproduce the rather wide electromagnetic showers observed in the real data, very low energy photons(far below the pair production threshold) and (Compton) electrons must also be traced. Such a detailed simulation by GEANT3 can produce good results but is cputime costly(more than 5 seconds per event). Although some acceleration technique is in principle possible, no attempt has so far been made(a rather sizable project). In addition, it is unlikely that low energy (lateral) tails are essential for the physics analysis. It might be possible to modify the reconstruction code (DRUNK) to remove/hide the lower energy (de)tails.

The single-particle spectrum derived from the di-muon data should be applicable to the Bhabha data if the gas tube does not distinguish the particle types. The shower width and uncalibrated energy(sum of the pulse heights) in each readout layer are examined for the Bhabha electrons at 3.1 GeV. It is found that the shower width and energy profiles agree with the real data remarkably well(a great improvement over SOBER) when the cutoff energies for photon and electrons are set to $10keV$(see Fig.). However, energy(pulse height) weighted lateral shower profile may be a more meaningful object for data Monte Carlo comparison.

5.2 ESC (by J. C. Chen)

ESC has two identical endcaps, each made of 24 layers of gas-filled aluminum square tubes(laid along the vertical direction) interleaved with 23 layers of lead plates(thickness 1/2 radiation length). All layers are circular ring shape but of different sizes. The first 6 layers of tubes in each endcap are read from both ends. The rest 18 layers in each endcap are combined into 3 readout layers (7-10, 11-14, 15-24) and read from one end only. Within a readout layer, tubes are read individually except for the few tubes on the outer edge.

Like BSC, ESC geometry is described in detail by SimBES, and the showers are produced by GEANT3. However, lacking of the muon data and difficulty in Bhabha event selection make the data Monte Carlo comparison a difficult task. Presently, the underlying single-particle spectrum used in the ESC digitization code is that of SOBER's(double Gaussian).

Using the same (rather high in comparison to BSC's) cutoff energy as SOBER's($E_\gamma =$

Table 2: SimBES run performance: memory size

SOBER	GenBES	SimBES	XSimBES
6.6 MB	4.2 MB	4.9 MB	13.0 MB

$2MeV, E_e = 3.6MeV$), similar(to SOBER) results are obtained for the Bhabha events at 3.1GeV: shower energy and width profiles agree with the real data well, see Figure 11 and Figure 12. Note that the cutoff energy in ESC is 2 orders higher than that in BSC. It remains to be investigated how the low energy details dissapeared in the reconstructed ESC data. Again, energy(pulse height) weighted lateral shower profile may be a more meaningful object for data Monte Carlo comparison.

5.3 TOF (by L. Zhang)

The coding is still in progress. Since TOF produces measurement of high precision (time resolution about 200ps), demand on the data and Monte Carlo agreement is accordingly high. A major difficulty is presented by the multiple hits on a single scintillator counter. What is the frequency, the causes and effects of multiple hits? This has not been answered during the SOBER years. For SimBES, this is only a trivial task since the subdetectors can be turned on/off and materials can be changed by data cards. First it is found that about 30% of the barrel scintillator counters are multiply hit(row III in Figure 14). Then by taking out BSC (row II in Figure 14), and taking out both BSC and MDC (row I in Figure 14), it is found that the MDC (actually the aluminum wall of MDC) is responsible for the multiple hits on BTOF. The role of MDC aluminum shell and endplates is demonstrated visually by XSimBES. It is also found that only a small fraction of the multiple hits are widely separated (see (c) in the figures).

6 Summary

The graphical&interactive version XSimBES is finished, and will be released once the user guide is finished. XSimBES will be reckoned as an inevaluable visual learning tool for particle physics, apart from being a code developing and debugging tool.

With more than 90% of the coding finished(in terms of lines of source code), the batch version SimBES is close to finish.

SimBES has achieved its design goals: understandable and understood; maintainable and improvable; reasonable size and speed. In addition, SimBES simulates both BES-I and BES-II(the original goal was to simulate BES-II only).

Apart from being a simulation program, SimBES is also suitable for the detector design, upgrade and study, and provides more handles to physics analysis, thanks to the flexibility provided by the data cards and the well designed framework of the code.

While the coding of SimBES is close to finish, the detector study has just begun.

For ESC and BSC, SimBES has achieved comparable and even better agreement with the real data than SOBER, after only a few weeks' detector study.

The major missing part is the digitization code for MDC and MUO. Another (vitally) important missing item is a complete manual. The goal is that the whole SimBES program (about 10,000 lines of source code) can be fully understood by an expert within a reasonably short period (a couple of weeks to months) with the aid of the manual.

We appreciate the support of the IHEP Computing Center and the Data Group of Division 1: SimBES project has never encountered problems from the machine and software environment(GEANT3.21 and CERN libraries).

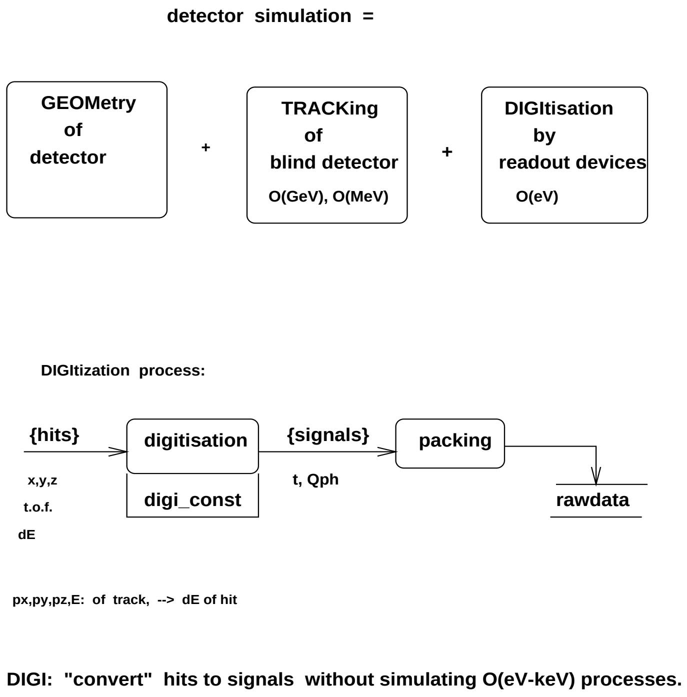

Figure 1: phases of detector simulation

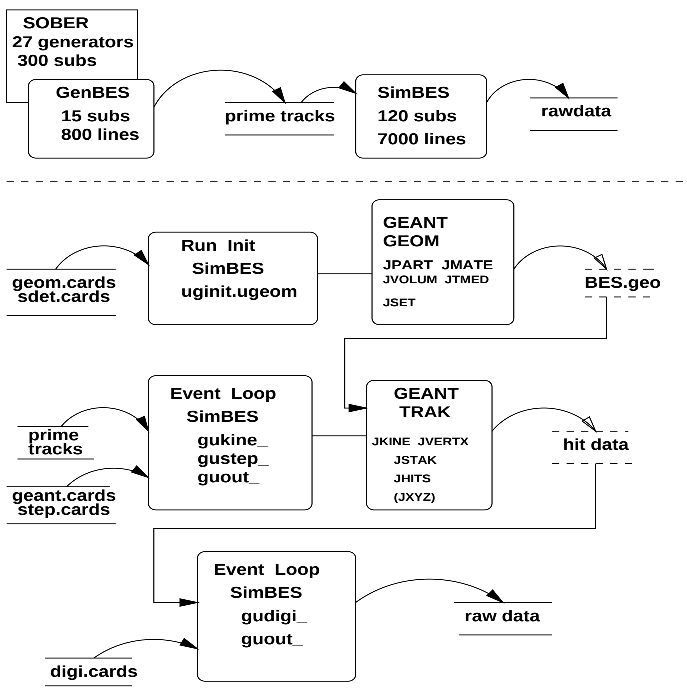

Figure 2: flow diagram of SimBES

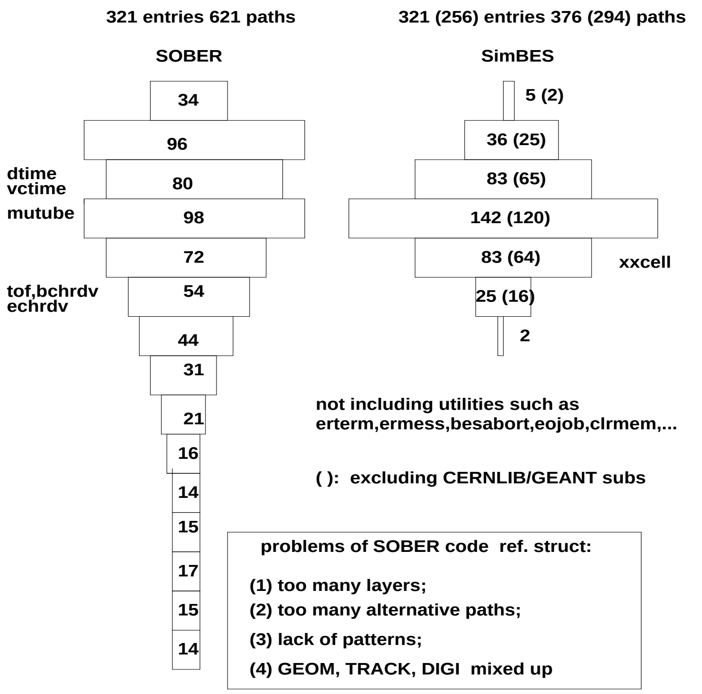

Figure 3: comparision of the code structures of SimBES and SOBER (a).

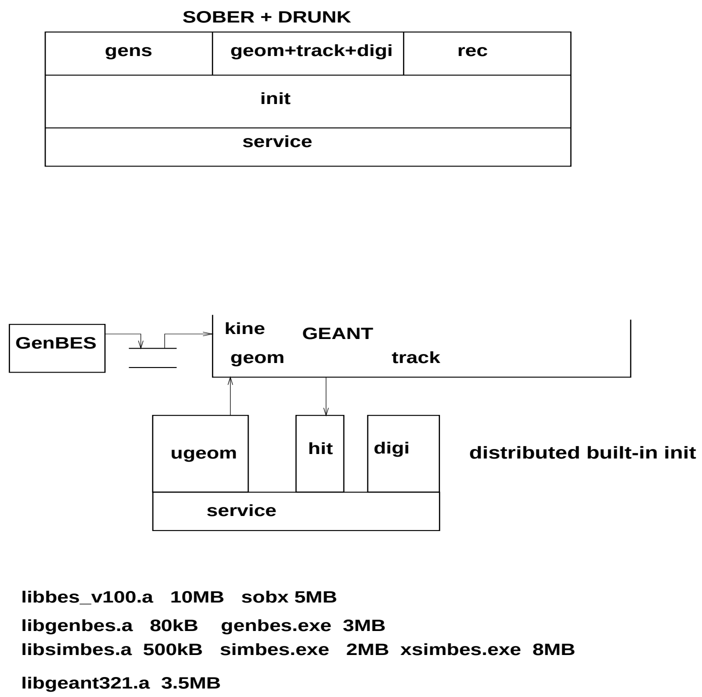

Figure 4: comparison of code structures of SimBES and SOBER (b).

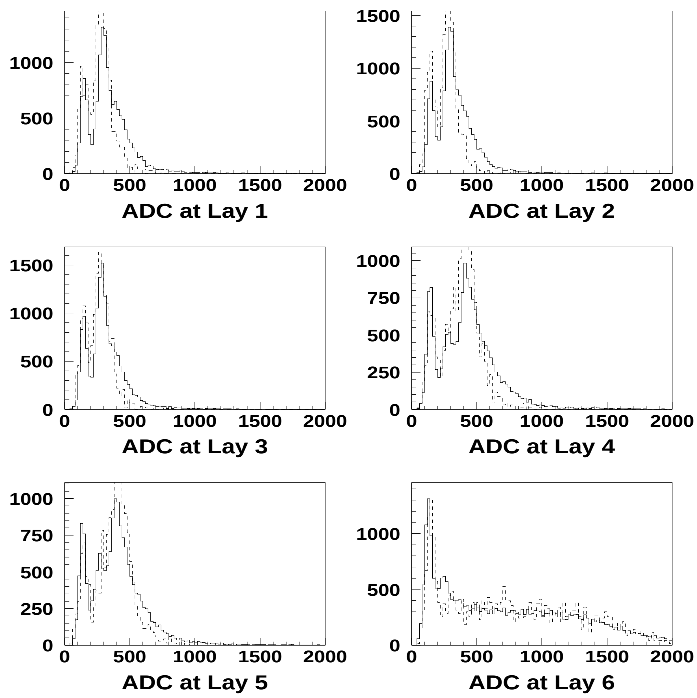

Figure 5: comparison of the single channel(both ends added) charge spectra of the dimuon(3.1GeV) data(solid line) and Monte Carlo(dashed line) in each readout layer of BSC.

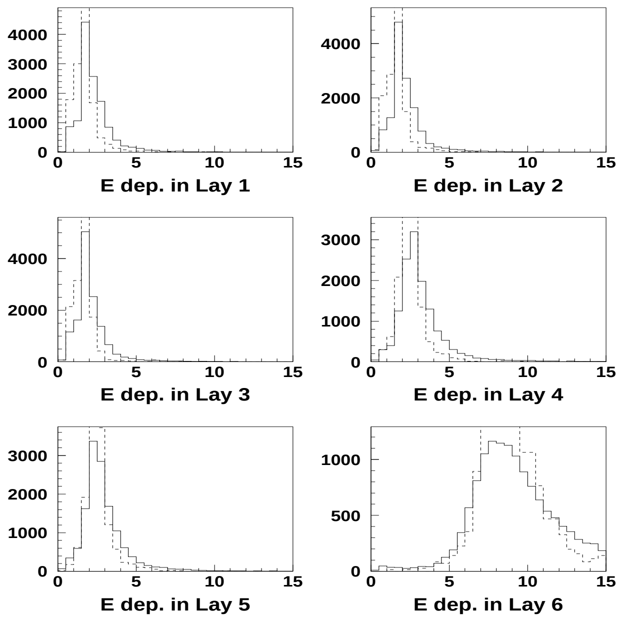

Figure 6: comparison of the cluster energy of the dimuon(3.1GeV) data(solid line) and Monte Carlo(dashed line) in each readout layer of BSC.

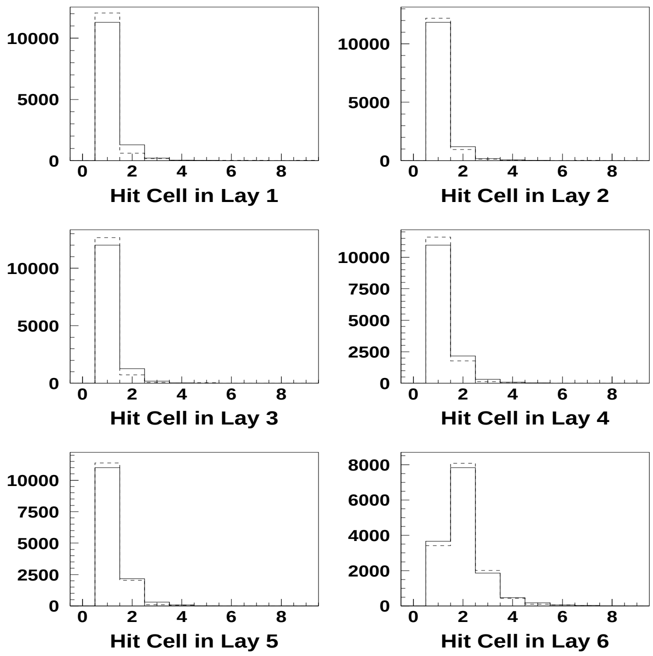

Figure 7: comparison of the cluster width(number of hit channels) of the dimuon(3.1GeV) data(solid line) and Monte Carlo(dashed line) in each readout layer of BSC.

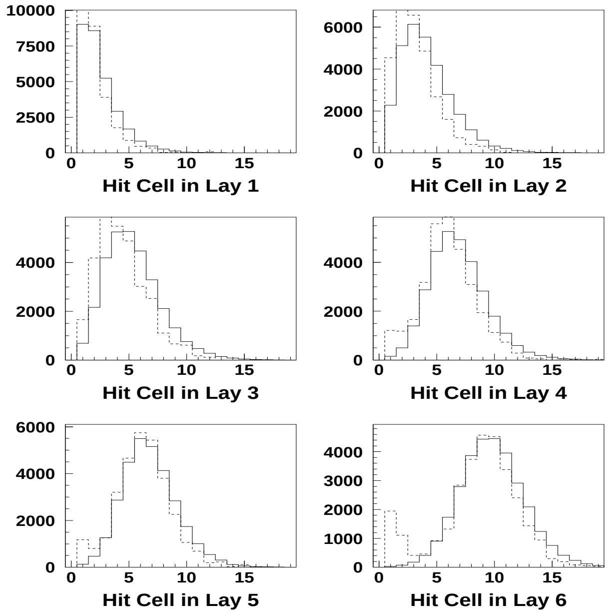

Figure 8: comparison of the cluster width of the Bhabha(3.1GeV) data(solid line) and Monte Carlo(dashed line) in each readout layer of BSC. cutoff energy=0.01MeV

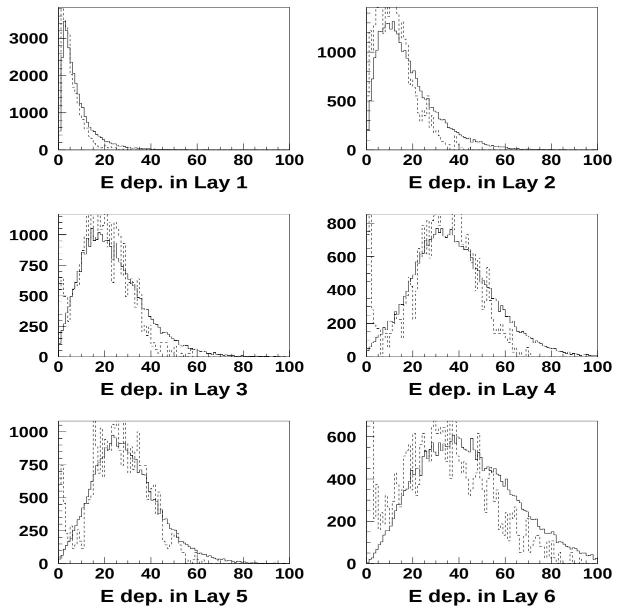

Figure 9: comparison of the cluster energy of the Bhabha(3.1GeV) data(solid line) and Monte Carlo(dashed line) in each readout layer of BSC. cutoff energy=0.01MeV. Note: a proper comparison needs MDC monte carlo data which is currently not available in SimBES.

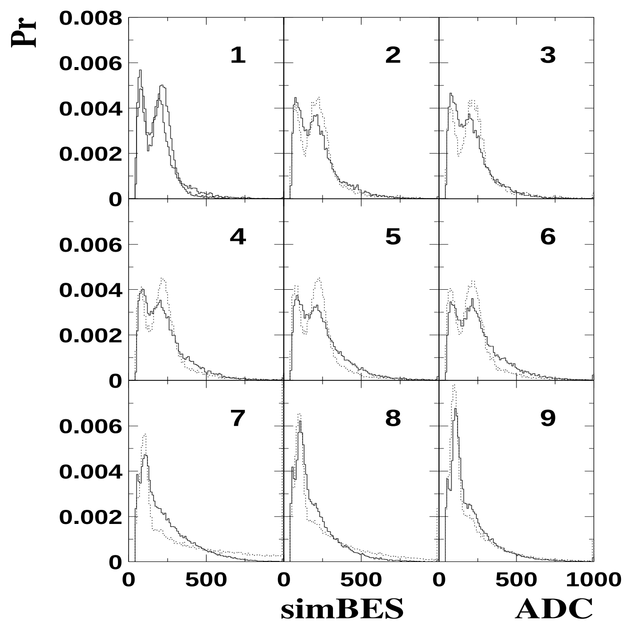

Figure 10: comparison of the single channel(both ends added) charge spectra of the Bhabha(3.1GeV) data(solid line) and Monte Carlo(dashed line) in each readout layer of ESC.

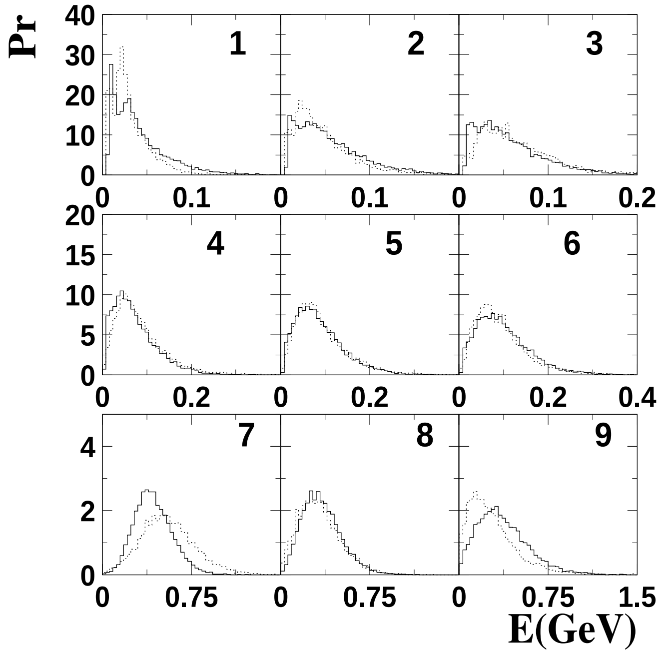

Figure 11: comparison of the cluster energy of the Bhabha(3.1GeV) data(solid line) and Monte Carlo(dashed line) in each readout layer of ESC.

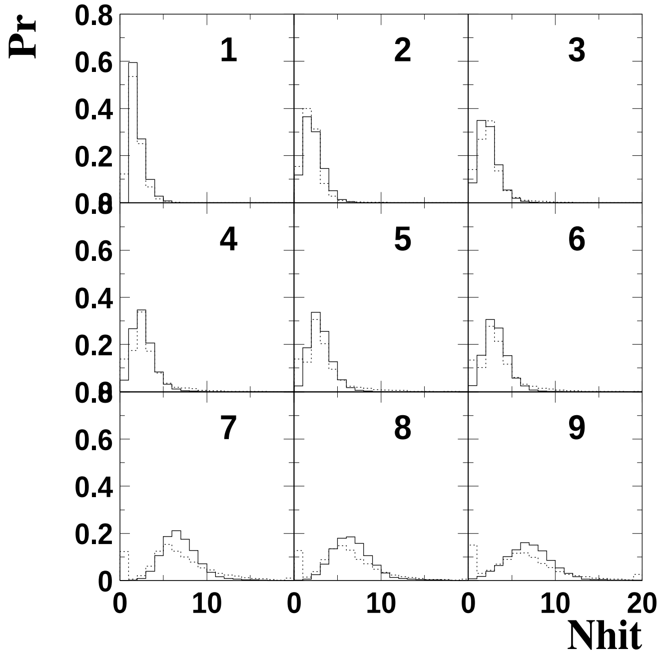

Figure 12: comparison of the cluster width(number of hit channels) of the Bhabha(3.1GeV) data(solid line) and Monte Carlo(dashed line) in each readout layer of ESC.

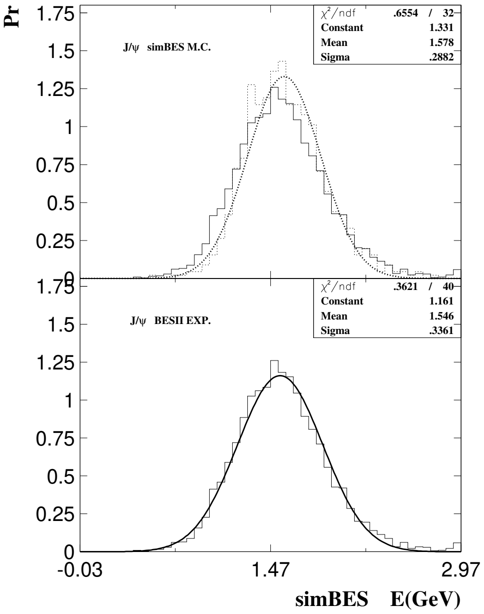

Figure 13: comparison of the total cluster energy of the Bhabha(3.1GeV) data(solid line) and Monte Carlo(dashed line) in ESC.

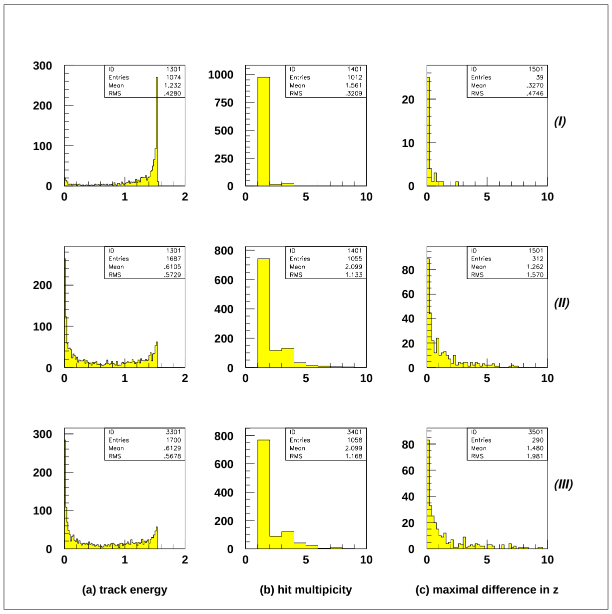

Figure 14: hits in the barrel TOF by 500 Bhabha (3.1GeV) Monte Carlo events. 3 rows for 3 different set ups of the geometry. (I) with BSC and MDC taken out of the geometry. (II) with BSC taken out of the geometry. (III) with every thing in. The 3 types of histograms are: (a): energy of the charged tracks. (b): hit multiplicity in a scintillation counter. (c): maximal difference in z in case of multiple hits.