Power Delivery Challenges for Multi-Core Microprocessors

Larry E. Mosley
Intel Corporation
MS SC1-19
2200 Mission College Blvd.
Santa Clara, CA 95054 USA
Tel: 1-408-765-0206
larry.e.mosley@intel.com

Introduction

In the past the power of the microprocessor has increased exponentially[1] and with this increase came an increase in the cost of power delivery. With power levels having exceeded 100 watts it had become clear that a change was needed and in recent years there has been a major emphasis on controlling the power of the microprocessor; however, higher power has also traditionally meant higher performance. In order to continue to improve performance one new development has been the introduction of multi-core microprocessors. These microprocessors have set new standards in performance per watt. Dual core processors are now common and quad core processors are ramping quickly. Looking to the future we can expect the number of cores to continue to increase. At some point we may reach a point of diminishing returns but microprocessors are being discussed having 100+ cores[2]. While the increase in the number of cores will continue to drive performance the power delivery will become more challenging. This paper will look at some possible multi-core scenarios and the challenges that may arise in looking for cost effective power delivery solutions. In addition the role that multi-layer ceramic capacitors may have for decoupling these multi-core microprocessors will be discussed.

Multi-Core

Background

Historically the microprocessor power (P) has increased approximately 20% per year (figure-1). This increase in power was primarily a result of an exponential increase in frequency (f) and was driven by the desire for ever greater performance; as can be seen from equation-1, the power is directly proportional to the frequency:

$$P = CV^2f \qquad [1]$$

where C is the amount of switching capacitance and V the applied voltage. As it became apparent that we could no longer continue this exponential increase in power a significant effort has gone into how to limit it; in the last several years the frequency, and power levels, have not seen any significant increase and in many market segments they have gone down. If we had continued the historic exponential increase in frequency we would be in the 15GHz to 20GHz range today; instead the frequency hit a peak at just around 4GHz a few years ago and now 3GHz is more typical. Since much of the performance gain in the past has come from increased frequencies we needed to find alternative methods to significantly improve performance. One of the most significant recent changes has been the move to multi-core microprocessors. As Justin Rattner (Intel Corporation Chief Technology Officer) has observed[2] "It's rather extraordinary that after decades of single core processors, the high volume processor industry has gone from single to dual to quad-core in just the last two years". This trend to increased numbers of cores is expected to continue. In early 2007 Intel Corporation demonstrated a Teraflop Research Chip (Polaris)[3] made up of 80 cores with each core measuring about 3mm^2. Again quoting Justin Rattner[2] "Moore's Law scaling should easily let us hit the 80-core mark in a mainstream processor within the next ten years and quite possibly even less." the 80-core mark in a mainstream processor within the next ten years and quite possibly even less." Within the next 5 to 10 years we could easily see a 100+ core processor. It's too early to know just what these centi-core processors will look like but it is expected that they will bring significant challenges in many areas, including power delivery. The next section will look at some of the possible power delivery challenges we will face especially in the area of decoupling.

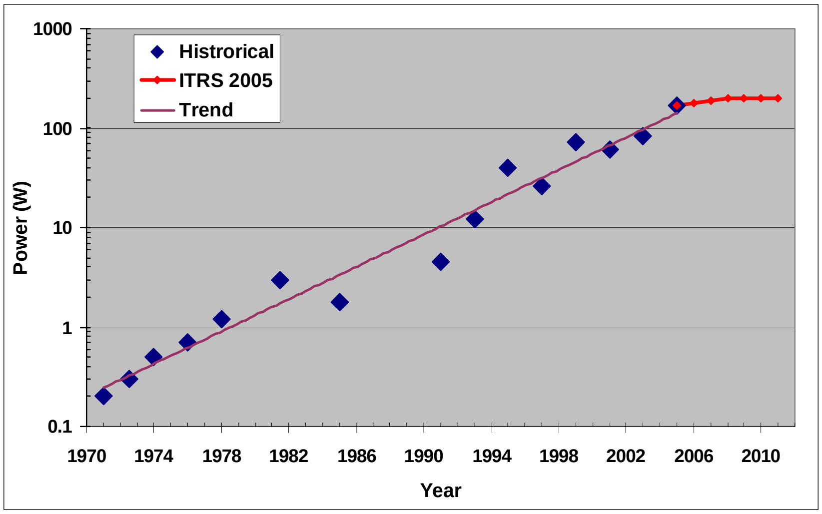

Figure-1: Historical power trends. Also shown is the ITRS 2005 prediction highlighting the trend to limit the power in future microprocessors.

At this point it is probably appropriate to state some disclaimers. As discussed above, the last few years has seen a radical change in the area of power demand and delivery; changes which were not foreseen even as late as the early part of this decade. No longer can we simply scale the power delivery systems. Today we have to look out into a hazy future and try and visualize what that future will look like. In the discussion that follows keep in mind that the scenarios discussed are only possibilities and will need to be refined over time. The goal of this paper is to stimulate thinking and discussion in this area which will hopefully lead to novel, cost effective power delivery solutions for the future.

Multi-Core Power Delivery

To aid in the discussion it is useful to define a baseline case: 130W (100A, 1.3V, 3GHz) and a 1cm^2 total core die area; total core die area is the sum of the areas of all of the individual cores, so for example a single core die would have a single 1cm^2 core and a quad core die would be made up of 4 0.25cm^2 cores. These numbers were chosen as they are fairly typical of the high end range seen today.

Figure-2 illustrates a series of multi-core die from a single core to 100 cores. The current industry trend appears to be a doubling of the number of cores about every 2 years. At this rate a 100 core processor could be a reality in the second half of the next decade. From a power delivery perspective, what might the decoupling needs look like for a 100 core processor? In order to answer this question one needs to make some assumptions about the nature of these cores. How will they be used? What power features are wanted or needed? The simplest case would be that all of the cores are on a single voltage rail and that the power is uniformly distributed among all of the cores. Based on our baseline an individual core would be about 1mm on a side and consume on average 1.3W of power. For this simple case the decoupling would be fairly straight forward and would be like decoupling a large single core die. At the other extreme one could envision that it would be desirable to have each individual core on its own voltage rail. By being able to control the voltage of each core, one could lower the voltage on those cores that don’t need to be running at maximum performance at that time, or even turn off some cores completely when they are not needed. This level of granularity would provide significant energy savings. In this extreme case one would need to provide decoupling to each individual core separately.

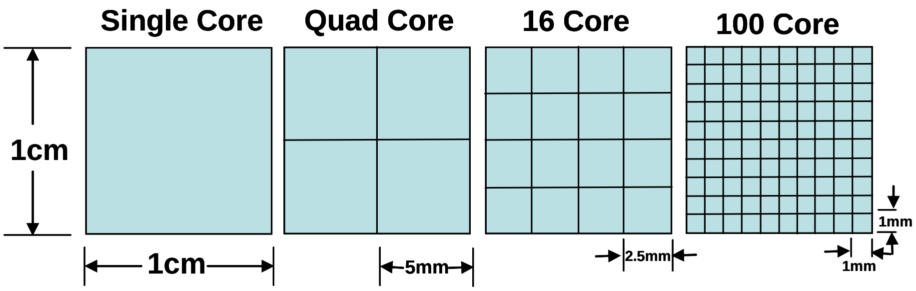

Figure-2: An illustration of the trend toward more and more cores on a chip.

The extreme case of having to decouple each individual core will probably not happen; rather there will most likely be something in between where some individual or cluster of cores will be on their own voltage rails, but this number will probably be limited. However the decoupling could still be just as challenging. Consider the example illustrated in figure-3. Here each core in the 3x3 cluster is assumed to be on a separate voltage rail. If one looks at trying to decouple the core in the center of the cluster, the challenge is just as great as trying to decouple each of the 100 cores individually.

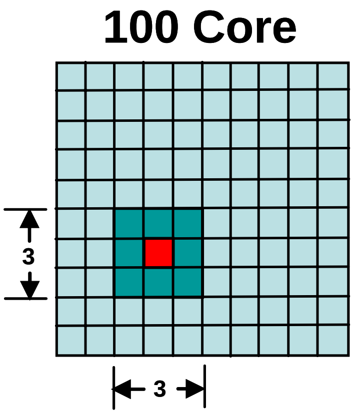

Figure-3: The 3x3 cluster of cores are each on an isolated voltage rail. The challenge is in providing decoupling to the core in the center of the cluster.

The cross section in figure-4a shows the die, package and land side capacitors (MLCC assumed) while figure-4b shows a similar cross-section except only for a single core. Both figures are roughly to scale, although not to the same scale. For figure-4a the die, representing the total core area of the die (excluding the cache), is 1cm long. In figure-4b the area marked die represents a single core and is 1mm long. The capacitor is representative of a 0402, shown lengthwise. As can be seen from the figures a 0402 (1mmx0.5mm) just fits lengthwise within the core area and since this does not take into account space for the solder pads or cap-to-cap spacing one could not use a 0402 to decouple the single core, so from a purely physical layout perspective one would be forced to go to at least a 0201.

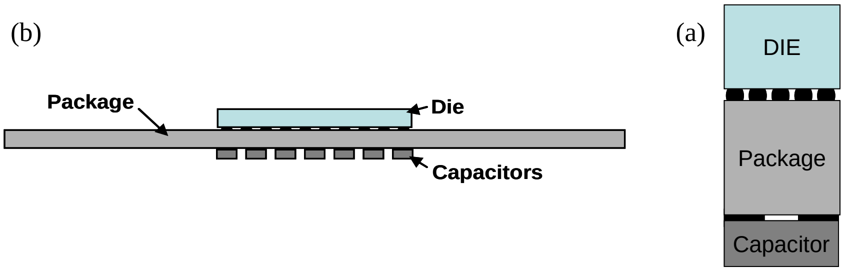

Figure-4: (a) Die, package, capacitor cross-section. (b) Cross-section of a single core ($1mm^2$) with package and capacitor.

Impedance

In this section the impedance, and in particular the inductance, requirements for decoupling these many-core processors will be discussed. In a previous paper[4] an attempt to predict the impedance requirements for future microprocessors was made. The figure from that paper is reproduced in figure-5 and includes an updated data point which can be used to measure the success of that prediction.

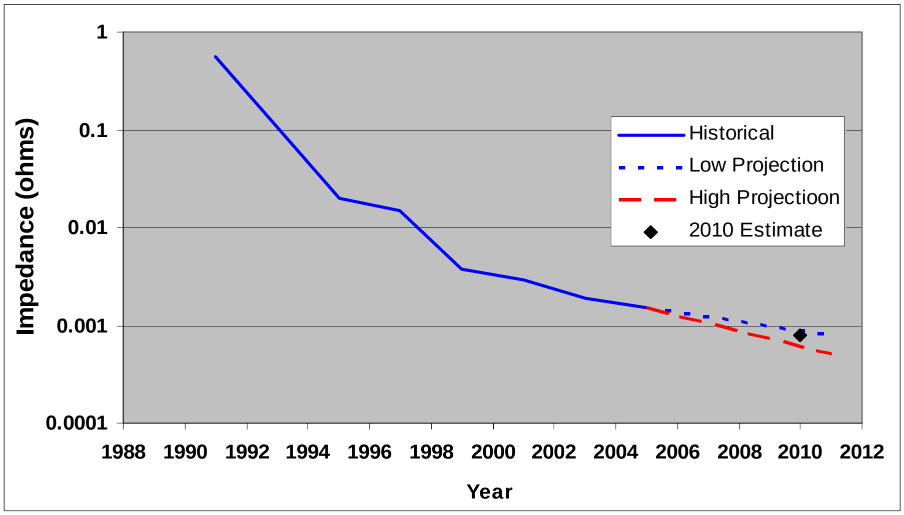

Figure-5: Impedance trends. The marker labeled "2010 estimate" represents the expected minimum impedance in 2010.

As can be seen in figure-5 a low and high projection was made for what the impedance requirements of the future might look like. The low projection represents a 20% reduction in the impedance per generation, approximately every 2 years, and the high projection represents a 30% reduction. The one updated data point in 2010 represents the best estimate for the minimum impedance in that timeframe and is about 0.8m . Looking at the projected impedance values for 2010 the low projection is about 0.9m and the high projection is about 0.63m ; the midpoint of these 2 values is approximately 0.76m which is close to, although slightly lower than, the estimated value. Assuming the same trend will continue into the future the impedance requirements should be on the order of 0.35m to 0.4m by 2016, which is around the time when we may start to see 100 core processors. This represents about a 60-70% reduction in impedance from today's values.

For the inductance requirements the discussion will be limited to first droop[5] only. As illustrated in figure-6, the first droop is a result of the interaction between the capacitance on die (C_{die}) and the total inductance from C_{die} to, and including, the first level of decoupling capacitance (C_1) off die, i.e. the land side capacitors previously described. The associated impedance is proportional to the square root of the ratio of the inductance (L) and C_{die} (see equation-2).

$$Z \sim (L/C_{die})^{1/2}.$$

[2]

The capacitance on die is not expected to change much in the future so any reduction in the required impedance will need to come from the reduction in the inductance and, due to the square root relationship, if the impedance needs to decrease by 60% then the inductance will need to decrease by about 85%. The effective inductance will need to be in the 0.4pH range; this assumes all cores are on the same voltage rail. To decouple a single core would need about 40pH; this is 40pH in a $1mm^2$ area and includes both the capacitor and package inductance. In the previous section it was suggested that layout considerations would limit the capacitor size to no more than that of a 0201. With improved layout rules one might get 2 0201 capacitors in a 1mm^2 area. Standard 2 terminals will not get us to 40pH, although reverse geometry might get us close, especially if there is a meaningful reduction in the package contribution.

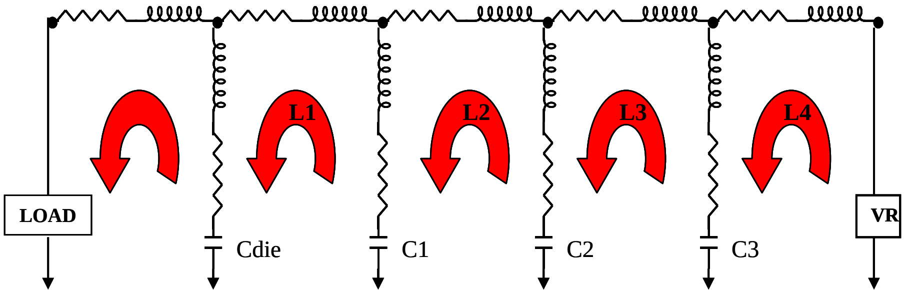

Figure-6: Simplified circuit diagram showing the various levels of decoupling between the voltage regulator (VR) and the die.

A further complication in decoupling a single core may come from the loss of C_{die}. The total on die capacitance is on the order of 0.1 F-0.2 F and is made up primarily of gate capacitance from both the core and cache. If we are decoupling a core on its own voltage rail most of this C_{die} will not be available to act as decoupling, therefore the inductance may need to be even lower, perhaps by a factor of 2 or more. In this case 2 0201's (reverse geometry) would not be sufficient.

Determining the amount of capacitance is a little more difficult. A typical rule of thumb is to have the amount of capacitance in the first stage of decoupling (C_1) to be at least 20x greater than the value of C_{die}; this is to minimize the noise that can result from the resonance between the 2 capacitors (C_1 and C_{die}). In addition C_1 depends on what the next stage in the power delivery network looks like. As one moves further from the die the amount of capacitance in each stage gets larger as does the inductance to that stage; typically the increase is at least an order of magnitude per stage. It is difficult to look out and say just what the power network for a 100 core processor will look like, but one thing that appears most likely is that the number of stages of decoupling between the die and voltage regulator (VR) will need to decrease in order to reduce the total impedance of the network. In the extreme case we may have only one stage of decoupling (C_1). The value of C_1 will depend on the impedance of the power network to the right of the first stage shown in figure-6, however based on all of the previous assumptions, it is estimated that somewhere between 1.5 F and 3 F per core (100 cores) will probably be needed.

Further Considerations

There are a number of other scenarios that could potentially be much worse than what has been previously discussed. Up until now it was assumed that the power is uniformly distributed across all of the cores. Consider the case where half the cores are turned off and the other half are on. One could allow the power of each core to double and still remain within the power envelope. This situation is somewhat analogous to having hot spots on a single core chip; hotspots are localized regions on the die having a power density much greater than the average. It is not unusual to have hot spots where the power density is 5 times greater than the average. The first droop inductance will need to scale with the increase in power per core so a further reduction of 2x or more in the inductance may be needed along with a corresponding relative increase in capacitance.

Another scenario might be that the power envelope is higher than the 130W assumed here. It may be desirable to increase the power to improve performance or add functionality. The introduction of the hi-k gate has dramatically reduced leakage however, as silicon geometries continue to scale, the leakage current may go up which could cause the power to increase and/or the noise margins to decrease. Both of these scenarios could require lower inductance, higher capacitance.

One additional consideration is the size of the cores. Above it was assumed that the total core area was $1cm^2$ and each individual core was $1mm^2$; however the total core area could be considerably less or the number of cores may go well above 100. Again higher power densities would require lower impedance.

MLCC's and Multi-Core

Over the last approximately 8 years we have gone from using 1206 to 0805 to 0603 capacitors on package. The primary driver has been to get to lower inductance, that is lower inductance per capacitor with more capacitors per unit area, while still trying to manage the cost. The industry has largely moved to 0402, and the 0201 volume is growing rapidly. It has been suggested by some that the right hand turn, i.e. putting a cap on the power as shown in figure-1, means we won't need lower inductance or more capacitance. However, even with the right hand turn, lower impedance will still be needed. As discussed above, even without the added complexity of many-core processors, the impedance trends are still seen to be decreasing. This is due to the continued trend to lower voltages[4]. If the power remains constant but the voltage decreases, then the current increases while at the same time our noise margin decreases, both of which requires a lower impedance; so both industry and impedance trends suggest that the move to smaller form factors will continue.

With the introduction of multi-core processors it would seem that the trend to lower impedance can only continue and may accelerate. How low a impedance will be needed will greatly depend on how the cores are decoupled; i.e. whether or not decoupling individual cores becomes a requirement. It was argued that in order to decouple an individual core, in a 100 core processor, that layout considerations alone could limit us to 0201's or smaller. Even for the simplest case where the power is assumed to be uniformly distributed amongst the cores it is questionable if 0201's, even reverse geometry, will provide a low enough inductance. If the power density for some cores is higher than average, whether it is due to non-uniform power distribution across the cores, smaller cores, or higher overall power, the inductance requirements could be 2, 4 or more times lower than for the case of uniform power distribution. Does one then consider 01005's or even smaller? It may prove very challenging to get a low enough inductance using discrete MLCC's for decoupling.

In addition to the question of can we get a low enough inductance, is can we get a high enough capacitance? Using a smaller form factor capacitor may lead to lower inductance but the capacitance per unit area, on the package, will decrease. For example in going from a 1206 to a 0603, the footprint area is reduced by about 4x however, because of layout design rules, one cannot place 4 times as many. Furthermore the effective volume of a 0603 is only about 1/10th that of the 1206; much of this is due to the smaller volume but also the cover layer thickness and stacking margins make up a greater percentage of the total volume of the 0603 than for the 1206. Both of these factors will get worse with smaller body sizes which means that there will be a need for higher capacitance density, i.e. more capacitance per capacitor but also in a smaller form factor.

The MLCC industry has historically increased the capacitance density by about 34% per year, mainly due to thinner metal and dielectric layers and by increasing the number of layers in a given body size. This has led to the capacitance having a roughly $1/d^2$ dependence, where d is the dielectric layer thickness, such that reducing the dielectric layer thickness by a factor of 2 increased the capacitance by a factor of 4. This assumes that the dielectric constant remains constant, however the thickness of the dielectric layer has gotten thin enough that the electric field seen across the dielectric layers has become large enough that the non-linearity of the dielectric material has become a factor[4,6].

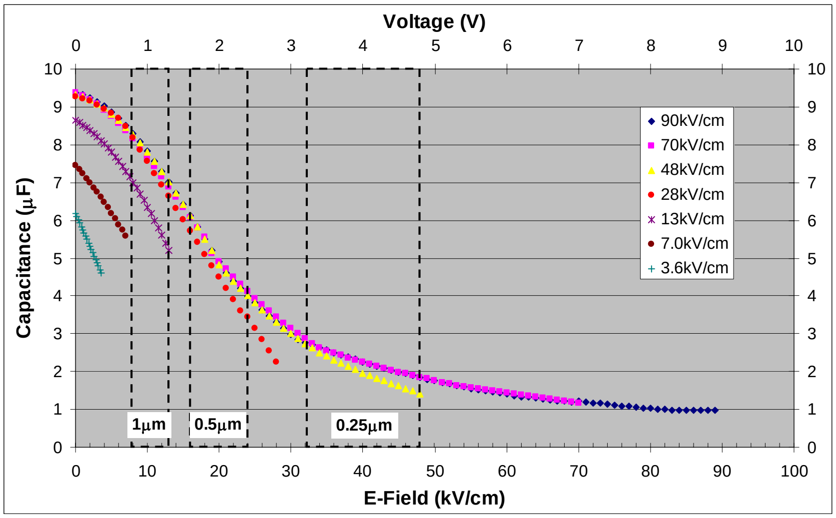

Figure-7 shows the capacitance as a function of electric field for a typical MLCC; the capacitor was nominally 10 F and the dielectric thickness was about 1 m; the measurements were made using a Sawyer-Tower circuit[6,7]. Each of the data sets shown is for a different peak electric field which is described in the legend. The dotted line boxes, labeled 1 m, 0.5 m and 0.25 m, represent the range of expected electric fields across the dielectric layers used in microprocessor decoupling for that particular dielectric layer thickness. Of interest for this discussion is the capacitance value at the peak field for each set of data. As can be seen from the figure, as the MLCC industry moves to thinner dielectric layers, the electric field across the layers will increase. With this increase in the field the capacitance value drops significantly, so that when one takes into consideration the actual operating conditions, the capacitance under high fields may be as much as 80% or more lower than what would be measured under industry standard conditions. In addition as we go to thinner dielectric layers the grain size will decrease. As the grain size decreases the dielectric material becomes more linear and loses its high dielectric constant[8]. Taking both of these factors into consideration raises the question “will we reach a point where the decrease in capacitance due to the loss in dielectric constant becomes comparable to, or perhaps even greater than, the increase in capacitance associated with the reduced dielectric layer thickness and increased number of layers. Whether this happens or not, it is clear that it is going to become increasing more difficult to try and maintain the historical rate of increase in capacitance using conventional MLCC technology. It is therefore questionable if we will be able to get enough total capacitance if we have to go to very small form factors.

Summary

The goal of this paper was not to try and define the power delivery requirements for many-core processors, although the temptation to try and put some numbers with some of the scenarios proved too great to completely resist. Rather the intent was to hopefully stimulate thinking and discussions on what unique power delivery challenges might arise with many-core processors.

A number of scenarios were looked at; scenarios that are entirely plausible although it remains to be seen if they are practical or not. The scenarios necessarily covered a wide range of possibilities, which is in keeping with the uncertainties associated with trying to look at this relatively new area of multi-core power delivery, and were centered on the base assumption of a 130W, 100 core, processor. They ranged from having all the cores on the same power rail to having each core on its own rail and while this later case is probably not likely, it was argued that having to decouple one core on its own voltage rail could be as complex as decoupling 100 cores each of which are on their own voltage rail.

It is anticipated that the need for lower impedance will continue even in the simplest scenario due to the trend toward lower voltage and lower noise margins and projecting out to the timeframe when 100 core processors may appear suggests that the impedance requirements may be in the 0.35m to 0.4m range. In considering a core on its own voltage rail it was found that the capacitor body size will probably need to be no larger than a 0201. For a uniform power distribution, it is questionable whether a reverse geometry 0201 can provide a low enough inductance. When one adds in the possibility of higher power densities from non-uniform power distribution, larger power envelopes, smaller cores or more cores, it is difficult to see how standard MLCC's will be able to provide a low enough inductance. Finally, with the trend to smaller form factors, it is unlikely that we will be able to get enough capacitance given the loss in dielectric constant due to high electric fields and smaller grain sizes. It is much too early to be able to have a clear idea of what the decoupling requirements will be for these many-core processors, but it is not too early to start looking at these what-if scenarios and trying to determine at what point we will need a new solution. One thing is clear however, the future of decoupling will provide many new and interesting challenges and opportunities.

References

1. P. Gelsinger; ISSCC 2001.
2. Justin Rattner, blogs.zdnet.com; February 12th, 2007.
3. Intel Corporation's Polaris Teraflop Research Chip demonstration at the International Solid-State Circuits Conference in San Francisco, Feb. 2007.
4. L. E. Mosley, CARTS USA 2006, pp 193-203.
5. Reference 4 contains a discussion on the various droops; first, second, etc.
6. L. E. Mosley and J. S. Schrader, CARTS USA 2007, pp 309-319.
7. L. E. Mosley, US-Japan
8. T. Tsurumi, Y. Mizuno, H. Kishi and H. Chazono, 3rd International Conference CICMT, 2007.