

Principles of Simulating Contact Between Parts using ANSYS

David H. Johnson, P.E.

Penn State-Erie

Erie, Pennsylvania, USA

Abstract

The ANSYS finite element analysis (FEA) program offers a variety of elements designed to treat cases of changing mechanical contact between the parts of an assembly or between different faces of a single part. These elements range from simple, limited idealizations to complex and sophisticated, general purpose algorithms. FEA analysts are frequently faced with modeling situations where changing contact cannot be assumed negligible and ignored. Finding the best choices for contact elements, element options, solver, and solution options can drastically improve the model's performance and reduce the analyst's frustration with a contact simulation model.

Contact elements can be grouped into four general categories based on increasing levels of sophistication or complexity:

- Point-to-point gap elements: CONTACT12, LINK10, COMBIN40, CONTAC52, CONTA178
- Point-to-line (or slide-line) contact elements: CONTAC26
- Point-to-surface contact elements: CONTAC48, CONTAC49
- Surface-to-surface contact elements: TARGE169, TARGE170, CONTA171, CONTA172, CONTA173, CONTA174

The order of these categories also reflects the history of contact element development over the past thirty years: from the initial cases developed for treatment of contact based on node point separation, to the most recent, general behavior elements which consider lower-order or higher-order, quadrilateral or triangular surfaces with contact checking occurring at the face integration points.

The international email discussion group, XANSYS, has seen many questions and answers regarding contact element usage over the past few years. New users have sought assistance and veteran analysts have offered advice and experience. This paper will review contact element usage, options, and capabilities, then, from a collection of XANSYS contact discussions and from personal experience, a summary of guidelines, tips, hints and tricks for contact element simulations will be presented.

Some of this information (Part 1) was originally presented at the Mallett Technology Inc., ANSYS User's Group Conference at the University of Maryland, College Park, MD on Oct. 2, 2001. A lively discussion on the contact element penalty formulation from that meeting, as well as additional figures and examples, are included in this new paper presentation (i.e., Part 2).

Introduction

Using contact elements in an FEA simulation is seldom a simple, painless experience. Changing contact between parts is a common phenomenon which, in some cases, can be treated with rigorous mathematical theory. Formulae for special cases can be found in machine design textbooks ^[1, 2, 3, 4] e.g., two spheres, two parallel cylinders, cylinders on a flat plate, gear teeth, roller bearings, etc. However, these theories only describe the stress in the contact region. Figure 1 shows an example of such a case.

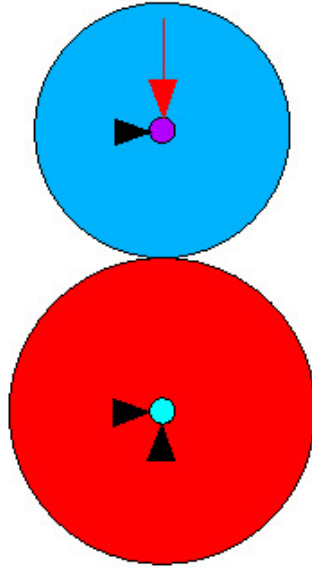


Figure 1 - Hertzian Contact Between Cylinders

Often, an analyst is interested in stresses away from the contact zone, and the contact behavior must be modeled properly to find the important stresses in the system. In those cases, the analyst would not be interested in the Hertzian or contact stress. Proper computation of forces and deformations in the contact zone is critical to determining the stress results throughout the model. As an example, consider a screw thread (or a gear tooth) in Figure 2. The maximum stress is expected at the root of the thread (or gear tooth) and not under the contact region. Some texts ^[5] briefly discuss the capabilities and procedures of using finite elements to model changing contact situations, but do not give the details needed for creating and solving such a model.

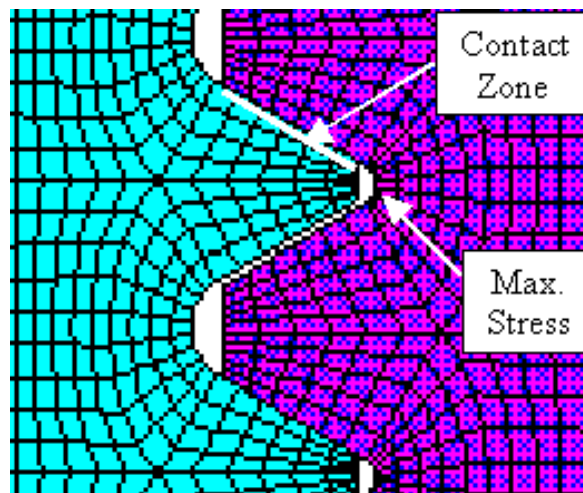


Figure 2 - Contact Between Threads

As an analyst approaches a problem that must include some form of contact behavior, numerous decisions must be made. To start with, non-contact approximations should be considered. In many cases, it may be possible to use coupled DOF's or to write constraint

equations (multi-point constraints) to capture some of the behaviors we would be inclined to treat with contact elements. The advantage of this approach is that the contact problem can be treated as a linear condition. The disadvantage is that the analyst must know in advance what the contact behavior will be. When that detail is unknown, a nonlinear solution with contact elements must be performed.

Selecting a Contact Element

Contact element choices include:

- 1) Point-to-point gap elements: CONTACT12, CONTAC52, CONTA178, COMBIN7, LINK10, COMBIN40

Point-to-point gap elements can be used in cases where little or no sliding occurs, and they should only be attached to the corner nodes of lower-order solid elements, i.e., they are not recommended with higher-order solids. Usually, it is desirable to control the mesh to produce matching node patterns on both solid faces to permit direct, node point to node point contact element definition. All of the nodes on one contact face of a solid element mesh would be connected by such a gap element to a corresponding node on the solid element mesh on the opposite contact face.

In operation, these elements are fairly simple: the deflection of the two nodes of a contact element are checked to see if they have separated or come closer together.

During model construction, the analyst supplies data such as:

- stiffness of a closed gap
- friction coefficient (if any)
- stiffness in the sliding direction
- the initial clearance or interference dimension and the initial open or closed status condition

COMBIN7, COMBIN40 and LINK10 elements are included in this list of point-to-point gap elements because they simply connect from one surface node to another, but these elements can have other complicated behaviors. COMBIN7 is a revolute joint element which has “stops” that act like the closing of a contact. COMBIN40 has gap and slider behavior, but also allows two parallel springs, and may have mass and damping for a dynamic simulation. LINK10, the tension-only (or compression-only) spar, can act like a rope or cable which carries load in tension but not in compression.

- 2) Point-to-line (or slide-line) contact elements: CONTAC26

The point-to-line gap element is available for two-dimensional problems only. The node point location on the flexible surface (that is, on the surface of a solid element mesh) is tested with respect to its proximity to some rigid target surface. The analyst defines:

- stiffness of a closed gap
- the geometry (shape and location) of the rigid surface
- friction coefficient (if any)
- stiffness in the sliding direction

- 3) Point-to-surface contact elements: CONTAC48, CONTAC49

The point-to-surface gap elements can be used in two- and three-dimensional systems. Large sliding behavior can be modeled, and dissimilar meshes on the solid element faces is handled very well. However, these elements are not compatible with higher-order solid elements

because the mid-side nodes of the quadratic elements are not used for the contact faces. With these contact elements, the analyst defines or selects:

- Stiffness of a closed gap (FKN)
- Formulation:
 - Penalty method
 - Penalty + LaGrange multipliers
- Allowable penetration tolerance (FTOLN)
- Pinball region size (PINB)
- Friction/sliding behavior

Most of these parameters will be addressed in the discussion that follows.

4) Surface-to-surface contact elements: TARGE169, TARGE170, CONTA171, CONTA172, CONTA173, CONTA174

The surface-to-surface gap elements overcome most of the limitations or restrictions of the other contact elements:

- Two- or three-dimensional systems
- Rigid or flexible target faces can be treated
- Large sliding is permitted
- Lower- and higher-order solid element faces are compatible with these contact elements
- The higher-order elements simulate curved surfaces without representing them as faceted, flat faces, i.e., the actual curvature is included.

Beyond the capabilities listed above, these elements have additional options, such as:

- Velcro (bonded, or “sticky”) behavior
- Automated (program computed) stiffness assignment for a closed gap
- Testing for contact at either the node points or at the integration points inside of the contact faces
- Ramped-on (gradual) treatment of initial penetration
- Consideration of shell and beam element thickness effects

Although these elements often have “default “ values for the contact parameters, the analyst can still specify:

- Stiffness of a closed gap (FKN)
- Formulation:
 - Penalty method
 - Penalty + LaGrange multipliers
- Allowable penetration tolerance (FTOLN)
- Pinball region size (PINB)
- Friction/sliding behavior
- Initial Closure (ICONT)
- Contact detection at the nodes or at the integration point within the element

Discussion

With regard to the possibilities listed above, we will focus on the more recent, more complex contact element characteristics and usage. Various behavior control parameters that were mentioned above will be defined and discussed for the CONTACT48/49 and for the CONTACT17x family of elements.

Pinball Region

During a contact element analysis, ANSYS checks each contact element to determine if it is in-contact, near-contact, or not near contact. A circular or spherical region around each contact (or target) element is used to make this determination.^[6,7] ANSYS has default values for the size of this pinball region, but the user can override the default to set specific levels for this parameter. More computation time is spent on contact elements in-contact or near-contact than on the elements determined far-from-contact based on the pinball size. Figure 3 gives an illustration where parts are “not near contact” based on the pinball region size.

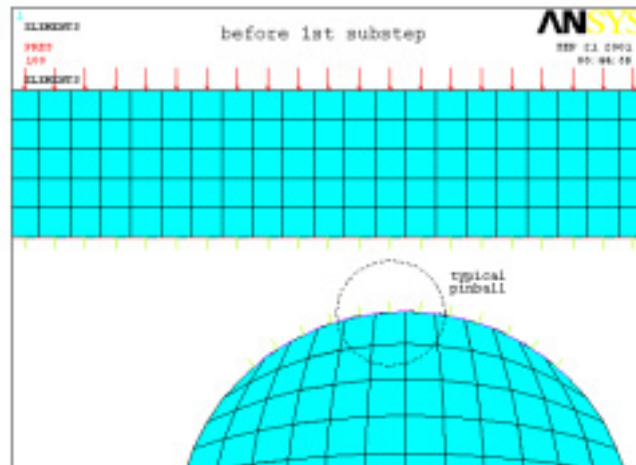


Figure 3 - Illustration of the Pinball Region

Static analysis solution failure can occur when no contact is detected between parts making the model unstable (appearing as insufficient constraint). Sometimes the contact elements do not engage because the faces are too far apart, i.e., outside of the pinball. This can also happen when too much penetration occurs and the contact element passes beyond the pinball zone. In both cases, XANSYS recommendations have included:

- Gradual loading over many substeps
- Applying the loads with imposed displacements (more stable) instead of imposed forces or pressure
- Use imposed displacements to “seat” the contact, then replace with the force (or pressure) load
- Increasing the radius of the pinball region can prevent these problems, but can slow the solution because more contact elements will appear to be nearly in contact

FKN Parameter: Stiffness of a CLOSED Gap

“All contact problems require a stiffness between the two contact surfaces. The amount of penetration between the two surfaces depends on this stiffness. Higher stiffness values decrease the amount of penetration but can lead to ill-conditioning of the global stiffness matrix and to convergence difficulties. Ideally, you want a high enough stiffness that contact penetration is acceptably small, but a low enough stiffness that the problem will be well-behaved in terms of convergence or matrix ill-conditioning.”^[9]

For the surface-to-surface contact elements, TARGE16x/CONTA17x, the element manual recommends “use the real constant FKN to define a normal contact stiffness factor. The usual factor range is from .001-100, with a default of 1.0. A smaller value provides for easier convergence but with more penetration. The default value is appropriate for bulk deformation. If bending deformation dominates, we recommend using a smaller value (0.01-0.1)”^[8]

Some XANSYS comments have included:

- For the TARGE16x/CONTA17x, when changing FKN: values between 0.1 and 10 usually suffice. Lower values (e.g., 0.1) are for bending-dominated problems and 1.0 for bulk deformation problems.
- For the TARGE16x/CONTA17x, the default value of FKN may be too small when using plasticity in a static model. The possibility of yielding in the contact zone may create “softer” FKN stiffness. In such cases, the user must override the default to increase FKN.
- With a difficult contact problem, start with a lower value of FKN. This gives excessive penetration, but faster runs. If the problem does not behave correctly, fix the modeling errors. If the problem behaves properly, then restart with higher FKN to get rid of the excessive penetration.
- If the surface of the model is faceted due to coarse meshing, high FKN values may help. This problem shows up as wild, localized displacements and stress concentrations on the contact boundary. There is a point where coarse faceting cannot be compensated and mesh refinement is indicated. Likewise, these localized high stress spots may be minimized by using the higher-order contact and targets that define truly curved surfaces, but the analyst must accept the larger model produced by higher-order solid and contact elements.

LaGrange vs. Penalty Formulation and FTOLN Parameter

“For surface-to-surface contact elements, ANSYS uses either the augmented Lagrangian method (default) or the penalty method. The augmented Lagrangian method is an iterative series of penalty updates to find the exact Lagrange multipliers (i.e., contact tractions). Compared to the penalty method, the augmented Lagrangian method usually leads to better conditioning and is less sensitive to the magnitude of the contact stiffness coefficient. However, in some analyses, the augmented Lagrangian method may require additional iterations, especially if the deformed mesh becomes too distorted. Use real constant FTOLN in conjunction with the augmented Lagrangian method. FTOLN is a factor based on the thickness of the element which specifies an allowable maximum penetration for the augmented Lagrangian method. If ANSYS detects any penetration larger than this tolerance, the global solution is still considered unconverged, even though the residual forces and displacement increments have met convergence criteria. The default for FTOLN is 0.1. You can change this value, but be aware that making the tolerance too small can cause an excessive number of iterations or non-convergence.”^[9]

From an XANSYS discussion:

There are two different ways to approach contact – the penalty method alone or a Lagrangian multiplier added to the penalty method. The penalty method modifies the present stiffness matrix by adding large terms to prevent too much penetration, i.e., a “penalty” is imposed. A higher

penalty stiffness leads to more likely cases of ill-conditioned matrices, and more difficult solution. On the other hand, the Lagrangian multiplier method augments (makes bigger) the stiffness matrix by adding extra DOF terms (in this case, contact pressure) to reconcile the contact. So the "downside" to Lagrangian methods generally is the fact that your stiffness matrix gets bigger due to extra DOF terms, but ill-conditioning is less likely.

From the Mallett Technology Inc, ANSYS User's Group Conference discussions:

Some experienced analysts feel that "better" contact performance and results may be found when the penalty formulation alone is used. An example cited was: a three-dimensional contact analysis, where there are many contacting faces with large motions possible. This model seemed to perform better with the penalty-only option with the TARGE16x/CONTA17x elements.

When using penalty-only contact elements the Lagrangian multipliers are not used; the penetration tolerance, FTOLN, has no effect on the solution; and the proper choice of the closed gap stiffness, FKN, becomes more critical.

Finally, an important postprocessing step involves checking penetration in the contact elements of a converged solution. A contact element face can "pass through" a target element face by a tiny amount. The amount of this "pass through" or penetration is related to the stiffness (FKN) and to the penetration tolerance (FTOLN). The analyst should, as a matter of routine, evaluate the amount of penetration and decide if it is at an acceptable level. Unacceptable levels of penetration are treated by reducing FTOLN and increasing FKN. Excessive penetration is expected to be most noticeable in displacement controlled, shrink fit, or initial interference simulations.

Using a SYMMETRIC Pattern of CONTACT Elements

It is common to define all contact elements on one surface and all target elements on the other surface. This is called an asymmetric contact pattern. However, there are times when this approach leads to an unconverging solution or, at best, to an excessively long solution. This may occur when the model is coarsely meshed, or when the surfaces are not easily described as either a target or a contact face.

ANSYS ^[10] guidelines for defining the contact faces in a model include:

For the CONTAC48/49 elements:

- If the contact region of one surface is flat or concave and the contact region of the other surface is a sharp edge or convex, the flat/concave surface should be the target.
- If both contact regions are flat, the choice is arbitrary.
- If both contact regions are convex (not sharp edges), the target surface should be the flatter of the two.
- If one contact region is a sharp edge and the other is not, the sharp-edge surface should be the contact surface.
- If both contact regions have sharp edges, or if the surfaces are wavy (alternately concave and convex), the choice of target surface depends on the configuration of the surfaces after contact. In such cases, symmetric contact modeling will usually be preferred.

For the TARGE16x/CONTA17x elements:

- If a convex surface is expected to come into contact with a flat or concave surface, the flat/concave surface should be the target surface.
- If one surface has a fine surface mesh and, in comparison, the other has a coarse mesh, the fine mesh should be the contact surface and the coarse mesh should be the target surface.

- If one surface is stiffer than the other, the softer surface should be the contact surface and the stiffer surface should be the target surface.
- If higher-order elements underlie one of the external surfaces and lower-order elements underlie the other surface, the surface with the underlying higher-order elements should be the contact surface and the other surface should be the target.
- If one surface is markedly larger than the other surface, such as in the instance where one surface surrounds the other surface, the larger surface should be the target surface.

When it is not obvious from the element guidelines (above), it may improve the solution convergence to create symmetric contact element patterns. In those cases or in those locations within the model, simply repeat the operation defining the contact pair, but switch your choices of contact and target surfaces as shown in Figure 4.

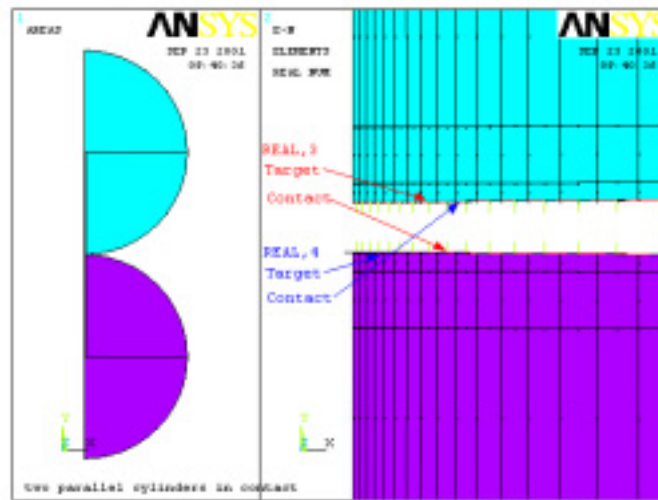


Figure 4 - "Symmetric" Pattern of Contact/Targets

Several XANSYS discussions have indicated that when unusual, incorrect stress distributions are observed under the contact surfaces (and this may be related to coarse meshing) that use of the symmetric contact method can smooth the stress results and give more accurate answers. XANSYS discussions have also pointed out that the contact pressure must be interpreted to account for the symmetric contact pattern. That means, the contact pressure may have two contributions at a specific contact location: from each of the symmetric contact pairs that cover that location.

ICONT Parameter

For the TARGE16x/CONTA17x, we “use the real constant ICONT to specify an initial closure factor. If the initial gap is less than this tolerance (ICONT times the depth of the underlying solid, shell, or beam element), ANSYS considers the model to be in initial contact. If you do not specify ICONT, then ANSYS provides a small value (less than 0.03), depending on the geometric dimension. Specify a very small value if you want a negligible initial contact closure.”^[8]

Discussions on XANSYS have indicated that:

- If ICONT is too high, the contact stress or pressure may be uneven
- In a few cases, the default value of ICONT may pull parts together which the analyst actually wants to keep separated.

An example of parts being pulled together when that is not desirable can be illustrated with two cylinders coming in contact (see Figure 1). The model is constructed using one-half symmetry, Figure 5. The geometry is developed so the initial contact should occur only at one point. A load at the center of the smaller cylinder and constraint at the center of the larger cylinder, push the parts together.

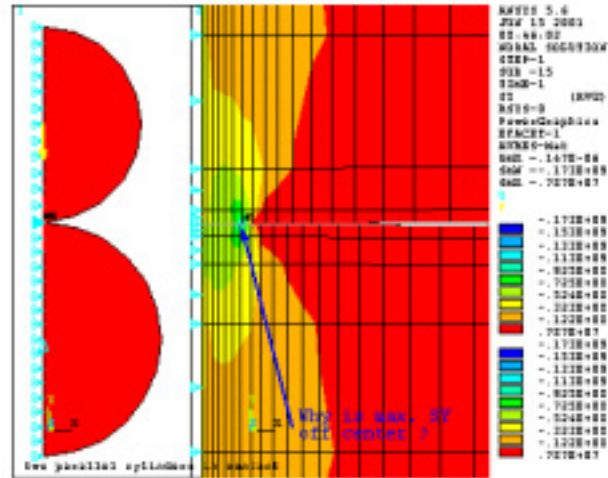


Figure 5 - Example Showing ICONT Problem

On the right image in Figure 5, the problem with contact pressure is presented, i.e., the peak contact pressure does not occur where it should, but is offset from the center of the contact zone. The ICONT setting assumed that the initial contact occurred over several elements, and not just at one point as planned. Geometrically, several contact pairs are “near” contact initially and fall under the threshold of the default ICONT value. Therefore, initial contact is spread over several faces, rather than happening at a single point.

The desired response is shown in Figure 6 where only by changing the ICONT parameter to a very small value ($1e10^{-8}$), the expected results are achieved, the peak contact stress occurs at the center of the contact zone.

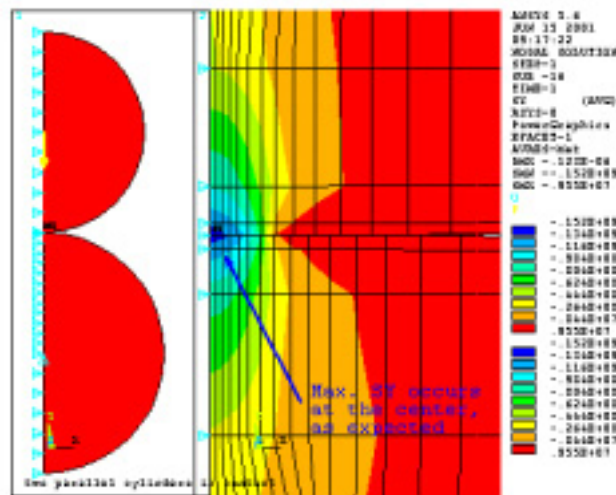


Figure 6 - Resolution of ICONT Problem

“Bonded” Contact

Using the TARGE170/CONTA174 pairs to bond dissimilar meshes. XANSYS members identified the appropriate element options (KEYOPTS) and solution settings as follows:

```
ET,n,CONTA174
KEYOPT,n,2,1
KEYOPT,n,6,0
KEYOPT,n,7,0
KEYOPT,n,8,1
KEYOPT,n,9,1
KEYOPT,n,11,0
KEYOPT,n,12,5 ! Keyopt 12 set to 5 is a bonded always condition.
```

To solve problem as a static linear model:

```
SOLCON,OFF,OFF
NSUBST,1,1,1
NEQIT,1 ! ANSYS will issue a warning when you do this, but it can be ignored as long as
        there are no other nonlinearities, e.g., plasticity or large deformations.
LNSRCH,OFF
NLGEOM,OFF
```

“Bonded” Contact has been found to be a very useful feature for many situations. In fact, ANSYS Inc. has added a section on bonded contact to the “Introduction to ANSYS – Part 2” seminar. This usage of contact elements is preferred over the use of constraint equations to connect dissimilar meshes because the contact elements will function correctly in a large deformation problem whereas the constraint equations are restricted to small deformation cases only.

An educational example of bonded contact in Figure 7 shows a simple 2D model of a cantilever beam. A square area near the fixed end is meshed with small elements to capture the high stress gradients. A second, unattached area is rectangular and represents the rest of the beam, away from the fixed end. These two meshes are connected at their common interface with a bonded-always, contact and target pair.

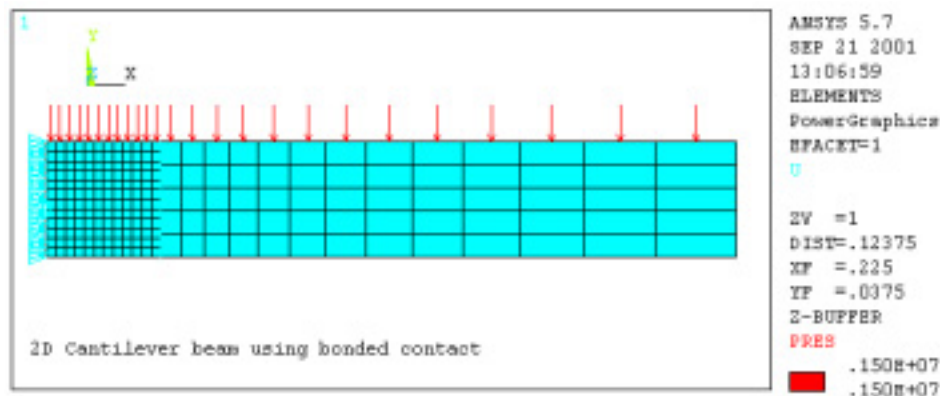


Figure 7 - Bonded Contact for Dissimilar Meshes

The model results presented in Figure 8 show smooth contours of the axial stress across the bonded contact interface. This indicates that bonded contact can be used without creating local discontinuity in the stress field as a constraint equation interface (CEINTF) may produce. The tip displacement of the cantilever beam according to theoretical formula should be 0.003 m. For the bonded contact model, the tip deflection at the neutral axis is 0.00312 m. Likewise, for the same

mesh, connected by CEINTF constraint equations, the tip deflection is 0.00312 m. Finally, for a very similar mesh, but having the areas “glued” together so no additional connection is required, the tip deflection is also 0.00312 m. The minor difference between theory and the three FEA solutions would be because the theoretical formula used does not include the shear deflection term and since the beam’s length is only six times its height, shear deflection may contribute a small, additional amount to the final deflection.

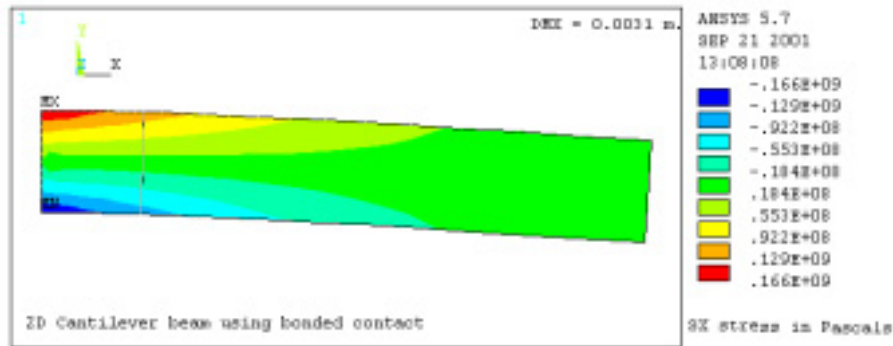


Figure 8 - Axial Stress Result with Bonded Contact

Solution and Debugging

Some common XANSYS tips regarding solution of models with contact include:

- With the surface-to-surface contact elements, large deformation effects should be included in the analysis.
- The sparse solver may be selected by default, but the PCG solver could be used when all element matrices are symmetric.
- Apply the loads gradually, ramped on.

When a solution fails due to large, rigid body motion, with the surface-to-surface contact elements try:

- Using CNOF (recommended) or ICONT (not-recommended) or PMAX, PMIN to close the initial gap
- Using keyopt(9) to ignore initial penetration (which may cause large reaction force)
- Using weak spring FKOP with no-separation option keyopt(12)
- Using small friction to prevent significant sliding

Anytime a solution fails:

- Look at the unconverged results to get a better idea of what's going on. In the postprocessor, use /DSCALE,1,OFF and PLVECT,U to see the displacements (as vectors) on the undeformed geometry. This may give insight as to where the difficulties lie.
- Look into the error file (jobname.err) that was produced during the analysis for hints as to the reason for solution failure.
- Consider turning off all other nonlinear behaviors while you debug the contact response.

- You can always start with a lower value of FKN (normal stiffness), plot deformed shape to determine if the penetration is excessive. If so, increase FKN by an order of magnitude and run again. Too small of a value of FKN may result in too much penetration, too large of a value of FKN may make convergence difficult.

Conclusion

As you undertake any analysis dealing with conditions of changing contact, study ALL the available documentation, including:

- the elements manual information on the contact elements you are considering
- the modeling and meshing manual information on creating the solid element mesh, as well as for creating the actual contact elements
- the theoretical manual information describing the behavior and restrictions or limitations of the contact elements you are considering
- the manual of verification problem examples which illustrate the usage of the elements and features you are planning to utilize.

Remember, it is often *extremely* useful to construct simple models, with known solutions, to learn the techniques and to understand the procedures for modeling situations of changing contact. Trying to learn contact element usage on a large, complex model is NOT an effective use of an analyst's time.

References

- 1 Juvinall, R. C. and K. M. Marshek, Fundamentals of Machine Component Design, second edition, John Wiley and Sons, 1991, pp. 322-329, 527, 581-587.
- 2 Shigley, J. E., Mechanical Engineering Design, second edition, McGraw-Hill, 1972, pp. 93-97.
- 3 Roark, R. J. and W. C. Young, Formulas for Stress and Strain, fifth edition, McGraw-Hill, 1975, pp. 513-522.
- 4 Timoshenko, S. P., and J. N. Goodier, Theory of Elasticity, 3rd ed., McGraw-Hill, 1970, pp. 409-420.
- 5 Adams, V. and A. Askenazi, Building Better Products with Finite Element Analysis, OnWord Press, 1999, pp. 160-167, 427, 447-448.
- 6 ANSYS 5.5 Elements Reference Manual, 10th edition, ANSYS Inc.
- 7 ANSYS 5.5 Theory Reference Manual, 9th edition, ANSYS Inc.
- 8 ANSYS 5.6 Elements Reference Manual, 11th edition, ANSYS Inc, November 1999.
- 9 ANSYS 5.6 Structural Analysis Guide, 5th edition, ANSYS Inc, November 1999.
- 10 ANSYS 5.7 Structural Analysis Guide, ANSYS Inc, 2000.