12th International Conference on Domain Decomposition Methods
Editors: Tony Chan, Takashi Kako, Hideo Kawarada, Olivier Pironneau, ©2001 DDM.org

29. Solving non-linear electronic packaging problems on parallel computers using domain decomposition

P.Chow[1], C.Bailey[2], C.Addison[1]

Introduction

Miniaturisation of electronic equipment, such as those found in a notebook computer, palm held devices, cell phones etc., requires high-density packing of electronic components onto printed circuit boards (PCB). To join the interconnections, solder materials are used to bond microprocessor chips and board during assembly. In the Reflow process case, the board assembly passes through a furnace where the solder bump initially in the form of solder paste, melts, reflows, and then solidifies to bond the interconnections. A number of defects may occur during and after this process such as, respectively, bridging of the liquid solder and cracking of solder joint, chip or board. With the increasing drive towards miniaturisation and smaller pitch sizes (gap between interconnection of solder bumps), these are serious issues to industry in manufacturing and component reliability in operation.

Finite Element Analysis (FEA) is used extensively in the electronic packaging community to calculate stress of solders and components, for reliability analyses [Lau93] [SYS97]. Computer simulations, together with some experiments, provides an effective design and optimization route to reducing these defects and in assessing solder and board integrity and reliability. For models to fully characterise the physical phenomena of the process that govern the integrity of the final joints requires the representing physics of:

- Heat transports with solder solidification involving latent heat evolutions
- Residual stress evolution involving thermal miss-match between materials.

Also, an integrated solution procedure is needed to solve governing equations of temperature, evolving solder shape, solidification, and stress, as they are interdependent. For example, stress analysis is dependent on temperature changes in solid regions. While for the solder joint formation, the solder material will initially, after heating, be liquid and when the board exits the furnace it starts to solidify and stress developments begin.

A microprocessor chip commonly has large number of interconnects that bonds to a circuit board. The general modelling practice is to take a Macro-Micro approach that simulates a single interconnect or assumes each interconnect behaves like a beam in the finite element analysis. In the Macro-Micro case, there is a data transfer between the models at each time step, see Figure 1. A detailed 3D model requires a sizeable mesh and long computing time, i.e. solving non-linear equations of thermal and mechanical systems; thereby, constraining the number of the number of cycles possible for the

[1] Fujitsu European Centre for Information Technology Ltd, UK, {chow, addison}@fecit.co.uk
[2] University of Greenwich, UK, c.bailey@gre.ac.uk

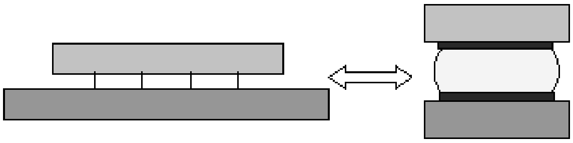

Figure 1: Solder Modelled as Beams (Macro) and Continuum (Micro)

design and optimization process. In the multiple chips case, it can easily leads to models with mesh sizes having millions of elements.

Parallel computing technology opens up the possibility of undertaking such detailed and large-scale analyses, and delivers the solution in a practical timeframe. In application areas such as automobile and aerospace, parallel computing has significantly reduced the time for analyses and increased the size of models (both of physical models and mesh sizes) that can be performed. Such success is also due to the advances in the Domain Decomposition method, now a key element in the majority of parallel models such as mesh or domain partitioning, linear and non-linear solver strategies, and matching and non-matching overlapping grids. Here, we show some parallel computations of 3D electronic packaging models that involves cooling, solidification, and residual stresses of solder joints and throughout the component during assembly. All the computations are performed on a Fujitsu AP3000 system using up to 12 processing elements, with the largest model completed having over 1 million elements.

Heat Transport Equations

The equations governing the physics of heat transport and solidification can be expressed as:

$$\rho c \frac{\partial T}{\partial t} + \nabla \cdot (\rho c \underline{v} T) = \nabla \cdot (k \nabla T) + S$$

where T, t, ρ, c, k, $\underline{v}$ and S are the temperature, time, density, specific heat, thermal conductivity, velocity vector, and source term, respectively. The equation for evolution of latent heat during solidification is represented by the source term, and expressed by:

$$S = -L\rho \frac{\partial f}{\partial t} - L\rho \nabla(\underline{v} f)$$

where L and f are the latent heat and liquid-fraction, respectively. The relationship between the liquid-fraction and temperature describes how the material (here it is the solder) solidifies between the liquidus and solidus temperatures range. For isothermal materials the latent heat release is instantaneous; this means liquidus and solidus temperatures are the same and translates to a vertical jump in the curve between liquid-fraction and temperature. Such numerical discontinuity needs to be addressed properly to maintain energy conservation, if not, it is possible to artificially gain or lose energies in the system. To fully conserve energy, the Enthalpy Source-Based method [VS91] is used to address such discontinuity.

Stress-Strain Equations

For stress analysis, the incremental equilibrium equations governing solid deformation are [ZT89] [TBC95]:

$$\Delta\sigma_{ij,j} = 0 \qquad (i, j = x, y, z)$$

where $\Delta\sigma_{ij,j}$ are the Cartesian components of the Cauchy stress tensor. The incremental stress $\Delta\underline{\sigma}$, $(\Delta\sigma_{xx}\ \ \Delta\sigma_{yy}\ \ \Delta\sigma_{zz}\ \ \Delta\sigma_{xy}\ \ \Delta\sigma_{xz}\ \ \Delta\sigma_{yz})$ is due to the elastic strain given by:

$$\Delta\underline{\sigma} = [D]\Delta\underline{e}^{el}$$

where $\Delta\underline{e}^{el}$ and $[D]$ are the elastic strain vector and elastic materials matrix respectively. The elastic strains are dependent on the total $\Delta\underline{e}$, thermal $\Delta\underline{e}^{th}$, and visco-plastic strain $\Delta\underline{e}^{vp}$ vectors given by:

$$\Delta\underline{e}^{el} = \Delta\underline{e} - \Delta\underline{e}^{th} - \Delta\underline{e}^{vp}$$

For small strains, the total strain, $\Delta\underline{e}$, is given by the gradient in displacements, which in matrix form is:

$$\Delta\underline{e} = [L]\Delta\underline{d}$$

where $[L]$ is the matrix of differentials and $\Delta\underline{d}$ is the displacement vector. The visco-plastic strains in this analysis are represented by the Perzyna [Per66] constitutive model give by:

$$\underline{\dot{e}}^{vp} = \frac{\partial e^{vp}}{\partial t} = \frac{2\lambda}{3\sigma^{eq}}\left(\frac{\sigma^{eq}}{\sigma^{y}} - 1\right)^{n} S_{ij}$$

where λ, σ^{eq}, σ^{y}, n and S_{ij} are the fluidity, von-mises stress, yield stress, strain rate sensitivity, and deviatoric stress, respectively. Within a time increment, Δt, the incremental visco-plastic strain is:

$$\Delta\underline{e}^{vp} = \Delta t\underline{\dot{e}}^{vp}$$

An Integrated Procedure

The solution procedure for the coupling of solidification and stress, plus others such as thermal convection (not included in the solder for the present study) is given in reference [BCF^{+}96], and is in the PHYSICA toolkit. Figure 2 shows the coupled solution procedure for transient analysis of temperature, solidification, and stress.

Within the time step loop the thermal variables, temperature and liquid-fraction, are first solved and the temperature changes, ΔT, calculated. To account for latent heat evolution during solidification, and other non-linearity, an iterative procedure is generally used. Next, the resulting changes of temperature and liquid-fraction are used in the stress calculations.

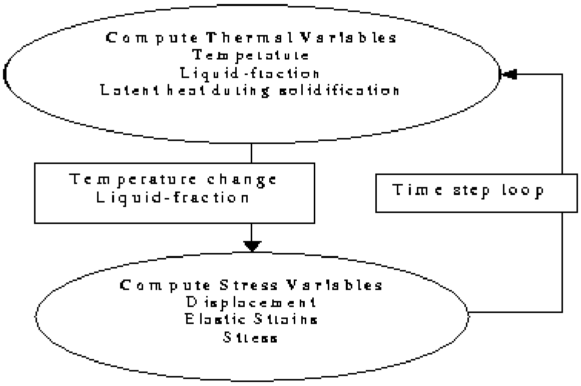

Figure 2: Coupled solution procedure in PHYSICA

Based on temperature changes the incremental displacements are calculated. Using the new displacements and current total stress, σ^o, the incremental total and viscoplastic strains can be calculated. The incremental elastic strain and stress can then be obtained for the time step. This incremental stress will update the total stress ($\sigma^n = \sigma^o + \Delta\sigma$) that will change the values for viscoplastic strain. Due to non-linearity and coupling, an iterative procedure is commonly used.

After the thermal and stress variables have been solved within the time step, the values for these become the old values for calculations at the next time step. As cooling progresses the liquid solder region solidifies and the resulting "solid" elements becoming eligible for stress calculations. The solution procedure continues until the simulation finishes.

Software & Parallel Model

The PHYSICA toolkit [Phy] [CCB$^+$96] from University of Greenwich is used for the study. It has an open single-code component-based software framework [CBM$^+$99] for coupled and Multiphysics applications. The code is 3D, unstructured mesh, with analysis models for fluid flow, heat transfer, solidification, elastic/visco plastic, combustion and radiosity. PHYSICA's parallel model (see refs [CBM$^+$99] [MCJ97]) is based on the Single Program Multiple Data (SPMD) paradigm, where each processing element runs the same program on a sub-portion of the model domain. The mesh, representing the model domain, is partitioned using the graph-partitioning tool JOSTLE [Jos] into sub-domains that are minimized for data exchanges between the overlapped region. Message passing is then used to perform any data exchange needed between these sub-domains on each processing element (PE).

For parallel codes to scale well for performance, the non-scalable portions needs to be eliminated - if not possible, it will be a point of concern in the solution procedure's critical path. Some common examples of non-scalable parts are, reading and writing to files (parallel input and output are currently system dependent, if available), and global summation operatives commonly found in popular linear solvers. In the version

Table 1: Parallel performance for electronic package case

CPU time in minutes			
PE	Solution time	Total time	Speedup
1	17.40	17.92	
4	6.00	6.75	2.65
8	4.87	5.62	3.19

of PHYSICA used for this work, the embedded JOSTLE is scalar and it is a critical point in the overall scalability. Also, by default, the whole mesh is first read into memory for JOSTLE to perform the partitioning before distribute to the PEs for processing. For larger model sizes this non-stop processing can be inappropriate due to memory demands by JOSTLE and PHYSICA - together, the optimal performance configuration of the computing system can degrade significantly. In this work, the mesh partitioning and analysis are executed separately for all the large models; first partitioning the mesh and save the PE index for each mesh element to file, then the analysis phase reads the save index data and distributes the mesh element to each PE for processing. For multiple run cases with the same number of PEs, this ”partition and save” approach may, in some instances, be more advantages than the non-stop approach; since for any amount of multiple runs the partitioning only occurs once. A parallel version of JOSTLE is underway to address the non-stop processing and other related matters.

Parallel Results

Table 1 shows the computing times for a model solve in electronic packing (consisting of 21,413 vertices, 57,577 faces and 18,150 elements) in CPU minutes for 1, 4 and 8 PEs. For this model the total time for 8 PEs is a speedup of 3.19 over a single PE, the solution speedup (without initial setup, such as mesh partitioning, and reading and writing to files) is 3.57. This means the non-solution portion takes about 11 and 13 percents of the total CPU time, respectively for 4 and 8 PEs, compared to 0.03 percent for single PE.

Figure 3 shows one quarter of a chip bonding to a PCB example being modelled during the reflow process, and Figure 4 showing an enlarged view of the solder bumps with two different attachment materials at top and bottom. The model consists of 273,504 vertices, 1,133,207 faces and 425,890 elements. Figure 5 shows the solidification fronts of the solder bumps during cooling phase of the reflow process. The corner solder bump is solidifying at a rate faster than its neighbours as indicated by the solidification front in dark colours. Figure 6 shows the magnitude of visco-plastic strain and deformation throughout the solder bumps at the end of reflow when all the solder bumps are solid. Again the corner solder bump has a higher amount of strain than all the other solid bumps. The deformation, as shown by the inclining solder bump, is board contracting more than the chip because of different thermal coefficients in the material properties.

Table 2 gives the computing times from 2 up to 12 PEs in CPU hours. The model

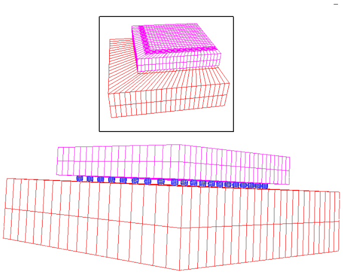

Figure 3: Chip bonding to PCB

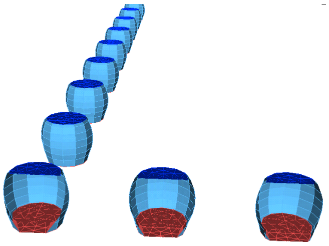

Figure 4: Solder bumps

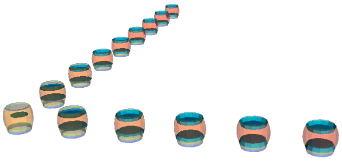

Figure 5: Solidifying solder bumps

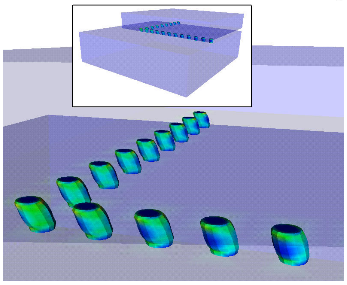

Figure 6: Solder bumps deformation

is too big for single PE on the AP3000, as it reports out of memory. The CPU runtime for 2 PEs is under 7 hours and 12 PEs is under 2 hours. This gives a speedup factor of about 8 for 12 PEs, representing a saving of about 5 hours in analysis time or an extra 1 to 2 cycles in the design and optimization process. For lower PE runs the speedup factor moves nearer to the linear scaling mark.

To get an idea of a single PE runtime, the same model was run on a Sun Enterprise 10000 (E10000) with 2GB memory in scalar mode. With UltraSPARC processors inside the AP3000 and E10000 systems, U170 and U250 respectively, a total CPU time of 15.34 hours was reported on the E10000 with solution time being 15.24 hours. If we put the E10000 result with the 12 PEs of AP3000, it represents a saving of over 13 hours in analysis time or giving an extra 5 to 6 cycles in the design & optimization process. In terms of speedups, it represents a factor of 9 (compared to 8) for analysis time and 11 (compared to 9) for the solution period. Figures 7 and 8, respectively, show graphs of parallel performance for total and solution times; the triangle markers indicates an idea of the true speedup if the 1 PE time had been possible. These estimates are obtained by substituting the E10000 single PE result in the calculation for speedups.

From the performance graphs, it is encouraging to see the curve for total time shows there are potential gains for this model case by adding more PEs (12 PEs is the highest we have access to at present). A downside is the non-solution portion of the analysis time is also increasing with PEs, some 19 percent (or 20 minutes) for the 12 PEs case. To see how larger models may fair, a similar problem with model size of

Table 2: Parallel performance for a chip bonding to a PCB example

CPU time in hours			
PE	Solution time	Total time	Speedup
2	6.281	6.748	
3	4.233	4.621	2.921
4	3.261	3.648	3.699
5	2.703	3.070	4.397
6	2.350	2.701	4.997
7	2.082	2.488	5.425
8	1.816	2.153	6.268
9	1.684	2.023	6.671
10	1.530	1.848	7.305
11	1.460	1.762	7.658
12	1.379	1.698	7.951

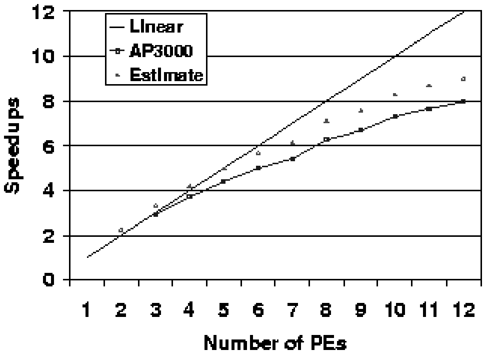

Figure 7: Total time performance

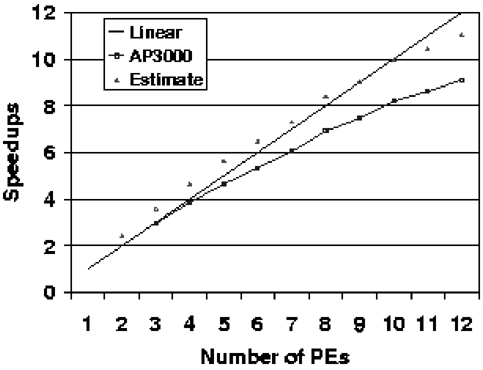

Figure 8: Solution time performance

1,205,997 vertices, 3,504,048 faces and 1,149,312 elements was conducted on 12 PEs. The parallel performance reports an analysis time of 6.01 hours with a solution period of 4.94 hours, representing some 18 percent or 1 hour for non-solution activities. This is encouraging, as the percentage figure has not altered significantly.

Conclusion

The above results indicate significant reduction in analysis time for electronic packaging applications, even for a model mesh size of 18,000 elements. As for larger models with elements in the millions, such as multiple chips on board cases, it can exceed the memory capacity of today's workstations. Parallel computing with domain decomposition offers a solution to run and deliver the analysis within a design and development timeframe.

The numerical experiments conducted indicate some 20 percent of the analysis time on 12 PEs are in non-solution activities, such as data retrieving and saving to files and setup period for parallel computation. Therefore, there is great potential in reducing this figure even further and improving parallel performance by removing the present scalar events in the procedure's critical path such as having parallel IO and parallel mesh partitioning. Memory usage is lot higher in mechanical analysis section than in the thermal section; a ratio of about 2 to 1 has been observed - this is primarily due to the segregated method used in thermal analysis section as to the full-system employed in the mechanical.

References

[BCF$^+$96]C. Bailey, P. Chow, Y. Fryer, M. Cross, and K. Pericleous. Multiphysics modelling of the metals casting processes. *Proceedings of Royal Society of London A*, 452:459–486, 1996.

[CBM$^+$99]P. Chow, C. Bailey, K. McManus, C. Addison, and M. Cross. A single-code model for multiphysics analysis-engine in parallel and distributed computers with the physica toolkit. In Choi-Hong Lai, Petter E. Bjørstad, Mark Cross, and Olof B. Widlund, editors, *Domain Decomposition Methods in Sciences and Engineering*, pages 396–402, Bergen, 1999. Domain Decomposition Press.

[CCB$^+$96]M. Cross, P. Chow, C. Bailey, N. Croft, J. Ewer, P. Leggett, K. McManus, K.A. Pericleous, and M.K. Patel. PHYSICA – a software environment for the modelling of multiphysics phenomena. *ZAMM*, 76:101–104, 1996.

[Jos]JOSTLE. University of Greenwich, London, UK (www.gre.ac.uk/˜wc06/jostle).

[Lau93]J. Lau. *Thermal stress and strain in micro-electronic packaging*. Van Nostrand Reinhold, New York, 1993.

[MCJ97]K. McManus, M. Cross, and S. Johnson. Issues and strategies in the parallelisation of unstructured multiphysics codes. In *Parallel and distributed computing for computational mechanics*, 1997.

[Per66]P. Perzyna. Fundamental problems in viscoplasticity. *Advanced Applied Mechanics*, 9:243–377, 1966.

[Phy]PHYSICA. University of Greenwich, London, UK (physica.gre.ac.uk).

[SYS97]H. Sakai, N. Yasuda, and K. Seyama. Reliability study of flip chip organic bga/csp packages. In *Semicon West 97, BGA Technology Conference*, 1997.

[TBC95]G. Taylor, C. Bailey, and M. Cross. Solution of elastic/visco-plastic constitutive equations: a finite volume approach. *Apply Mathematical Modelling*, 19:746–760, 1995.

[VS91]V. Voller and C. Swaminathan. General source-based methods for solidification phase change. *Numerical Heat Transfer*, 19:175–190, 1991.

[ZT89]O.C. Zienkiewicz and R. Taylor. *The finite element method*. McGraw-Hill, 1989.