Proposal Name

Development of Particle-Flow Algorithms and Simulation Software for the ILC Detector(s)

Classification (Detector)

Calorimeter: simulation software and algorithm development.
Note: Much of the proposed work, especially those on simulation software, span across all detector subsystems, and should really be considered in the broader premise of "Full Detector". In the absence of such a category, we have chosen to put this proposal under "calorimeter", since our involvement in simulation and other software tasks stem from our interest in developing new algorithms for calorimetry.

Personnel and Institution(s) requesting funding

G. Blazey, D. Chakraborty, J. G. Lima, J. McCormick, R. McIntosh, V. Zutshi.
Northern Illinois Center for Accelerator and Detector Development, Northern Illinois University [1]

Collaborators

S. Magill et al., *Argonne National Laboratory*,
N. Graf et al., *Stanford Linear Accelerator Center*,
C. Milstene et al., *Fermi National Accelerator Laboratory*,
R. Frey et al., *University of Oregon*,
J. Yu et al., *University of Texas at Arlington*,
G. Wilson et al., *University of Kansas*,
U. Mallik et al., *University of Iowa*,
The CALICE collaboration.[2]

Project Leader

Dhiman Chakraborty
`dhiman@fnal.gov`
(815)753-8804, (630)840-8569, (630)452-6368

Project Overview

The Northern Illinois University(NIU)/Northern Illinois Center for Accelerator and Detector Development (NICADD) group is interested in calorimeter R&D for the proposed ILC.[1] We propose to develop, in simulation and in prototype, designs for a hadron calorimeter (HCal) optimized for jet reconstruction using particle-flow algorithms (a.k.a. "energy-flow algorithms"). Simulation/algorithm development and hardware prototyping are envisaged as the two main components of our efforts. This proposal addresses the first component while the second is the subject of a separate proposal.

An e^+e^- linear collider is a precision instrument that can elucidate Standard Model (SM) physics near the electroweak energy scale as well as discover new physics processes in that regime, should they exist. In order to fully realize the potential anticipated from a machine of this type, the collection of standard high energy physics detector components comprising an experiment must be optimized, sometimes in ways not yet realized at current experiments. One such example is the hadron calorimeter which will play a key role in measuring jets from decays of vector bosons and other heavy particles such as the top quark, the Higgs boson(s), etc. In particular, it will be important to be able to distinguish, in the final state of an e^+e^- interaction, the presence of a Z or a W boson by its hadronic decay into 2 jets. This means that the dijet mass must be measured within ~ 3 GeV, or, in terms of jet energy resolution, $\sigma(E) \approx 0.3\sqrt{E}$ (E in GeV). Such high precision in jet energy measurement cannot be achieved by any existing calorimeter in the absence of a kinematically overconstrained event topology. Similar precision in measurements of jet and missing momentum will be crucial for discovery and characterization of several other new physics processes as well as for precision tests of the Standard Model. Such ambitious objectives place strong demands on the performance of the calorimeters working in conjunction with the tracking system at the ILC, and requires development of new algorithms and technology.

The most promising means to achieving such unprecedented jet energy resolutions is through particle-flow algorithms (PFA). A PFA attempts to separately identify in a jet its charged, electromagnetic, and neutral hadron components, in order to use the best means to measure each. On average, neutral hadrons carry only ∼11% of a jet's total energy, which can only be measured with the relatively poor resolution of the HCal. The tracker is used to measure with much better precision the charged components (∼64% of jet energy), and the electromagnetic calorimeter (ECal) to measure the photons with $\sigma(E) \approx 0.15\sqrt{E}$ (∼24% of jet energy). The left panel of Fig. 1 shows the momentum distribution of particles in a representative multijet physics process of interest. On average, only a small fraction of a jet's energy is carried by particles with momenta greater than 20 GeV. The right panel of Fig. 1 shows the precisions of energy measurement by a calorimeter, and momentum measurement by a tracker, of single charged hadrons, as functions of their momenta in one nominal detector design. Measurements from the tracker are at least two orders (one order) of magnitude more precise than those from the calorimeter for particles below 20 GeV (100 GeV). A net jet energy resolution of $\sigma(E) \approx 0.3\sqrt{E}$ is thus deemed achievable by using the HCal only to measure the neutral hadrons with $\sigma(E) \approx 0.6\sqrt{E}$. However, this will certainly require extensive and simultaneous optimization of detector design and tuning of algorithm parameters.

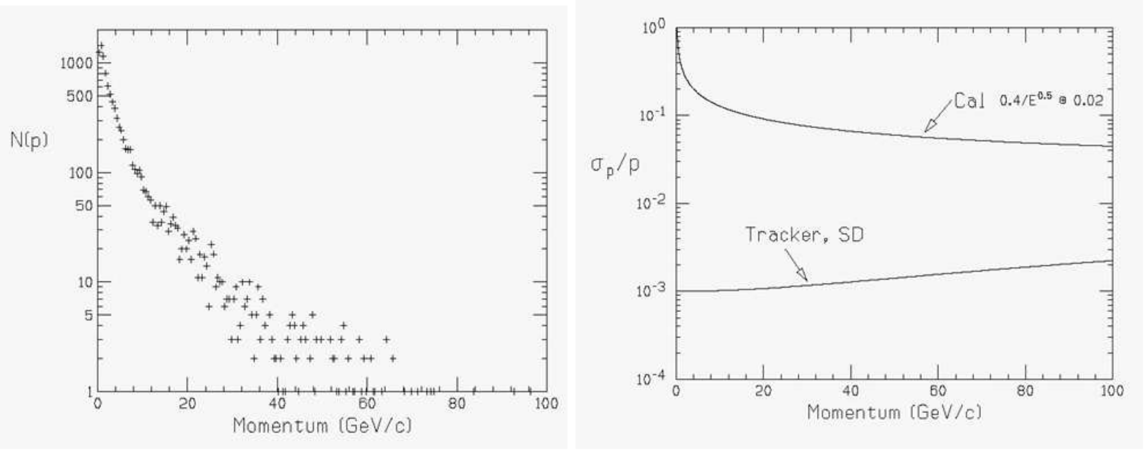

Figure 1: Left: the momentum distribution of particles in $e^+e^- \to ZZ \to 4$ jets events at $\sqrt{s} = 500$ GeV. Right: the fractional energy (momentum) resolution of an excellent calorimeter and that of a good tracker as functions of particle momentum.

A calorimeter designed for PFAs must be finely segmented both transversely and longitudinally for 3-d shower reconstruction, separation of neutral and charged clusters, and association of the charged clusters to corresponding tracks. This requires realistic simulation of parton shower evolution and of the detector's response to the particles passing through it. Accurate simulation relies heavily on analysis of data from beam test of prototype modules. The detector optimization requires the simulation, visualization, and analysis packages to be highly flexible, which calls for careful design and implementation of the software itself.

Very large numbers of events will have to be simulated to evaluate competing detector designs vis-a-vis ILC physics goals. Characterization of signatures arising from processes predicted by some extensions of the SM will require simultaneous coverage of broad ranges of undetermined parameters. Parametrized fast simulation programs will thus have to be developed once the algorithms have stablized. Parametrization of PFAs will require much work, and is one of our key objectives.

Broader Impact

The impact of the proposed research is not limited to calorimetry at the ILC. Lessons learnt from this experience are likely to prove valuable at other future Experimental High Energy experiments as well.

We have had good success so far in involving young students from different disciplines in our R&D activities, which we consider an important aspect of our research program. Indeed, much of the initial coding for the full-detector simulation program and most of that for the test-beam module simulation program described in the next section were done by MS students, Mr. McIntosh and Mr. McCormick, respectively. While Mr. McIntosh has subsequently decided to continue with our group, Mr. McCormick, a Computer Science major, has taken up a joint appointment co-funded by NICADD and SLAC to work full-time as a staff programmer on ILC detector simulation software development. Our group represents an excellent mix of individuals from different national origins, and we seek participation from the widest possible spectra of gender, ethnicity, geographic origin, and levels of physical ability.

Inter-institutional collaboration, within the US as well as across national boundaries, is a fundamental requirement for a project of this nature and magnitude to succeed. We have come to appreciate early the value of sharing not only the results, but all source code, usage instructions, and data files freely. All of the results described in the next section are carefully documented on our web page, with complete instructions for installing and running the simulation software, examples, and sample data files.[1] Furthermore, we are committed to entertaining requests for simulation jobs from any potential user.[3] Indeed, our group is responsible for maintaining the production version of the current ALCPG standard detector simulation program, and the primary provider of benchmark samples, as well as special requests for the entire American LC community. This reflects our strong commitment to fostering the spirit of collaborative networking and inter-institutional partnership among all users. Some of the funds requested in the 2nd and 3rd years will be used to enhance our capability to produce large volumes of simulated data in response to the anticipated increase in demands from the users.

NIU has an active outreach program with a dedicated coordinator. Activities include visits to schools and civic organizations in the region and hosting open house events to promote public awareness and interest in science. The presentations, which usually include demonstrations, highlight the importance of fundamental scientific research and give the audience a glimpse of techniques used in experimentation and interpretation of data. Attendance at these presentations exceeds 10,000 per year. Members of our team also volunteer in Fermilab's "Ask-a-Scientist" program and similar events offered through NIU's outreach web site.

Results of Prior Research

Members of NIU, ANL, SLAC, and UTA began collaborating on PFAs, simulations, and software development efforts in January, 2002. Several other groups have since joined the effort, and links have been established with European colleagues who had been active in this area already. The results that emerged through discussions at our regularly scheduled meetings have been presented at the Calor conferences, ECFA and ACFA meetings, the American LC workshops, and at the International LC Physics and Detector Workshops.

Toward the optimization of the HCal design, the NIU and ANL teams have started investigating both analog (cell energy measurements) and digital (hit density measurements) methods as functions of the cell size. Our preliminary findings suggest that with sufficiently small cells, the digital method yields a more precise measurement of the hadron energy, i.e., fluctuations in hit density are smaller than those in the sampled energy of a hadronic shower. Use of local hit density in lieu of the deposited energy to weigh the calorimeter hits results in superior energy resolution and lateral containment of single hadron showers. Two independent approaches to the implementation of a PFA have helped improve our understanding of the optimal cell sizes and geometry for best charged/neutral hadron shower separation in jets within the context of some specific overall detector parameters.

We will now briefly summarize our HCal optimization and algorithm development efforts. The HCal must be optimized to achieve, with due consideration of costs, benefits, and risks, the best balance between the reconstruction and energy resolution of neutral-hadron-initiated clusters in a jet, and the ability to separate them from the charged components. This is intimately related to the first step in the development of a particle-flow algorithm as described below. The elements are highly inter-related, and must be optimized simultaneously. All figures in this section were generated using GEANT4-based full-detector and test-beam simulation programs and reconstruction algorithms developed by our group.

HCal absorber/active media properties: The detector simulation and analysis of physics events within the Java Analysis Studio (JAS)-based software environment developed at SLAC, is flexible in the choice of absorber and active media type and thickness within the limits of the HCal volume. Our group has put together a GEANT4-based detector simulation package called LCDG4 to work within this environment, and produced many data sets spanning a range of cell shapes and sizes, and event types (single particles as well as benchmark physics processes). Teams from NIU, ANL, SLAC, Kansas, and Iowa, are studying a wide variety of events simulated with this package. We will optimize the HCal by comparing dense materials (e.g. W) to less dense ones (e.g. Stainless Steel) as absorbers, and scintillator- vs. gas-based devices (e.g. RPC, GEM) as active media. Containment of hadronic showers, density of hits, single-particle and jet energy resolutions will be used as performance measures. Substantial progress has been made in this direction already. The left panel of Fig. 2 shows the energy resolution as a function of single $\pi^{\pm}$ energy, estimated using hit density weighting, for two different lateral segmentations of the scintillator option, and the proposed segmentation for a realistic RPC design. The right panel of Fig. 2 shows the density-weighted angular widths of single-hadron showers as functions of their momenta in reasonably realistic scintillator- and gas-based designs. The more realistic gas-based geometry and the scintillator design under consideration give comparable results.

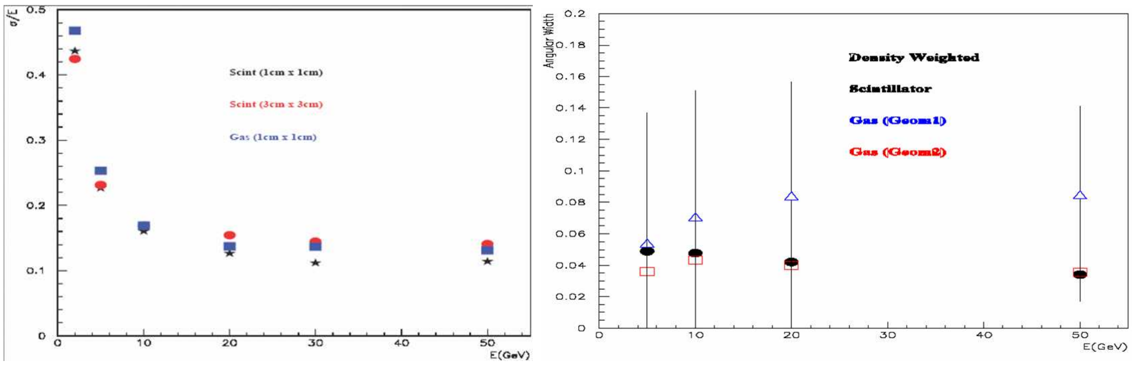

Figure 2: Comparisons of scintillator vs. gas as the HCal active medium. Left: the fractional energy resolution of single $\pi^{\pm}$ using density-weighting clustering in scintillator and gas-based geometry. Scintillator tiles of 1 cm^2 (stars) is not a practical proposition, but it is studied to understand the dependence of energy resolution on lateral segmentation of the active layer for a given choice of technology. Even the realistic 9 cm^2 scintillator option (circles) offers a somewhat better resolution than a 1 cm^2 gas configuration (squares) under this particular weighting scheme. The two are comparable at higher energies. Right: the density-weighted angular width of single $\pi^{\pm}$ showers as function of their momenta, in HCals with 9 cm^2 square scintillator tiles (circles) and those with 1 cm^2 square gas-based cells (triangles for “Geom1” and squares for “Geom2”). The “Geom2” configuration is fairly close to the RPC design currently under consideration. Other density-weighting schemes remain to be investigated.

HCal transverse granularity/Longitudinal segmentation: We plan to optimize the 3-d granularity of cells for the most promising PFAs vis-a-vis the active medium technology (see the left panel of Fig. 2) The methods developed here are generalizable to different total detector geometries, i.e., SiD (which uses silicon wafers for tracking), LD (which uses TPCs for tracking), etc. The basic performance measure here is the ability to separate showers initiated by charged and neutral hadrons - the key to any PFA. Indeed, if all the cell energies in a jet could be fully separated and correctly assigned to the parent particle, then a jet energy resolution of $\sigma(E) \approx 0.15\sqrt{E}$ (E in GeV) could be achieved. It turns out that the limiting factor in the overall jet energy resolution is the confusion term arising from imperfect association due to finite granularity and misassignment. From the reconstruction algorithm's point of view, it is this term that poses the biggest challenge.

Analog vs. digital readout for the HCal: The question of optimal 3-d granularity is intimately related to that of the readout method, which needs to be evaluated by comparing jet resolutions with both analog and digital readout. At the extreme, "digital" readout means a single-bit "yes/no" decision on whether or not a minimum ionizing particle (MIP) has passed through a given cell. Since such digital measurements are less susceptible to Landau and path-length fluctuations than full (12-15 bit) analog measurements, hit counting has smaller spread than energies samples in the active medium. We have shown that for small cell sizes (< 10 cm^2), the number of cells hit is a superior estimator of energy of single hadrons below 20 GeV (where more than 95% of the particles in typical jets in multijet events of interest are concentrated) to traditional analog sampling. Since the spatial spread of a shower increases in a less-than-linear proportion to its energy, the advantage gradually disappears at higher energies. We find that a semi-digital readout allowing multiple thresholds to classify the hit status of a cell can be effectively used counter this non-linearity. This supports the contention that the dynamic range sacrificed in order to achieve finer segmentation is not likely to hurt us at all.

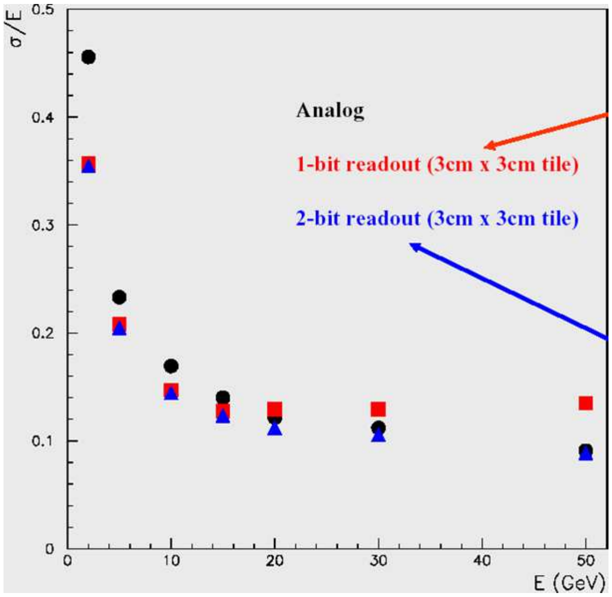

Figure 3: The fractional resolution of single $\pi^{\pm}$ energy using full analog (circles), 1-bit digital (squares), and 2-bit digital (triangles) read-outs. We find that 3 thresholds (i.e. 2-bits) is optimal.

It may be prudent to consider both the best analog and the best digital version of the HCal for eventual evaluation with test beams provided both prove potentially capable of meeting the energy resolution requirement. Such testing could spur future advances in readout technology.

Particle-flow algorithms: For the first time in calorimeter development, it is necessary to take into account the reconstruction algorithms in designing the detector(s). How good the jet energy resolution will be depends ultimately on how well the PFA is formulated and tuned. We have implemented, as the first step of a PFA, an algorithm that produces "pre-clusters" of calorimeter cells using energy and local density as weights. The user can choose the parameters such as thresholds, neighborhood definitions etc. These preclusters serve as a quasi-geometry-independent set of objects for the subsequent steps. The second step is to extrapolate the tracks and match them to preclusters whenever appropriate, so that the energies of all charged-hadron-induced preclusters can be replaced with the corresponding track momenta. The third step is to identify the photons by shower-shape analysis in the ECal. The final step involves combining the track momenta with photon and neutral hadron energies to produce high-precision jet energy measurements. Our goal is to minimize the dependence of implementations of the high-level algorithms on detector geometry details. This will greatly facilitate portability of the algorithms across regional boundaries and detector design choices. In addition to the preclustering algorithm from NIU, we already have separate preliminary codes for identification of track segments minimum-ionizing-particles (Iowa), propagating the tracks through the calorimeter taking energy loss into account (FNAL, NIU), photon reconstruction (SLAC, Kansas), and reconstruction from these of PFA jets (ANL, NIU). We are now working to combine these pieces into a fully functional jet-finding algorithm.

The PFA developed at NIU, performs full jet reconstruction using Monte Carlo "truth" for track matching. The result of this is shown in Fig. 4. We see that this PFA affords a 40% improvement in jet energy resolution compared to a traditional purely calorimetric measurement.

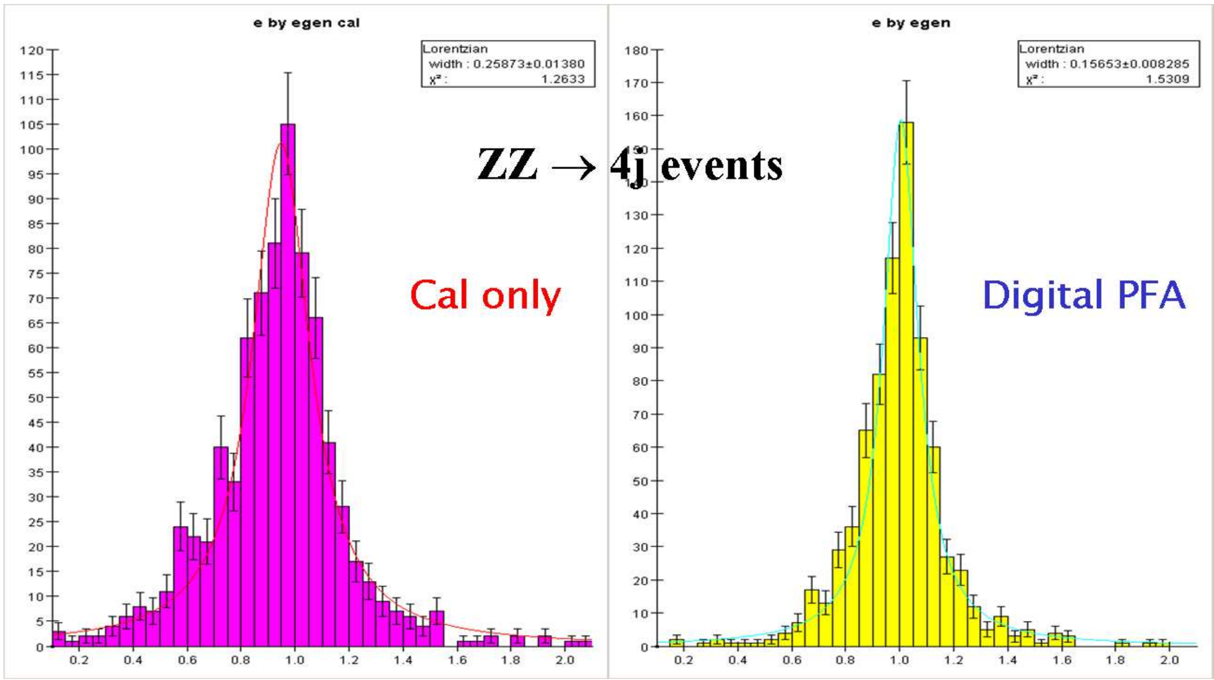

Figure 4: The estimated jet energy normalized to the true energy in $e^+e^- \rightarrow ZZ \rightarrow 4$ jets events at $\sqrt{s} = 500$ GeV events using purely calorimetric measurement (left) and the PFA developed at NIU (right). The calorimetric estimation uses traditional analog energy measurement, while the PFA uses semi-digital (2-bit) measurement in this case. The PFA can be used in analog mode as well. No kinematic fitting based on event topology is used in either measurement.

The NIU group has also made significant contributions to LC detector simulation software during the past 3 years. We ported and have been maintaining all of the current American software on the Linux platform. Since mid-2002, we have been processing simulation requests from several groups engaged in LC R&D, on Linux farms at NIU and FNAL. We organized a workshop at NIU/NICADD in November, 2002 (`http://nicadd.niu.edu/ws/`), to bring the groups together, chart a plan, and set out in an organized manner. This was followed by similar workshops at SLAC in 2003 and at ANL in 2004. Following are the simulation software projects where we have made substantial progress and hope to remain commited:

Simulation of full detector concepts: We have developed, in close collaboration with our colleagues at SLAC, a stand-alone GEANT4-based simulation package. This package, named "LCDG4", fully complies with the model put forth by the ALCPG simulation group, and adds several useful functionalities to it.[4] It produces "raw" hit output in the globally accepted LCIO format and supports projective geometries in θ, ϕ, as well as non-projective ones with cells of constant linear dimensions. After extensive testing, this package has been adopted as the standard for ALCPG. Figure 5 shows an example event display from LCDG4.

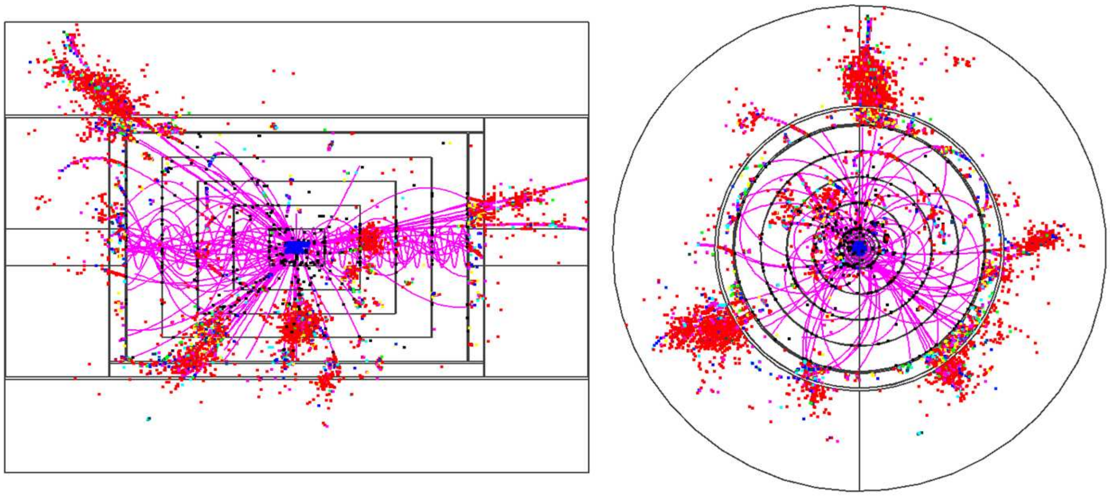

Figure 5: Side view (left) and end view (right) of an $e^+e^- \to t\bar{t} \to e + \nu +$ 4 jets event at $\sqrt{s} = 500$ GeV in the SD detector simulated using the LCDG4 program developed at NIU.

Simulation of test-beam prototype modules: Further, as members of the CALICE collaboration (CAlorimeter for the LInear Collider with Electrons,[2]), and in active cooperation with our European colleagues, we have produced a GEANT4-based simulator for the detector prototype module that is expected to be exposed to test beams at Fermilab over a period of 3 years starting in mid-2005. This program, called "TBMokka" is built on an alternative simulation framework called "Mokka", developed independently by our European colleagues. Figure 6 shows an example event display from a stand-alone GEANT4-based program to simulate a test-beam module (this program is used as a cross-check for TBMokka).[5]

We are well on our way to combining the best features of LCDG4 and Mokka. The new "next generation" simulation program incorporates a more advanced XML-based geometry description system. Like its predecessors, it supports run-time geometry specification, but offers a wider range of options and easier changes to geometries using GDML (Geometry Description Mark-up Language), a Geant4-specific extension of XML (eXtensible Mark-up Language).

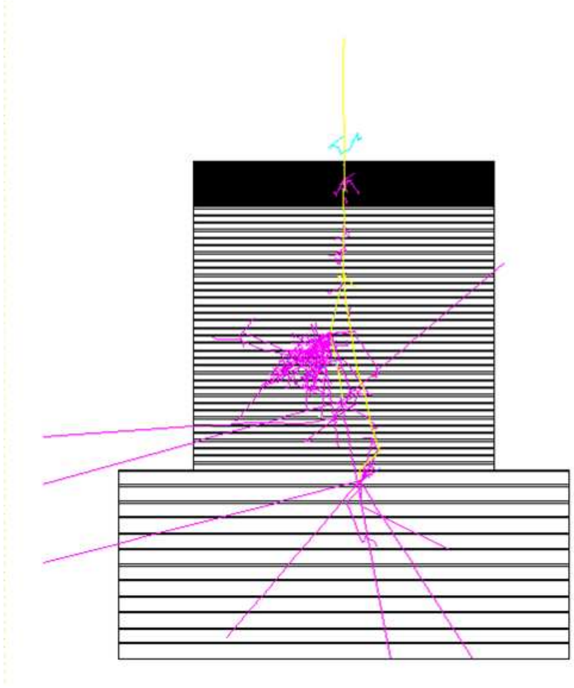

Figure 6: Side view of a 2 GeV π^+ passing through a near-final test-beam prototype module simulated using the GEANT4-based simulation program developed at NIU.

Simulation of the signal extraction process following energy deposition: In another endeavor, we have designed and coded the first version of a package, called "DigiSim", to simulate the conversion of energy deposits produced by GEANT4 to electronic read-outs.[6] This package offers the user a simple, flexible, and standard way to simulate the effects of thresholds, noise, cross-talk, inefficiencies, attenuation, and timing, that are involved in signal collection, propagation, and conversion (digitization). In essence, it allows the user to model an arbitrary transfer function from the energy deposited at the cell to the corresponding "raw data". DigiSim can be used either in a stand-alone mode to produce a persistent output, or as an on-the-fly preprocessor to the reconstruction program. In stand-alone mode, it produces output in the same format as that envisaged from the real detector (except, of course, the simulation output also contains the "Monte Carlo truth", which the real data does not). No high claim to the performance of an algorithm can be substantiated without a realistic accounting of the above-mentioned detector effects. Thus, DigiSim plays a vital role, and has been warmly welcomed by the user community worldwide. We expect it to be used for the simulation of both the various test-beam prototypes and full-detector designs.

A summary of recent results from the above activities can be found in the 3 presentations made by the project leader, on behalf of the NIU/NICADD group, at the ILCD05 meeting at Paris.[7]

Among the members of our group we have adequate experience in calorimeter hardware, electronics, reconstruction software, and algorithm development. We anticipate close collaboration with other groups with similar interests. Active links have been established with ANL, SLAC, FNAL, DESY, and several university groups including the CALICE member institutions. We have produced, with groups across the world as signatories, a preliminary "requirements document" for the simulation software suite for the ILC detector(s).[8]

The steady progress that we have achieved so far has been made possible by funding received received for this purpose during the past 3 fiscal years from DOE and NSF, in addition to generous, but less specific, funding from the Department of Education. In FY 2002 we received $45K from the DOE under its Advanced Detector Research program. An exploratory grant of $8.5K was awarded by the NSF in following year. Finally, last year (FY2004) we were awarded $35K for our simulation software and algorithm development activities.

Activities outlined in this proposal are synergistic with the proposals for hardware prototyping of different technology choices. We will continue to remain in close contact with the groups involved in hardware development for the ECal and the HCal.

Facilities, Equipment and Other Resources

The funding requested here will be augmented by the following support, totaling approximately $500K, from other sources:

(a) NIU/NICADD personnel,
(b) ANL/SLAC personnel,
(c) Computing hardware and support provided by NICADD,
(d) 40-CPU Fermilab Linux farm (run by NIU personnel). These machines are relatively old, with per-CPU-capacity roughly a fifth of those requested in this proposal.

FY2005 activities and deliverables

During the first year we will concentrate on two things. First, design and development of well-structured PFAs that can be easily ported across detector design details, as described before. Both analog and digital versions (for the hadronic section) of the algorithms, which give encouraging preliminary results, will be further investigated and optimized. Second, completion of the DigiSim package so the algorithms can be tuned on input that closely resembles real data. Beam test will provide an opportunity to understand not only the detector hardware, but the simulation and reconstruction software as well.

The first year deliverable will be a first version of a class of particle-flow algorithms based on full simulation and reconstruction of the calorimeter and the tracking system. In addition, the standard GEANT4-based simulation facility (farm+server) will be available for to the entire ILC community through a web-based request form.[3] Also, the TBMokka simulator will be finalized during this period.

FY2006 activities and deliverables

Apart from extensive tuning of the algorithms, comprehensive studies of critical physics processes will have to be carried out in order to understand the impact of the calorimeter performance on the physics program of the Linear Collider. These studies will employ both the analog and digital versions of our PFAs. The second year deliverables will be further development of PFA-based jet-reconstruction and a detailed assessment of physics reach vs calorimeter performance for the ILC with a clear statement on the desirability of a digital or analog option for the hadronic calorimeter.

FY2007 activities and deliverables

In the third year we will embark on the development of parameterized simulations of the particle-flow algorithms. The technology and geometry are expected to have been narrowed down by that time setting the stage for such parametrized fast simulation for extensive physics studies. By the end of the third year we expect to produce, in collaboration with other groups, a fast simulation program based on PFAs. In addition, extensive benchmarking of critical physics processes, as well as evolution of pattern-recognition and reconstruction algorithms will continue.

Budget justification

The first year's activities revolve around the development of particle-flow algorithms and simulation. This will involve NICADD staff members (not included in the budget shown here), and 1.0 FTE post-doctoral associate. Optimization and detailed performance studies of the algorithm will be carried out in the second year by 1.0 FTE graduate student and 1.0 FTE post-doc with additional support from NICADD staff. During the third year, the development of parameterized simulations will be supported by 1.0 FTE post-doc, together with 1.0 FTE graduate students. Communication of progress and exchange of ideas through international workshops and conferences will be crucial for our endeavor to have a global impact. We estimate four domestic trips at $1.5K each and two international trips at $2.0K each during the first year, and 1.5 times as many in the second and third years. The equipment cost accounts for a 20-CPU Linux mini-farm + file server which will be needed in early FY06 to augment the allocation from Fermilab, as the simulation service enters a serious production phase. This additional capacity will have to be doubled toward the end of FY 2007.

Fringe benefits to personnel at NIU's mandated rate of 52% of salary, and indirect costs at the off-site rate of 26% (instead of the usual 45%, since the requested personnel will work in offices at Fermilab allocated specifically for ILC R& D) are included in the requested amount.

Three-year budget, in then-year K$

Item	FY2005	FY2006	FY2007	Total
Post-doctoral Associates	41.0	42.2	43.5	126.7
Graduate Students	0	21.0	21.8	42.8
Undergraduate Students	0	0	0	0
Total Salaries and Wages	41.0	63.2	65.1	169.3
Fringe Benefits	23.8	24.5	25.2	73.5
Total Salaries, Wages and Fringe Benefits	62.7	85.5	90.3	238.5
Equipment	0	25.0	23.0	48.0
Travel	10.0	15.4	15.9	41.3
Other direct costs	0	0	0	0
Total direct costs	74.8	128.1	129.2	332.1
Indirect costs (26% of non-equipment)	19.4	26.8	27.6	73.8
Total direct and indirect costs	94.2	154.9	156.8	405.9

References

[1] The NICADD web page, presentations, documents: `http://nicadd.niu.edu/` and `http://nicadd.niu.edu/research/lcd/`,

[2] `http://polywww.in2p3.fr/flc/calice.html`,

[3] `http://nicadd.niu.edu/~jeremy/lcd/simreq/`,

[4] `http://nicadd.niu.edu/~lima/lcdg4/`,

[5] `http://nicadd.niu.edu/~jeremy/lcd/tbeam/`,

[6] `http://nicadd.niu.edu/digisim/DigiSim.html`,

[7] `http://polywww.in2p3.fr/actualites/congres/ilcd2005/`,

[8] The ILC detector Simulation Requirements document: `http://forum.linearcollider.org/` → "Full Simulations" → "Simulation Requirements Document".