

## **STATUS REPORT**

**Project Name:** “Calorimeter Reconstruction and Particle Flow for the Silicon Detector Concept”

### **Personnel and Institution requesting funding**

Matthew Charles, Usha Mallik, and Niels Meyer  
The University of Iowa

### **Project Leader**

Usha Mallik  
[usha-mallik@uiowa.edu](mailto:usha-mallik@uiowa.edu)  
(319) 335-0499

# Calorimeter Reconstruction and Particle Flow for the Silicon Detector Concept

Matthew Charles, Usha Mallik, and Niels Meyer  
The University of Iowa

Usha Mallik's group at SLAC is working on shower reconstruction for the highly segmented calorimeters of the Silicon Detector Concept. This work is a key component of a Particle Flow Algorithm currently under development with the goal to reach the excellent energy resolution needed for high precision measurements at the International Linear Collider.

## 1 Introduction

The International Linear Collider (ILC) is the highest priority future large project in particle physics and will complement the expected LHC findings with high precision measurements. To match this physics goal, the detector(s) at the ILC must reach excellent energy resolution of the order of  $30\%/\sqrt{E}$  [1].

Several detector concepts have been proposed. Three of these were discussed in detail at Snowmass 2005: the Large Detector Concept (LDC) [2], the Global Large Detector (GLD) [3], and the Silicon Detector concept (SiD) [4]. Each of these three assumes that the anticipated resolution can only be achieved by reconstruction of individual final state particles using a Particle Flow Algorithm (PFA). Of the three, the SiD uses the most fine-grained detectors: a high precision silicon tracker, a Silicon-Tungsten sampling electromagnetic calorimeter (ECAL) with approximate cell size  $4\text{mm}\times 4\text{mm}$ , and a hadronic calorimeter (HCAL) with approximate cell size  $1\text{cm}\times 1\text{cm}$ <sup>1</sup>.

In this proposal, we describe the corner stones of a PFA for the SiD, the contributions of our group towards its realization, and the status of our various projects. Only the immediate goals for this year are described; these are critical to achieving a realistic detector design.

---

<sup>1</sup>The final HCAL technology for the SiD has not yet been chosen; RPC-, GEM-, and scintillator-based detectors are being considered.

## 2 Particle Flow Algorithm

Energy deposition in calorimeters occurs via statistical processes. In general, the physics of electromagnetic showers is well-understood and shower development can be modelled straightforwardly. For photons and electrons of a given energy, there is a good degree of consistency across showers in the Silicon-Tungsten SiD ECAL, both in terms of the shape and of the total energy deposited in the active elements. By contrast, hadronic showers are more difficult to model and display a great deal of variation in structure; the energy resolution of hadronic showers in the SiD projective RPC calorimeter is expected to be significantly worse than that of electromagnetic showers in the ECAL [5]. A more accurate way to measure the energy of charged hadrons is to use momentum information from the tracker and a particle ID hypothesis. For the limiting case of an algorithm with perfect pattern recognition, the jet energy resolution is then dominated by the neutral hadron energy resolution.

In practice, the resolution is also degraded by an imperfect algorithm assigning charged energy to neutral particles and vice-versa: this is often referred to as the “confusion term” in the energy resolution. One of the main goals of a particle flow algorithm is to minimize this confusion term. This is only possible if individual particles within jets can be resolved, which in turn implies a high-granularity (imaging) calorimeter. The precision of the detector will increase the experimental challenges (e.g. large numbers of channels, calibration, and noise). A realistic PFA is essential to explore the mutual dependencies between detailed detector design, technical issues, costs, and the final energy resolution, and to guide the concept towards a final design optimized for physics output.

There are many technical challenges to achieving such good resolution: hardware, software, and algorithmic. One year ago, the SiD collaboration had barely the beginnings of a PFA. Since then there has been tremendous progress: implementations of many of the components are now available such that Steve Magill of ANL was able to present  $Z$ -pole results with a proof-of-concept PFA at Snowmass [6]. The challenge is now to develop and tune a PFA which takes full advantage of the fine granularity and which reaches the target energy resolution in high-multiplicity events at a center-of-mass energy of 500 GeV and above, and which is sufficiently general to allow comparisons of different detector designs.

## 3 Status Report

Substantial improvements in the PFA implementation are expected from more advanced calorimeter reconstruction techniques, where the Iowa group has built up competence since 2003. Through 2005, the post-docs Matthew Charles, Wolfgang Mader and Niels Meyer have contributed to the project, sharing their time between the Linear Collider effort and the group's commitment to the BaBar experiment at SLAC. Niels joined the group in February after graduating from the University of Hamburg; Wolfgang will leave the group in December.

### 3.1 Software Tools

In order to undertake detailed studies of cluster structure, it is essential to have algorithms which can identify general patterns such as track segments or dense energy deposits. In the previous proposal, work by Wolfgang on minimum ionising particle (MIP) segment finding was reported, including reconstruction of long-lived  $K_S^0$ . A preliminary version of a Minimum Spanning Tree (MST) clustering algorithm was also described; this takes as input a metric providing a definition of distance between two hits, and then clusters hits according to a threshold parameter on this distance. Since then, Wolfgang has formalized the code for both the MIP-finding and MST algorithms and committed it to the `hep.lcd` CVS repository.

Upon joining the group in February, Niels began by updating and generalizing the MST interface so that it could combine not just individual hits but also clusters, in anticipation of the need to associate secondary neutrals and other fragments with their parent cluster. He also added a generalised decision-maker interface for more flexible steering and user interaction.

Niels and Matthew have also worked to convert the existing code to the new `org.lcsim` software framework. This is essentially complete for the MST and track segment finding algorithms, as well as the decision-maker interface. The code is available in the `org.lcsim` CVS repository. Matthew is currently adding his structural algorithm (see Sec. 3.4) and has begun contributing to the SLAC group's work on geometrical routines (principally the two-way conversion between calorimeter cell channel ID and spatial location) which are still missing in the new software framework.

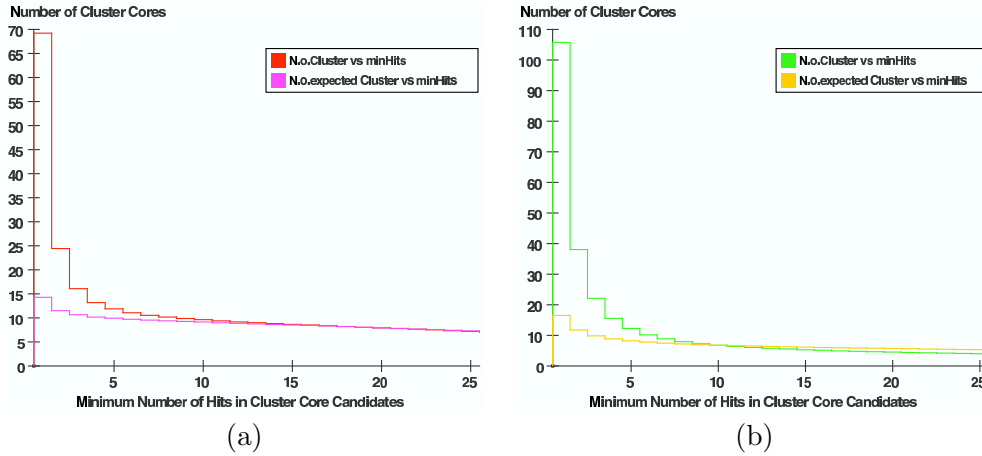


Figure 1: The mean number of clusters per  $Z^0$  event as a function of the minimum number of hits  $n$  in the cluster, for (a) the ECAL, (b) the HCAL. The first trace shows the number of reconstructed clusters with at least  $n$  hits in the calorimeter; the second trace shows the number of MC particles with at least  $n$  hits in the calorimeter, extracted from truth information. The excess of reconstructed clusters at small  $n$  is primarily due to a large number of small secondary clusters.

### 3.2 MST Studies with $Z$ -Pole Events

The MST algorithm links contiguous groups of hits into clusters if the three dimensional distance between hits is used as its metric. This is particularly effective for hadronic clusters, which frequently have many secondary tracks emerging at wide angles; a simple cone-based clustering method would have reduced efficiency in such cases.

Wolfgang studied the MST with this metric in detail using hadronic events simulated at the  $Z$ -pole. The threshold parameter of the algorithm is varied, and output clusters with at least  $n$  hits are identified as cluster cores, the leading contributions of a shower. By comparing the reconstructed and expected number of cluster cores, see Fig. 1, the optimal combination of threshold and minimum size of a cluster core was obtained for the electromagnetic and the hadronic calorimeters (3 cm and 5 hits for the ECAL, and 10 cm and 8 hits for the HCAL). The results from this study were presented at the LCWS workshop at SLAC [7].

### 3.3 Electromagnetic Showers

Niels studied the reconstruction of electromagnetic showers, starting with the MST algorithm and parameters obtained as described above. Photons were identified with a simple selection based on cluster size, shape, composition and position in the calorimeter. The selection is tuned to accept one cluster per shower, as shown in Fig. 2a.

The efficiency and purity of electromagnetic clusters reconstructed in  $K_S \rightarrow \pi^0\pi^0$  events were studied. Because the photon clusters are typically close to one another, it was found that a large value of the threshold parameter (e.g. 3 cm, as obtained for hadronic  $Z$ -pole events in the ECAL) results in reduced purity: clusters frequently contain energy deposits from two different particles, even in cases where the showers could easily be separated by eye on an event display. Furthermore, these merged clusters are inconsistent with a single-photon shower and therefore fail the shape cuts mentioned above, leading to a low efficiency. On the other hand, a very small threshold value results in improved purity at the expense of energy collection efficiency. The solution is a reconstruction in two passes: Identifying the shower cores with a very tight MST threshold, and assigning remaining energy deposits (fragments) to the cores.

Using this approach, it has been shown that showers from two photons can be resolved if the separation between the photons at the calorimeter surface is 3cm or more (see Fig. 2b). The cores found in this way contain the majority of the total energy deposition, as illustrated in Fig. 2c. Currently, strategies to enhance the energy collection efficiency based on a two dimensional distance definition between fragments and the principle axis of the shower core are under study. Preliminary results, produced in the `hep.lcd` framework, were presented at Snowmass in August 2005 [8, 9].

### 3.4 Hadronic Showers

Reconstruction and identification of hadronic showers is central to the PFA approach. There is a great deal of variation between individual showers: designing a general algorithm to reconstruct them is not straightforward. Matthew and Usha began to tackle this problem by studying a number of single-particle and low-multiplicity events in detail, attempting to understand their structure in such a fine-grained detector. Based on their observations, Matthew developed the following method.

The components of hadronic clusters may be categorized as (a) dense clumps, (b) track segments, (c) a halo of less dense hits following a hard

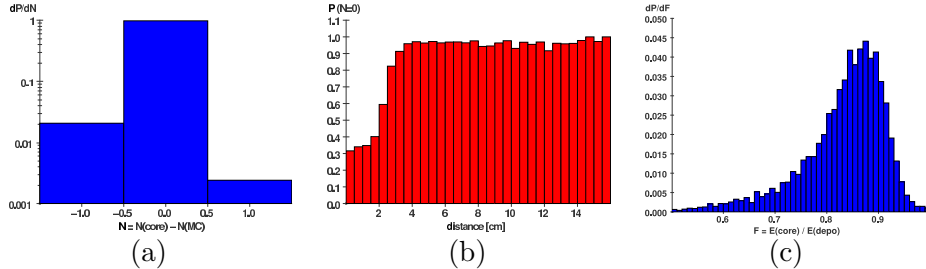


Figure 2: Performance of the photon-finder in a sample of  $K_S \rightarrow \pi^0\pi^0$  events. Plot (a) shows the difference between the number of photons produced and the number reconstructed. Plot (b) shows the ratio of reconstructed clusters to actual photons as a function of the photon separation on the calorimeter surface. Plot (c) shows the fraction of the energy recovered when a core is found.

interaction, and (d) displaced secondary fragments. Code based on the MIP-finder and MST tools was written to identify components (a) and (b), and a cut-based selection was developed to determine whether a given pair of components were directly linked. By linking together these basic components, the “skeletons” of hadronic showers are reconstructed; components (c) and (d) can then be added to recover the remaining hits. In this way, hadronic showers can, in principle, be reconstructed with high efficiency and purity even in very dense environments. Preliminary results from this algorithm, produced in the `hep.lcd` framework, were presented at Snowmass in August 2005 [10, 11].

After Snowmass, the algorithm was revised to use a likelihood selector to identify correct (or incorrect) links in the place of the cut-based selection. Further geometrical criteria were also added to the selector. In order to assess the algorithm’s performance, modular code which used truth information (i.e. cheating) at each stage of a full PFA was written. The performance when using truth information throughout was evaluated on a sample of approximately 400 hadronic  $Z$ -pole events simulated for a version of the SiD detector with a sampling scintillator HCAL. The (non-cheating) reconstruction algorithm for identifying and linking components (a) and (b) was then substituted for the corresponding module and the performance re-evaluated on the same sample of events. The result, an energy sum distribution with RMS 3.4 GeV, was—within the statistics—indistinguishable from the energy resolution achieved when using the full truth information.

Adding a simple, non-cheating module to associate the fragments in categories (c) and (d) without using truth information worsened the resolution to an RMS of 4.3 GeV. These results, produced in the `hep.lcd` framework, were presented to the SiD group in September 2005 [12] and are shown in Figure 3.

The algorithm has now been converted to the `org.lcsim` framework; this required rewriting the bulk of the code. Several new features have been added in the process to make a complete PFA; in particular, charged tracks are now matched to clusters in a much more realistic fashion, extrapolating helices from the interaction point and looking for a consistent track segment near the entry point to the ECAL (or, failing that, another nearby cluster). A status report was presented to the SiD group in November 2005 [13], including preliminary results from the likelihood selector which are shown in Figure 4. Initial energy sum plots look promising, but the code is still being debugged and tested.

From the results obtained since Snowmass, this method of reconstructing hadronic clusters seems promising, and performed well at finding and identifying the main body of the clusters in hadronic  $Z$ -pole events. The critical challenge—for this and other clustering algorithms—will be the association of fragments, principally secondary neutrals. This is where effort will be focused next, using experience and tools from the algorithm described above.

## 4 Future Plans (Deliverables)

The next steps will be to improve and extend the algorithms presented in sections 3.3 and 3.4. Code to assign fragments to nearby cores will be developed both for electromagnetic and for hadronic showers. The different algorithms will then be integrated into a single PFA with the following general structure:

1. Find the cores of electromagnetic showers
2. Make initial assignments of fragments to the electromagnetic cores
3. Find the core components of hadronic showers and link them (where appropriate) to form cluster skeletons
4. Make initial assignments of fragments to the hadronic showers
5. Extrapolate charged tracks to the calorimeter surface and associate them with clusters

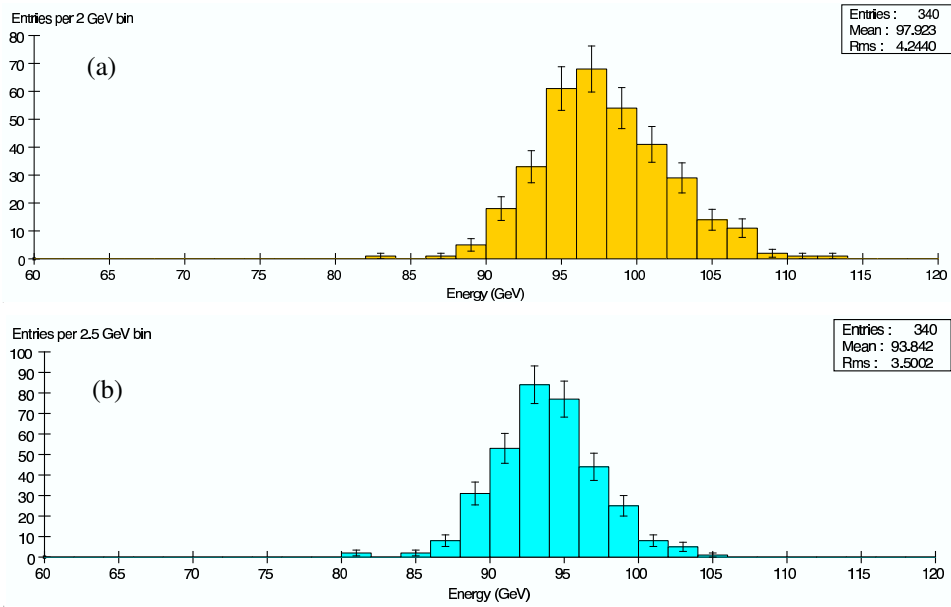


Figure 3: Energy sum plots for  $Z$ -pole events, showing (a) the reconstructed energy sums without cheating, and (b) the reconstructed energy sum for a PFA with perfect pattern recognition. The true energy sum is 91.0 GeV, but the correct sampling fractions were not available for this simulated detector; as a result, the overall energy scale is off and the resolutions are worse than could be achieved with full calibration. The RMS values correspond to  $37\%/\sqrt{E}$  for (a) and  $45\%/\sqrt{E}$  for (b).

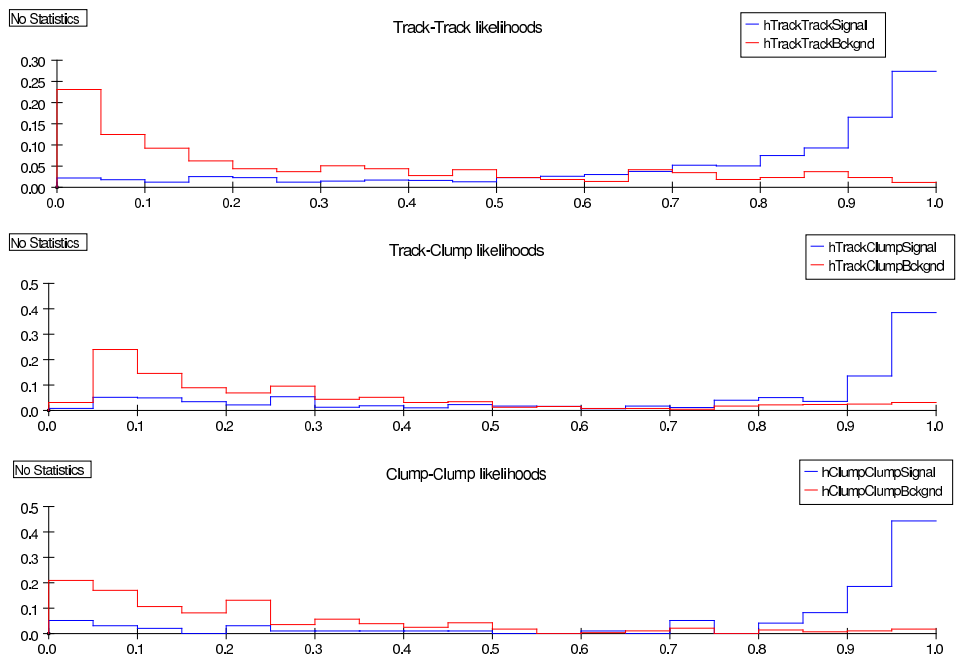


Figure 4: Likelihood distributions for links between components of hadronic clusters, obtained in a sample of hadronic  $Z$ -pole events.

6. Refine the assignment of fragments and hadronic components (especially if the assignments are ambiguous, or if the energy of a charged cluster is inconsistent with the track momentum)

Once this is accomplished, additional steps to further improve the algorithm will be considered—for example, handling of  $K_S$  and other long-lived particles, merging of  $\gamma$  pairs into  $\pi^0$ s, and using event information to improve the cluster assignments iteratively. Throughout the development, an important goal is to keep the algorithm as general as possible so that it can be applied to other detector designs.

## 5 Resources and Budget

The group has devoted the effort of 1.5 post-docs to the Particle Flow Algorithm for the SiD concept, described in the proposal. Last year the financial support to the Iowa group (Task A) for ILC R&D in the form of a supplement was \$31,500 (\$50,000 the year before). The base program of the group working on the BaBar experiment has historically been supported with two post-docs (and students). In anticipation of the level of ILC activity, and the increased necessity to establish a PFA, the group had hired Dr Wolfgang Mader in early 2004; he could not be supported beyond December 2005, and is leaving. The ILC activity at the current level has so far been supported by the base program primarily. This is not practical any longer. A minimum personnel support of one post-doc is needed to sustain the R&D activity of the group to maintain the steady and continuous progress. The rest (half a post-doc) can be supplied from the base funding. Additionally, there are frequent regional workshops and a few LCWS meetings where travel is necessary for active workers. Travel for the post-docs and the PI to some of these is included. The total estimated cost includes 32.7% fringe benefit for the personnel, travel and 26% off-site indirect cost rate, and is explained in detail below.

## References

- [1] Worldwide Linear Collider Study, 2002 Report on International LC Detector R&D, <http://physics.uoregon.edu/~lc/randd.pdf>
- [2] LDC concept webpage, <http://www.ilcldc.org/>
- [3] GLD concept webpage, <http://ilcphys.kek.jp/gld/>

- [4] SiD concept webpage, <http://www-sid.slac.stanford.edu/>
- [5] J. Repond, SiD Calorimeter Overview,  
[http://alcp2005.colorado.edu:8080/alcp2005/program/detector/SID/jose\\_repond20050815231213.ppt](http://alcp2005.colorado.edu:8080/alcp2005/program/detector/SID/jose_repond20050815231213.ppt)  
*Presentation at the 2005 International Linear Collider Physics and Detector Workshop, Snowmass, Colorado, 14-17 Aug 2005.*
- [6] S. Magill, PFA Development for a LC Calorimeter,  
[http://alcp2005.colorado.edu:8080/alcp2005/program/detector/PFA/steve\\_magill20050824151757.ppt](http://alcp2005.colorado.edu:8080/alcp2005/program/detector/PFA/steve_magill20050824151757.ppt)  
*Presentation at the 2005 International Linear Collider Physics and Detector Workshop, Snowmass, Colorado, 14-17 Aug 2005.*
- [7] W. Mader, MIP Reconstruction Techniques and Minimum Spanning Tree Clustering, SLAC-PUB-11359, *Contributed to 2005 International Linear Collider Workshop (LCWS 2005), Stanford, California, 18-22 Mar 2005*
- [8] N. Meyer, Electromagnetic Showers with the MST Algorithm,  
[http://zebu.uoregon.edu/~rayfrey/LC/SiD-cal\\_Snow05/meyer.ppt](http://zebu.uoregon.edu/~rayfrey/LC/SiD-cal_Snow05/meyer.ppt)  
*Presented at the SiD Calorimetry session, 2005 International Linear Collider Physics and Detector Workshop, Snowmass, Colorado, 14-27 Aug 2005.*
- [9] N. Meyer, Electromagnetic Showers with the MST Algorithm,  
<https://wiki.lepp.cornell.edu/wws/pub/Projects/CalIowaPfa/niels-20050822-Snowmass.pdf>  
*Presented at the Particle Flow Algorithm session, 2005 International Linear Collider Physics and Detector Workshop, Snowmass, Colorado, 14-27 Aug 2005.*
- [10] M. Charles, Dissecting the Structure of Hadronic Clusters,  
[http://zebu.uoregon.edu/~rayfrey/LC/SiD-cal\\_Snow05/mcharles.pdf](http://zebu.uoregon.edu/~rayfrey/LC/SiD-cal_Snow05/mcharles.pdf)  
*Presented at the SiD Calorimetry session, 2005 International Linear Collider Physics and Detector Workshop, Snowmass, Colorado, 14-27 Aug 2005.*
- [11] M. Charles, Dissecting the Structure of Hadronic Clusters,  
[http://alcp2005.colorado.edu:8080/alcp2005/program/detector/PFA/mat\\_charles20050825234612.pdf](http://alcp2005.colorado.edu:8080/alcp2005/program/detector/PFA/mat_charles20050825234612.pdf)  
*Presented at the Particle Flow Algorithm session, 2005 International*

*Linear Collider Physics and Detector Workshop, Snowmass, Colorado,  
14-27 Aug 2005.*

- [12] M. Charles, Status of Particle Flow Studies,  
[http://www.slac.stanford.edu/~mcharles/talks/  
2005-09-21\\_pflow/mcharles-short.pdf](http://www.slac.stanford.edu/~mcharles/talks/2005-09-21_pflow/mcharles-short.pdf)  
*Presented to the SiD Calorimetry group, 21 Sep 2005.*
  
- [13] M. Charles, Update on cluster association algorithms,  
[http://www.slac.stanford.edu/~mcharles/talks/  
2005-11-17\\_pfa/slides-mcharles.pdf](http://www.slac.stanford.edu/~mcharles/talks/2005-11-17_pfa/slides-mcharles.pdf)  
*Presented to the SiD Calorimetry group, 17 Nov 2005.*

BUDGET JUSTIFICATION / EXPLANATION PAGE  
The University of Iowa  
(U. Mallik)

	Detail	Year Total
<b><u>SALARIES AND WAGES</u></b>		
Asst. Research Scientist		
months/year:	12.00	
rate/month:	3,833.33	46,000
<b>Total Salaries and Wages</b>		<b>46,000</b>
 <b><u>STAFF BENEFITS</u></b>		
Asst. Research Scientist	32.7%	15,042
<b>Total Staff Benefits</b>		<b>15,042</b>
 <b><u>TRAVEL</u></b>		
Domestic:		
Attendance and participation in spring & fall scientific conferences		
Trips	2	
Persons	1	
Days	3	
Air fare	750 /trip	1,500
Subsistence	150 /day	900
Car rental	50 /day	300
Registration	300	600
Subtotal		3,300
 Attendance and participation in summer scientific conference for Post-doc and PI		
Trips	1	
Persons	2	
Days	7	
Air fare	750 /trip	1,500
Subsistence	150 /day	2,100
Car rental	50 /day	350
Registration	300	600
Subtotal		4,550
 Foreign:		
Scientific collaboration at Bangalore, India		
Trips	1	
Persons	1	
Days	7	
Air fare	2,300 /trip	2,300
Subsistence	260 /day	1,820
Subtotal		4,120
 <b>Total Travel</b>		 <b>11,970</b>

	<u>Detail</u>	<u>Year Total</u>
<b><u>OTHER DIRECT COSTS</u></b>		
Computer maintenance and repair		100
Telephone, data and modem lines for project activities		100
Photo and printing services of research reports/papers		100
Postage, express delivery and other expenses		<u>100</u>
<b>Total Other Direct Costs</b>		<b>400</b>
<b><u>TOTAL DIRECT COSTS</u></b>		<b><u>73,412</u></b>
<b><u>FACILITIES AND ADMINISTRATIVE COSTS</u></b>		
26% of Modified Total Direct Costs (MTDC) for off-campus research activities, per rate agreement negotiated with DHHS.		
Total Direct Costs	73,412	
Less Exclusions	<u>0</u>	
MTDC	73,412	
F&A Rate	26.0%	
<b>F&amp;A Costs</b>		<b><u>19,087</u></b>
<b>TOTAL ESTIMATED COSTS</b>		<b><u><u>92,499</u></u></b>