Discussion paper on the use of MAPS in a LC electromagnetic calorimeter

June 8, 2004

1 Introduction

The CALICE collaboration is studying a tungsten-silicon electromagnetic calorimeter (ECAL) for use at a future linear collider (LC). Although the detector design is by no means fixed, the current ideas are described in section 2. Section 3 describes the possible use of MAPS as the sensitive detector. Section 4 lists the outstanding issues which need to be answered.

The timescale for a LC is still very uncertain, but a working schedule has the machine turning on for physics in 2015. The ECAL will need several years to assemble before this date and, together with prototyping, it is assumed the TDR must be ready around 2009. This implies there is around 5 years of R&D to define the basic concept of the ECAL.

2 Current CALICE design

The physics of the LC requires accurate energy reconstruction for hadronic jets, which in terms of the ECAL means tracking single MIP particles through the ECAL. This requires high granularity of the sensitive detectors in the ECAL and it needs to be effectively a "tracking calorimeter".

The sensitive detectors being considered by CALICE are silicon wafers with simple 1×1 cm^2 diode pads and it is assumed that each wafer will have something like 16×16 pads. These wafers will be mounted directly onto a PCB, which will made as wide an one wafer and as long as feasible, ideally the full length required of 1.6m. The signals from the silicon wafers will be AC-coupled to an ASIC, also mounted on the PCB, which will probably provide shaping, digitisation and some level of buffering. Each wafer will be read by one or two ASICs. At the end of the PCB will be circuitry to collect the data from all the wafers and send them out by fibre to off-detector receivers.

The main requirements for the diode/ASIC system are set by the shower densities expected. In the core of the highest energy showers at the LC, the density will be up to 100 MIPs/mm^2. Hence, over a 1×1 cm^2 diode pad, the maximum expected signal corresponds to around 10^4 MIPs. The diode pads also have to be sensitive to single MIPs, which defines the dynamic range as $10^4 \sim 2^{14}$. For reasonable efficiency, the signal/noise for a single MIP needs to be at least 4/1, which means an equivalent digitisation range of 16 bits is needed. In practise, it is assumed this will be handled with a multi-gain amplifier.

To perform well, the ECAL must also minimise the gap between the tungsten converter sheets. This constrains the thickness of the wafer/ASIC/PCB structure and makes cooling a significant problem. Hence, low power is a major aim for all the electronics on the PCB.

A very large amount of silicon wafers is needed for the ECAL. The overall design has a cylinder of length around 3m, an average radius of around 1m and has 40 layers of silicon wafers. There are also endcaps. In total, the area of silicon needed is around 3000 m^2. With each wafer being 256 cm^2, this is approximately 10^5 wafers. Ease of fabrication and assembly are major issues but the biggest consequence is the cost. Even with an optimistic price in the future of £1/cm^2, this is still £30M for the silicon alone. Clearly, there is a major incentive to improve the performance so as to be able to reduce the required number of layers.

The radiation levels are very low compared with LHC, being estimated to be at least a factor of 10^3 lower. Correspondingly, the event rate for interesting events is also much lower at around 1Hz or less.

One final complication is the LC technology choice since the beam collision timing depends on this. Currently there are two main designs being considered. The superconducting ("cold") design, originating from DESY, has a bunch structure of up to 5000 beam crossings within 1ms, i.e. a 200ns spacing, with a 199ms gap of no beam until the next bunch train. The "warm" design, from SLAC/Japan, has much shorter bunch trains and crossing times, with 300 beam crossings within a 400ns train, i.e. only a 1.4ns spacing. To obtain the 4/1 or better MIP signal/noise required, the ASIC shaping times being considered are of the order of 100ns. This would not allow differentiation between crossings within the warm machine bunch train and so would require all background to be integrated over the whole train and hence be superimposed on top of the single event of interest. It is not clear if there is a need for cosmics; if taken between bunch trains then this could substantially reduce the amount of time available to read out the data from each train. A decision on the LC technology is expected by the end of 2004.

2.1 Possible use of MAPs

The issues which it would be worth considering if MAPS can improve are the following:

1. Reliability of bulk fabrication and assembly.
2. Improved mechanical thinness and/or cooling.
3. Better timing resolution for the warm machine design.
4. Better tracking and hence reduction of cost.

Clearly, one possible approach would be to have MAPS with pixels the same size as the diode pads and then integrate the ASIC circuitry into the MAPS. This would potentially benefit items 1 and 2 above:

- The electrical connection of the wafers to the PCB in the CALICE design is quite simple; the pads and power connections are assumed to be bump-bonded directly to the PCB. The ASIC will probably require a more complicated connection and so would be a source of assembly effort (and associated failure probability).
- The major heat load in the CALICE design comes from the ASICs, which are small compared with the wafers they read out and so the heat sources are lumped. This leads to local heating and large temperature differentials across the PCB. With the readout circuitry spread more evenly over the surface of the MAPS, this would be reduced. In addition, the structure surface could be smoother as the ASICs stick out by several mm from the PCB surface and make any cooling pipework complicated. MAPS might make the mechnical integration of the cooling and electronics easier and would potentially make the whole structure thinner.

It would be interesting to know if such an approach would be straightforward; an obvious concern is that the ASIC design on the MAPS could need significant effort. In addition, unless the MAPS can be made to be a similar size to the diode wafers, any improvements in assembly may be outweighed by the larger number needed to be assembled.

However, it would be interesting to consider a conceptually simpler approach which helps in terms of items 1 and 2 (as above) but also might improve items 3 and 4 also. Here, no analogue readout of a pixel is made, but each is read out as one bit, indicating whether a signal of a MIP (or above) was seen in the pixel. Each pixel would need a discriminator and good enough signal/noise to make the output clean. Clearly, resolution is lost if the probability of more than one MIP within one pixel is high, so the pixel size must be small enough to handle the highest density expected, i.e. 100 MIPs/mm^2. This indicates a size of 100×100 μm^2 for the pixels is the maximum and possibly something half the size in each dimension, i.e. 50×50 μm^2 would keep the probability of multiple MIPs in a pixel low enough. With pixels of this size, a MIP signal/noise of at least 10/1 might be achieved, giving a clean signal if the discriminator threshold can be kept stable. This size is assumed in the following. Possible benefits are then:

- An obvious consequence of the much finer pixel size is that the tracking is correspondingly more accurate. This could reduce the number of layers of silicon needed and hence the cost.
- Retrieving the signal quickly could be possible, allowing the possibility for the warm machine that beam crossings within the bunch train can be distinguished, or at least than the integration time can be reduced from the whole train of 300ns to something significantly smaller, like 30ns or better.

Another obvious consequence of the much finer pixel size is that the data volume would increase enormously. Each 1×1 cm^2 diode pad, which was read out as a 16 bit value, has been replaced by a 200×200 array of pixels, each reading out as one bit, which results in 40 kbits. The CALICE design has 30M diode pads, so the raw data volume per ADC sample is 60 MBytes. The MAPS would result in a raw data volume per sample of 150 GBytes. Clearly due to the low event rate, most channels have no signal so a huge amount of zero suppression will be possible in both cases. However, the increase in raw data rate might make this simplest scheme unfeasible. One approach then might be to sum the bits over $2^n \times 2^n$ arrays of pixels and report out a digital sum of the pixels in that area. The data size for each sum would be $2n$ bits, rather than a single bit, but the number of sums would be $1/2^{2n}$ of the number of pixels, giving a volume reduction of $2n/2^{2n}$. E.g. summing over a $32 \times 32 = 2^5 \times 2^5$ pixel array would give an effective pixel size of 1.6 mm, have a data size of 10 bits and would reduce the effective number of pixels by 1024, meaning a raw data volume per sample of around 1.5 GBytes. The effective pixel size is still much smaller than the diode pads. (The advantage of this over a 1.6×1.6 mm^2 pixel is that no ADC is required and the signal/noise and readout speed will be better.) Such digital addition is presumably trivial compared with a 16 bit ADC design. One could even think of performing sums only when required; in cases where there are very few hits, individual pixels could be reported out, while in dense showers, sums could be sent instead and even the value of n used could be density dependent.

Another issue is the threshold required. This would have to be stable and, given the number of channels, would need to be uniform at least for all channels on a single MAPS. Having trim DACs on the MAPS would be operationally complicated given the number of wafers in the system.

3 Outstanding issues

1. What are feasible data rates output from MAPS? What sparsification algorithms would be possible to implement? How much buffering (in terms of bytes of memory) could be feasibly built into MAPS?
2. How stable and uniform would a discriminator level be?

3. What is the highest possible speed of readout of $50 \times 50\ \mu\mathrm{m}^2$ pixels? What time resolution on the energy deposits could be obtained?
4. How much improvement in jet reconstruction could be obtained using $50 \times 50\ \mu\mathrm{m}^2$ digital pixels compared with $1 \times 1\ \mathrm{cm}^2$ analogue pads? How does this translate into reducing the silicon area without degrading the performance?
5. What is the likely power consumption of such MAPS? Is it feasible to only power up during bunch trains, given the need for pedestal, gain and discriminator stability?
6. What would be the cost per cm^2 of MAPS?
7. What component size could be made with MAPS?