Engineering Design I: Methods and Skills
Topic Readings

Steven H. Collins

February 23, 2015

Chapter 6

Conceptual Overview of Assemblies

In the previous chapters, we have primarily discussed the design of single continuous components. Of course, most engineered products comprise assemblies of many parts. Here we will discuss some basic concepts for assembly design, such as the reasons for using multiple components and the effects of connecting parts together in different ways.

6.1 Why (not) to use an assembly?

There are many possible motivations for designing a system with multiple parts. For example, we may need relative motion between components, such as that provided by a bearing between a shaft and a support structure. It might also be impractical to manufacture a desired structure as a single part. For instance, the desired geometry might be very complex or there may be disparate size and complexity scales involved, such as in a design with a long simple beam and a small multi-featured element on the end. We might also wish to use different materials in various places, again making manufacture of a single structure impractical. Or, we may wish to be able to disassemble and reassemble the resulting system for modularity, portability, maintenance, or repair.

Designing a structure with multiple separate components also introduces new potential problems to identify and resolve. In general, increasing complexity, such as by adding more components and connections, increases opportunities for something to go wrong. Connection points tend to be weak points and result in less efficient load carriage even when designed well. This is because they tend to introduce stress concentrations, tend to concentrate load on catalog parts that are imperfect for a given application, tend to involve contact loading, and often rely on materials with imperfect properties. For instance, two metal plates connected by bolts would be weaker for the same mass or heavier for the same strength than a single continuous plate if the designs for each scenario were optimized. As another example, welding or gluing these plates together would require use of material with lower specific strength (resulting from melting and cooling in the case of the weld). Putting together assemblies also takes time, which can add to product cost. And connection points tend to be imprecise, leading to potential issues with tolerances across multiple connected parts.

As you design a machine, keep these trade-offs in mind when you choose whether to make something from multiple parts and how to connect those parts together.

6.2 Common Connections

We will discuss various ways of making rigid and articulating joints in a future chapter, but let's list some here without detail to stimulate our creativity. Fixtures are commonly achieved with machine screws, bolts and nuts, setscrews, rivets, retaining rings, pins, keys and keyways, welds, and adhesives. Parts can be designed to engage with these elements and each other via normal forces (which is best wherever possible), via shear forces (generally to be avoided), or by friction forces (which can have advantages over shear in many applications).

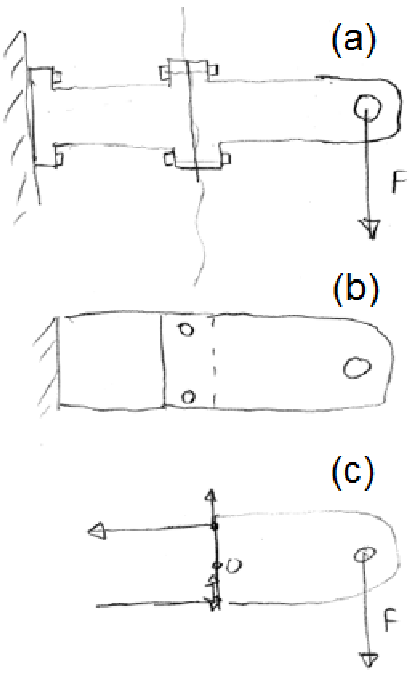
Articulations are commonly created by ball bearings or plain bearings (a.k.a. bushings), in either radial or thrust (axial) configurations. Degrees of freedom can also be introduced using intentional compliance, e.g. springs. A single rotational degree of freedom is often the simplest and most effective way of allowing motion at a connection between parts, and it is usually worth investigating this approach first. Multiple rotational degrees of freedom, e.g. through a ball joint, are also possible, but can make it difficult to control part motions. Linear slides provide a single translational degree of freedom, but these tend to be large and heavy and to have problems with friction, binding, and tolerances. It is advisable to avoid linear degrees of freedom where possible. Multiple linear degrees of freedom, e.g. gantry systems, tend to multiply, rather than add, such negative attributes.

6.3 Assemblies and Loading

Different ways of attaching the same parts can result in very different loads at connections, which makes geometry of attachments very important.

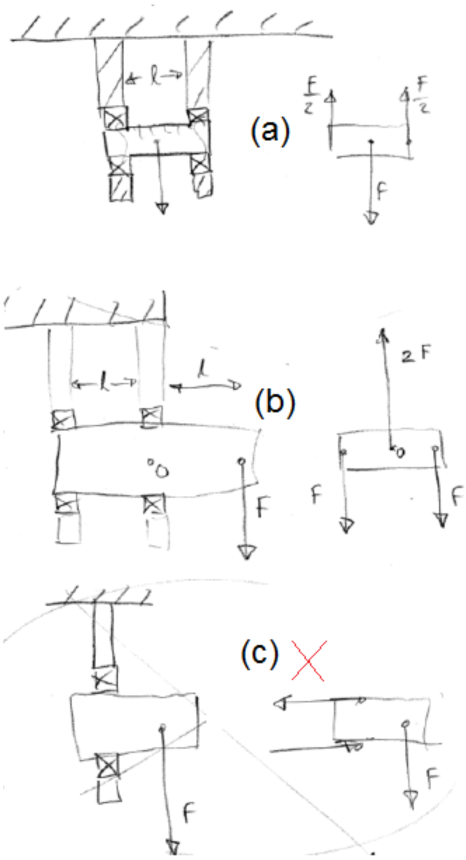
Example 1: A two-part beam

Consider the case of an end-loaded cantilevered beam constructed in two parts of equal length depicted at right. Connecting the beam sections using screws (or nuts and bolts) that are aligned with the beam axis (a) results in a normal contact force between beam sections at the bottom and tensile loading of the screw at the top. This design has the advantages of better qualitative loading, lower forces (because of the wide spacing of the reaction points), and tighter tolerances along the beam axis. However, it requires more complex geometry and manufacturing because of the flanges with holes along the beam axis. If holes are instead placed orthogonal to the beam axis (b) and pins or loosely-tightened screws are used to connect the beam parts, this results in shear loading of the screws or pins. This design has the advantage of simpler manufacturing, but causes less desirable loading, more beam mass (because of the overlapping sections), higher reaction forces and stresses, and poorer tolerances along the beam axis (because of the need for a slip fit between the screws or pins and the holes). These problems can be reduced by instead using large nuts and bolts to press the two plates together tightly, allowing transfer of load through friction near the contact between the two plates. Prior to tightening, the alignment of the two plates could be set as desired, possibly resulting in good tolerances. However, the bolts would be under higher loads to generate sufficient normal forces, because coefficients of friction tend to be less than one. The bolts would therefore be heavier. The free body diagram in (c) helps us understand the loading implications.



Example 2: Mounting a shaft with bearings

Consider the case of a shaft supported by bearings housed in a support structure as depicted at right. If the shaft is supported such that the load application point is between the two bearings (a), this is known as 'simple support'. Simple support is strongly preferred over cantilever support, where the bearings are both on one side of the load application point (b), because of the implications for bearing load and shaft tolerances. By performing static analysis, we can see that in this case simple support results in the peak bearing load to be ½·F, while cantilever support results in a peak bearing load of 2·F. This means we would need a bearing that was four times bigger! This is a serious issue, since bearings often contribute substantially to assembly mass.



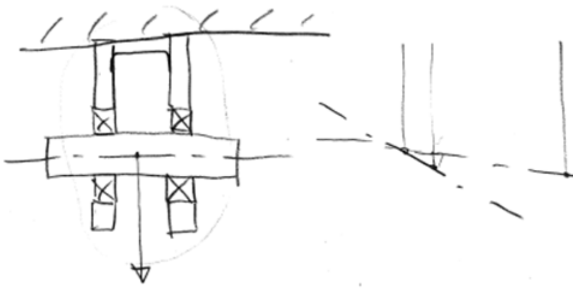
Bearings are not typically designed to take axial loads, which would be caused if we only had one bearing for this shaft (c). A ball bearing would break under such conditions, while a bushing would bind or wear badly.

6.4 Assemblies and Tolerances

Every manufactured part has a range of expected (and hopefully allowable) error. When parts are put together, these errors can combine to cause larger overall errors. If several parts are connected end to end, their tolerances are said to 'stack', such that the total expected or allowed error in absolute position of the final component is equal to the sum of the expected errors of all the parts in the series. Often, you will care about errors in relative positions of certain features of parts in certain directions, and putting things together in a particular way can minimize those critical errors in particular.

Example: Mounting a shaft

Consider again the case of supporting a shaft with bearings, depicted at right. The support structure might be constructed of different parts for each bearing, and the location of the holes in these parts might have some error. The bearings themselves might have some 'slop' or 'play', meaning they may allow some relative movement in the radial direction. Taken together, we could calculate the maximum angle of the shaft with respect to the desired horizontal orientation. A quick calculation will show that spacing the bearings further along the shaft axis will result in lower expected error in angle. Better yet, we might manufacture the bearing support element as a single block, with both holes milled in the same process, resulting in lower expected error in vertical position of the bearings independent from axial spacing.



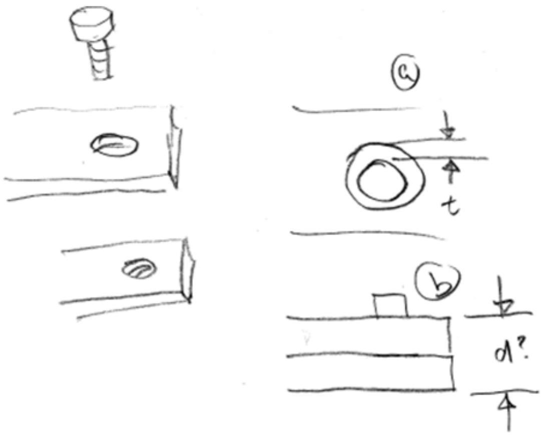
6.5 Assemblies and Constraint

As we saw when using static analysis during the design process, rigid interconnected parts can be *perfectly constrained*, where only one set of reaction forces and moments are possible, *over-constrained*, in which case there are many possible sets of reaction forces based on static analysis and further model information (such as compliance) is required, or *under-constrained*, where parts cannot be in static equilibrium and will move (unless the applied loads have zero magnitude). Note that, confusingly, the systems of equations that describe these situations have the opposite nomenclature, with 'under-constrained' equations describing the family of solutions to over-constrained rigid bodies.

Perfectly constrained assemblies are desirable, but are often difficult to achieve for practical reasons. Over-constraint occurs commonly, and can have some surprising results for the way parts are loaded.

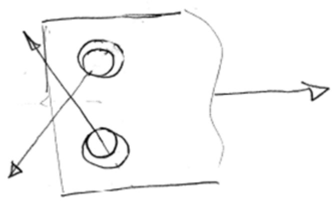
Example 1: Rigid joint constraint

Consider a bar with a through hole and a bar with a threaded hole that are connected by a screw. In the radial direction (a), a gap is required to allow the screw to pass through the hole. In this direction, the assembly is under-constrained before the screw is tightened, and there is a relatively large amount of expected positioning error, likely several tens of thousandths of an inch. If the screw is tightened securely, the two bars will be locked into a relative radial position (although the tightening process could itself rotate the bars). The combined thickness (b) is better constrained; as the screw is tightened, the only errors expected in d are due to the manufacturing processes that produced the bars, typically on the order of thousandths of an inch.

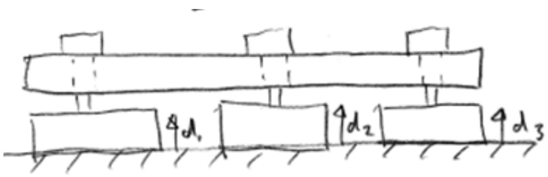


Example 2: Tolerances and Load

Consider the plate at right, which is loaded in tension and supported by two pins (or lightly-tightened screws). Depending on our assumptions, this part is either under- or over-constrained. If we assume that the holes have been designed to be large enough to accommodate the pins despite expected errors in pin size and hole size and location, the fixture is under-constrained and some motion can occur. Perhaps surprisingly, errors in this geometry can also lead to a force multiplier (as pictured) since the contact point between the pins and the plate might not be in the ideal location, resulting in a wedge-like effect. Once again, tightened bolts would allow frictional load transfer and resolve both these issues, but probably a design with normal force transfer would be an improvement. If the holes are not designed with sufficient 'clearance' to allow the pins to fit in despite expected errors, this system is over-constrained. During the pin insertion process, the pins and plate are likely to need to stretch (strain) to fit together, resulting in residual stresses in both, plastic deformation, and/or failure to assemble.



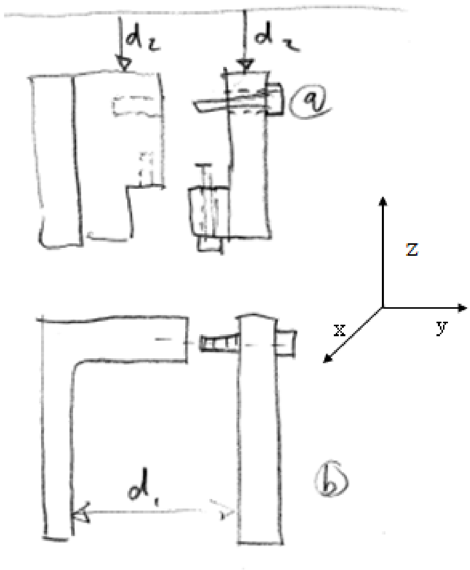
Example 3: Over-constrained rigid joint, with induced bending



Consider the situation above, where a bar is fixed to three separate supports using screws as shown. This fixture is over-constrained, since just two (or even one, depending on assumptions) of these points would fully define the vertical position and orientation of the part. In an ideal case, more connections might reduce stresses at each one by distributing the load. If, however, the three supports' heights are slightly different due to small manufacturing errors, this may require the bar to bend to touch the tops of all the supports. As we have seen, bending can introduce high stresses. If expected errors are high and the bar is stiff, more connections might actually weaken or damage the part.

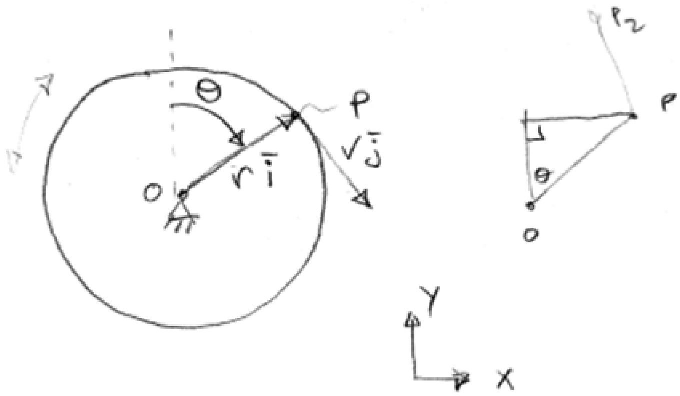
Example 4: Indexing surfaces and constraint

Consider the two sets of example parts being mounted together with screws at right. In (a), normal contact on both the vertical and horizontal surfaces constrain the position of the mounted part in both the y and z directions. This would be a good way to ensure precise positioning of the mounted part. In (b), we see a potentially less precise mounting pattern for a different scenario. The separation of the lower edges, d_1, will depend not only on the y position of the top of the mounted part, which is well-constrained by a normal contact, but also on the angle of the part. This orientation is set by a narrow contact surface, and small y-direction errors on this surface would result in significant errors in d_1. Since the screw is large compared to the contact surface, errors in screw position, orientation, and head shape could also have significant effects on d_1.



6.6 Assembly Kinetics and Dynamics

Now that things have begun to move around, it is possible that those motions will themselves induce loads on the components of the assembly. Let's briefly review the aspects of rigid body dynamics that are relevant to our problem, focusing on two simple cases: constant, high rotational speed and zero speed with high acceleration.



For an point mass rotating around O at a distance of r as shown in the figure above, with the Cartesian ($\vec{x}$ and $\vec{y}$) and polar ($\vec{i}$ and $\vec{j}$) coordinate systems defined, at any time instant, the position vector of the object is

$$\vec{ri} = r \cdot sin\theta \cdot \vec{x} + r \cdot cos\theta \cdot \vec{y}.$$

Thus the velocity is

$$\vec{v} = \frac{d}{dt}(\vec{ri}) = \dot{\vec{ri}} = r(cos\theta\dot{\theta}\vec{x} - sin\theta\dot{\theta}\vec{y})$$

with magnitude (that can be calculated using the Pythagorean theorem)

$$|\dot{\vec{ri}}| = |v| = r\dot{\theta},$$

and the acceleration is (making sure to apply the product rule)

$$\vec{a} = \frac{d^2}{dt^2}(\vec{ri}) = \ddot{\vec{ri}} = r \cdot \Big((cos\theta\ddot{\theta} - sin\theta\dot{\theta}^2)\vec{x} + (-sin\theta\ddot{\theta} - cos\theta\dot{\theta}^2)\vec{y}\Big)$$

with magnitude

$$|\ddot{\vec{ri}}| = r \cdot \sqrt{\ddot{\theta}^2 + \dot{\theta}^4}.$$

From these equations, let's consider two special cases:

- **Special Case 1: Accelerating from zero velocity:**

$$\dot{\theta} = 0, \ddot{\theta} \neq 0, \Rightarrow$$

$$\vec{a} = r(cos\theta\ddot{\theta}\vec{x} - sin\theta\ddot{\theta}\vec{y}) = r\ddot{\theta}\angle(90^o + \theta) = r\ddot{\theta}\vec{j},$$

which means the acceleration is purely tangential, with magnitude $|\vec{a}| = r\ddot{\theta}$.

- **Special Case 2: Steady-speed rotation:**

$$\ddot{\theta} = 0, \dot{\theta} = const \neq 0, \Rightarrow$$

$$\vec{a} = r(-sin\theta\dot{\theta}^2\vec{x} - cos\theta\dot{\theta}^2\vec{y}) = r\dot{\theta}^2\angle(-\theta) = -r\dot{\theta}^2\vec{i} = -\frac{v^2}{r}\vec{i}$$

which means the acceleration is purely radial, pointing from the object to the center of circular motion, with magnitude $|\vec{a}| = {v^2}/{r}$.

For both cases, we have

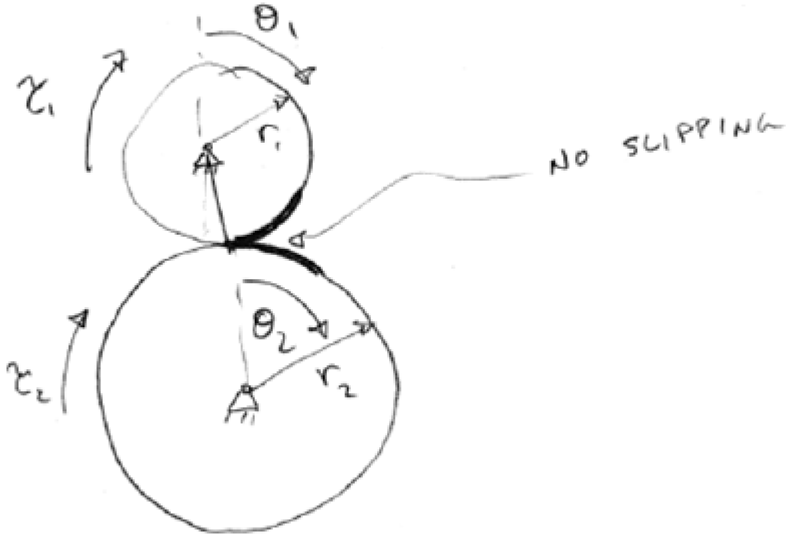
$$\vec{F} = F_{radial}\vec{i} + F_{tangential}\vec{j} = m\vec{a}$$

How do we use this information? Well, in order to accelerate a part of your assembly, the rest of the assembly must exert a force on it, and so there must be reaction forces that produce this net force. Another way to think about these forces that arise as a result of acceleration is as if they were due to an external force acting on the center of mass of your part, kind of like a special gravity. You can then use static analysis methods to analyze the stresses in your components. When things spin or accelerate quickly, such dynamic loads can become significant and can lead to component failure. This effect is often combined with fatigue, since rotational motions are often repeated many times per second and can cause full load reversal on each cycle.

6.7 Gearing Elements

In designing assemblies, we often wish to use leverage or gearing to increase forces at particular reaction points or to alter speeds of different elements relative to one another. You might use gears, timing belts and pulleys, linkages, or other leveraging elements to accomplish this. Let's refresh our memory as to how gearing works in rigid body systems.

For two circular elements geared together as in figure below:



Assume that there is no slipping on the contact point. The arc length covered by the contacts of the two elements in the same period is the same:

$$L_a = -r_1\theta_1 = r_2\theta_2,$$

from which we can see that the positions are related as

$$\theta_2 = -\frac{r_1}{r_2}\theta_1,$$

and since r is constant in time, the **angular velocity relationship** is

$$\dot{\theta}_2 = -\frac{r_1}{r_2}\dot{\theta}_1.$$

Newton tells us that the forces experienced at the contact point by the gears, F_1 and F_2, are equal and opposite, or:

$$F_1 = F_2$$

By the definition of torque, we also know that

$$F_1 r_1 = \tau_1 \text{ and } F_2 r_2 = \tau_2,$$

therefore, we get the **torque relationship**

$$\tau_2 = \frac{r_2}{r_1}\tau_1$$

So, by selecting r_1 and r_2 appropriately, we can set the torque or velocity of the second gear given an input torque and velocity of the first gear. This ratio of gear radii (or pulley radii or sprocket radii or lever lengths) is referred to as the **gear ratio** of the system. It is usually denoted as N:1, said ”N to one”, where N is the ratio of the torque ”out” of the gear set to the torque ”in” to the gear set. If we consider τ_2 as the output in the example above, the gear ratio would therefore be r_2/r_1. If, for example, $r_1 = 2$ and $r_2 = 20$, the gear ratio would be 10:1.

6.8 Acknowledgments

Thanks to Juanjuan Zhang, Rachel Jackson, Myunghee Kim and Kirby Witte for help in editing this chapter.