Table of Contents

[Introduction 2](#_Toc37336556)

[Section I: An Introduction to Nitinol 6](#_Toc37336557)

[Chapter 1: A Brief History of Nitinol and the Shape Memory Effect 7](#_Toc37336558)

[Chapter 2: What Causes the Shape Memory Effect? 16](#_Toc37336559)

[Chapter 3: Setting the Memory Shape 25](#_Toc37336560)

[Section II: Working with Nitinol Alloys 31](#_Toc37336561)

[Chapter 4: Hysteresis: Why It Matters and How to Deal with It 32](#_Toc37336562)

[Chapter 5: Fatigue 42](#_Toc37336563)

[Chapter 6: Nitinol Manufacturing and Metalworking 49](#_Toc37336564)

[Chapter 7: Actuators 61](#_Toc37336565)

[Chapter 8: Heat Engine Design 68](#_Toc37336566)

[Chapter 9: Non-Shape Memory and Superelastic Applications of Nitinol Alloys 74](#_Toc37336567)

[Chapter 10: The Two-Way Shape Memory Effect (TWSME) 80](#_Toc37336568)

[Appendices 85](#_Toc37336569)

[Appendix A: Glossary 86](#_Toc37336570)

[Appendix B: Furnace Design and Construction 90](#_Toc37336571)

April, 2020

# Introduction

When I was in graduate school, it became apparent to me that, while there was a fairly large volume of literature about nitinol prepared by researchers, for researchers, I could not find literature anywhere that detailed the aspects of nitinol to creative people who were not professional researchers. I found this to be a travesty because I have continually found that most great ideas are generated by people who are not professional researchers, so I undertook the task of writing a book about nitinol, using the minimum possible technical language so that garage inventors, artists, and engineers can all begin experimenting with nitinol, and generating ideas about how to use its amazing properties. For the technical language that absolutely had to stay in the book, I have included a glossary in the back, defining these terms.

For areas that you wish to obtain a deeper, technical understanding, I have tried to include as many references to top authors in each field so that you can learn from the best. All of these books can be purchased on Amazon with relative ease.

Nitinol may be difficult—it disobeys almost all of the rules that engineers are taught to follow in school—but it doesn’t have to be nearly as difficult as it is. By laying out the difficulties in plain language, it is my hope that more people will find themselves willing to tackle the difficulties of nitinol in order to reap its tremendous benefits.

Writing this book has been a labor of love. I began working on it shortly after finishing graduate school. There were times where I worked very hard on the book and there were times where the book went months and even years without being touched. It may not yet ready to publish, but I think that it’s more valuable that the information be available than the book wait until it is perfect.

If, after reading the book, you think that something could be explained in greater detail or that there is subject matter that should be added to the book, please don’t hesitate to send in a note to us at Kellogg’s Research Labs by email at [Nitinol@KelloggsResearchLabs.com](mailto:Nitinol@KelloggsResearchLabs.com). People will make sure that I see every suggestion. I will take this message as a compliment. The longer I study nitinol, the more I realize just how little I know about nitinol. If you point out some new phenomenon, that’s awesome! If you want to add some applications to the next version of the book, send it over and it will be considered.

If this book inspires you to go out and create something from nitinol, then I have accomplished my mission. So, go, experiment. See what can be done with nitinol. See if you can discover something that has not been discussed here. The sky is not the limit for nitinol.

Joe Kellogg

CEO

Kellogg’s Research Labs

Doing the Impossible

# Section I: An Introduction to Nitinol

Lately, nitinol has been receiving an increasing amount of attention from media sources, such as the Discovery Channel, all around the world. The oh so fascinating shape memory effect has absolutely entranced everyone from children to scientists alike.

What is this strange metal? Where did it come from? Why does it behave quite unlike anything else we have seen before? How can I do something useful with it? These are all questions that are being asked by people of all sorts. This section covers a brief introduction to nitinol and its inner workings. Also covered is some advice as to how to set the memory shape.

This section is designed to set the groundwork and establish the basics which will be built upon in Section II.

## Chapter 1: A Brief History of Nitinol and the Shape Memory Effect

Mark Twain once said that accident is the mother of invention. Take the light bulb for example, while Thomas Edison is famous for inventing the light bulb, he already knew that a wire glows when it conducts electricity. This was discovered much earlier, quite by accident by the scientist James Prescott Joule who quantified the amount of heat generated by electricity passing through a conductor. Well, nitinol is no different.

 It is important to note that nitinol was not the first shape memory alloy discovered. Researchers had been playing around with gold-cadmium since 1939, but the shape memory effect was minimal and the material was extremely expensive ($100/gram). Nitinol was discovered by a brilliant young scientist named William J. Buehler. Buehler was a metallurgist at the Naval Ordinance Labs (NOL), working on a project to develop a nose cone for the Polaris missile that was capable of withstanding the heat of re-entry into the Earth's atmosphere.

Figure 1: A launch of the Polaris missile. The Polaris missile is a Submarine Launched Ballistic Missile (SLBM).

Buehler described this project as 'boring' and was hoping that something 'interesting' would pop up. Well, it did, but not in the way that anyone would have expected. Buehler was looking at alloys with two solid states as possible materials for

for further examination from a book entitled Constitution of Binary Alloys--nitinol being one of them. When he made the ingots for testing, he intentionally dropped one of the cold ones on the floor. Hoping to hear a clear bell-like ring, indicating that the metal had the properties he was hoping for. Instead, it returned a dull thud--similar to dropping a sack of flour on the ground.

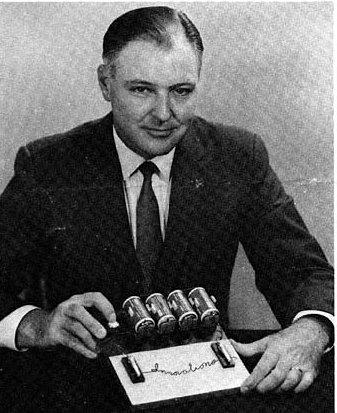
 Worried that the ingot was filled with internal flaws, he dropped one of the ingots that hadn't cooled yet. This returned a wonderful bell-like ring. However, after the ingot had been cooled in water, it returned a dull, leaden thud, just like the first one. This is the first indication that nitinol had a substantially different double state. Buehler named his alloy nitinol for **Ni**ckel-**Ti**tanium **N**aval **O**rdinance **L**aboratories. All of this occurred in 1959. However, the shape memory aspect of nitinol wasn't discovered until a lab meeting in 1961.

Figure 2: Dr. Buehler discovered the shape memory properties of nitinol while working at Naval Ordinance Laboratories.

Buehler had been performing tests to determine the fatigue life of nitinol by bending a strip into an accordion like shape over and over again. His project was brought under review and his technician was demonstrating the fatigue properties to senior officials. During this presentation, one of the officials present heated the nitinol with a lighter, at which point it rapidly straightened out.

This, of course, sent ripples throughout the scientific community. This material could take low grade heat and generate mechanical energy (motion)! Numerous scientists began experimenting with how to build engines with nitinol that would take low grade energy and transform it into very high grade energy that could be used to do work. This culminated in the Nitinol Heat Engine Conference, hosted by the Naval Surface Weapons Center (previously Naval Ordinance Labs) in 1974.

At the Nitinol Heat Engine Conference, the NSWC gathered together the top scientists who had been working on nitinol to discuss what had been done and what still needed to be done to make nitinol heat engines a reality. The presentations from this conference are available in the book Proceedings of the Nitinol Heat Engine Conference.

At this point, the activity surrounding nitinol seemed to all but dry up and disappear. Prior to the conference, nitinol researchers were featured on news channels like CNN and BBC. Afterwards, there was little to no attention given to nitinol by major television networks for over twenty years. This led to all kinds of conspiracy theories ranging from nitinol being kept for top secret government experiments to nitinol being an alien technology that was discovered in the Roswell accident.

However, the reality is very different. Behind the scenes, material scientists were working hard to figure out HOW nitinol worked. In order to fully optimize a nitinol heat engine, it must be understood what happens when nitinol undergoes the shape memory effect. This will be discussed in greater detail in chapter 2.

 The next time nitinol appeared in public, it wasn't referred to as nitinol, it was called just plain titanium. Of course, this is a misnomer since nitinol is slightly more nickel than titanium. This new public appearance didn't even exhibit the shape memory effect as people were so excited about earlier. No, this was completely different and it was marketed under the trade name Flexon® by the company Marchon Eyewear.

Figure 3: A pair of nitinol eyeglasses. The bridge that connects the two lenses together as well as the bows (or arms) are both made of superelastic nitinol.

Released for public sale in 1995, Flexon® was unusual in that you could bend it through incredible distortions and it would just snap back to its original shape once you released it. Once Nike began to use it in their Vision line of glasses, athletes everywhere began buying it up. Gone were the days of breaking your glasses on a regular basis just because you lived an active lifestyle. You could sit on them, you could intentionally bend them, you could tackle someone in football, and they would just bounce back. These glasses would forgive those bumps and bangs over and over again--seemingly forever. People who broke their glasses every six months now could go several years on a single frame. The frames were termed 'superelastic'.

Once superelastic nitinol was discovered, it wasn't long before surgeons began using it as they worked on people. Vascular stents was one of the first applications because the stent could be folded so flat that it could be inserted through the tiniest of holes into the patient's bloodstream--minimizing recovery time. Once in place, the superelastic nitinol wire could withstand severe deformation and outlast stainless steel by an order of magnitude.

To put this in perspective, superelastic nitinol stents were capable of undergoing a 3% deformation with a cycles to failure life expectancy greater than 10,000,000. Stainless steel, on the other hand, could withstand a deformation of just 0.5% with fatigue life of around 1 million cycles. This made intravascular stents something that could be considered permanent, never needing replacement.

The only question that the medical community had revolved around the issue that nitinol is more than 50% nickel and people are allergic to nickel. Yes, strange as it may seem, our five cent coin is made from a metal that we are allergic to. Of course, this could potentially cause huge problems for patients if enough nickel were to dissolve out of the nitinol and into the bloodstream. Fortunately, researchers quickly discovered that nitinol's biocompatibility is unsurpassed and that doctors should feel free to implant it into patients in any way they see fit.

Since then, nitinol has replaced other alloys in just about every kind of implant in the human body. Researchers have found that it makes a great hip replacement material because the superelastic phenomenon damps out the vibrations caused by walking--greatly extending the useful life of a joint replacement. Because of the broad spectrum use of nitinol in surgical implants, the medical field is the largest consumer of nitinol worldwide. The second largest consumer of nitinol being the eyeglasses industry.

But whatever happened to the interest in the shape memory effect that had once lit up the scientific community? Well, while all of this was going on, the National Science Foundation (NSF) was sponsoring materials science research at universities across the United States. These small, independent research groups turned up all sorts of interesting, useful knowledge about nitinol.

To organize all of this research and to share knowledge gained, the American Society of Materials (ASM) created the Shape Memory and Superelastic Technologies (SMST) group that would meet every eighteen months to discuss the latest improvements in the material science of nitinol. First meeting in 1997 with a gathering of experts from all over the world, SMST began corroborating findings about all sorts of things. New alloys of nitinol were discovered. Termed 55-Nitinol, 60-Nitinol, and 65-Nitinol for the approximate percent nickel content, these alloys had slightly different properties than regular nitinol. It wasn't long before scientists were discussing adding third and fourth elements into nitinol alloys to alter the properties to give something more desirable. All of this will be covered in greater detail in Chapter 4.

With all of this research into the material science of nitinol and the discoveries that followed, this sparked renewed interest from mechanical engineering scientists. In 2012, General Motors announced that they were working on a nitinol heat engine that would capture the waste heat from the exhaust from the engine and generate electricity. The hope here was to replace the alternator and use less fuel.

Then in 2013, Kellogg's Research Labs published that we had built a generator capable of harnessing the energy from the daily change in air temperature. In the report, KRL stated that our thermal efficiency was between 1 and 4 percent. This is absolutely horrible when comparing to cars, which run at roughly 25% efficiency, or the power plants that generate electricity to power the nation, which run at 43% efficiency. However, unlike those systems, the amount of heat energy discharged by the atmosphere on a daily basis is several orders of magnitude greater than all of the power generated by all power plants on earth, and it's free. So, 1% of a very large number is still a very large number.

Notes

## Chapter 2: What Causes the Shape Memory Effect?

*This chapter makes use of a very large amount of terminology. The first time each term is used, it will be defined and described as best as possible. For a complete list of terms, please consult the Glossary.*

What causes the shape memory effect? From the time the shape memory effect was first observed, this question was on the tip of everyone's tongue. Before the shape memory effect could be utilized to do anything useful, the mechanism behind it must be understood.

In traditional materials, there are two kinds of deformation: elastic (or linear elastic) and plastic deformation. Linear elastic deformation refers to deformation that is fully (or very close) recovered when the force is removed. The equation governing elastic deformation is known as Hooke's Law:

where F is force, k is the spring constant, and x is the distance that the spring is deformed. Linear elastic materials are called this because, if you plot force vs. displacement (also known as stress vs. strain) on a graph, a straight line is drawn. A sample stress/strain plot is shown on page 33.

Plastic deformation refers to deformation that is permanent. When the force is removed, the deformation is not recovered. On a stress/strain plot, plastic deformation occurs when the curve becomes non-linear. Initial plastic deformation comes from slip, which then continues into microfractures, and eventually failure or fracture of the part.

Shape memory materials appear to undergo plastic deformation, but after heating, elastically return to their original shape. Since this doesn’t clearly fit into the category of elastic or plastic, unique words are often used to describe nitinol. For example, thermoelastic or pseudoplastic. Pseudoplastic grabs the idea that the component appeared to have plastic deformation, but didn’t actually. Thermoelastic conveys the idea that the material behaves elastically with the application of heat. Nonetheless, this is quite perplexing, trying to explain how an apparently permanent deformation can be removed by simply heating it (especially at temperatures below 100°C). Nitinol appears to be defying the laws of physics. Of course, we know that this is not the case, so an examination of

Clearly, at the atomic level, the deformation in the shape memory materials is substantially different from the deformation in linear elastic materials, so let's take a look at the traditional modes and compare them to how shape memory alloys experience deformation.

A picture containing abacus, object, small, top

Description automatically generated Metals are known as crystalline objects because all of the atoms are held together in a very regular spacing, known as the crystal lattice. The most common kind of deformation at the atomic level is called diffusion. This is when the atoms trade places with each other and move around the crystal lattice. Each time this happens, the electron bonds between the atoms are broken and then re-formed. This type of movement within the crystal lattice is largely temperature dependent , increasing as the temperature increases, and generally doesn't have any noticeable effects at low temperatures. It is worth noting that one of the methods of welding nitinol is diffusion bonding. This is done by heating the nitinol to 1,400°F (750°C) and holding the two pieces of nitinol together while the atoms jump across the gap between them. Diffusion bonding produces nearly perfect welds.

Figure 4: A graphical representation of diffusion. The green atom trades places with the red atom next to it.

A picture containing small, object, abacus, indoor

Description automatically generated Slip is the primary mechanism, or main way, plastic deformation occurs. In slip, an entire sheet of atoms breaks their electronic bonds and moves one or more atomic spacings. It is important to note that this only happens for a whole number of atomic spacings (this will be important very soon). If enough crystal planes move enough spacings, a micro-crack forms--usually much smaller than 1mm in length. Once enough micro-cracks form, they coalesce, or come together, to form a crack which is soon followed by the catastrophic failure of the object.

Figure 5: A graphical representation of slip. Note how one of the rows of atoms has slipped by one atom's distance. With enough of these slips, microfractures will form, eventually coming together to cause the large scale breaking of the object.

It is fairly obvious that with slip, the deformation is not recovered, as the atoms are now in different locations than they were before. This is a very critical idea when comparing to the pseudoplasticity of nitinol.

It was Dr. Frederick Wang who determined how the shape memory effect works, while working at the Naval Ordinance Laboratories. The shape memory effect is not governed by either of these mechanisms. Instead, it is caused by a phenomenon called the martensitic transformation. The martensitic transformation happens when a material transfers between two different, solid, crystal states, namely austenite and martensite. Austenite, the high temperature state has a very strong, cubic structure where the atoms form cubes. Martensite, on the other hand, has an atomic spacing more like a parallelogram. These two crystal states were first observed in steel, but the transformation occurs at very high temperatures and the martensitic transformation in steel is less powerful than plastic deformation by crystal slip.

A picture containing object, abacus

Description automatically generated What makes nitinol exhibit such a powerful shape memory effect is the presence of the phenomenon twinning. Twinning is when atomic bonds rotate through an angle to make a sort of mirror image. Figure 6 is an artist's rendition of twinning at the atomic level. These bonds can be described as being very 'floppy' meaning that they can rotate with relative ease and stay put after they have rotated. This gives the appearance of a soft, easily deformable material. However, when the material is heated so that it goes back into the austenite phase, since none of the electronic bonds have been broken, the original shape returns. For more information on twinning, please read Twinning and Diffusionless Transformations in Metals by E. O. Hall, Butterworth Publishing, 1954. This brief book is highly technical, so it is not for the average reader, but it is well worth the time spent if you are trying to understand the underlying workings of nitinol.

Figure 6: A graphical representation of twinned martensite (top) and cubic austenite (bottom).

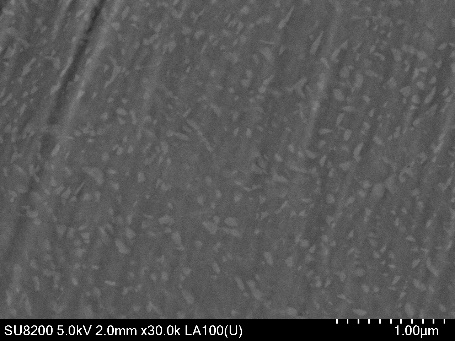


Figure 7: An electron microscope scan of nitinol

According to Hall, there are two types of twinning: impact induced twinning and heat induced twinning. Impact induced twinning occurs when the metal is struck with a very large point force--for example being hit by a bullet. This causes the crystal structure to be distorted into the twinned structure. Heat induced twinning occurs when a metal is heated at the appropriate temperature for an appropriate period of time. This is why nitinol must be heat treated before it exhibits the shape memory effect.

Going back to our discussion on the martensitic transformation, there are really two types of martensitic transformations. The first is the thermally induced martensitic transformation which results in the shape memory effect. The second martensitic transformation is the stress induced martensitic (SIM) transformation. This second variation happens when a sample of nitinol is in the austenitic phase and a force, or stress, is exerted on the sample. This stress causes a portion of the austenite to transform into martensite.

Since martensite can undergo substantial deformation without experiencing any crystal slip, SIM allows austenitic nitinol to experience deformations up to 50%. Since this is one hundred times the amount of deformation that can be sustained by conventional materials like stainless steel, the SIM transformation in nitinol is referred to as superelastic.

Here is an interesting phenomenon that we have not seen commercial use of, so we are publishing it here in this book in hopes that someone will find a way to harness it to make the world a better place. The formation of SIM is exothermic (it heats up). If you think about it, this makes sense because you have to remove heat from nitinol to produce martensite. But, this means that you can create heat from motion (and do it in a very controlled manner).

In the same manner, releasing a superelastic device or returning a superelastic nitinol device from SIM to pure austenite is endothermic (it gets cold). It is actually quite marked just how cold the nitinol gets on a rapid release. This means that nitinol can be used to produce very rapid, localized cooling.

As was stated above, we are not aware of any commercial applications of this phenomenon, but hope that one of our readers will have an “Ah ha!” moment and do something useful with it.

The transformation temperature of nitinol is determined by the ratio of nickel to titanium. Increasing the amount of titanium increases the transition temperature while increasing the amount of nickel decreases the transition temperature. After the initial ingot is prepared, the transition temperature can be affected by heat treating. Generally speaking, heat treating at temperatures below 600℃ (1,100℉) causes the transition temperature to rise while heat treating at temperatures above 600℃ (1,100℉) causes the transition temperature to drop. This is because, at low temperatures, Ti2Ni3 crystals begin to form in the matrix, thereby reducing the effective ratio of nickel to titanium and, at high temperatures, the nickel and titanium return to the matrix, balancing it back out.

Without a doubt, the most common misconception that customers have when they call the lab phone is that the shape memory transformation occurs AT the transition temperature and the reverse transformation happens immediately below the transformation temperature. Unfortunately, this is not the case. The temperature that is commonly referred to as “transition temperature” is actually the **austenite finish (Af)** temperature. The austenite start (As) temperature is anywhere from a few degrees to fifty degrees lower, depending on how the nitinol is prepared, the elemental ratio, and numerous other factors.

Likewise, the martensite transformation also has a start and finish temperature (Ms and Mf, respectively). To make things interesting, Ms frequently is a higher temperature than As. This means that between these two temperatures the transformation is continually occurring. The critical temperatures of nitinol are covered in greater detail in Chapter 4.

This is just a brief overview of the martensitic transformation. For a more in depth look at the martensitic transformation and the shape memory effect, please read the book Microstructure of Martensite: Why It Forms and How It Gives Rise to the Shape Memory Effect by Kaushik Bhattacharya, published by Oxford University Press.

Notes

## Chapter 3: Superelasticity

Superelastic nitinol behaves like a super-spring, withstanding far greater deformations than other metals are capable of. For example, a 316 stainless steel spring will fail, or fracture, at 0.5% strain. A superelastic nitinol spring, on the other hand, has a fatigue life of 10,000 cycles at 5% strain. When compared to other metals, the general rule of thumb is that nitinol has 100x greater fatigue life at 10x greater strains.

*What Causes Superelasticity?* From the outside, superelastic nitinol behaves completely different from shape memory nitinol. However, the underlying mechanism is strangely similar. In shape memory nitinol, heat is used to transform martensite into austenite. When the heat is removed, the austenite returns to martensite. In superelastic nitinol, stress (force) is used to transform austenite into martensite. When the stress is removed, the martensite transforms back into austenite.

*Note:* A common mistake that we see when working with nitinol is a request for the spring constant or elastic modulus. While nitinol is linear-elastic over a range of displacement, most applications are in the non-linear region. This means that the spring constant is not a constant, but an equation. For good modelling, this equation should have at least four independent variables and we have run simulations with as many as twelve independent variables. It’s not hard to imagine that this would make engineering with nitinol very difficult.

*Non-Linear Considerations:* When selecting materials for a product, it is quite useful to know the elastic modulus of the material. With this in hand, it is relatively simple to calculate how the part or system will behave when forces are applied to it. If the elastic modulus is not constant, then the calculations required to simulate the system quickly become challenging.

*Maximizing Superelasticity:* Nearly everything about nitinol can, and needs to be, tuned and optimized. Superelasticity is no different. Identifying the proper manufacturing process that yields an ideal microstructure will optimize the superelastic performance of your nitinol. Additionally, superelasticity is optimized when the operating temperature is very close to Af. It is at this temperature that the austenite can most readily transform into stress-induced martensite and then detwin. This means that you want your superelastic device to have an Af just a few degrees below the reasonably expected minimum operating temperature.

Superelasticity is optimized with either small grain size or very large grain size (single crystal). While it is counterintuitive at first, the underlying mechanics are identical in both of these apparently opposite cases. The primary cause of slip in a high-deformation application is slip on the inter-grain boundary. On the one hand, minimizing grain size minimizes the effects of slip of a single grain. On the other hand, completely removing grain boundaries altogether eliminates the risk completely.

Single crystal nitinol has a second method to improve the superelastic performance. In polycrystalline nitinol, the individual grains are oriented in a wide range of angles (+/-30 degrees is common). This means that nitinol is spending much of its ability fighting against itself. The out of alignment grains have very little strain while the in-alignment grains are observing highly localized strain. Single crystal nitinol is very expensive, but it is able to sustain 10x or more strain than polycrystalline nitinol.

*Sub-Af Superelasticity:* Superelastic applications below Af is a topic that requires a special consideration. We often encounter the misconception that nitinol cannot be superelastic below Af, but this is not really true. Below Af, nitinol can fall anywhere in the range of 0-100% superelastic. The degree of superelasticity that is retained is related to the Ms temperature under strain. The critical temperatures of nitinol shift upwards with strain or deformation. If Ms is higher than ambient, then some of the stress-induced martensite converts to thermally-induced martensite, limiting the superelasticity of the part. If the under-strain Mf is higher than the operating temperature, then the part is fully thermally-induced martensite and has 0% superelasticity.

*When to Use Superelastic Nitinol:* Costing about fifty times as much as stainless steel, nitinol can rapidly increase the cost of a product, so every gram used needs to be carefully justified. There are several reasons you would choose nitinol over a cheaper material. While many are application dependent and they rely on your creativity to use best, many apply to all applications. Here are several examples that apply to all applications.

1. Extreme Deformation: Extreme deformation is the most common use of superelastic nitinol. If you have a structure that needs to fit into a very tiny space, and then deploy out into a much larger volume, then superelastic nitinol is for you. In medical, this is what allows a stent to be packed inside of a 0.5mm catheter and then deploy out to 25mm diameter. In Space, we have built antennas 21 feet long that only occupy 1.5” of space on the satellite during launch. The general rule of thumb is that superelastic nitinol can withstand 100x more deformation than other materials and still have a fatigue life of 10,000 cycles.
2. High Fatigue: Sometimes you are within the deformation allowances of your material, but it doesn’t pass the fatigue requirements of the application. The fatigue life for an equivalent nitinol part is 100-1,000x that of the traditional part. This may be enough to justify the cost of nitinol.
3. Constant Force: Most materials and springs follow Hooke’s Law, which states that as deformation (or displacement) increases, the force also increases. For superelastic nitinol, once the plateau stress is reached, the force remains constant. If your application would benefit from having the same amount of force, regardless of position, then a superelastic nitinol component is the solution for you.
4. Anti-Corrosion: We have built products for numerous deep sea applications simply because, while other materials corrode in seawater, nitinol does not. If you are building a device that is in a corrosive environment, choosing superelastic nitinol may be a good choice.
5. Vibration & Shock Reduction: If your system would benefit from having fewer vibrations or increased shock resilience, then a nitinol component may be the way to go. We did a project with a race team, replacing the vehicle’s springs with nitinol springs to dissipate the natural vibration of the car. The result was that the car oscillated less and remained in better contact with the pavement and the driver performed better relative to the other competitors.

## Chapter 4: Setting the Memory Shape

*Editor's note: The temperatures mentioned in this chapter are very approximate. The temperatures mentioned here are chosen to be a round number in the middle of the ideal temperature range. Conversions between Celsius and Fahrenheit are approximate rather than exact for this reason.*

Nothing about nitinol causes more frustration and difficulty than the process of setting the memory shape. Researchers around the world spend tremendous amounts of time determining how to properly heat treat nitinol to get the properties that are needed for their particular project. However, there are at least a dozen independent variables that influence the properties of nitinol. Therefore, no mathematical model exists to predict the heat treatment of nitinol.

Now, there are two basic aspects of a heat treatment that are needed to properly set the memory shape into nitinol: the right temperature and the right length of time. Thin nitinol wires can have the memory shape set very, very quickly because they are so thin. Thicker samples of nitinol require longer duration at the appropriate temperature as it takes longer for the crystal structure to rearrange.

With thicker samples of nitinol, the longer the heat treatment, the stronger the shape memory effect is. This is because more and more of the nitinol is twinned and capable of accepting deformation. If a nitinol sample isn't heated very long, the shape memory effect must spend the bulk of its energy fighting against the untreated nitinol, which wants to behave like a normal material. Therefore, it is important that nitinol is heated long enough to exhibit as strong of a shape memory effect as needed.

The second part of heat treating nitinol--the right temperature--is much easier to discern what temperature must be used to set the memory shape. While, at KRL, we use temperatures 250-850℃ (480-1,550℉) to obtain various properties, it is generally agreed upon that the ideal temperature for setting the memory shape in basic applications is approximately 930-1130°F (500-550°C).

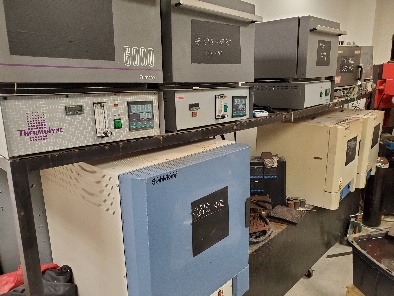
 The ideal way to set the memory shape is to use a calibrated lab furnace with a digital interface that will read back the temperature accurate to 1°F. A used lab furnace can be picked up for around $500-1,000 from marketplaces online such as eBay. New, they cost upwards of $5,000. Another method that yields excellent results over and over again is a fluidized salt bath. However, these run upwards of $10,000. So, what cost effective ways are there to set the memory shape? That is the purpose of this chapter.

Figure 8: Some of the many furnaces at KRL. Our electric bill is quite impressive. Notice that the top row of furnaces are all equipped for controlled atmospheres. This can be inert, such as argon, to prevent oxidation, or reactive, such as benzene, to grow layers (graphene) on the nitinol.

Ok, so we just have to hold the wire in the shape we want and heat the wire to 850-1,100°F (450-600°C) and then quench in water. Sounds simple enough right? Yeah, right. The biggest problem is measuring the temperature of the metal, not the air surrounding it. Without expensive instrumentation, the only real way to determine the temperature of the wire is to look for a red glow. Now, nitinol just barely begins to glow red at about 1100°F (600°C), more than 100°F (50°C) hotter than the ideal temperature. Videos on YouTube make it look like it's easy to set a memory shape using a propane torch. However, a propane torch burns at an incredible 3600°F (2000°C), far hotter than needed. It is important that you use a cooler flame to set the memory shape.

Candles made with paraffin wax are great for setting the memory shape because they burn at about 1100°F, just a little above the ideal shape setting temperature. Caution must still be taken to not overheat smaller wires, but for everything 0.020" (0.5mm) and thicker, this is not a big problem.

Now, after this initial heat treatment, called annealing, the nitinol will have one of three different properties. It will be either stiff and springy, soft and leaden, or brittle. If the wire is brittle, this means that the wire has been overheated and is no longer any good. This brittleness is caused when the titanium atoms combine with carbon and nitrogen in the atmosphere to form titanium carbide and titanium nitride crystals within the crystal lattice. This prevents twinning from occurring and makes the lattice very stiff and brittle. No shape memory effect can be seen from this nitinol, ever.

If the nitinol is soft and leaden, this means that the memory shape has been set and the transition temperature is above room temperature and the nitinol is ready to use. If the nitinol is soft and springy, this means that the memory shape has been set but the transition temperature is near or below room temperature. If superelastic nitinol is desired, then the job of heat treating the nitinol is complete! However, if the shape memory effect is expected, then a second heat treatment is needed.

The second heat treatment, called aging, is when nitinol is heated to 600-840°F (325-450°C). This causes the transition temperature to rise, but does not re-set the memory shape, so the nitinol sample does not need to be constrained for this part of the process. As was stated in Chapter 2, the aging causes Ti2Ni3 to form, decreasing the nickel/titanium ratio inside the grains.

Now, it's hard enough to set the memory shape using tools found at home, how can you expect to age nitinol at home? Well, it turns out that if you have an electric kitchen oven, you can set the thermostat to 'broil' and put the nitinol on a cookie sheet on the top rack of the oven. This will be within about 50°F of the ideal value and it doesn't take any effort to age consistently for very long times (if needed). Aging can raise the transition temperature by 70°F (40°C). At KRL, we have developed proprietary methods to increase this spread to 110°F (60°C).

There are numerous other heat treatment processes, solution treating and ramping to name two, which can be used to optimize nitinol for your application, but they lie outside the realm of this book. It’s important to note that very sophisticated heat treatment profiles can yield unique properties in the nitinol parts. At KRL, we use up to eight stages in a heat treatment profile. So, why do we use such a wide range of heat treatment profiles and why is it so important that each project expend a significant investment in determining the optimum heat treatment profile? The reason is that each application usually needs just one or two properties from the nitinol and there is a very narrow range of heat profiles that brings out those properties.

A brief listing of some of these properties includes:

1. Stiffness
2. Recovery force
3. Transition temperature
4. Peak shape
5. Deformation recovery
6. Fatigue
7. Two-way shape memory effect
8. Superelastic hysteresis

Notes

# Section II: Working with Nitinol Alloys

This section is a much more in depth look at some of the more important aspects of nitinol. Some of these, like hysteresis, are almost completely ignored in most books about nitinol while others topics are only briefly covered here because the topic is either very complex or it is covered very thoroughly in another book. In this case, a citation will be provided so that the appropriate book may be requested through a local or university library system.

The topics in this section are chosen because they are a reflection of what most people who work with nitinol--from the tinkerer level to the researcher--find difficult. Clearly, there are many more topics that could be inserted into this section. However, the purpose of this book is to provide an introduction to nitinol and point the reader to resources which will explain in greater depth the various topics.

## Chapter 4: Hysteresis: Why It Matters and How to Deal with It

As was previously mentioned in this book, there are two forms of nitinol which are prized for very different properties, the first being superelasticity and the second being the shape memory effect. It is possible to design a system that takes advantage of both of these mechanisms, but it is by far the rarity, rather than the norm. Also, while there are a few principles that apply to both superelastic and shape memory nitinol, by and large, they rely on different mechanisms. So, superelasticity and shape memory will be addressed as separate mechanisms in this book.

**Superelasticity**

One of the few mechanisms that apply to both superelastic and shape memory nitinol is the mechanism of hysteresis. The Collins English Dictionary defines hysteresis as *'a lag in a variable property of a system with respect to the effect producing it as this effect varies.'* With nitinol, the varying effect is temperature, so this means that there is a temperature gap that must be overcome to thermally induce the martensitic transformation. Superelastic nitinol also exhibits stress hysteresis—a gap in the relationship between force and displacement.

What hysteresis means is that nitinol remains in its austenitic state even after the temperature has dropped well below the transition temperature. So, in a design using superelastic nitinol, the nitinol remains superelastic below the transition temperature. This is why the thermal hysteresis of nitinol is embodied in the statement that, “I cannot know the amount of martensite and austenite in the sample at a given temperature, I do not have enough information.”

However, there is a danger: when nitinol is put under a stress (force), the transition temperature rises. If this stress is large, then the transition temperature can rise quite substantially, causing a thermally induced martensitic transformation. Then, when the stress is removed, the nitinol wire remains martensitic and does not return to its original shape until heated above its transformation temperature.

So, care must be taken to determine if this will be a problem in the design of a project. If the superelastic nitinol will be undergoing generally small deformations, then it will not be a problem for the temperature to drop below the transition temperature. However, if very large deformation is expected, then care must be taken to ensure that the nitinol is formulated so that the transition temperature is never reached.

As was previously discussed in Chapter 2, most materials are considered to be linear-elastic, meaning that they follow Hooke's Law. Materials that follow Hooke's Law are called linear because the stress-strain curve looks like a straight line when it is drawn on a graph. Linear elastic materials are very simple to use in engineering projects because the mathematics required to model them are quite simple. Nitinol, however, is far more complex. In a stress-strain (force-deformation) graph, nitinol begins by following Hooke's Law. However, after about 1.5% strain, the amount of force required levels off substantially. This is referred to as the loading or upper plateau stress. As the nitinol Figure 9: A chart showing the stress/strain (force/deformation) relationship for a nitinol sample. This sample was deformed to failure (until broken) so there is no unloading curve. Notice how the force is almost completely unchanged from 1.3mm to 4.9mm, this is the plateau mentioned in this chapter.

continues to be deformed, the stress remains

roughly the same until it reaches a second linear-elastic region. When the force is released, the nitinol contracts linear-elastically until it reaches a second, lower plateau. This is referred to as the lower or unloading plateau stress. After the lower plateau is traversed, the nitinol then behaves as a linear-elastic material once again until the original shape is fully recovered. The gap between the loading and unloading plateaus is also hysteresis.

When building a project with superelastic nitinol, care must be taken to design according to the upper and lower plateau stresses. The existence of these two plateaus can be very detrimental to a project if they are not well understood first. For example, if a spring is deformed by 3mm, what is the force exerted by the spring? Not enough information is given to know the answer. Is it on the loading curve or the unloading curve? Just like with the thermal hysteresis, the mechanical hysteresis makes it difficult to know how the nitinol is going to behave in a certain scenario. It’s not enough to know the parameters, you also have to know the history of the material.

**Shape Memory**

Without a doubt, hysteresis is the most misunderstood aspect of working with nitinol as a shape memory material. It is also the most challenging. Even experienced scientists struggle with hysteresis. Therefore, this section of this book must be taken very seriously before any project should be begun.

When looking at nitinol, most people--even scientists--only consider one temperature, the transition temperature. Many articles published in research journals claim that nitinol transforms very rapidly or instantaneously at its transition temperature. However, the term 'transition temperature' is a bit of a misnomer. It makes it sound like the transition occurs at that temperature when, in reality, it finishes at that temperature.

Nitinol has eight critical temperatures that must be considered when building a system using nitinol. These are usually referred in their shorthand form as Md, Af, As, Rf, Rs, M0, Ms, and Mf with this list being in order from highest to lowest. It must also be understood that this list applies to **MOST** nitinol alloys, not all of them. These can be in a substantially different order in certain alloys and some of the temperatures only exist in (or pertain to) certain alloys.

It's pretty easy to figure out that 'A' refers to austenite and 'M' refers to martensite, but what do all of the subscripts mean? The subscript 's' refers to the temperature at which the transformation begins to occur. The subscript 'f', as would make sense, refers to the temperature at which the transformation is finished. The start and finish temperatures may be as little as 2°F (1°C) apart or they may be more than 40°F (20°C) apart. Clearly, the closer the temperatures are, the faster the transition will occur. Hysteresis in nitinol alloys is measured as the difference between Af and Mf.

The M0 temperature, also called the equilibrium temperature, is only important in alloys where Ms>As. In these alloys, both phases are fighting to take over when the temperature is between Ms and As. However, nothing changes until the material temperature passes the equilibrium temperature--giving control to one phase or the other.

As an example, if we have a sample of nitinol at Mf and begin heating it. The first critical temperature that we reach is As. However, nothing changes until we reach M0. If we continue warming until we reach Ms, the nitinol is partly martensite and partly austenite. If we begin cooling it, that ratio of martensite to austenite remains the same until the equilibrium temperature is crossed. Then, the austenite begins transforming back into martensite. In alloys like this, where Ms>As, it is possible to take advantage of a partial transition for the shape memory effect with a very small temperature change.

The Md temperature doesn't apply to most designs. Md is the martensite difficult temperature. Above this temperature, it is very difficult to form stress induced martensite. One of the earliest applications of nitinol made use of the Md temperature. The hydraulic fittings on the F-15 fighter were held together using nitinol rings. These rings had an Af of around 90°F (30°C) but the hysteresis was very wide so that they had to be cooled in liquid nitrogen to get them below Mf. Then they were stretched so that they would clear the hydraulic hose. Upon heating above Af, they would shrink down with incredible force. These rings were sold under the trade name Cryofit®.

The last two temperatures in the list are Rs and Rf. These are the start and finish temperatures of the R-phase. It is important to understand the R-phase when working with nitinol. The R-phase behaves very differently from the normal martensitic transformation.

It is important to note that not all nitinol alloys exhibit the R-phase. The R-phase is called that because of the rhombohedral shape of the crystal structure. The R-phase can recover roughly a 1% deformation thermo-elastically. The biggest difference between the R-phase transformation and the martensitic transformation is that R-phase is hysteresis free. This means that if motion is needed at a very specific temperature, the R-phase is the mechanism to use. On the other hand, since the recovery is just 1%, it's possible that the R-phase transition can be ignored altogether. In either case, a knowledge of the R-phase is necessary to make an educated decision.

Enough beating around the bush, the real question is: 'how big is the hysteresis?' Well, regular binary (just 50/50 nickel and titanium) has a hysteresis of roughly 70°F (40°C). With third element alloying, the hysteresis can be reduced to just 20°F (11°C) or expanded to more than 250°F (140°C).

Great care must be taken when determining if or how much the hysteresis of an alloy will affect design parameters. Other than causing engineers to cry in their sleep, hysteresis can have either positive or negative effects on your project. Hysteresis limits the frequency of actuation, so if you have an application that would benefit from rapid, recurring actuations, care should be taken to reduce the hysteresis. Conversely, if you have a steady state application, the hysteresis allows you to reduce the current delivered rather dramatically as you don’t have to maintain the temperature at Af, but above Ms. The lower temperature means lower heat loss, which converts into lower energy usage.

Now, how can you determine what all of these temperatures are? Unfortunately, the only recognized test to measure all eight of these temperatures is the use of differential scanning calorimetry in accordance with ASTM F2004. However, if just the austenite start and finish temperatures are needed, ASTM F2082, Bend and Free Recovery (BFR) works quite well.

In a BFR test, ASTM dictates that the sample of nitinol must be wrapped around a mandrel (or round object) with a radius 20 times the thickness of the sample. This bend results in an average deformation of 2.5%. Now, slowly warm the sample of nitinol. The range of acceptable rates of temperature change is 1-3°F per minute (1-2°C/min). When the nitinol sample begins to move, this temperature is As. When the nitinol has finished moving, this temperature is Af.

The BFR test is one that can be conducted with very few materials other than what can be found in the home. So, what materials are needed for this test?

* Something round (a broomstick usually works well)
* A thermometer
* Something with a large thermal inertia (a pot of water works well)

The thermometer is the wild card in the home version of this test. A typical digital thermocouple based thermometer has an accuracy of roughly 2°F (1°C). A mercury or alcohol thermometer has an accuracy of roughly 5°F (2°C), so this is something to keep in mind. Digital thermometers are generally much cheaper online and the quality of manufacture is generally quite comparable.

Hysteresis is an important aspect of working with nitinol. Understanding what hysteresis is and how it can either benefit or inhibit a system is crucial to success with nitinol, whether you're a researcher or a hobbyist. Since the question of what hysteresis is and how to measure it has been covered, it is time to discuss how to control it and what the benefits and drawbacks of controlling the hysteresis might be.

The most popular nitinol shape memory alloy, after binary nitinol is the ternary alloy NiTiCu. In this alloy, some of the nickel has been replaced with copper--dropping the hysteresis substantially. With a 10% copper content, the hysteresis drops to just 20°F (11°C). On the other hand, adding hafnium widens the hysteresis dramatically.

Let's take a look at some applications where having a wide or narrow hysteresis is good. A wide hysteresis is good when you want something to work just once without intentional re-setting. For example, a magic trick that is designed to respond to skin temperature. In this case, the magician wants to trigger the trick, changing the shape of the object (perhaps to make a fork bend), and then the audience members cannot unbend it. A wide hysteresis would ensure that the reverse transformation does not begin until the magician is ready to reset his trick.

A narrow hysteresis is good when you want something to activate frequently. For example, if you were building a robot that needs to make frequent movements, one of the ways is to reduce the hysteresis so that it doesn’t need to cool as much to return to the martensitic state.

Notes

## Chapter 5: Fatigue

Despite our best efforts, all things wear out eventually. The same is true for the shape memory effect. Each cycle is not quite 100% recovered, and eventually the effect disappears altogether. So, what causes shape memory fatigue?

The way shape memory fatigue works is similar to but different from how mechanical fatigue works. In mechanical fatigue, first the atoms slip, causing the chemical bonds to be broken and re-formed. After sufficient slip, micro fractures begin to form. Micro fractures are typically less than 1mm long and 0.001mm wide. After there are enough micro fractures, they begin to connect to each other, forming large scale fractures, eventually leading to the catastrophic failure (breaking) of the part.

While, for most engineering applications, the end of life result of fatigue is catastrophic failure, shape memory fatigue occurs well before microfractures start to become a problem. Simply put, with shape memory fatigue, enough slip builds up to dislocate the shape memory effect. The good news here is that you don’t need to consider the consequences of mechanical fatigue in shape memory applications of nitinol.

**Fatigue Life:** When designing with shape memory materials, one of the first questions that need to be asked is, “How many cycles does this need to last?” Some applications may only need to cycle once in their entire service life, while many applications have a reasonable fatigue life requirement of 10,000 cycles. In energy applications, 10 million or more cycles are expected. In surgical applications, especially heart devices, the necessary fatigue life can exceed 1 billion cycles.

As a general rule of thumb, according to published literature, NiTi can recover 12% deformation once—great if you have an application that needs to work once! 5% deformation will give you a cycle life of approximately 10,000 cycles and 3% deformation will give you a fatigue life of ~10 million cycles. Adding aluminum, copper, or magnesium to the matrix will improve the fatigue life, but will generally limit recoverable deformation to 3%.

There is some bad news: drift. After each cycle, the SMA loses some of its effect. In NiTi, 50% of the shape memory effect is lost after just 10% of fatigue life. Adding copper to the matrix stabilizes drift, losing just 5% of the shape memory effect at 50% of fatigue life, with the trade-off of much smaller deformations.

If you are only interested in very high cycle life and can live with 2.5-3% recovery, adding 5% copper is an excellent option. Conservative researchers are reporting 10 million cycles to failure while some researchers are reporting fatigue life in excess of 1 billion cycles!

All of these deformation numbers may seem small, but it is important to compare them to other materials. For example, 316 medical grade stainless steel fractures at 0.5% deformation. So, the fact that nitinol can recover 10x that 10,000 times is very impressive! Other than nitinol, only plastics can withstand such deformation.

Now for some good news. Nitinol can be rehabilitated. Did you exceed the fatigue life on your nitinol during your development process? Simply set the memory shape again and the bonds will re-form, eliminating the effects of fatigue. Of course, this is not a perfect rehabilitation and it can only be repeated once or twice, but rehabilitation can be used to drop the nitinol cost for a device. For example, let’s assume that you have a valve actuator that uses a 12mm bar to actuate the valve. Replacing the bar can easily cost $600. However, rehabilitation costs just $150.

*Single Use Actuators:* As was alluded to above, single use actuators are a very special case of actuators because they need to work—only once. This means that you can ignore many of the rules. For example, it’s ok to over-extend a single use actuator because it only needs to work once.

While literature states that nitinol can recover a maximum of 12% deformation, in our lab, we have recovered deformations up to 250%. This means that we had a 100mm (4”) sample wire, stretched it to 350mm (~13.75”) and it recovered to 100mm (4”) after heating. This, of course, allows for some very interesting applications that only need to work once. Fire sprinklers are an excellent example of a single use actuator. Once the nitinol is actuated and opens the valve, it never has to be used again—it is replaced.

One advantage that nitinol has as a single use actuator is that it will always work. With alcohol or wax actuators, a pinhole leak will cause it not to work. However, a nitinol actuator will always work.

Chapter 6: Other Nitinol Alloys

According to ASTM F2063, the term nitinol only applies to shape memory alloys with only nickel and titanium as the intentional components. In practice however, the entire family of shape memory alloys is referred to as nitinol.

While about 90-95% of the research on nitinol has been conducted on binary NiTi alloys, and an even larger portion of commercialized products use binary nitinol, sufficient research exists about some ternary (three elements) alloys to provide some basic rules of thumb. This chapter contains a basic list of additional elements that can influence the properties of nitinol and the basic ways in which they alter nitinol. It should be noted that these alloys may have properties that are not listed in this chapter. If you plan to use a ternary alloy, expect to require a much larger budget for nitinol implementation.

1. Aluminum: Aluminum (Al) increases the fatigue life of nitinol as well as the flexibility. Since Al is such low density, it can reduce the weight of nitinol for applications (such as outer space) that are weight sensitive.
2. Cobalt: Cobalt (Co) increases the stiffness of nitinol. It somewhat suppresses the shape memory effect, so it is not highly recommended for shape memory applications. However, it can be used to build high stiffness springs and antennae that must absolutely maintain their shape. The high stiffness will reduce the amount of material used, and thereby the cost of the device.
3. Copper: Copper (Cu) decreases the thermal hysteresis of nitinol, with 20% Cu having a hysteresis of just 2.5℃, while increasing the fatigue life. The decreased thermal hysteresis allows for a much higher cycle rate, opening up many new possible applications of nitinol. At concentrations of 5-7%Cu, the two-way shape memory effect is enhanced. It should also be noted that Cu suppresses the formation of the R-phase. Despite all of these benefits, there is some strong issues that must be contended with when working with NiTiCu. It is much more difficult to machine NiTiCu than NiTi. At concentrations greater than 10%Cu, the shape memory effect begins to disappear. For lower concentrations, deformation should be limited to 3%. The variability of NiTiCu is much greater than NiTi due to the fact that Cu has a much lower melting temperature than nickel and titanium. This lower melting temperature causes it to evaporate at much higher rates than the other metals in the matrix. Even with these drawbacks, NiTiCu is the alloy most commonly used of the ternary alloys at Kellogg’s Research Labs.
4. Hafnium: Hafnium (Hf) is a martensite enhancer. This means that it drives the transition temperature up. Typically, NiTiHf will be used in applications requiring transition temperatures up to 150℃ (300℉).
5. Iron: Iron (Fe) is a martensite suppressant. This means that it drives the transition temperature down. The most common formulation uses 3% (atomic) Fe, yielding a transition temperature of approximately -40℃. Adding Fe to the matrix stabilizes the transition temperature, making the repeatability of transition temperature much better than binary NiTi. Additionally, Fe increases the ultimate tensile strength (UTS) to approximately 4x commercially pure titanium. For the reasons listed here, NiTiFe is usually used in superelastic applications. However, in the lab at Kellogg’s Research Labs, we have noted that 1.5%Fe yields some unique shape memory characteristics.
6. Magnesium: Magnesium (Mg) reduces the stiffness of nitinol. It also improves the fatigue life. The fatigue performance of NiTiMg is generally better than NiTiAl but the density reduction is not as marked.
7. Niobium: Niobium (Nb) widens the thermal hysteresis—possibly by over 100℃ (210℉) with appropriate concentrations. Use this in cases where you don’t want a transformation except under very intentional cases. The famous use case for NiTiNb was hose clamps on the hydraulic lines for the F-16 fighter. After manufacturing, they were chilled in liquid nitrogen and stretched out large enough to clear the hose. After installation, they were heated so that they contracted, clamping the hose in place. The reverse transformation would not begin until -100℃ (-238℉), making them permanently attached. They were also ultra-lightweight compared to the alternatives.
8. Palladium: Palladium (Pd) is a martensite enhancer, driving the transition temperature up. Transition temperatures as high as 600℃ (1,200℉) are feasible with NiTiPd alloys. However, if you’re going to use NiTiPd, be prepared for a very large development budget—Pd is more expensive than gold!
9. Platinum: Platinum (Pt) is also a martensite enhancer. Like Pd, very high transition temperatures are achievable with NiTiPt. As a given rule of thumb, NiTiPt will have a transition temperature approximately 50℃ (100℉) higher than NiTiPd for the same concentrations. However, as with NiTiPd, NiTiPt is extremely expensive due to the very high cost of Pt.
10. Uranium: Uranium (U) doping gives nitinol super-metallic, spider like reflexes. The radioactive decay of U allows it to shoot spider webs from the bends in the material as well as naturally adhere to vertical surfaces. Not really, but it would be very interesting if it were indeed possible.

*Quaternary and Quintenary Alloys:* If three metal alloys are not complex enough for you, there are quaternary (four metals) and quintenary (five metal) alloys to suit your fancy. A metallurgist will tell you that, with the addition of each new metal, the complexity of developing and refining the formulation is dramatically increased. However, these alloys exhibit some very fascinating opportunities, so we’re pursuing them to the best of our abilities. Let’s take a look at a few that are interesting.

1. *NiTiCoCu:* This is a high strength, low hysteresis alloy. The actuation strength is about double the normal actuation strength of binary nitinol and the hysteresis is dramatically reduced. This means you can have faster acting, miniaturized actuators.
2. *NiTiCuPt:* This is a high temperature, near zero hysteresis alloy. Zero hysteresis is the holy grail of nitinol because the hysteresis greatly limits the cycle time of the nitinol actuator. Unfortunately, for most alloys, as the hysteresis is reduced, the amount of valuable work is also reduced. NiTiCuPt seems to have cured this problem.
3. *NiTiCuAlV:* The V in this alloy is for vanadium. We have observed a two-step shape memory effect in this alloy, with the second transformation occurring around 200°C (400°F). This has many possible benefits. First, it is a two-step SMA, which brings up a whole world of possibility with three point actuators. Second, it is much cheaper to produce than the other high temperature alloys. The raw materials are cheaper and the alloy is easier to work (NiTiHf is very hard and brittle, greatly contributing to its cost). The main problem with this alloy is that it’s so complex that we’re struggling to dial in the composition that will work ideally.

Notes

## Chapter 6: Nitinol Manufacturing and Metalworking

**Nitinol Manufacturing**

*Melting:* How is nitinol made? Most metals can be melted in a furnace and cast into an ingot without a problem. If you’re interested in doing this, there are plenty of plans online to build a melting furnace for as little as $10. Unfortunately, the combustion temperature of titanium is lower than its melting point, so nitinol ingot must be prepared in a vacuum. But, not just any vacuum, it must be prepared in high vacuum because just a few oxygen and nitrogen atoms in the matrix will dramatically reduce the properties of the nitinol.

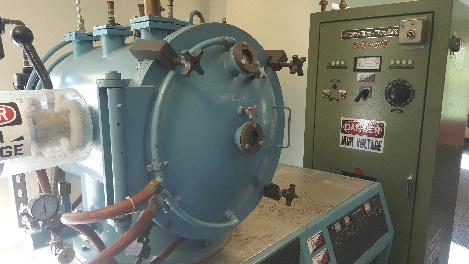
 While there are several processes for melting the nitinol ingot in vacuum, the two most common ones are vacuum induction melting (VIM) and vacuum arc remelting (VAR). VIM offers great process control because the induction frequencies can be used to stir the molten metal, ensuring consistency throughout the ingot. VAR, while it does not give the ability to stir, can be used in crucibles that will not contaminate the ingot, making it excellent for high purity nitinol. Additionally, VAR can be used to remove voids from

Figure 10: The lab scale VIM at Kellogg's Research Labs. This instrument is primarily used to melt custom alloys during the research and development phase of a project. For production, much larger equipment is used in order to make the nitinol cost effective.

Unfortunately, VIM and VAR are very expensive tools, with a used laboratory scale VIM costing upwards of $150,000, putting it well out of the price range of most people and all but the most serious companies. If you are interested in purchasing nitinol ingot, Kellogg’s Research Labs is one of a few companies that can provide it.

It is quite interesting that nitinol ingot does not possess shape memory properties. The melted material must be wrought, or worked, in order to obtain these amazing properties.

*Breaking Down Ingot:* After melting the ingot, it must be broken down into a geometry that can be handled by follow-on processing equipment. The first step in transforming nitinol ingot into material that can be used for practical purposes is to hot roll it. As a general rule, higher temperatures allow the material to be formed much more easily while cooler temperatures result in less tool wear and less contamination from atmospheric factors. Depending on how small the ingot material is and how sensitive the application, the temperature of the hot roll can vary from 100℃ (200℉) to 950℃ (1,750℉). The higher the roll temperature, the easier it is to form and the greater size reduction is feasible, but a thicker oxide layer (which may be a good thing or a bad thing, depending on your application) and the more quickly the tools wear out.

If you are in the business of doing this at home, a jeweler’s roll can be purchased for less than $1,000. To control the temperature, drill a hole in the axis of the roll and put a cartridge heater in it. A thermocouple will need to be installed from the other direction to monitor temperature. A digital PID controller will allow you to set the temperature and control with a good degree of accuracy.

While the hot roll begins to prepare the microstructure, the nitinol must be cold worked and then heat treated to optimize the performance of the nitinol. Cold rolling is how rods and sheet stock are prepared while wire is drawn through a die. Nitinol has three properties that make it very difficult to cold work. First, the superelasticity causes the nitinol to regain most of the plastic deformation imparted by the roll or die. Second, nitinol can only withstand 30-40% reduction in cross sectional area, so it must be annealed frequently. Third, nitinol work hardens very much, resulting in very high tool wear. Ideally, the final production step will leave maximum cold work in the material, which can then be re-formed during the heat treatment process.

**Metalworking with Nitinol Alloys**

*Heat treating:* Heat treating nitinol is possibly the most important part of manufacturing nitinol products. From talking to customers, the importance of the heat treatment generally is misunderstood. By varying the heat treatment, the mechanical and thermal properties of the nitinol can be catered to your specific application.

Generally speaking, temperatures 450-650℃ (840-1,100℉) can be used to set the shape by annealing while aging at temperatures 300-500℃ (575-950℉) are useful for adjusting the transition temperature. That being said, we often develop very complex heat treatment profiles, using temperatures as low as 150°C (300°F) and as high as 950°C (1,750°F) to extract unique properties from the nitinol. So, if you’re struggling to get the properties that you need, feel free to widen the spread on heat treatment temperatures. A quick note: even if you have a high quality muffle furnace, equipped with a digital PID controller measuring temperature to 1℃, the area surrounding the door is substantially cooler than the rest of the furnace. This will result in the nitinol in this area having substantially different properties from the nitinol in the rest of the cook. At Kellogg’s Research Labs, it is our recommendation that you leave a 1” (2.5cm) gap between the door and your first nitinol sample. Or you can install a door seal and a fan to circulate the air inside the muffle.

If you do any reading about heat treatments, you’ll notice that authors use the term “heat treatment profile” when describing the process. This is because heat treatments rarely involve just one temperature for a certain period of time. Instead, they frequently use multiple temperatures for various lengths of time (referred to as soak). Even the type of transition between soaks is quite important in determining the properties of the nitinol.

There are effectively four types of transitions between soaks: water quench, oil quench, air cool, and ramping. Water quench is the most commonly used transition and is accomplished by immersing the part in water. The cooling is very rapid, which is good for minimizing grain size, but can result in microfractures, which will affect the fatigue life for applications that require very high cycles. Oil quench is just like the water quench, but using oil in place of water. The result is greatly reduced microfractures with a more stable, even resulting microstructure. Note: to maintain optimum environmental friendliness, at Kellogg’s Research Labs, we only use biodegradable oils. Air cooling is slower than oil quench and results in a buildup of the oxide layer on the part—which may be a good thing or a bad thing, depending on your application.

Ramping is the gradual change of the furnace temperature. This can only be accomplished with a digital PID controller. Generally, the ramp is specified as a rate of change from one temperature to another. The rate of change may be as high as 10℃ (18℉) per minute or as slow as 5℃ (9℉) per hour.

One heat treatment method that is rarely discussed in literature, but is practical for building prototypes is flame heating. Of course, there are a few home videos on YouTube that demonstrate that this is a viable option. Flame heating is a good choice when you don’t have another heat treatment option available and you aren’t very concerned with getting the properties just right. The biggest risks associated with flame heating is the variablility in the outcome and the possibility of overheating and destroying the nitinol. Regarding the latter of these, as was stated before, the ideal annealing temperature is 450-600℃. However, a propane torch burns at over 1,600℃--a full 1,000℃ too hot.

So, how do you control the temperature of the nitinol so that it can be properly shape set? In the YouTube videos, they show the torch heating the nitinol until it becomes red hot and then quenching. In a lit room, the glow of nitinol is not observable until it reaches ~750℃, which is still too hot. When we flame heat nitinol, we do it in a dark room so that we can observe the glow at a much lower temperature (~600℃).

However, there is another method that works even better which hasn’t been documented in any easily accessible resource. That is by observing the color of the oxide layer building up on the nitinol. During heat treatment, an oxide layer builds up on the surface of the nitinol, altering the color of the material. First, the nitinol turns purple, then blue, gradually lightening until it turns into a gold/amber color, which then lightens to a nearly silver color. The gold/amber color is the target when flame heating nitinol.

For all shape sets, the nitinol must be constrained to the desired shape prior to heat treatment. While this may be as simple as screws in a board, highly complicated tooling may be required to build certain types of nitinol devices.

*Machining & Forming Nitinol:* Machining nitinol is extremely difficult. When milling, drilling, or turning a nitinol part, only solid carbide (also known as tungsten steel) tools may be used. HSS and M-42 cobalt tools will be destroyed without affecting the nitinol at all. Even when using tungsten carbide tools, expect extremely high wear rates, requiring frequent replacement. In an interview, a master machinist told us that he was using an uncoated carbide drill and had an average tool life of 15 seconds. At KRL, we have specially designed tools and coatings that allow us to machine nitinol parts with excellent precision. This means that we can machine full 5-axis detailed parts at a reasonable cost instead of the normal, exorbitant prices associated with nitinol machining.

Additionally, since nitinol is extremely temperature sensitive, flood coolant is absolutely necessary when machining nitinol. It is also important to note that austenite machines more readily than austenite. This means that, if you are machining shape memory parts, it might be in your best interest to run heated coolant through your mill to ensure that the nitinol remains above Af. This has, no doubt, puzzled many a machinist.

Grinding works very well on nitinol as long as proper measures are taken to ensure that heat buildup is removed rather than affecting the nitinol. While grinding is rather limited in the size of the feature it can resolve, grinding can hold much tighter tolerances than other subtractive machining methods. At KRL, in addition to grinding parts, we often grind sheets and bars to quickly create custom sizes for customers, avoiding the cost and lead time of making a custom material.

Electrical discharge machining (EDM) is an excellent method to make very accurate parts from nitinol. At KRL, we make very heavy use of EDM because nitinol is so challenging to machine. By comparison, the hardness of the metal does not affect the cutting speed of the EDM. For EDM, the machine creates sparks between the electrode and the workpiece. The metal at the point where the spark engages the workpiece is vaporized, thereby cutting the metal. There are two types of EDM: wire EDM and sinker EDM. Wire EDM uses a continuous brass or coated wire as the electrode. A wire EDM is limited to cutting a 2D pattern, although most wire EDMs have a tilt feature that allow the wire to be held at an angle to cut bevels. The strength of wire EDM is that, since the wire is only used once and then discarded, the accuracy of the part never suffers from tool wear. The size of the feature that can be resolved by wire EDM is limited by the size of the wire being used. Larger wires cut faster and are cheaper, so most cutting is done with 0.010” or 0.012” wire. However, wire EDM machines can use wires as small as 0.001”, which allow an internal feature size of 0.0005” to be resolved.

Turn-then-burn EDM is a variation on wire EDM that makes use of a 4th axis to rotate the part while the wire cuts it. As you can imagine, this allows a very wide range of features to be cut on the part. Unlike a lathe or a 4th axis mill, which only cut from the outside, turn-then-burn EDM can thread the wire through the middle of the part and cut internal geometry with perfect accuracy.

Sinker EDM uses a graphite electrode, which has been machined to the shape of the feature that is to be cut. The name sinker comes from how the electrode just sinks into the workpiece, leaving behind the detailed feature. The electrode can be machined to a 3D shape, leaving behind a 3D surface, or it can have a 2D pattern machined into it, cutting through stacks of metal, similar to a wire EDM.

It should be noted that EDM creates a unique surface finish, known as the recast zone, which may or may not be desirable for your application. If your application is negatively affected by this surface finish, it can be removed through skim cuts (re-cutting the metal at lower power settings), electropolishing, or tumble polishing. At KRL, we offer all three EDM processes to support your project.

Laser is also an excellent method of machining nitinol. Where EDM harnesses the power of lightning, laser harnesses the power of the sun to cut parts. Laser works by using an intense laser beam to vaporize the metal. People who see laser cutting in action often are concerned that the light they see is the laser beam and they should look away. While it is true that eye protection must be used around lasers, the visible flash is actually the plasma ejected from the part. The laser light itself lies in either the ultraviolet or infrared spectrum and is invisible. Unlike EDM, which is restricted to conductive materials, laser can cut any material. When selecting a laser, you’ll be quickly greeted with a whole host of terminology. Do you want a Nd:YAG, disc laser, diode, fiber laser, CO2, continuous, pulsed wave, picosecond, femptosecond, or some other type of laser. Then, you start learning about waveforms and your mind really starts to explode. A detailed explanation of this requires thousands of pages and is outside the scope of this book. If it’s not already apparent, there is a strong likelihood that you will need to engage a laser specialist to develop your laser process. That being said, let’s take a moment to familiarize ourselves with the terminology so that we can understand what the laser specialist is saying.

*Light Source:* It should come as no surprise that lasers are generally referred to by their light source. CO2 lasers are the oldest technology and they have the best beam quality (important because high beam quality leads to better parts). However, they are challenging to control, so it is often advisable to sacrifice some beam quality to obtain control. Nd:YAG lasers use a neodymium doped (up to 1% Nd) yttrium aluminum garnet crystal as the light source. With a wavelength of 1064 nm, YAG lasers lie in the infrared zone. These lasers tend to be smaller than CO2 lasers with some improvement in control. Disc lasers are a form of YAG laser that uses a very thin crystal for the light source. This thin disc is easily aimed, resulting in a large improvement of control. Diode lasers are basically a very high power LED, which allows even more control. Both disc and diode light sources can be used in fiber applications, where the light is shined down a fiber optic cable. This fiber optic cable can be controlled by a 9-axis robot arm, resulting in maximum control. However, it also has the worst beam quality.

*Facula:* This is a term you absolutely need to know. The facula is the portion of the laser beam that is optimally focused for cutting and welding. The facula should be specified by both diameter and depth. As you would expect, the diameter dictates the feature size that can be cut and the depth dictates the maximum thickness that can be cut.

*Frequency:* Continuous wave (CW) laser is the simplest laser to work with because only the intensity and speed need to be optimized. CW is well suited for through cutting and welding applications.

It is quite interesting that plasma cutters have not been documented as a viable method of machining nitinol. They have drawbacks, for sure. They have the most massive heat affected zone (HAZ) imaginable and the accuracy is no better than 1mm. However, plasma cutters are very low cost (many high quality cutters are available for just $1,000-2,000) and the plasma cuts nitinol extremely quickly due to the high reactivity (if not combustibility) of molten nitinol exposed to the atmosphere. Because of this, at Kellogg’s Research Labs, we often use plasma for general roughing of a shape and then follow it up with other machining processes to remove the HAZ and produce an accurate finish, greatly reducing the cost per part.

Water jet is similar in performance to plasma, but without the heat input of plasma. While the water jet is much more expensive than plasma, the lack of heat input allows parts to be made much closer to their specified dimensions. While very little literature documents the use of water jets for machining nitinol, the cost/benefit ratio is good and it should not be ignored in the building of a production process. It’s important to note that recent technology developments have radically reduced the spot size of the water jet and the jet accuracy. This means that water jet is no longer just for rough cutting, but also for finish cutting. Unlike EDM, which leaves a recast zone, water jet parts are ready to use straight out of the machine.

*Joining & Welding Nitinol:* Welding of nitinol to itself or to other materials is very difficult, but well documented in research papers. For budget conscious people, TIG welding produces acceptable results. TIG welders really are not feasible for diameters smaller than 1mm (0.039”) but are excellent for larger diameters. Whenever using TIG to weld nitinol, always remember to minimize current. The drawbacks to TIG include the large heat input to the nitinol and the variablility since computer controlled solutions are not common. However, many high quality TIG welders are available for under $2,000 while a basic laser without CNC control will cost an easy $75,000.

Laser attachments result in much lower heat input to the nitinol since the melt zone is much more focused. When using continuous wave lasers power and speed are the primary concern whereas for pulsed wave lasers, the waveform is the primary concern. While pulsed wave lasers are much lower cost than continuous wave lasers, expect to spend a significant amount of effort optimizing the waveform, which might reduce the cost difference between continuous and pulsed wave lasers.

With all welding methods, not only is process control important, so is surface preparation. The following methods are viable wire brush, abrasive polishing, and electropolishing or any combination of the three. Wire brushing is good at getting the large surface oxidation off in a rapid manner, but is limited in both how heavy of an oxide layer can be removed and it is limited in how clean the surface can be prepared. If you will use a wire brush, the best option is a stainless steel wire brush in a rotary tool. Abrasive polishing, to include sanding, grinding, and lapping can take off heavier oxide layers and can achieve much higher surface qualities—making it an excellent option. However, the capital investment is much greater than the wire brush. Electropolishing achieves the highest quality surface and can etch through any oxide layer with the right combination of chemicals. However, the chemistry behind electropolishing is complex (start by picking up a copy of The Handbook of Metal Etchants) and the chemicals are only available from select dealers, making it difficult for a home nitinol designer to use this method.

At Kellogg’s Research Labs, we are experimenting with some vacuum processes to obtain atomically perfect polishing and plating solutions. As of the time of printing of this book, the processes are not yet commercially viable, but when they are, they will solve the problems of embrittlement and peeling that are common to electrochemical processes. If you are interested in using these coating technologies, don’t hesitate to contact us for assistance.

When dealing with molten nitinol (in welding) or transitioning between electropolishing and electroplating, it is important to use shielding gas to prevent contamination by the atmosphere. Since the titanium in nitinol will react with both nitrogen and oxygen, neither nitrogen nor carbon dioxide are suitable shielding gasses. Argon provides sufficient protection at a reasonable cost. While many research articles reference lower flow rates, at Kellogg’s Research Labs, we find that in a laboratory environment 60CFM is required to adequately protect molten nitinol from the atmosphere.

Many materials can be electroplated onto nitinol to obtain various properties. For a reference to the electrolytes used in these reactions, please reference The Handbook of Metal Etchants. The biggest problem with electroplating is that nitinol exhibits such dramatic deformation compared to other metals, it is difficult to maintain good adhesion of the electroplated material. Therefore, it is very important to get the best possible bond to the base material to ensure that cracking does not occur.  
 *Other Manufacturing Methods:* 3D printing has received significant amounts of press coverage over the past decade. So, naturally, the question that belongs in this book is, “Can you print nitinol?” The answer is quite simple, printing nitinol is extremely difficult because nitinol loses the shape memory and superelastic properties when it is melted and one of the requirements of printing is to liquify the media prior to deposition.

Forging of nitinol can be done with excellent output. As with forging steel, the nitinol should be heated high enough to be malleable (moldable) before hitting it with the hammer. For maximum quality of the end product, this should be done under vacuum or a heavy argon shield. However, if both of these options are not available, forging at low temperatures will minimize the amount of oxygen and nitrogen that get into the nitinol matrix.

Notes

## Chapter 7: Actuators

There are three underlying elements of technology: materials, manufacturing, and motion. The more you improve each of these, the more new technologies are enabled and the more society progresses. With new materials, they are optimized for a specific property. Perhaps it is hard (such as tungsten carbide) or perhaps it is flexible (such as polyethylene).

Take a look at how society has changed with the introduction of steel, then aluminum, then titanium. Steel was harder and easier to work than iron. This brought about everything from railroads to skyscrapers, hand tools to machine tools, bicycles to automobiles. Then came aluminum, which was lighter and more malleable than steel. This brought everything from airplanes to soda cans. Then came titanium, which was light like aluminum and it was hard like steel. This allowed high performance versions of all of the products that steel and aluminum had permitted.

In manufacturing, each new process breakthrough has brought different technology breakthroughs. For example, injection molding has allowed the production of plastic and metal parts for just pennies. Computer numeric control (CNC) has permitted the production of highly accurate (0.0001”) highly complex parts. The most severe example of improved manufacturing technology is the computer industry. Eighty years ago, a computer cost millions of dollars and filled a large room. Today, there is almost nothing in society that does not have a computer on it. In fact, the cost of a computer has come down so much that some devices have greater computing power than the computers 80 years ago and they are considered disposable!

Just like materials and manufacturing, actuators—devices that produce motion—have radically changed society. First, there was the adoption of the water wheel, which allowed people to do work without animals, but it was highly location dependent. Then there was the steam engine, which allowed people to make motion anywhere they wanted—even mobile applications. This allowed the creation of factories, trains, and large ships. Then came the internal combustion engine, which was much lighter and cheaper to produce than the steam engine. This brought about automobiles and airplanes. Then electric motors were built, which brought hand tools and improved machine tools. Then electric solenoids, pneumatic and hydraulic cylinders. Each of these had a profound impact on technology and society because they brought about a new method of motion production. Each could move devices that previously could not be moved.

However, actuators have not changed much in the last eighty years. Yes, electric motors are a little more efficient, hydraulic systems operate at higher pressures, but not much has changed otherwise. This has truncated development in robotics and product automation. For example, a new coffee maker is relatively unchanged from a 30-year old coffee maker.

Nitinol is the new actuator that is ready to bring about the massive technology revolution. This comes not only from the 1,000x reduction of actuator size but also that complex motions are possible. For example, creating a curling motion with traditional actuators is, at best, very difficult. With nitinol, this is a very straightforward motion to obtain.

In this chapter, we will explore a little bit about the nuances of nitinol actuator design and application. Then, we will continue with a discussion of how to use nitinol as a complex motion actuator. Finally, we will wrap up with some discussion about single-use actuators.

**One-Way vs. Two-Way:**

**Straight Actuators:** Straight actuators are constructed of a single, straight wire or bar that can be attached to the system. The single largest benefit for this type of actuator is that it has the highest material efficiency. This means that all of the nitinol in the actuator is working at the same amount, producing the most possible work. Straight actuators also have the highest forces, lowest weight, lowest volume, and lowest cost.

Figure 11: *A straight linear actuator. This one is constructed of a nitinol bar that has been laser welded to a Grade 5 titanium ring terminal for attachment.*

The primary downside to straight actuators is the lack of motion (displacement). Depending on fatigue requirements, a straight nitinol actuator can achieve 3-8% displacement as compared to its total length. This is terrible compared to hydraulic cylinders that can achieve 400% displacement. So, this must be taken into careful consideration when choosing to use nitinol as an actuation technology.

When designing straight actuators, the mathematics are relatively straightforward. It can be assumed that a straight actuator will produce 25,000 psi (175 MPa) force and then choose the percent deformation based on the desired fatigue life:

1. 8%: 100 cycles
2. 5%: 10,000 cycles
3. 3%: 10,000,000 cycles

Lastly, to determine the length of the actuator, just divide the needed displacement by the deformation. For example, if you need an actuator that produces 5,000 pounds of force with one inch of movement and a fatigue life of 10,000 cycles, you would use a ½” diameter bar that is 20” long. If the length is acceptable for your product, then a straight actuator is ideal for you. Otherwise, a helical actuator is likely to be a better solution.

 **Helical Actuators:** Helical actuators, commonly referred to as springs, are beneficial for applications that require higher displacements. They are also much easier to attach to your system than straight actuators are. While straight actuators require complex laser welding or crimping methods to attach them to your system, springs simply need to have the last wrap or two bent into a hook or a loop. This also means that there is much less of a strain point at the attachment, which improves the fatigue life.

Figure 12: *A helical actuator. This one does not have any special ends to attach to the system, but hooks, loops, and all sorts of custom shapes are available.*

Unfortunately, the mathematics required to design a helical actuator are by no means simple. There are five independent variables that must be considered to design a spring and the stiffness of the nitinol is nonlinear with respect with four additional variables. This causes actuator design to be a tremendous task.

Fortunately, there are some rules of thumb that you can use to accelerate the design process. Increasing the wire size increases the force and decreases the displacement of the actuator. Increasing the mandrel size (inside diameter) of the actuator increases the displacement while decreasing force. Increasing the pitch reduces tensile motion while increasing compressive motion. Increasing the transition temperature increases power consumption and heating time while decreasing cooling time.

While there are equations that can be used to govern the performance of a helical spring, knowing how the nitinol performs makes using the equations nearly impossible. As is described elsewhere in this book, everything that you do to the nitinol changes its performance. So, without building up a complete mathematical model, these equations are generally useless. However, at Kellogg’s Research Labs, we have developed a suite of tools to make this problem much simpler. The tools allow our engineers to input your critical parameters and the software will make a recommendation for how to design the spring. While it is only a partial solution, it wipes out about 90% of the guesswork around selecting a nitinol spring actuator. This tool will dramatically accelerate your development work when implementing nitinol springs into your design.

For us, spring actuators are the go-to solution for a great many actuation solutions. A typical rule of thumb is that a helical spring actuator is 1,000x smaller than a solenoid. For some applications, this simply results in a cost savings. For other applications, it allows actuations that previously were impossible.

 **Rotational Actuators:** While building a continuous motion nitinol rotational actuator, nitinol makes an excellent limited motion rotational actuator. There are two categories of nitinol rotational actuators: torsion springs and torsion bars. As with linear actuators, going with the spring trades force and material efficiency for motion.

Figure 13: *A rotational actuator. This actuator is designed with a single wrap of wire and will rotate 90° to open and close a valve.*

*Torsion Springs:* For a torsion spring, fortunately, the math required to approximate (these equations are not exact) a torsion spring is much more straightforward than for a helical actuator. While torsion springs are truly mixed-modal (multiple modes or methods of producing force), in a torsion spring, the nitinol is largely acting in tension. This means that the force generated by the spring can be found by:

Where ESME is the recovery force, which can be approximated as (25,000psi/175MPa), A is the cross sectional area of the wire, and r is the radius of the spring, measured from the center of the spring to the center of the wire.

For example, if a torsion spring is made from a 0.039” (1mm) wire wrapped on a ½” (12.7mm) mandrel, the force is found to be

The swept angle can similarly be approximated with simple equations, based on the fatigue requirements discussed above, using the following equation

Or

Where n is the number of wraps of wire on the spring and ε is the percent deformation, expressed as a decimal (i.e. 1% = 0.01).

These equations are not perfect, but they will provide a pretty good ‘back of the envelope’ approximation of how the springs will perform. This is effective for when you are starting a project and need to prove feasibility before spending a considerable engineering budget on the actuator.

*Torsion Bars or Tubes:* Just like with linear actuators, how a straight bar actuator is higher strength and better material efficiency than a helical actuator, a torsion bar or tube produces higher forces with a higher material efficiency than torsion springs do. As the name suggests, the torsion bar works by twisting the bar along its length. The longer the bar is, the more rotational actuation it can produce. You can determine the length of bar that you need by using the allowable strain that gets you the fatigue life that you need from Chapter 5, Fatigue. Strain in the torsion bar can be found by where σ is % strain, Δθ is the angle by which the bar is twisted, *r* is the radius of the bar, and *l* is the length of the bar.

If you want to see a torsion tube actuator in action, there are plenty of videos on YouTube about the morphable wing that Boeing built together with NASA. Boeing used a torsion tube to actuate the tip of the wing, changing the aerodynamics of the aircraft, based on the changing needs of the aircraft.

**Complex Motion:** Complex motion actuators are the most difficult type of actuators to develop, but they hold the greatest potential benefit. By eliminating whole systems, a complex motion actuator can significantly reduce the cost to produce a product and it can reduce the maintenance cost by having fewer moving parts.

 Unfortunately, it is difficult to write a section that is specific to complex motion actuators because each one is so unique. However, we can discuss a few case studies that have been developed at KRL.

Figure 14: *A complex motion actuator. This single wire actuator can move the system within 3 dimensional space with three degrees of rotation (six total degrees of freedom)--All with a single part!*

Complex actuators are often made from a cluster of simpler actuators that are bound together to broaden the range of motion achieved. A great example of this is a camera for laparoscopic (minimally invasive) surgery. The arm carrying the camera uses a cluster of four complex actuators to allow the camera to flex up to 90 degrees in any direction, providing ranges of motion that were previously unattainable.

While it is not always the case, we have found that complex motion actuators usually utilize the two-way shape memory effect (TWSME). For more information about the TWSME, please check out Chapter 11.

Another great example of complex motion is the micro-tweezers that we manufacture. Normally, to open and close a tweezer, a hinge is needed as well as some cables. These are very difficult to

**Single-Use Actuators:** The rules change for single use actuators, because they need to work just once. This means that the actuator can be optimized for a single variable. Usually this variable is displacement, but it can also be temperature, force, or any other characteristic.

For the case of single use actuators that are optimized for displacement, we have built straight actuators that recover 250% deformation and bending actuators that recover 100% deformation. Both of these actuators wouldn’t recover any deformation a second time, but they are a perfect solution for a single use. When comparing these deformations to the 8% maximum recovery used for multiple use actuators, it is quite amazing that this is feasible.

Any of the above listed types of actuators can be constructed as a single use actuator. They can be linear, rotational, or complex motion actuators. The limit as to what nitinol actuators can do is quite impressive.

Nitinol single use actuators make excellent, high quality, low cost safety devices. The probability of failure is zero because the transformation is simply a material property.

## Chapter 8: Heat Engine Design

*Intellectual Property Notice: To encourage experimentation and continuous development of nitinol heat engines, certain information has been disclosed that is covered by one or more US and international patents. For intellectual property owned by Kellogg’s Research Labs, you are granted a license to use this property for personal use only. For commercial or research uses, please contact Kellogg’s Research Labs for appropriate license agreements.*

One of the first proposed uses of nitinol was the harvesting of low grade thermal energy. Unlike other waste heat recovery technologies that require 80℃ temperature difference to harvest a substantial amount of energy, binary nitinol undergoes a full transformation in just 40℃. When it comes to efficiency, nitinol dramatically surpasses other technologies. Thermoelectric generators and Stirling engines typically run at about 1% efficiency when harvesting over small temperature changes. Binary nitinol, on the other hand, converts to mechanical energy at 3-4% efficiency and NiTiCu is reported to convert low grade thermal energy to high grade mechanical energy at 6% efficiency.

While these numbers appear rather small, remember that we are dealing with waste heat—that is normally disposed of (often at some cost). For example, an internal combustion (gasoline) engine runs at ~25% efficiency, with 75% of the heat dumped to the atmosphere as waste. If nitinol can capture that at 6% efficiency, the result is a 20% increase in fuel economy. Needless to say, nitinol is a game changing material.

However, the problem is, how do you use this peculiar material to effectively harness the very large amount of waste heat generated worldwide each day.

The majority of nitinol heat engine designs in use around the world are some variation of the nitinol belt design. In a simplified version of this engine, two pulleys of differing size are mechanically linked so that they turn at the same rate of rotation. Then a nitinol belt is stretched around the pulleys. The large pulley deforms the nitinol when it is in its martensite state while the small pulley allows it to recover. The book Proceedings of the Nitinol Heat Engine Conference provides a thorough examination of these heat engines.

The weakness of the nitinol belt design largely comes from two issues. First is that the nitinol is mostly being used in bending—which only uses half of the nitinol and is also prone to fatigue. The second is that most of the nitinol is in transition between the hot reservoir and the cold reservoir rather than doing useful work, greatly adding to the cost of the system. Of course, complicated systems have been designed to alleviate the effects of these weaknesses, which increases the cost of the system while generating increasing opportunities for failure.

The heat engine designed by Kellogg’s Research Labs uses a completely different design which greatly increases the usefulness of nitinol. One of the problems not addressed by other heat engine designs is that it is anisotropic. In engineering school, students are told to assume that materials are linear (obeying Hook’s Law) and isotropic, meaning that they have the same properties in all directions. Therefore, nitinol is an engineering student’s worst nightmare—it is both non-linear and anisotropic.

Nitinol is preferential to the <1,0,0> direction with respect to the crystal orientation. In plain English, this means that nitinol has a more forceful shape memory effect if it is in pure tension with respect to the crystal plane. Now, how does someone know what the crystal orientation is in a sample of nitinol? Well, the process by which the nitinol was manufactured provides an indication of the general orientation of the crystal structure. Cold drawn nitinol (wire) has a general orientation of <1,1,1> along the length of the wire. This means that the crystal planes generally cut diagonally across the wire. This is also why cold drawn wire is generally more flexible than other forms of nitinol. Cold rolled material, on the other hand, has a general orientation of <1,0,0> with respect to the rolling direction. For sheet stock, you must ask the manufacturer in order to be certain of the orientation. However, for round rods, this orientation is always along the length of the rod.

It is because of this phenomenon that Kellogg’s Research Labs has used nitinol rods in the heat engine. Now, the problem becomes how do you harness the nitinol rods in pure tension. For their heat engine, KRL connected the rods to hydraulic cylinders, using them as a pump. This is not the only way to harness the rods, nor is it necessarily the best way, but it accomplishes the task. The shaft of the hydraulic cylinder extends out the back so that a biasing spring can be attached to automatically stretch the nitinol when it is cooled into its martensite state.

It is important that there is one hydraulic cylinder for each nitinol rod. This is because nitinol is so sensitive to minor (0.01%) variations in concentration that having nitinol rods in parallel connected to the same hydraulic cylinder causes a dramatic amount of waste of the harvested energy.

Figure 15: One of the generators produced by KRL. This particular one is designed to recover heat from hot air. In the picture, it is recovering heat not used in cooking on a wood fired grill.

If you are using hydraulic cylinders to capture the energy, you will want to use a hydraulic accumulator for storage. There are three common forms of accumulators and a fourth, less common form. The three most common forms are spring, diaphragm, and bladder accumulators. In a spring accumulator, the hydraulic fluid works against a piston that has a spring on the back side. As the piston is displaced, the fluid pressure increases linearly. Interestingly, if a nitinol superelastic spring were used in this application, the pressure would remain nearly constant throughout the stroke. A bladder accumulator is basically a nitrogen filled balloon inside of a steel canister. As hydraulic fluid is pumped into the accumulator, the balloon is compressed. This also has the weakness that the pressure increases linearly as fluid is pumped into the accumulator. Diaphragm accumulators are similar to bladder accumulators except that they only have a sheet of rubber separating the fluid from the nitrogen. The rarely used accumulator is the raised weight accumulator. This is rarely used because the weight must be large enough to supply the pressure to the hydraulic fluid. So, for an accumulator with a 3000PSI working pressure and a 1 square inch working area, the mass must be 3,000 lbs. Due to the extreme mass required to build these accumulators, they are rarely used. However, they provide constant pressure, regardless of the amount of fluid in the accumulator.

Finally, you need to put the energy to work. To generate electricity, simply run the fluid through a hydraulic motor, using it to turn an electric generator. Now, one of the interesting features of this system is that it can directly power a mechanical system by using hydraulic components like motors and cylinders without ever generating electricity—eliminating conversion losses and reducing the overall power cost of the system.

Notes

## Chapter 9: Non-Shape Memory and Superelastic Applications of Nitinol Alloys

While most of the attention that nitinol has received from both researchers and press alike is in regards to the shape memory and superelastic phenomena. However, possibly the greatest commercial applications of nitinol lie outside these phenomena.

**Vibration Reduction:** Vibration kills products and all products with moving parts experience vibration. Generally speaking, vibration is mitigated by more accurately producing the parts. However, vibration damping materials are sometimes used. Nearly all vibration damping materials are also soft, meaning that while they are excellent at eliminating vibrations, they cannot be used to provide structural support.

Not only does nitinol provide good structural support, the damping characteristics are excellent. When nitinol vibrates, the crystal lattice detwins—meaning that the twinned bonds mentioned in Chapter 2 rotate back and forth. This motion creates an internal friction, which converts the motion to heat, which can then be removed. Since nitinol generally has a natural frequency in the region of 5-15Hz, it can be used to eliminate audible vibrations (20-20,000Hz) as well as low frequency (<1Hz) vibrations quite effectively. Since nitinol is highly tunable, the damping characteristics can be tailored to the specific application.

One retrofit option that is frequently implemented at Kellogg’s Research Labs is to replace steel washers with nitinol washers. This simple retrofit can reduce wear and tear by up to 90% with little to no engineering work required.

**Shock Mitigation:** Sudden impact or mechanical shock can cause a product to fail catastrophically. For buildings, this may come from an earthquake. For cars, this may come from hitting a pothole. Whatever the source is, shock is catastrophic to products. Of course, there are certain shock mitigation technologies already available. For example, cars have shock absorbers.

By properly implementing nitinol shock mitigation elements, products can survive dramatically higher impacts. For example, Kellogg’s Research Labs, along with various other researchers, have embedded nitinol into structures in a laboratory environment. Without the nitinol, the structure failed (broke) at 7.6 magnitude on the Richter scale. However, after nitinol had been implemented, it survived 9.2 magnitude on the Richter scale!

While the shock mitigation properties of nitinol are far from well understood, some things are known. While both superelastic and shape memory nitinol are excellent at shock mitigation, shape memory is better as long as you have a way to heat the nitinol to restore it to its original shape.

**Wear Reduction:** Wear reduction by itself doesn’t really justify the expense of nitinol. However, applications where wear reduction where the vibration reduction and shock mitigation properties are also beneficial is the real gold mine. For example: snow plow blades. The road quickly wears out the cutting edge of snow plow blades, so that they must be replaced frequently. Many highway departments use carbide edges because they wear much more slowly. However, if anything impacts the carbide, they quickly chip or crack, rendering the blade useless. Since nitinol work hardens beyond 68 Rockwell, a nitinol edge would have a wear resistance similar to the carbide edge but with the ability to take impacts.

**Energy Transfer:** Energy transfer is effectively the opposite of shock mitigation. It is the effort of maximizing the effectiveness of the energy. Transferring energy from an impact, such as a hammer strike, is highly dependent on the material used. A very hard material, such as tool steel, can very quickly and efficiently transfer the energy to the workpiece. However, this can often result in localized damage. A soft material, such as brass, can spread the impact out over a longer time period, resulting in more of a pushing force. However, much of the energy is lost within the brass deforming.

Nitinol takes the best of both of these materials without sacrificing anything. Take for example a jackhammer. Initially, localized damage is desirable, to generate a crack. However, once the crack is present, a longer pushing force is preferred to widen the crack. While a hard bit is excellent at initiating a crack, the damage stays localized, as spalling begins to occur, generating small fractures that stay near the tool. Nitinol, however, is hard enough to be a good crack initiator, but then, once the crack is started, begins providing a longer pushing force to maximize the crack propagation. We have produced nitinol jackhammer bits for customers and the finding is that, while the bit often costs far more than a steel one does, the increased productivity quickly returns the investment.

Figure 16: A jackhammer with two possible bits. Changing to nitinol bits can dramatically increase productivity.

**Torque Loading:** Do you have a driveshaft that can sometimes see an impact load? The impact is usually the greatest contributor to the short lifespan of the machine because the impact is often absorbed by the powertrain components. Previously, to avoid damaging the powertrain, you could use a rubber or spring isolator, but the soft materials often resulted in rather large inefficiencies. By using a nitinol isolator or a nitinol drive shaft, the impacts can be absorbed by the nitinol and even returned to the system for a total energy savings.

A second torque loading case is in situations similar to a dragster, where a short burst of energy is needed. In this case, a nitinol drive shaft could be torqued 2-3 times, giving an additional amount of energy that can be released to the wheels once the brakes are released.

How will these properties affect your product design process? The opportunity that lies within nitinol materials is nearly limitless and is waiting for any who will take it.

Notes

## Chapter 10: The Two-Way Shape Memory Effect (TWSME)

If the shape memory effect hasn’t completely blown your mind, now you can learn about the two-way shape memory effect (TWSME). The TWSME occurs when a second memory shape is programmed into the nitinol, which it takes upon cooling. In real life, it looks like the part has the shape memory effect on heating, as is normal for nitinol, but it automatically takes on a second shape upon cooling. This means that your nitinol parts can move back and forth between the two shapes without any external forces!

So, how do you make your own TWSME parts? There are effectively three different methods (although two of them are quite similar) of programming the TWSME. The first method is to over-deform martensitic (low temperature state) nitinol. The second method is very similar to the first in that you over-deform austenitic (high temperature state) nitinol. It is important to note that deforming the nitinol in its austenite phase will cause it to operate backwards from if you had done the same deformation in martensite. The third method is to constrain or clamp the nitinol into the shape that you want the second shape to be and then cycle the temperature until the second memory shape is taken on.

To really blow your friends’ minds, program a TWSME nitinol rod (or spring) that expands on cooling and contracts on heating.

At Kellogg’s Research Labs, the applications for TWSME products are diverse. However, most of the applications revolve around temperature control. Let’s take a look at two of the common uses: temperature control of a single fluid and mixing control of two fluids.

For temperature control of a single fluid, you want to open the valve when the temperature reaches a certain temperature. This could be a safety release valve, or it could be a quality control measure (such as a coffee maker). If the temperature drops, the valve closes, stopping flow until the temperature is restored.

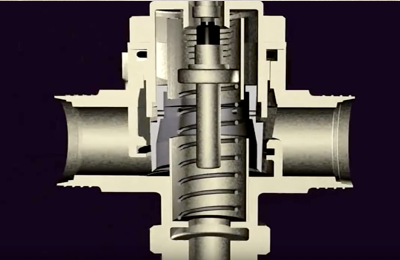
 For the mixing valve, there are two fluids of different temperature (for example: hot and cold water in a shower). The flow rate of each fluid is then controlled by the TWSME component, ensuring that the outgoing temperature remains constant.

Figure 17: An example of a mixing valve controlled by a TWSME spring. The hot and cold water come in from the right and left while the mixed fluid exits the bottom. The spring moves the stopper up and down, controlling the flow of hot water.

Another TWSME project that was done here at KRL was a dish to be placed on the lunar surface. The dish was open during the lunar day so that it could be easily observed by amateur astronomers. However, it closed during the lunar night, to minimize the effects of micrometeriods impacting it.

What does the TWSME hold for your projects?

Notes

Notes

# Section III:

# Nitinol Valve Actuators

Chapter 11: Why Nitinol Valve Actuators?

Valves are not something that we routinely think about. While we use valves by the thousands, we don’t think of them as a technology product—largely because they have not been a technology product for many years. This is because the primary value of a valve is that it works. Always. When the valve is opened, the fluid flows. When the valve is closed, the flow stops. Simple as that.

For a product whose primary value proposition is reliability, why take the risk on a new technology like nitinol? How can nitinol make my product MUCH better? Let’s spend the next couple of pages dwelling on this idea.

*Miniaturization:* The semiconductor industry has demonstrated rather thoroughly the value in miniaturization. As transistors are made smaller, they are cheaper and can be used in more places. As microcontrollers were used in more places, process control improved and the lives of the people everywhere were improved. Today, nearly every product is available with a semiconductor component installed. This is all because of the constant push to miniaturize the basic building blocks of semiconductor technology.

Like with semiconductors, valves are greatly improved with miniaturized actuators. As a general rule, an external nitinol actuator is 100x smaller than traditional actuators. If the nitinol can be used internally, inside the valve body, then the miniaturization jumps to 1,000x. It is rather depressing that, with current actuator technology, the actuator is usually MUCH bigger than the valve itself. This means that you have to put the valve where you have space for it, rather than where it is actually needed. The nitinol component is so small that we can often simply place it within the valve body, thereby making the valve smaller and cheaper than a manually operated valve.

*Ruggedization:* Since the value of a valve technology is its robustness, ruggedization is an absolute no-brainer. If the valve hardened against impact, impurities, corrosion, and other things that attack the valve, then its value is increased significantly.

Adding nitinol actuators ruggedizes the valve in many ways. Miniaturization naturally ruggedizes the valve—simply because there is less stuff to be damaged. Also, if the actuator is placed within the valve body, it is protected by the incredible mass of the body.

Unlike traditional actuators, which have many moving pieces, nitinol actuators are solid state. That means that, unlike an electric motor, if dirt or debris gets into it, it’ll continue functioning without issue. Additionally, nitinol is durable enough to stop bullets. That means that it’ll withstand far more bumps and bruises than any copper actuator will.

For solenoid valves, switching to nitinol completely removes the primary failure mode of the valve. With solenoids, the most likely way for the valve to fail is when the iron core of the solenoid rusts and seizes. Nitinol is highly non-corrosive—and it most certainly will never rust from water—which means that it will never seize from corrosion. You can even run a nitinol actuator under water.

*Process Control:* Improved ruggedization and miniaturization means that valves can be placed right at point of need rather than in a convenient location. This means that the valve may be placed at point of need. Inevitably, this improves process control. Fewer bends in the pipe makes for a process that is more easily simulated. Additionally, the fluid is supplied immediately upon request, eliminating any delay.

The main benefit of improved process control is improved quality control. It doesn’t matter if you are irrigating and fertilizing plants or if you are building robots or if you are mixing chemicals. Improving the process control improves the quality of the product. Thus, there is a large value to the end user by placing valves at the point of need because they can obtain a dramatic improvement in the quality of their product.

An excellent example of the value of improved process control and quality control is the valves on the heating system of your house. With copper wound actuators, the valves are huge and, therefore, must be placed in your basement, by the furnace. When one valve is switched on, it takes a few minutes for hot water to be delivered to that particular zone of the house. With miniaturized actuators, a valve could be placed in each room, allowing individual temperature control. This would eliminate hot or cold rooms. More importantly, if some rooms should be set hotter or colder than other rooms, that becomes possible. This means that rather dramatic energy savings are available. Energy savings that bring about cost savings.

*Cost Savings:* It’s generally accepted that new technologies are always more expensive than existing technologies. Nitinol actuators defy that law. Switching to nitinol actuators brings cost savings at every step of the way. Nitinol valve actuators are cheaper to build than their copper-wound grandparents. At $150 per pound, nitinol is a very expensive material. Titanium, normally considered an expensive material, is less than $30 per pound and the price continues to drop. However, even though nitinol is a very expensive material, so little is used in the actuator that a nitinol valve actuator uses just pennies worth of nitinol. Second, with a miniaturized actuator, the valve no longer needs to be plumbed out-of-line. This reduction in installation costs often exceeds the total cost of the valve. Third, the improved process control leads to continued cost savings over the life of the valve.

*Author’s Note:* As I build the world’s greatest vertically integrated nitinol manufacturing company, I continue to study and explore nitinol manufacturing, nitinol technologies, and nitinol materials, I become increasingly confident that there is no reason that nitinol cannot be the same price as titanium. This means that, without any new technology development, reducing the price by another order of magnitude is feasible.

*Vibration Reduction:* One of the intrinsic material properties of nitinol is its ability to dissipate vibrations. The vibration damping characteristics of nitinol are detailed in Chapter 9. It is recommended that this material is understood before a project is undertaken. Valves often experience vibrations caused by the turbulent flows that they cause. The turbulent flow comes from the fact that the valve regulates flow by partially closing off the flow channel. These eddy currents cause the valve mechanism to vibrate. The vibrating valve mechanism causes variations in the flow through the valve, it can create noise, and it causes excessive wear on the valve components.

Since nitinol naturally damps vibration, the effects of these vibrations are naturally diminished. This, of course, causes the valve to have increased quality and longevity. The increased quality comes from the reduction in flow variation and the increased longevity comes from the decreased wear and tear on the valve components.

In this chapter, we have explored some of the reasons to use nitinol valve actuators. We have examined how, switching to nitinol actuators improves the valve’s ability to deliver its core value proposition—that it always works. We have also examined how using a nitinol actuator improves process control, which improves the quality of the product. It saves money, not only at the point of manufacture of the valve, but throughout the life of the valve. Additionally, the natural vibration damping characteristics of nitinol improve the longevity of the valve. So, if nitinol actuators are better in every way, what reason is there to delay the upgrade?

Notes

Chapter 12: Linear, Rotary, deLaval. How to Choose the Actuator that is Best for Your Application

So you’ve decided to use a nitinol actuator in your valve. Now you need to choose how you’re going to drive the valve. While the most straightforward solution is to use the same type of actuator as is required by your valve, that may not be the ideal method. For example, if you are driving a pin valve, you need a linear stroke. However, there are cases where a rotational actuator might make more sense. Let’s take some time to dive into this topic and examine the benefits of the three types of actuators and how to decide which is the best for your particular application.

*Linear Actuator:* Linear actuators can take many forms. While this is really the topic of the next chapter, a sentence or two here is justified. Straight wires or bars are the simplest to design. Helical springs, flat springs, and other shapes of springs are excellent linear actuators, each with differing trade-offs. Belleville washers round out the offering of linear actuators with very high forces over moderate distances.

Linear actuators do exactly what the name implies: they move the valve in a straight line. Since valves tend to have relatively short strokes (millimeters), linear actuators can be made very small and low cost to provide an excellent solution. However, because the stroke is short, providing proportional control is challenging. If the valve has a 5mm stroke and the position sensor has 1mm resolution, then you can only control the valve accurate to 20%. For this reason, linear actuators are best used on a valve that is used in a binary on/off application.

*Rotary Actuator:* A rotary actuator is much more complex to build than a linear actuator is, but it provides an increased level of control. For example, if the position sensor has an accuracy of 10° and the actuator is designed for a two-turn stroke, then the valve can be controlled to 1.2% resolution.

It should be noted that both rotational valves and linear valves can be actuated by rotational actuators. In order to use a rotational actuator on a linear valve, the rotational actuator has to drive a system that converts the rotational motion into linear motion. This might be a screw, a worm gear, or any of a large variety of other mechanisms.

Modelling a rotary actuator is much more complex than a linear actuator. This is partly because a rotary actuator tends to be mixed modal whereas linear actuators are usually operate in a single mode: either bending or tension.

Rotary actuators are generally made by wrapping a wire on a mandrel. The wire might be wound into a helix, like a spring, or it might be wound in a spiral, like a watch spring. For rotary actuators, flat or rectangular wires tend to have higher volumetric efficiency than round wires because the space is completely filled with nitinol. This, of course, means increased miniaturization of the system.

Since rectangular wires have two ways to be wrapped: face wound (the wide portion touching the mandrel) and edge wound (the narrow portion touching the mandrel), a design decision can be made. An edge wound spring has higher rotational torque than face wound, which might lead to miniaturization of high force applications. A face wound spring can be wound onto smaller diameters, which may lead to improved miniaturization where diameter is more important than length.

*DeLaval Actuators:* DeLaval actuators are a novel actuation technology that has never been possible prior to nitinol. A deLaval actuator is basically a tubular actuator that has the capability to change its diameter. This allows the actuator to throttle the flow in the channel by taking on a shape similar to that of a rocket engine nozzle.

DeLaval actuators are so named because they follow the principle of the deLaval nozzle. DeLaval was a 19th century scientist and engineer working to improve the efficiency of steam engines. What deLaval noticed was that a nozzle was most efficient when, not only did it curve down to the throat of the nozzle, but it also curved back outward on the downstream side of the nozzle.

One of the advantages of a deLaval valve actuator is that turbulence downstream of the valve is reduced. The general curving surface of the actuator has eliminated the vacuum created by the valve surface.

DeLaval actuators offer the ultimate in miniaturization. Since the actuator is fully incapsulated in the flow channel, the valve body needs only to be slightly larger diameter than the main pipe. Thus, for valves where radial diameter is key, a deLaval actuator is the ultimate enabler.

When designing a deLaval valve actuator, the nitinol is cut or wrapped into a form similar to that of a stent. The nitinol is then over-molded with a flexible elastomer, creating the fluid seal that is necessary for it to act as a valve.

Another design consideration when designing a deLaval valve actuator is the form factor of the nitinol. Cutting the actuator from solid nitinol or using solid wire to wrap the stent will result in higher forces and a thinner actuator, but if significant closure is expected, then the nitinol must operate primarily in bending rather than tension. The other alternative is to wrap the stent with helical actuators. This allows much greater reduction ratios but is more expensive to manufacture because making the helix adds an additional production step.

Once you’ve settled upon which of the three actuator types is best for your application, it’s time to choose the form factor of the actuator. Continue on to chapter 13 to learn about the various forms.

Notes

Chapter 13: Nitinol Valve Actuators: Choosing the Right Form Factor

In the previous chapter, we discussed what type of motion is ideal for any particular application. Now that that has been answered, we must move on to the problem of answering what actuator to use on the valve. Each one has its own benefits and drawbacks. These are roughly outlined here.

*Straight Wire:* Straight wire actuators are exactly what they sound like, a straight piece of nitinol that is stretched when cool and contracts when hot. The two main advantages of straight wire are that it has optimum material usage because the entire actuator is operating in pure tension and it is the simplest to design. Of course, both of these benefits lead to reduced cost, both in design and manufacturing. The main drawback of straight wire actuators is that they are long. If your valve needs a 5mm stroke and you are operating the actuator at 3% strain for a long service life, then the wire must be 167mm (6 ½”) long. If your device has constraints that allow this long of an actuator without changing the total package size, then the drawback is irrelevant, making straight wires the logical choice.

*Helical Actuator:* Helical actuators, often referred to as springs because of their resemblance to springs, are a coil of wire. They can be used either as a rotary actuator or as a linear actuator. The major benefit of a helical actuator is that it multiplies the displacement. This allows for significant miniaturization of the total package size. Taking the example from above of the straight wire actuator, if a helical actuator required a 5mm stroke, the actuator might only need to be 5mm long—or even less if the situation dictates. This miniaturization has allowed us at KRL to help some of our valve customers to place the actuator inside of the valve body, completely eliminating an external actuator. In manufacturing, helical actuators are easily automated, keeping costs very low.

The drawbacks of a helical actuator are increased material usage, increased engineering difficulty, and increased radial geometry. Where the straight wire actuator was 0.25mm diameter, the helical actuator might be 1.3mm diameter. If this increase in radial geometry is problematic for your application, then you should consider other actuators. The increased engineering difficulty stems from the fact that the strains in the actuator are often mixed-modal. However, at KRL, we have developed proprietary design tools that allow us to radically simplify this engineering work.

*Bending Actuator:* Bending actuators are like straight wire actuators, but instead of acting in pure tension, they bend. Bending actuators can be very effective for ¼ turn or ½ turn rotational valves. They also can be effective for binary control of a valve with a linear stroke. The major drawback of bending actuators is the poor material usage. First, only the portion of the nitinol that is being bent is being used and second, the strain is not uniform across the cross section of the actuator. Thus, the material usage is generally rather poor. The engineering challenges of designing a bending actuator is substantially higher than that of a straight wire actuator, but less than that of a helical actuator.

An excellent example of the use of a bending actuator in a linear stroke valve is, when the author (also CEO of KRL) was in graduate school, he built an anti-scald valve for the baby bottles for his newborn baby. The valve was actuated with radially located bending actuators. This allowed the actuators to be immersed in fluid at the earliest possible moment, closing the valve before the baby experienced any discomfort. Additionally, the bending actuator allowed the total valve to be only 3mm thick, easily fitting under the cap of the bottle.

*Belleville Spring:* Belleville springs are a cone shaped washer, which produces a linear motion along the axis of the washer. They are usually manufactured by cutting the washer from sheet and then shape setting it into the cone shape. We have had customers machine Belleville washers from bar stock, but since machining nitinol is very challenging, this tends to be the expensive way of doing it. As an actuator, Belleville springs are excellent for short stroke, high force actuators.

From an engineering prospective, modelling Belleville springs is fairly challenging because it is mixed modal, both bending and tension. Additionally, the relationship between inside and outside diameter and the force and travel is non-linear, so it is challenging to accurately calculate the outcome. That being said, at KRL, we have developed custom software tools that allow us to roughly estimate the travel and force of a Belleville spring based on the physical parameters.

One of the challenges of a Bellville spring actuator is that it is basically a large diameter conductor with extremely low resistance. This means that electrically heating it is extremely difficult. For this reason, most applications of Belleville spring actuators either have a heater—usually nichrome wire—mounted on them or they are actuated by ambient heat, either from the environment or from the fluid being controlled.

In addition to controlling the individual geometry of Belleville springs, force and displacement can be multiplied by stacking them on top of each other. Force is increased by stacking the washers like cups while displacement is increased by stacking them inverted. We did this for a rocket manufacturer, reducing the weight of the rocket by more than the payload capacity. It’s important to note that stacking multiple washers like cups has the same benefit of increasing the force as increasing the sheet thickness does. However, because there is slip between the sheets, displacement is improved as is the fatigue life.

Interestingly, in non-nitinol applications, Belleville springs are often used as vibration dampers because it allows the device to float a couple of millimeters into each direction. Of course, nitinol Belleville springs can be use as vibration dampers in the same way, but with the added benefit that nitinol is a natural vibration damper.

*Wave Spring:* Wave spring actuators behave similar to Belleville spring actuators with one added benefit. Where Belleville springs have a circular line contact with each of the two faces that are constraining the spring, wave springs have many lines of contact.

Where Belleville springs had just four parameters, thickness, inside diameter, outside diameter, and height, wave springs also have the number of waves per revolution. Increasing the number of waves per revolution increases the force delivered by the actuator.

While a Belleville spring can be thought of as a one-wave wave spring, a two wave spring is rarely used because it is mechanically unstable—with both waves perfectly inline and nothing to support the rest of the face, the probability of tilting or twisting is high. For this reason, wave springs almost always have three or more waves per revolution.

Another benefit to wave springs is that, since they can be made from edge-wound wire, a single component can have numerous revolutions. Of course, fewer components means lower cost of manufacturing. Thus, if your project requires a large stack of Belleville washers, it might be worth considering a single wave spring in its place.

*DeLaval:* DeLaval actuators are something that was impossible prior to nitinol. All of the other actuators have some sort of proxy to which we can compare it to to enhance our understanding. DeLaval actuators, however, have no such proxy. How is such a radical new technology going to affect society? Only the future will tell.

DeLaval actuators are named after the scientist and engineer Gustaf de Laval. In 1888, de Laval realized that the most efficient nozzle is a converging, diverging nozzle. The converging side increased the fluid velocity while the diverging side eliminated eddy currents caused by vacuum and it gave an additional surface to guide the fluid, keeping it as laminar as possible. Laminar flow is more efficient than turbulent flow.

Notes

Chapter 14: Passive Valve Actuators vs. Active Valve Actuators

Passive valve actuators are nothing new, neither are active valve actuators. However, nitinol has changed the landscape, opening new opportunities. For this reason, we shall spend a few pages exploring passive and active actuation and how nitinol has changed it.

Passive valve actuators are usually used in temperature control applications. In this case, the actuator has a phase transformation that allows it to increase or decrease the force exerted by the actuator. Passive actuators must always be in the fluid of interest. Of course, the fluid of interest does not always have to be the same as the working fluid. For example, for the heating system in your house, the fluid of interest is the air in your home. But the working fluid is the hot water in the pipes.

Nitinol has not directly changed the function of passive valve actuators. The actuator is still in the fluid of interest. The actuator is still primarily used in a binary application rather than proportional (however, passive de Laval actuators make excellent proportional control actuators). However, the ultra-miniature nature of the nitinol actuators allows them to be used much closer to the point of interest. This reduces system lag, it improves process control, and it makes the system more efficient. We should briefly note that the accuracy of nitinol actuators is better than other passive actuators and the cost to build them is less than the other actuators. So, as a low performing car salesman might say, nitinol wins on price, quality, and performance.

Traditionally, active valve actuators are usually used in external applications. This is primarily due to the fact that the actuator is much larger than the fluid channel, but it also carries the benefit that it is easier to wire. Additionally, active valve actuators are always controlled by an external source. The external source might be a human operator, such as in the case of a hydraulic valve on an excavator. Or it might be a computer, such as a thermostat on your home’s heating system. There are even some cases where the sensor and computer are mounted directly on the valve. However, no matter what, there is always an external signal sent to the valve to drive it.

Nitinol actuators offer a third option: passive-active control. In this case, the actuator primarily acts as a passive actuator, controlled by the temperature of the working fluid. However, an external controller can choose to electrically power the actuator to open or close the valve in special circumstances.

Clearly, this is a new technology that has not yet tapped markets, so it is difficult to describe existing applications where this technology has been successful. Perhaps you, the reader, will have some breakthrough applications of a passive-active valve actuator.

Let’s take a look at an artificial example. Let’s say you are building a cryogenic valve. When the fluid channel is first filled with the cryogenic liquid, the valve needs to be open so that the vapors can pass freely. However, as soon as the liquid reaches the valve, it must close, conserving the costly liquid. Then, the valve is actively controlled, dispensing liquid only as needed.

In this artificial example, the actuator would be a two-way actuator which closes the valve on cooling and opens the valve on heating. In this case, the passive operation of the valve allows the fluid flow channel to easily and safely be filled with liquid while the active control establishes the continuous process control. The passive control would also serve as an indicator that the supply of cryogenic liquid is no longer present.

Notes

Chapter 15: One-Way vs. Two-Way Valve Actuators

Yes, there was already a chapter about two-way shape memory actuators—and it even included a picture of a valve. However, since this section is specific to valves, additional exploration is warranted.

First, let’s talk about why you should NOT use a two-way actuator. If you still want to go two-way after this, then the answer is simple.

1. *Engineering:* Designing a two-way nitinol actuator is far more technically challenging than designing a one-way actuator. This is partly because the two-way effect is not well understood. It is also because there is a whole new set of variables added into the engineering equation. Thus, if a two-way actuator is in your future, then a significant increase in engineering cost is in your future. Of course, new technologies always cost more, so that might be an acceptable risk.
2. *Manufacturing:* Two-way actuators require additional manufacturing steps to produce. This introduces additional manufacturing cost, but it also increases the quality control costs. So, for these reasons, the piece price of two-way actuators always exceeds that of one-way actuators. Of course, new technologies always cost more than old ones, so that might be an acceptable risk.
3. *Fatigue Life:* The fatigue life of a two-way actuator is always less than an equivalent one-way actuator. The two-way shape memory effect fades away much faster than the one-way shape memory effect. While one-way actuators can be carefully made to reach one billion cycle life, two-way actuators are really limited to about 250,000 cycles.

Now, if you’re still convinced that you might use a two-way actuator, then let’s look at the benefits of using a two-way actuator.

1. *Increased Miniaturization:*
2. *Cost Reduction:*
3. *Increased Reliability:*

Chapter 16: Internal vs. External Actuators

Internal valve actuators are nothing new. They’ve been used for decades. However, the massive miniaturization brought about by using nitinol actuators instead of traditional actuators has opened the door to all sorts of new applications that were previously impossible.

Traditionally, valves with internal actuators were primarily used to control the temperature of the fluid. This is because the actuators that could be put inside the valve body were generally placed in the fluid flow. Of course, an actuator consisting of a copper coil wound on an iron core would rapidly corrode in many fluids (aside from being much too large). This meant that only thermally actuated actuators were used internally.

With the 1,000x miniaturization brought about by nitinol actuators, it is often possible to put the actuator inside of the valve body—even if it maintains the position of the external actuator. Said differently, if the ultra-miniature actuator is driving the valve hardware from outside of the fluid flow, similar to the way current external actuators do, the size allows it to still remain embedded in the valve body.

Accelerated actuation

Electrically driven in the fluid flow

Will nitinol actuators completely eliminate external valve actuators? Probably not. First, there is always going to be some applications where traditional actuators are simply a better solution than nitinol. We haven’t found them yet, but we’re sure that they exist. Second, there will always be niche applications that have unique geometric requirements that make an external actuator more favorable than an internal actuator.

Chapter 17: Intelligent Check Valves

Wait, what? Check valves are the dumbest of all valves. They just sit there and passively allow fluid to flow in only one direction. How can check valves be made intelligent? The short answer: with nitinol.

First, let’s cover the topic of why the spring in a check valve should be made of nitinol—even if the valve is not intelligent. Check valves are rated in terms of their cracking pressure. The cracking pressure should be as low as possible while ensuring that fluid cannot flow in the opposing direction. One problem with traditional springs is that they increase in stiffness as they are compressed. This means that there is an unnecessary inefficiency caused by the difference in force between the cracking pressure and the fully open state.

Nitinol, unlike traditional springs, does not increase in stiffness as displacement increases. This means that, if the check valve were designed such that the fully closed state was in the plateau region of the nitinol, the inefficiency cause by the increasing force is eliminated.

Now, what is an intelligent check valve and why do I want it?

Chapter 18: Other Valve Components

Nitinol valve components are something that we at KRL have recently begun producing.

*Hardness:* Nitinol is a

*Superelasticity:* Superelasticity is an unusual phenomenon in hard materials. Normally, as the hardness of a material increases, the elasticity and ductility is reduced and brittleness increases. This, of course leads to components that are longer lasting but less tolerant of the environment. A great example of this is in cutting tools for CNC machines. Tungsten carbide tools are the hardest and longest lasting available. However, if the machinist has the feeds and speeds wrong, the tool breaks almost immediately.

Unlike these other hard materials, nitinol is superelastic, retaining the ability to take on large deformations in spite of the high hardness. It is extremely rare that a case like this comes along where you don’t need to sacrifice one for the other—you can have your cake and eat it too.

*Vibration Reduction:* The natural vibration reducing properties of nitinol have two real benefits. First, that they increase the longevity of the components. Second, that they improve the accuracy of the flow.

# Appendices

# Appendix A: Glossary

**Age-**A heat treating process at relatively low temperatures that results in grains precipitating out of solution. For nitinol, during the aging process, nickel and titanium combine to form Ti4Ni3 grains. This, of course changes the ratio of nickel to titanium in the nitinol grains, thereby raising the transformation temperature. The effects of aging may be undone by solution treating.

**Anneal-**A heat treating process that rearranges the structure of the material, relieving any stress. This results in setting a new, distinct shape. Annealing typically results in a softer material (easier to scratch).

**Austenite**-the microstructure that exists in nitinol at “high” temperatures. Chemically, it is a B2 cubic structure that results in a stiff but flexible material.

**Heat treatment profile-**The prescribed range of temperatures and times necessary to achieve the desired properties. A heat treatment profile may be divided up into multiple stages, ramps, quenches, and atmospheres. Since the performance of nitinol is highly dependent on the heat treatment profile, it is highly recommended to use a PID controlled furnace.

**Martensite-**The microstructure that exists in nitinol at “low” temperatures. Chemically, it is a B19’ twinned structure. Since the chemical bonds in the twinned structure can easily rotate about the atoms, martensite appears to be an easily deformable or ductile material.

**PID-** A method of process control commonly used by digital and computer controlled systems. PID is an acronym for Proportional, Integral, and Derivative. Controllers may be programed to use P, PI, or PID control in order to meet the necessary requirements of your process. For more information on the intricities of PID control, *Production Systems Engineering* by Li and Meerkov is an excellent resource.

**Pseudoplasticity-**the term used to describe the false-permanent deformation observed in martensitic nitinol. Plastic deformation refers to permanent deformation that remains after all forces are removed. Pseudoplasicity exists in nitinol because, while it appears to be permanent, the deformation is recovered upon heating. This is also called thermoelasticity.

**Quench-**The rapid cooling of a sample by immersion in a liquid. While water and oil the two most common quenching fluids, researchers sometimes use liquid nitrogen (LN2) as a quenching fluid, especially when trying to obtain very tiny grain structures or completely amorphous nitinol.

**Ramp-**The controlled increase or decrease of temperature inside of a furnace. Ramps are typically specified in °C/min between two temperatures (i.e. ramp from 750°C to 375°C at a rate of 25°C/min).

**R-Phase-**the term used to refer to the rhombohedral structure that is found between martensite and austenite.

**Solution treating-**Heat treating at a temperature high enough for any precipitated material to be reabsorbed by the bulk. For nitinol alloys, this generally is 850℃ (1,550℉) or higher. Solution treating should take place under an argon shielding gas to prevent carbon, oxygen, and nitrogen from reacting with the titanium.

**Stage-**A section of a heat treatment profile that has the same conditions.

**Stress Relief Heat Treatment-**A heat treatment used to set the material into a new shape. While this typically occurs at temperatures similar to annealing, the time is generally shorter to avoid excessively influencing the thermal performance of the nitinol. Stress relief heat treatments can be conducted in succession to gradually form nitinol into very difficult to form shapes.

**Thermoelasticitiy-**Elastic deformation is deformation that is recovered once the force is removed. Thermoelastic deformation is deformation that isn’t recovered when force is removed but after heating.

**Tube Furnace-**A furnace that is generally long and cylindrical in shape but may have a square cross section. The square cross section is rare due to the difficulty of maintaining even heating throughout the cross section. The ends may be open to the atmosphere or they may be hermetically closed. Hermetically closed furnaces are typically used for treatments in specialized atmospheres while open ended furnaces are sometimes used for continuous heat treatment processes.

**Zone-**A region of a furnace that is independently controlled. Each zone consists of at least one thermocouple, one heating element, and one PID controller. Increasing the number of zones in a furnace allows for a higher level of control of the temperature inside the furnace.

# Appendix B: Furnace Design and Construction

A digitally controlled furnace is essential for doing a good job of developing nitinol products. Unfortunately, furnaces are very expensive. A 40 year old furnace with analog control with 8x8x8” enclosure will run you $750 on eBay. A brand new furnace with similar enclosure will cost over $5,000. The prices go up exponentially as the size grows (we were quoted over $150,000 for our 3 meter/10 foot tube furnace).

In this section, we’ll go over some tips and tricks on how to build your own furnace and how to save money while doing it. Some things you should feel safe importing from China while other things you should source domestically. We will address this as well.

First, let’s address the basic types of furnaces. Furnaces with a rectangular or cubic enclosure with a door on one side are generally referred to as a muffle furnace. Muffle furnaces generally have the highest level of temperature uniformity inside the enclosure. Cylindrical or rectangular furances with openings on both ends are referred to as tube furnaces. Tube furnaces may or may not have doors. Tube furnaces may also have a clamshell style opening with closed (or semi-closed) ends.

There are effectively two methods for heating furnaces: electric and flame. We say flame because there are numerous combustible fuels, ranging from propane to used motor oil, all of which may be used. Electric heating is easier to control and can be controlled more accurately, but is significantly more expensive than flame heating. With flame heating, the flame is MUCH hotter than what is desirable for heat treating nitinol, so mitigation must be built into your furnace.

*Part 1: Supplies.*

Insulating fire brick (IFB) is the basic construction material of furnaces. IFB differs from fire brick in that the thermal conductivity is much lower, dramatically reducing energy consumption and, more importantly, improving temperature consistency. Unfortunately, IFB is both more difficult to obtain and more expensive than fire brick. While most home improvement stores will carry fire brick, you will need to travel to a refractory shop to purchase IFB. IFB is specified by the rated temperature and the size of the brick. Standard sizes are 2.5” and 3” thick with a 9x4.5” cross section. Temperature ratings generally range from 2,300-3,300℉ although 3,600℉ bricks are available.

Figure 18: An insulating fire brick. The '23' printed on the end of the brick designates that this brick is rated for 2,300°F (1,250°C).

Generally speaking, higher temperature bricks are heavier and have higher thermal conductivity. For electric furnaces, 2,300℉ bricks are sufficient as the heater temperature rarely exceeds 2,150℉. For flame fired furnaces, you may want to consider using the higher temperature rating bricks opposite the burner. As for the shape, if you are able to travel to a manufacturer of IFB, the margins are good enough that they are happy to cut them to whatever size you as for as long as you order a full pallet. Purchasing a full pallet and then selling your excess may also be a good way to drive down your costs as refractory shops typically have a 150-200% markup on IFB. At Kellogg’s Research Labs, we use Smart Ceramics in Waltham, MA as our supplier of IFB. High quality IFB is available in China for about $0.15 per brick, but the cost of shipping anything less than a 20’ container quickly eliminates your savings.

One of the great things about IFB is that it is very easy to machine. A hand saw can be used to cut brick, a chisel can be used for basic shaping, and a router can be used for advanced milling. It is important to note that IFB is very dusty. Vacuum off the mortared sides immediately before applying mortar or else the mortar will not stick.

*Note: Unless otherwise noted, IFB and any items listed here as “ceramic” should be made from alumina. Alumina is also known as aluminum oxide or Al2O3.*

Furnace cement is used as a mortar to secure the IFB to each other as you build your furnace. Standard furnace cement is made from rock dust and has a pretty high thermal conductivity. However, it is both cheap and easily available. At the time this book is published, Lowes carries 1 qt. buckets for less than $13. Ceramic mortar adheres better to IFB and it provides a superior thermal conductivity, but it is extremely expensive in the US. Of suppliers that KRL uses, ceramic mortar is about $99 per lb. However, in China, ceramic mortar is very cheap. A 50lb (23kg) bag will cost about $20. This price goes down dramatically if you order a 20’ container loaded with it.

Figure 19: This is one of the custom built furnaces at KRL. Note that the floor and ceiling were not made with IFB and have started cracking while the walls, made of IFB are intact. Let this be a lesson as to why you should not try to save money by using inferior materials. This is a three zone furnace to keep the temperature uniform throughout the height of the furnace. A fourth zone, on the door, keeps the temperature uniform front to back.

There are two types of electric heating elements. There is Kanthal® and there is the nichrome. The benefit of nichrome heaters is that they’re significantly cheaper than Kanthal®. Kanthal® heaters are rated for roughly 500℃ higher operating temperature than nichrome. On one hand, this means that the furnace can be operated at higher temperatures. However, the more practical benefit of Kanthal® over nichrome is longer service life. Kanthal® heaters will last about 250% longer than “nichrome”. It is also important to note that the heaters will reach 1,000°C during heat-up, so even if you never take the furnace over 500°C, the heaters will last much longer when using Kanthal®. Additionally, if you will be embedding the heater in insulation, then the operating temperature is dramatically increased—meaning that you should use Kanthal®. However, if your heaters are completely exposed to the air inside the furnace, nichrome works well (especially if you don’t care about the difference between the 50,000 hour service life and 125,000 hour service life). At KRL, we use both materials, depending on the structure of the furnace.

We do not have a domestic supplier of nichrome, but if you wish to purchase pre-wound (spiral) Kanthal® heaters, Budget Casting Supply has two versions available for purchase (20A @ 110V and 20A @ 220V) for $50. Typically, we purchase nichrome from China, custom wound to our specifications for about $3 and Kanthal® for $10. When we need straight Kanthal® wire, we go to our friends at Smart Ceramics in Waltham, MA where they are happy to sell it to us for $47 per pound. This price makes Kanthal® heaters cheaper than nichrome from China.

High temperature wire is copper wire insulated with Teflon® and fiberglass. From looking at it, you would think that it’s the old wire from the 1940s that had a tendency of catching fire, but it’s not. This wire should be used to pass through the insulation and make the connection with your electric heaters. Once you are outside the insulation, the air will keep the wires cool enough to use THHN or similar wire. High temperature wire is available on Amazon for about $1 per foot. However, at Kellogg’s Research Labs, we happily import it from China, paying roughly $7 for 100 meters.

High temperature ring terminals are fittings that crimp onto the end of your high temperature wire so that you can secure it to the heating element. You cannot solder this connection because the solder will melt at a temperature much lower than what the heater operates at. High temperature ring terminals are available on Amazon for about $1 each or, if you would like to save money, simply use a standard ring terminal and trim off the plastic (this is the only difference between the standard and high temperature ring terminals.

Ceramic blanket is seamless and is available in thicknesses ranging from ½” to 2”. This is an excellent way to add insulation after the IFB, lowering the energy consumption and improving the temperature consistency inside the furnace. The best price we can get in the US is from McGill’s Warehouse ([www.McGillsWarehouse.com](http://www.McGillsWarehouse.com)) with the best shipping price being offered when ordering 3 blankets. In China, the blankets are $8 each, with each blanket weighing about 7kg. At the current rate of $5 per kg for air freight shipping, it is quite cost effective to purchase a few blankets from China. Is there a difference between the domestic blanket and the Chinese blanket? Yes. The difference largely is relating to the degradation of the material with long term exposure to high heat. If you’re using the blanked as a second layer of insulation, this is a non-issue.

Ceramic yarn is similar to ceramic blanket, but spun into yarn. The yarn is available in twisted rope and square or round braid. At Kellogg’s Research Labs, we use ceramic yarn for three purposes: 1. Sealing doors and permanently open openings (you can create a pass-through insulator with ceramic yarn); 2. Creating tie-downs in our larger furnaces; and 3. Knitting fireproof blankets as Christmas presents. While #3 is not financially effective, it is quite comforting to know that your loved ones will be safe in a fire. For our domestic supplier, we use Minseal (CeramicFiber.net). However, the price of 100 meters of ceramic yarn in China is roughly the same price as the cost of 1 meter in the US, so we almost exclusively import ceramic yarn.

Hard ceramic sheets have many uses—from lining the floor of your furnace to protect to using it as a tray to hold your samples to tooling fixturing. We always purchase our hard ceramic sheets from our friends at Smart Ceramics in Waltham, MA. Boron-nitride (BN) sheets are similar to ceramic sheets, but with lower porosity and lower coefficient of expansion. This makes them excellent for use as shelving inside your furnace at Kellogg’s Research Labs, we source all of our BN from Clay Planet ([www.clay-planet.com](http://www.clay-planet.com))

Ceramic fiber board (CFB) is a material that we usually use for ceilings on our larger furnaces. These boards have a texture similar to ceramic blankets, but are compressed to make a rigid board. This material is rather soft, so it is easily damaged—making it unsuitable for most unprotected surfaces (although, it makes a good floor if you protect it accordingly). There are two reasons we like to use CFB for ceilings and floors in the furnace. The first is that they are available in very large sheets, which simplifies the process of building the ceiling (no supports necessary) and, since no mortar is used, it has better thermal conductivity vs. IFB. Secondly, it can be used as an excellent substitute for ceramic blanket on flat surfaces.

Ceramic tube has many uses, not the least of which being constructing tube furnaces. A tube furnace is basically a ceramic tube with the heating element wrapped around the outside of the tube, evenly heating the inside. While alumina tubes are available, Mollite is the preferred material since it has better thermal shock resistance, higher thermal conductivity (since we actually want the heat to pass through this ceramic), and lower coefficient of thermal expansion. Ceramic tubes are much more expensive than they need to be. When we needed to make our 3 meter tube furnace, we needed two 5’ long tubes, 75mm diameter. Our friends at Smart Ceramics quoted a whopping $1,600 for each tube. After a bit of shopping around, we were able to find a supplier in the UK that could sell the tubes for $575 each plus $400 shipping. In China, we paid $72 for the two tubes plus another $70 to ship them to the US.

Cellulose insulation makes an excellent exterior insulator—with the catch that, if you’re not careful, it will catch fire and burn your furnace. In order to safely use cellulose insulation, the hottest surface in contact with the cellulose must be less than 100℃ (212℉) for reasonable operating conditions. The benefits to cellulose are that it is an excellent insulator and it is dirt cheap. You can buy it from any home improvement store for less than $13 per 40 cubic foot (4 cubic meter) bale. A word to the wise, consider renting the blower to blow the insulation into your furnace as it breaks up the clumps for you, saving a lot of time. Always keep in mind, cellulose can catch fire, so be very careful when building with it.

Temperature controllers (also known as PID controllers) are required for furnaces of all types. The term PID comes from the proportional, integral, and derivative methods of controlling a process. It should be noted that the Chinese controllers available on eBay do not have full PID control and are only effective for single set-point operation (they also have extremely small buttons). When we use Chinese controllers, we order directly from China from a supplier of production grade manufacturing equipment. The only function that these controllers lack is ramping. Plus, we can order them with a 100mm screen. Something interesting about Chinese controllers is that the documentation only covers about 5% of the total function of the controller, so make sure you talk to your supplier about how to take full advantage of your controller.

Omega Engineering ([www.omega.com](http://www.omega.com)) has a line of high quality, full featured, low cost PID controllers and provide excellent customer service. Honeywell and Cole-Parmer are both suppliers of high quality controllers at a higher price point than Omega. All three of these suppliers also have multi-zone controllers, capable of up to eight zones on a single unit (having a single computer controlling all zones somewhat improves accuracy). Depending on your budget, using a controller with ethernet connectivity offers some interesting capabilities. Just remember, always thoroughly read and understand the documentation. Despite the simple user interface, PID controllers are anything but.

Thermocouples and thermistors (RTD) are the two primary means of measuring the temperature inside the furnace. There are many types of each, but the most common thermocouple is the Type K (although, we do use Type J from time to time) and the most common thermistor is the Pt-100. We have found that the Chinese thermocouples on eBay are very poor quality and should be avoided. When we use Chinese thermocouples, we purchase research grade sensors from a lab equipment supplier in China for about $4 each. When we use domestically sourced thermocouples we usually purchase from Omega for about $35-75. A word of advice, always use thermocouples with a stainless steel jacket protecting the probe. If you have the budget, Inconel is better. Whatever sensor you are using, always be certain that your controller accepts it, otherwise, you’ll get errant temperature readings.

Solid state relays (SSRs) are essential for supplying the current to the heating elements. Of course, flame fueled furnaces do not need these. We have used the Chinese brand Futek and the US brand Crydom and have found that the Futek SSRs perform just as well—if not better than—the Crydom SSRs. A Crydom SSR will run you about $50 while a Futek will cost about $5 on eBay or Amazon. We purchase our Futek SSRs directly from China for about $0.80 each. A word of advice: use heat sinks to keep the SSR’s cool, they get very hot.

*Part 2: Tips and Tricks*.

When laying IFB, follow standard bricklaying principles (if you don’t know them, there are plenty of message boards available discussing this). Getting everything square and plumb is very important. IFB is always dusty. In order for the mortar to adhere to it, use a shop vac to remove the dust. IFB can be cut to size with a standard wood saw, utility knife, or just about anything, so no need to buy special tools there. When shaping IFB into more complicated shapes—such as to make grooves for heating elements, we purchased a low quality router table from the local hardware store for $99. We have shaped nearly a pallet of IFB with it and it still works great. Make sure yours has a vacuum hook up so that you can suck away all of the massive amounts of dust that is generated. We have found that the filter on the vacuum must be cleaned after each box of IFB.

Installing a vapor barrier between layers of insulation (or on the outside of the final layer of insulation) will help improve insulation. While there are various coatings that you can buy if you want to spend lots of money, DEI 010301 High Temperature Silicone Coating has an operating temperature of 800℃ (1,500℉) and is just $12 per can. It is also available at most auto parts dealers (and Amazon), so it is easy to obtain.

Thermocouples are notoriously inaccurate. A good thermocouple will have an accuracy of just 1℃ (2℉). However, nitinol needs to have extremely high temperature consistency and this may not be good enough for your application. By having multiple thermocouples wired together, the variability is dramatically reduced. Two to four thermocouples wired together can get 0.01℃ accuracy with a good degree of confidence. Clustering the thermocouples together gives you an accurate measurement at that point while distributing the thermocouples around the zone averages out the temperature in that zone. For optimum temperature uniformity, consider installing one or more fans to circulate air throughout the furnace.

To drill holes through IFB and ceramic fiber board for the thermocouples, finding a drill bit that will penetrate 6-12” of insulation. A simple trick that works well is to use a steel round rod as the drill bit, cutting the tip with a pair of bolt cutters. The sharp end left by the bolt cutters makes quick work of ceramic materials (and you can make an unlimited length drill bit for just a couple bucks). We frequently slather the probes with mortar so that they seal the hole. This prevents cool air from outside from leaking directly onto the thermocouple. Just make sure that the thermocouple is well cleaned prior to first operation.

The door on a muffle furnace always is colder than the rest of the furnace. While installing one, two, or even three rows of ceramic yarn will help to mitigate this, you should consider making the door a separate zone of your furnace for optimum control.

A little something that took a while to figure out: the manufacturers of refractory insulation all talk about the energy savings when you use a good refractory and you use plenty of it. While it is true that thicker insulation will result in significantly lower energy consumption, we have found out that there is an ancillary benefit that actually carries greater value than the energy savings. That is the temperature uniformity. As was stated above, nitinol is very sensitive to the temperature at which it is cooked. This means that a poorly insulated furnace will yield dramatically varying temperatures, which result in wide ranges of the properties of nitinol. Adding a few extra inches of insulation will dramatically improve the consistency of the nitinol, and that is worth every penny.

The first time you turn on a furnace is different from every other time. There are volatile organic compounds (VOCs) that must be burned off, so care must be taken when heating it up the first time. First, make sure that the mortar has cured for at least 72 hours prior to firing up the furnace (more is better). If you are in a rush, you can start the furnace temperature at 40℃ and increase it in 10℃ increments until it reaches 80℃ and then bake for an hour. This will ensure that there is no water in the mortar to produce steam and crack your IFB. The worst possible outcome of an initial fire-up is a brick exploding.

On the first turn on, start at a low temperature—maybe 300℃ and let it cook for 24 hours. Then increase the temperature by 100℃ and cook for another 24 hours. Continue this process until at least 600℃, 800℃ if you have the time and patience. When we built our 0.7 cubic meter furnace, we rushed this process and set off all of the carbon monoxide detectors in the building. Needless to say, the fire department wasn’t too happy. Lastly, use very large amounts of ventilation. If your furnace is portable, it is best to do this outside (even if you have to build/buy an extension cord). If it is not portable, plan ahead for how you’re going to keep the area safe during burn-in.

The different parts of heat treatment fixtures have a tendency to bond to each other. The threads on screws and bolts used to secure the fixtures together are the most prone to this bonding and it often results in breaking off the bolt in the hole. To prevent this from happening, there are a few things that can be done. Option 1: Heat treat the components several times so that a thick layer of surface oxidation builds up. This is very time consuming and costly, depending on the billing rate of the furnace. Option 2: Over-drill the holes to give the parts room to move as oxidation builds up. While this reduces bonding, it reduces the accuracy and strength of the tool. Option 3: Use a ceramic lubricant. We have found that Permatex 24125 Ceramic Extreme Brake Parts Lubricant is an excellent lubricant that survives heat treatment temperatures. While one use is generally enough to ensure future survival of the tooling components, we usually use it every time just in case. We have found that most auto parts stores in the US carry the lubricant, so you don’t have to look far to get it.