

Dimensional instability of materials: how critical is it in the design of optical instruments?

Roger A. Paquin
Advanced Materials Consultant
Tucson, Arizona

Abstract

Dimensional instability exists to some extent in all components no matter what the materials may be. So the question is not "how can we eliminate instability?" but rather, "how can we reduce it to a tolerable level?" The maximum allowable dimensional instability will vary with application and depends on the particular component and its role in the optical instrument. The purpose of this paper is to provide the basis for deciding how much can be tolerated and for making intelligent choices in the selection of materials and processes for components that will achieve stability design goals with which to meet optical instrument performance specifications. This basis is a better understanding of the causes of instability and methods for minimizing instability.

After a discussion of tolerable levels of instability, four types of dimensional instability are defined: thermal, temporal, cycling and hysteresis, with examples given for each. The principal causes of these instabilities: external stress, changes in internal stress, microstructural changes and inhomogeneity/anisotropy of properties, are explained in some detail along with a discussion of material types and properties. Most importantly, methods for minimizing the instabilities are shown. This discussion includes specific recommendations for commonly used materials including: processing techniques to minimize instability, specific problems observed in some materials and how to avoid the problems, and some general guidelines regarding the effects of fabrication methods on stability.

It is most important to realize that increasingly tighter specifications for optical instruments mean that the optomechanical designer must work concurrently with other engineering disciplines, particularly materials and processes engineers, to insure the desired thermal and temporal stability of the product.

1. INTRODUCTION

All materials are dimensionally unstable to some degree. In preparing to design and fabricate dimensionally stable components, it is important to realize that this implies controlling the sources of dimensional instabilities to a level such that any dimensional changes that occur are kept within specified tolerances. To be able to accomplish this requires an understanding of the sources of these instabilities. The primary purpose of this paper is to impart such an understanding.

The key to building stable instruments is knowing the requirements. Stable materials can then be chosen from which components can be fabricated utilizing methods that minimize introduction of dimensional instabilities. But what exactly is dimensional instability? My definition is simply the dimensional change that occurs in response to internal or external influences.

The problem is to control this dimensional change, a distortion or strain in the component of interest, to levels that will not compromise instrument requirements. If we require stability on the order of machining tolerances, strain of approximately 10^{-3} , there is no serious problem. If parts must maintain dimensional tolerances to parts per million, e.g., micrometers/meter, then care and consideration must be given to materials selection and processing steps. This is the realm of precision optical instruments. If nanotolerances are required, e.g., nanometers/meter, there is little help available, although this has recently become the subject of increasing study. For nanostability, the principles of controlling the sources of dimensional instability must be diligently applied – and then hope for the best.

Once a tolerance has been specified to bound the amount of allowable instability, the next step is determining the potential sources of instability and controlling them to meet the specification. The balance of this paper explains the common types of instabilities and their sources, and gives suggestions as to how they can be controlled. While this paper is, of necessity an overview, I recommend for further study Marschall and Maringer's excellent book¹ on the subject, although it is unfortunately now out of print, and the two volumes of SPIE Proceedings dealing specifically with dimensional stability, volumes 1335² and 1533³.

2. TYPES OF INSTABILITY

The evidence of dimensional instability is all around us and takes many forms, from the warping of a 2x4, which is readily apparent, to the figure change in a mirror subjected to a change in environment. I have chosen to classify instabilities into four categories which I call:

- temporal instability,
- thermal/mechanical cycle instability,
- thermal instability and
- hysteresis.

Each of these can have magnitudes ranging from sub-nanostrain to very large numbers, as the following examples show.

Temporal instability is the change that takes place in a component as a function of time in a fixed environment. It is a permanent change. For example, Figure 1 shows the behavior of two sets of nominally similar one inch gage blocks tested at NBS over a period of roughly thirty years⁴. The best sample, No. 2 in set B, changed only one microinch in 22 years, while the worst changed $24\ \mu\text{in}$ in 25 years. Note that the worst sample was changing at the rate of 10^{-6}in./in./yr . That is a very small amount and yet it is totally unacceptable for the application. This kind of behavior is usually associated with relaxation of residual stress.

Thermal/mechanical cycling instability is the change measured in a fixed environment after exposure to a variable environment. It too is a permanent dimensional change. A common example is the change that takes place in fiber reinforced composites as shown in Figure 2 for boron reinforced aluminum⁵. The figure shows the behavior of three samples which changed length by over 0.02% on the first cycle, but the amount of change decreased with each succeeding cycle, approaching an asymptote. Later I show

this kind of behavior in other materials. For composites, the cause for the dimensional changes is usually internal micro-cracking and in single phase materials it is usually some other form of internal stress relief. Similar behavior has been observed with mechanical cycling and vibration⁶.

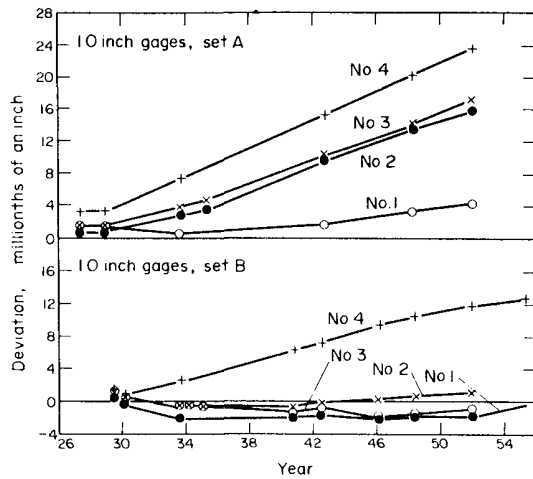


Figure 1. Example of temporal instability of two sets of nominally similar one inch steel gage blocks⁴.

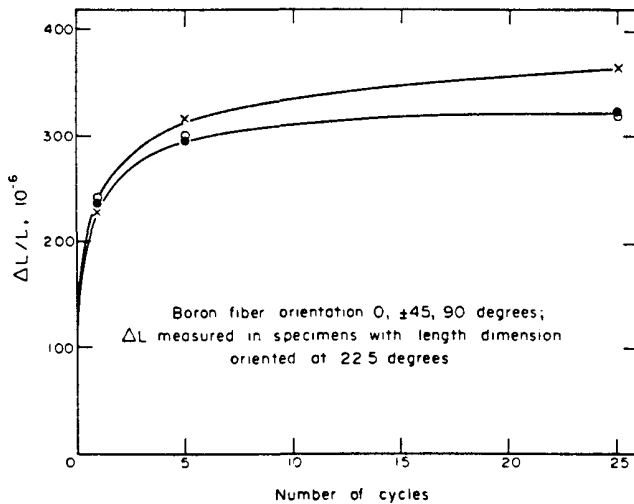


Figure 2. Example of thermal cycling instability of boron/aluminum composite samples thermally cycled from -55°C to $+55^{\circ}\text{C}$ ⁵.

Thermal instability, as I choose to define it, is the dimensional change measured in a fixed environment after a change from another fixed environment, independent of the environmental path. This dimensional change is reversible upon returning to the original conditions. Figure 3 shows evidence of just such a change. This beryllium (Be) mirror was made from an experimental billet produced in the late 60's which had a substantial amount of thermal expansion inhomogeneity⁷. It was interferometrically tested

many times over a period of ten years and always exhibited the same distortion when heated as is shown in the figure, and always returned to the same optical figure at room temperature, within the 0.02 wave accuracy of the instrument. This behavior has been virtually eliminated in modern Be materials.

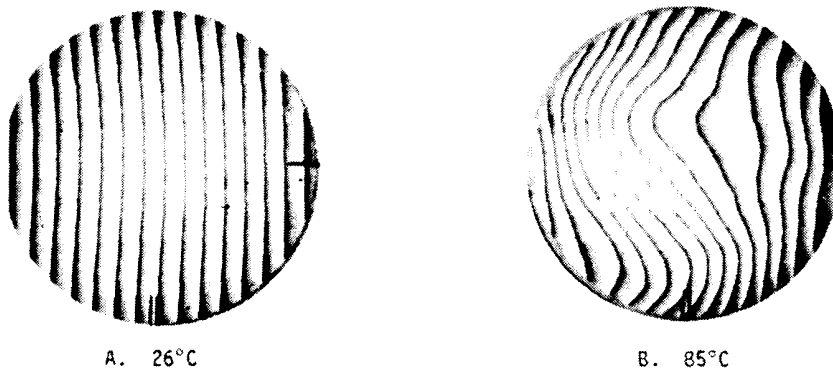


Figure 3. Optical interferograms of an electroless nickel coated experimental beryllium alloy mirror (ca. 1968) showing a reversible thermal instability of approximately two waves⁷.

What I call hysteresis is the change measured in a fixed environment and dependent on the environmental path to another fixed environment. This type of distortion can be permanent or reversible. For example, in Figure 4, the length of Zerodur on cooling from 300°C to 20°C depends on the cooling rate⁸. This is typical behavior for glasses containing MgO. But note that the curves are parallel below 150°C indicating that the temperature range of sensitivity is between 150°C and 300°C. This behavior has been eliminated in a new version of this material called Zerodur M.

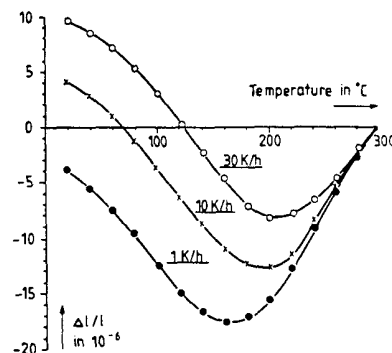


Figure 4. Thermal length contraction of Zerodur for three cooling rates from 300°C to 20°C illustrating a hysteresis type of dimensional instability⁸.

These are the major types of dimensional instabilities that can be encountered in components for optical instruments. I will show later that almost all observed instabilities can be placed into one of these four categories and that the sources of these instabilities are readily understood so that their effects can be minimized.

3. SOURCES OF DIMENSIONAL CHANGE

The sources of dimensional changes such as those described above can be attributed to one or more of the following factors:

- externally applied stress,
- changes in internal stresses,
- microstructural changes and/or
- inhomogeneity/anisotropy of properties.

I will describe each of these in some detail in the following sections.

3.1 External Stress

When an external stress is applied to a component, if it behaves according to Hooke's law, it should deform elastically no matter how long the stress is applied and return exactly to its original shape when the stress is removed. But this being an imperfect world, where all practical materials are imperfect, there are other responses to externally applied stress⁹ as shown in Figure 5.

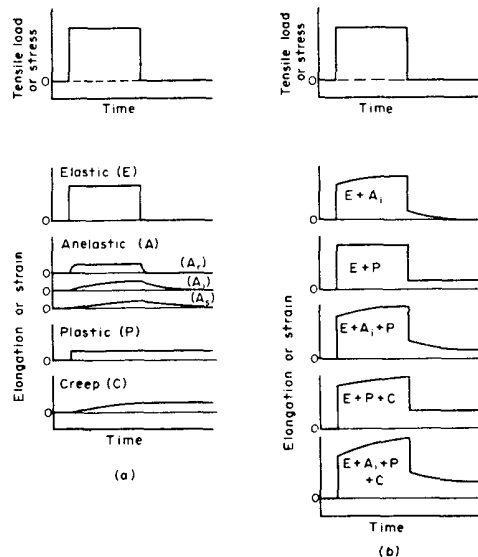


Figure 5. Schematic illustration of types of response to an instantaneously applied and removed tensile stress: (a) individual types of response, and (b) combined responses⁹.

If a load is applied suddenly, held for a length of time and then released as shown at the top of the figure, the elastic response is as shown for curve (E), a typical response for glasses, ceramics, and brittle metals. A time dependent type of elastic response is shown in (A) and is called anelastic strain. The three curves show rapid, intermediate and slow response times. For this type of behavior there is no strain when the load is first applied, but it increases toward an asymptote with time; when the load is removed, the strain asymptotically returns to zero. Anelastic behavior has been observed in some glass ceramics at low temperature, but is more commonly observed in polymers, and

particularly in polymer matrix composites where both the anelastic effect and mechanical properties are effected by moisture content and temperature. This behavior has been summarized by Wolff¹⁰. A combination of elastic and anelastic behavior is shown in the E+A curve. Plastic strain is permanent and does not decrease as the load is removed, as shown in P. The most common behavior for metals, a combination of elastic and plastic response to stress is shown in the E+P curve. Time dependent plastic strain is called creep, C in the figure. Other possible combinations are also shown.

There are a number of material properties that are important to dimensional stability. A listing for a number of optical and structural materials is given in Table 1. Among the important thermal properties are: coefficient of thermal expansion (CTE), more precisely called instantaneous linear coefficient of thermal expansion (α'), specific heat (c_p), thermal conductivity (K) and thermal diffusivity (D). Thermal figures of merit are the steady state (CTE/K) and transient (CTE/D) distortion coefficients. The important mechanical properties are elastic modulus (Young's Modulus, E) or stiffness represented by the slope of the stress vs. strain curve as in Figure 6; Poisson's ratio, the relationship between tensile (or compressive) and shear strain; yield strength (at 0.2% offset), the stress to cause 2×10^{-3} permanent or plastic strain; and ultimate or fracture strength. Fracture toughness (K_{Ic}) is related to the energy to propagate a crack. Specific stiffness (E/ρ) is a measure of structural efficiency. Particularly important for precision applications are microyield strength, the stress to cause 1×10^{-6} plastic strain (one microstrain); and microcreep strength, which has no acceptable definition other than that it is less than the microyield strength and is a constant stress that produces microstrain after some period of time.

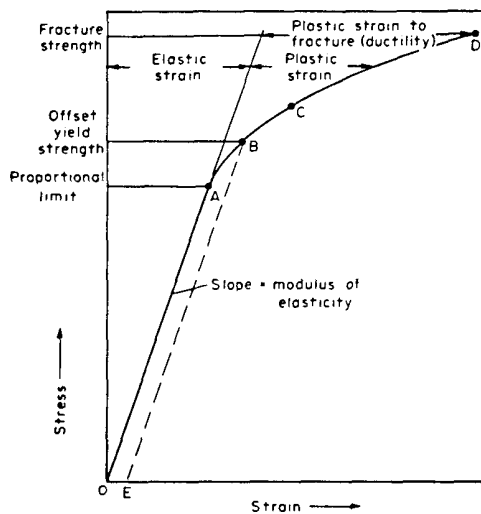


Figure 6. Schematic stress-strain diagram to illustrate typical material behavior¹¹.

Microyield behavior can not be inferred from the macro-mechanical properties of either yield strength or modulus. For example, in Figure 7 the behavior of I-400 Be and 2024-T4 aluminum (Al), metals with approximately the same yield strength, are compared. The Be sample exhibits a microyield strength of only 6.8 ksi, but with increased

Table 1. Properties of selected Materials

MATERIALS	P		E		E/P		MYS*		K _{IC}		CTE		K		Cp		D		DISTORTION COEFF			
	lb/cum	g/cc	Msi	GPa	10E8 in	Ksi	Ksi	Ksi	Btu/hr-ft-F	W/cm-K	Btu/lb-F	ppm/F	Btu/hr-ft-F	W/cm-K	Btu/lb-F	W-sectg-K	sqm/hr	sqm/sec	um/W	CTE/K	TRANSIENT	
PREPARED	Small	Small	Large	Large	Large	Large	Large	Large	Large	Large	Small	Small	Large	Large	Large	Large	Large	Large	Small	Small	Small	
Fused Silica	0.079	2.19	10.4	72	1.32	[1-2]	<1	0.28	0.8	0.178	0.5	0.014	0.8	0.014	0.745	0.0085	4.72	3620	59			
ULE	0.8	2.21	97	67	1.21	[1-2]	<1	0.008	0.76	0.183	0.015	0.013	0.77	0.013	0.77	0.0077	4.32	115	19			
Zerodur	0.019	2.53	13.3	92	1.46	[1-5]	<2	-0.05	0.91	0.194	-0.09	0.016	0.812	0.016	0.812	0.0076	4.27	-570	-12			
Beryllium	0.087	1.85	42	287	6.3	2	>10	6.3	112	0.436	71.4	1.93	1.825	1.93	1.825	0.572	319	590	20			
T-70A	0.087	2.7	9.9	68	1.02	10	>20	13.1	96.5	0.214	23.6	1.67	0.896	1.67	0.896	0.69	385	1410	34			
Aluminum 6061-T6	0.323	8.95	17	117	0.53	0.5	>20	9.2	230	0.092	9.2	3.99	0.385	3.99	0.385	1.155	644	410	14			
Copper	0.29	8.02	28	193	0.97	7	>20	9.6	9.4	0.12	17.3	0.162	0.5	0.162	0.5	0.04	22.5	10600	428			
Stainless Steel 304	0.291	8.05	20.5	141	0.7	3	>20	0.7	6	0.12	1.26	0.104	0.5	0.104	0.5	0.026	14.3	1190	48			
Invar 36	0.294	8.13	21.5	148	0.73	5	>20	0.17	6	0.12	0.3	0.105	0.51	0.105	0.51	0.025	13.8	290	12			
Super Invar	0.16	4.43	16.5	114	1.03	10	>15	4.9	4.2	0.135	8.8	0.073	0.56	0.073	0.56	0.029	16.2	12100	303			
Ti6Al4V	0.084	2.33	19	137	2.26	[>10]	<2	1.39	85.7	0.17	2.5	1.48	0.71	1.48	0.71	0.892	497	170	3			
Silicon	0.104	2.89	47.9	330	4.61	[>40]	<5	1.47	94.8	0.17	2.64	1.64	0.7	1.64	0.7	0.81	455	161	3			
SiC	0.112	3.11	54.1	373	4.83	[>40]	<5	1.49	85	0.16	2.68	1.47	0.68	1.47	0.68	0.7	393	180	4			
RB-30% Si	0.116	3.21	67.4	465	5.81	[>40]	3	1.33	110	0.17	2.4	1.9	0.72	1.9	0.72	0.82	462	128	3			
RB-12% Si																						
SiC																						
CVD																						

* for brittle materials, the values in parentheses are fracture strengths

stress yields little more, whereas the Al alloy resists yielding for a high microyield strength of 38.8 ksi but then continues yielding readily.

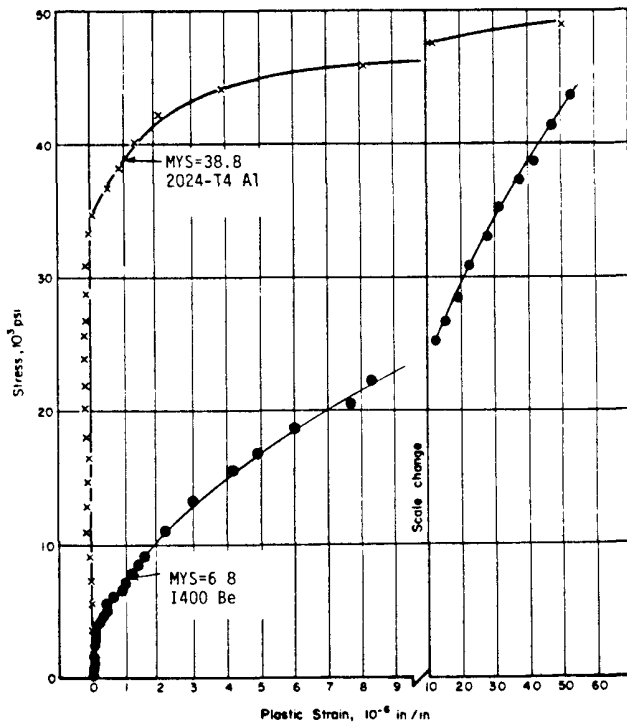


Figure 7. Comparison of a beryllium and an aluminum alloy that have similar yield strengths but very different microstrain behavior¹².

Microyield strength is strongly dependent on the prior history of the material. If it has been annealed, microyield will be lower than in almost any other condition. Conversely, if there has been prior straining, either through intentional or inadvertent application of stress, the microyield will be raised as demonstrated for a type of stainless steel in Figure 8. While prestraining produces a stronger material, it also leaves a level of residual stress which may be detrimental. Residual stress is discussed further in the next section.

Since high microyield strength (MYS) is a desirable property, and since many materials have relatively low MYS, it is important to know that there are methods for increasing it. Prestrain, as mentioned above, is one method, but it has its disadvantages. Many aluminum alloys, after rolling to plate form, are stretched a few percent to both straighten and level the stress through the thickness of the plate, which also increases MYS; but this process also seems to lower microcreep strength. Since the process of microyielding occurs, at least in the early stages, by movement of dislocations, anything that pins or prevents dislocation movement will increase MYS. By reducing the grain size of a material, dislocations are more readily pinned, as they are when particle or fiber reinforcement is added to a single phase material. Multi-phase materials almost always have higher MYS than similar single phase alloys. Thermal treatments that precipitate a second phase or produce a metastable phase tend to increase strength, and al-

loying a pure material usually produces dislocations and lattice strains that likewise increase MYS.

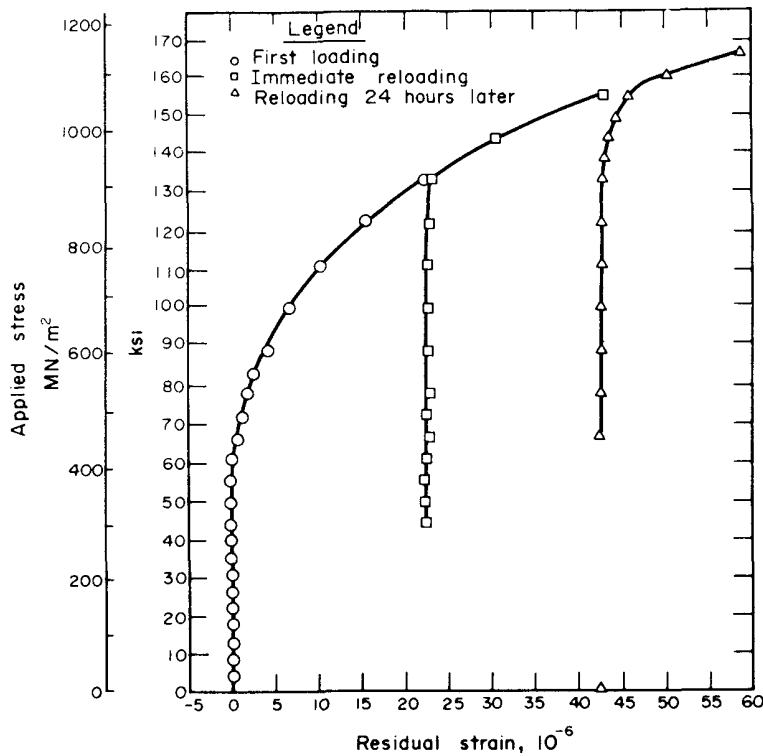


Figure 8. Effect of prestrain on microyielding behavior of 440 C stainless steel¹³

3.2 Changes in Internal Stress

While external stress is applied and removed from a component and is readily observed and measured, internal stress is not obvious. A component can be free of external attachments, even floating in a zero-g environment, and have internal stresses. They are in equilibrium and consist of balanced tensile and compressive stresses.

There are two types of internal stress called short range and long range. The spatial extent of these is, as the name implies, microscopic and macroscopic respectively. The long range internal stress is better known as residual stress. To illustrate both types, consider a component machined from an inherently anisotropic metal like Be. If the component was annealed prior to machining, the bulk of the material will consist of Be grains that vary in crystallographic orientation, and therefore in CTE at the grain boundaries. This produces short range internal stress at the grain boundaries as further described below. At the surface of the part, the machining will have plastically deformed a surface layer within which the residual or long range stress will be approximately the yield strength, approximately 40ksi for Be. Below the deformed layer there would be a partially deformed transition layer where the residual stress level would decrease rapidly from 40ksi (probably tensile), through zero to a low level compressive stress.

Short range internal stress can also result from unequal amounts of distortion between neighboring crystals in plastically deformed material. It primarily arises from inhomogeneous CTE: in a two phase material, between adjacent crystals with anisotropic CTE as described above for Be, or in a matrix with a dispersed phase or reinforcing particles, whiskers or fibers. Table 2 lists the theoretical maximum values of thermally induced microstrain due to CTE mismatch between adjacent grains of a few non-cubic materials^{13,14}. In practice, the average values are approximately one third of these calculated maxima. You can see that for Be, one third of 63.8 psi/°C over a 100°C temperature change results in a short range stress of over 2000 psi, a value that exceeds the microyield strength of some Be alloys.

Table 2. Theoretical Maximum Values of Short Range Internal Stresses Due to Thermal Expansion Anisotropy^{14,15}

<u>Material</u>	<u>Lattice</u>	<u>psi/°C</u>
Zinc	Hexagonal	177.
Calcite	Rhombohedral	165.
Cadmium	Hexagonal	91.4
Beryllium	Hexagonal	63.8
Quartz	Rhombohedral	43.1
Indium	Tetragonal	32.6
Magnesium	Hexagonal	2.74
Graphite	Hexagonal	0.94

Long range internal stress, residual stress, is usually the result of processing operations such as forming, heat treating, welding, machining or plating. As you would expect, the level of the stress is dependent on the severity of the operation, as shown in the following examples. Figure 9 shows how the temperature of the quench water effects the residual stress in Al alloys. The yield strength also drops when the water quench is less severe. Polymer quenchants are available that can provide the low residual stress of a boiling water quench with the strength of the cold water quench.

The effects of grinding parameters on residual stress in steel are shown in Figure 10. Here we see that abusive conditions, heavy feed, high speed and no coolant, produce high compressive stress with a deep layer, while a gentle cut produces a shallow layer with a small compressive stress¹⁷. Similar effects take place in glasses and ceramics. Sharp milling cutters generally produce compressive stresses with conventional feeds and speeds, but a surface layer with high tensile stress occurs when the tools are dull¹⁸.

The introduction of residual stress, or any change in the balance of the stress will cause changes in dimensions of the component. This means that removal of a stressed layer will cause dimensional changes as demonstrated in Figure 11. When two specimens with surface residual stress are acid etched, they both shrink, the one with the higher stress (deeper cut) shrinking more. This principle is used in the manufacture of dimen-

sionally stable components where after heavy material removal operations the surfaces are acid etched to restore unstressed dimensions to the part²⁰.

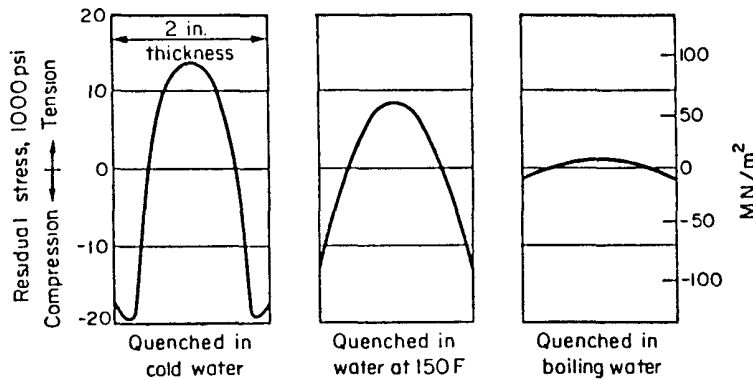


Figure 9. Residual stresses in specimens of 7075-T6 aluminum plate quenched in water at different temperatures¹⁶.

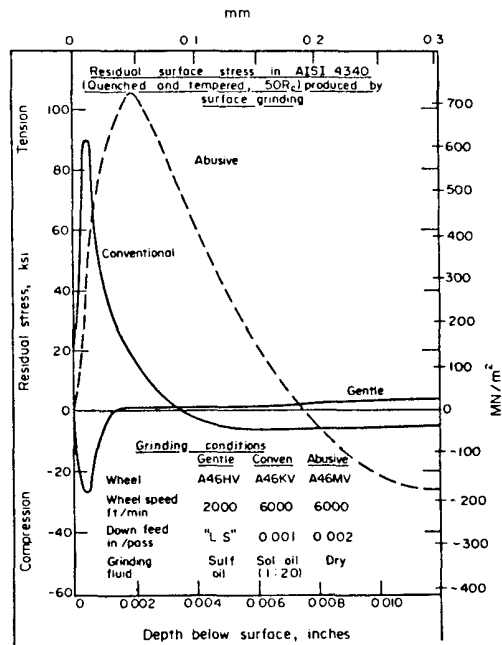


Figure 10. Residual surface stress in AISI 4340 steel produced by surface grinding¹⁷.

Residual stress can also decrease spontaneously with time with a related change in dimensions. This effect is called stress relaxation and the decrease in stress is proportional to the stress level as shown in equation (1), where s is stress, t is time and τ is the relaxation time.

$$(-ds / dt)\tau = s \tag{1}$$

But stress also decreases exponentially with time as shown by equation (2).

$$s = s_0 e^{-t/\tau} \quad (2)$$

Note that when time, t , is equal to the relaxation time, τ , the ratio of stress to original stress, s/s_0 , is equal to $1/e = 0.37$. This behavior is shown in Figure 12, where a Be mirror was fabricated with no treatment for stress relief after annealing the rough blank. Note the exponential shape to the optical figure change curve.

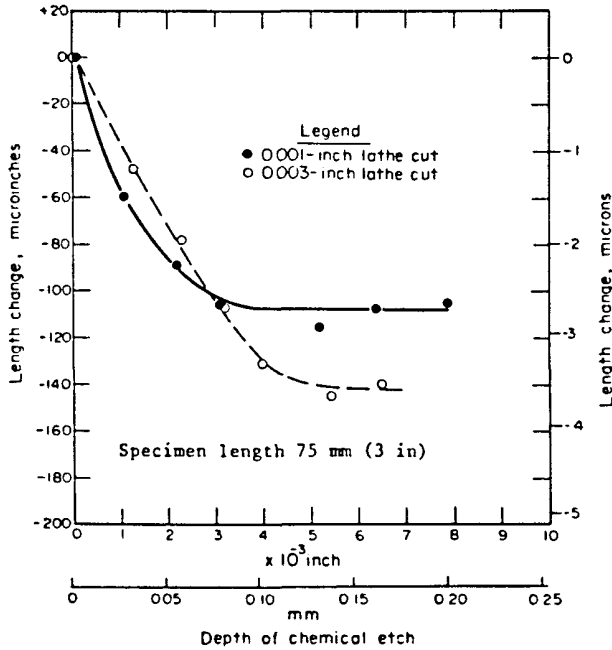


Figure 11. Dimensional changes in 3.0" long specimens of Ni-Span-C on chemical removal of residual stresses due to machining¹⁹.

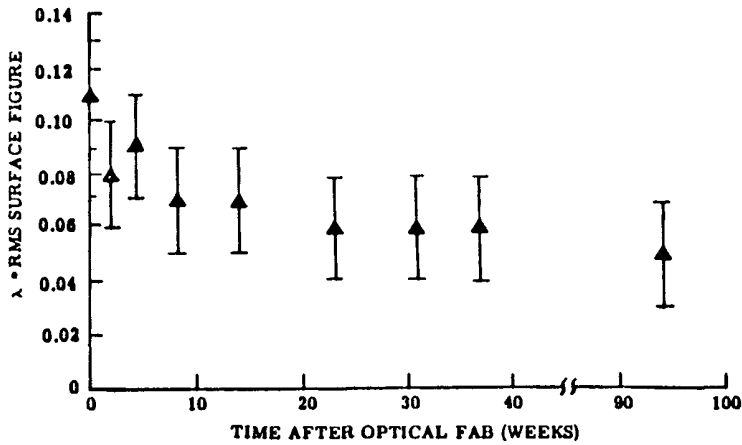


Figure 12. Temporal stability of a commercial 13" diam. beryllium mirror fabricated from vacuum hot pressed block.

Stress relaxation is also a thermally sensitive process, behaving according to the Arrhenius relationship of equation (3), where E is the activation energy, k is Boltzmann's constant and T is absolute temperature.

$$1/\tau \propto e^{-E/kT} \quad (3)$$

This phenomenon can be used to reduce the stress level with thermal treatment - stress relief. However the question is often raised whether an isothermal treatment for some reasonable time or thermal cycling between elevated and reduced temperatures is the more effective stress relief treatment. Much has been written on this subject as summarized in Chapter 6 of Marschall and Maringer¹, but the best answer is, "it depends". It depends on the crystal structure and purity of the material; it depends on the prior thermo-mechanical history of the component; it depends on the temperature, time and rate of change of temperature; and of course it depends on the level, type and distribution of the internal stresses.

One example of what can happen is shown in Figure 13 where both isothermal and thermal cycling treatments were given to Be specimens previously stressed to 11.2 ksi²¹. In this case, the low annealing temperatures of 100° and 190°C do very little but relieve peak stresses. The best treatment is a 600°C stress relief treatment, but this is a higher temperature than most designers would want to subject a semifinished component. A temperature of 400°C still only removes 40% of the stress, but note that thermal cycling from 400°C to either -70°C or -196°C provides a 55% reduction. The cycling is more effective than the isothermal treatment to the same upper temperature. This cycling effect may only hold for non-cubic materials with reasonably high expansion anisotropy as listed above in Table 2, or for materials with more than one phase such as the composite materials shown in Figure 2. There is conflicting evidence in the literature on the effect of thermal cycling on stress relief of homogeneous cubic or amorphous materials.

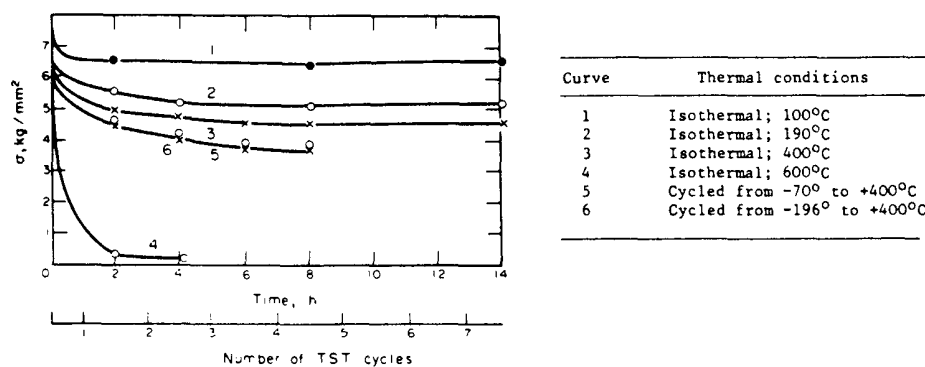


Figure 13. Comparison of the effects of thermal cycling and isothermal exposure on stress relaxation of pure beryllium²¹.

For those materials where thermal cycling is more effective than isothermal exposure at the same elevated temperature for reduction of residual stress, the proposed mechanism is as follows. These materials develop short range internal stress when the temperature is changed due to the expansion mismatch between adjacent grains and/or phases. When this stress is added to the long range, or residual stress, the yield, or microyield

strength is exceeded locally and plastic strain results. When the temperature is returned to room temperature, the plastic strain remains, but the level of the residual stress is reduced. Holding at the elevated temperature provides no further benefit unless it is close to the annealing temperature. When the temperature is then reduced, the sign of the short range stress is reversed, exercising the material further and providing more stress relief. In a similar manner, vibration, or mechanical cycling can provide stress relief although, as for thermal cycling, there are resulting dimensional changes.

For composites where the interface stresses can be substantial, thermal cycling is mandatory before dimensional stability to microstrain levels can be achieved. As many as 100 cycles are often required to achieve stability under constant moisture level conditions.

For reducing stress levels in critical components we then have a number of options:

1. Thermal treatments such as isothermal exposure or thermal cycling,
2. Mechanical treatments such as vibration or mechanical working,
3. Removal of surface material by chemical etching, controlled grinding and/or polishing or other stress-free methods and
4. Time.

3.3 Microstructural Changes

Microstructural changes in materials can result in both induced internal stress and dimensional change. The type of response depends on the material and the kind of microstructural change. The changes that can take place in materials used in optical and precision instrument systems include devitrification in glasses and glass ceramics; phase transformations, recrystallization and grain growth in metals and ceramics; and devitrification, phase transformations, recrystallization and grain growth in coatings. This cause of dimensional instability is quite common, but can not be covered adequately here. I will only discuss a few examples to illustrate the principle. The first example, illustrated in Figure 14, is the dimensional change that takes place in heat treatable aluminum alloys during precipitation heat treatment, also called aging or precipitation hardening. This example illustrates that a component which is aged from the solution treated and quenched condition to obtain maximum mechanical properties, normally 4 to 8 hours, will undergo a dimensional change due to the precipitation of the second phase. The change is small for 6061, a significant shrinkage for 7075 and a significant expansion for 2014. But notice that additional hours of aging, often performed for stress relief, induces additional significant dimensional change, particularly for the 2014 alloy. This behavior is typical of the 2XXX series Al alloys and must be taken into consideration when they are used. The 6061 alloy, most often used for precision optical structures and mirrors, changes the least, verifying its applicability for these applications.

The second example I previously cited in Figure 1. The change taking place in this material is typical of many tool steels where there is a sluggish phase transformation taking place from retained austenite to martensite, a transformation accompanied by a net

expansion. The relatively stable parts, numbers 1-3 in set B, were sub-zero cooled to drive the reaction to completion.

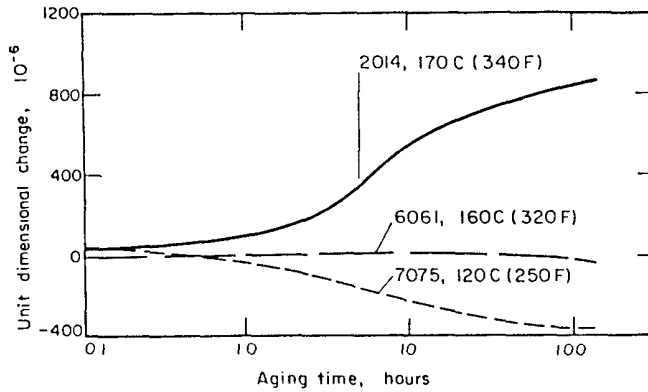


Figure 14. Dimensional change as a function of time at the precipitation heat treating temperature employed to produce the T6 temper for three aluminum alloys²².

Electroless nickel coatings are used extensively, both for polishable coatings and for corrosion protection of Al, Be and iron alloy components. The coatings are nickel-phosphorous alloys that may be amorphous and are thermodynamically unstable as deposited. The stable form of the coating is nickel phosphide (NiP) in a matrix of Ni, but in this form it is not polishable. Coatings can have significant residual stress depending on phosphorous content and substrate CTE, and annealing changes the stress state²³, as shown in Figure 15.

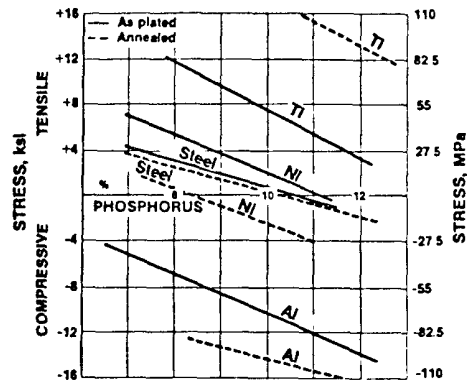


Figure 15. Stress of electroless nickel on aluminum, nickel, titanium and steel as a function of phosphorous content²⁴.

During thermal treatment (annealing) immediately after plating, hydrogen is driven off, adhesion improves, amorphous coatings become polycrystalline and diffusion of phosphorous toward formation of NiP begins. Since NiP has a lower specific volume and CTE than the quasi-stable Ni-P alloy, there is a shrinkage and a decrease in CTE for the coating during thermal treatment that also changes the stress state; the magnitude of these changes depends on the annealing conditions²⁵. For higher temperatures and/or

longer times, NiP starts to form as a precipitate in the coating. For machinability by diamond turning and/or polishability, the precipitate should not be allowed to form.

The annealing temperature used in the referenced studies is 190°C, with four hours at temperature for Be²⁶ and one hour for the other materials²³. Stress goes from tensile to compressive with increasing phosphorous content and increasing substrate expansion coefficient, while annealing changes the stress toward compression for higher expansion materials and toward tension for lower expansion materials. What this all means is that for a particular substrate material, to obtain zero stress in the coating at room temperature after annealing, the right phosphorous content must be chosen, e.g., 5% on Al and 11.5% for Be. For use at lower or higher temperatures, different phosphorous contents should be selected to minimize the stress.

3.4 Inhomogeneity/Anisotropy of Properties

Most materials, as fabricated, are neither completely isotropic nor homogeneous. By that I mean that they are to some level anisotropic, i.e., having some preferred directionality of properties, and/or inhomogeneous, i.e., having a spatial variation in properties. Anisotropy of properties exists in pure single crystals of materials. Inhomogeneity of properties occurs in bulk and is a function of raw material fabrication processes.

Cubic materials generally have anisotropy in their elastic properties. For example, the Young's modulus of elasticity of pure iron varies with crystallographic direction from 19.2 Msi to 41.2 Msi. Similar variations in shear modulus and Poisson's ratio are also present. Comparable variations in these properties occur in other cubic materials such as Cu, Ni, Si, beta SiC, etc. When these materials are used in polycrystalline form, the variations average out and are not noticed. In components that are fabricated from single crystals, or applications that have crystallographic texture such as plated or chemically vapor deposited (CVD) materials, there can be substantial elastic property anisotropy and this should be included in any detailed deformation modeling of such components. Thermal properties such as CTE are isotropic for cubic materials. Thermal conductivity, which is also isotropic, is affected by grain size and grain boundaries so that for plated or CVD materials there is anisotropy. For example, CVD SiC has a deposition texture and both elastic modulus and thermal conductivity have approximately 15% anisotropy.

In a similar manner to the cubic materials, there are variations in elastic properties in non-cubic materials, i.e., hexagonal, rhombohedral, tetragonal, orthorhombic, etc. However the thermal properties of the non-cubic materials are anisotropic. For example, the CTE of Be is 38% higher in the basal plane than it is in the axial direction of a Be crystal or grain. This anisotropy leads to the microstructural strains listed above in Table 2. Some materials such as graphite and quartz have negative CTE in some directions and positive in others. The effects of anisotropy of thermal properties, for the various crystalline materials that exhibit it, can be minimized with a fine-grain, randomly oriented microstructure or a pseudo-isotropic layup of fibers.

Inhomogeneity can be attributed to spatial variations in chemistry, grain size and/or grain orientation, and many other factors. In general, CTE inhomogeneity in glasses and glass ceramics is due either to compositional variations or can be related to annealing history; for metals, ceramics and salts, it is due to compositional or microstructural

variations where the latter can be due to crystal orientation differences and/or the presence of other phases. For composites, CTE inhomogeneity, at least on a micro scale, is a given due to the presence of multiple phases. The inhomogeneity can be caused by reinforcement orientation or from variations in the concentration of the reinforcement. Care should be exercised in selection of multiphase materials that will be used over a temperature range to insure dimensional stability.

Components having CTE inhomogeneity can behave in the manner shown in Figure 3 at the beginning of this paper. In this particular case, the cause is a combination of Be powder with basal plane cleavage, a poor distribution of grain sizes and inappropriate consolidation parameters. Current Be fabrication processes have virtually eliminated these types of inhomogeneities through the use of impact ground or spherical powder with well controlled chemistries combined with hot isostatic pressing of the powder²⁷⁻²⁹.

4. LOW EXPANSION MATERIALS

Low expansion materials have been used for many years for precision instrumentation to minimize thermal dimensional changes in service. For optics, fused silica and fused quartz have replaced borosilicate glasses in many applications, and in recent years ULE and the low expansion glass ceramics have been used extensively for more critical applications. Invar alloys have been used in precision metering applications for many years, especially in combination with fused silica optics because of the excellent CTE match at room temperature. More recently composite materials, particularly graphite-epoxy composites, have been used for precision structures since they can be tailored for near zero expansion over a fairly wide temperature range. However each of these material types has distinct advantages and disadvantages. S.F Jacobs gave a good history of CTE measurements including inhomogeneity, hysteresis and temporal stability³⁰, and his paper in this volume³¹ goes into much more detail in the coverage of this subject.

One very important issue that has been raised³² is the ultimate stability that can be obtained in precision structures, particularly those fabricated from low expansion materials. In many cases the temporal instabilities far exceed the thermally induced distortions. Another aspect of this problem relates to instabilities in composites, where there is a substantial spatial variation in properties and residual stress within any component. Stabilization by thermal cycling is critical and temperature limitations for interfacial stresses exceeding strength limitations must also be considered.

Borosilicate glasses have been used for large mirror substrates, e.g., the 200 inch Hale telescope primary on Mt. Palomar, but these glasses have relatively high CTE, 3.0 ppm/°C and are no longer considered low expansion glasses. Lightweight mirror blanks continue to be fabricated from these glasses due to the relatively low fabrication temperature compared to fused silica and the utilization of more sophisticated blank fabrication methods^{33,34}. Most glasses and glass ceramics are chemically inert, can be highly dimensionally stable and can be used over fairly wide temperature ranges, although each material has a range where it performs best. ULE and the glass ceramics have essentially zero expansion at room temperature and fused silica has zero CTE at approximately 160K. All of these materials are brittle and have very low fracture toughness and fracture strength in tension, have low Young's modulus and relatively high density. Fabrication can be difficult for complex shapes, and attachments require clever methods and thermal compensation. Zerodur can have an instability in the range

of 150°C to 300°C⁸, but use of Zerodur M or careful control of heating and cooling in this temperature range eliminates any change in dimensions or CTE. Zerodur has been used extensively for metering bar/optical bench applications.

The invar alloys can have near zero CTE over a limited temperature range when given a three step heat treatment³⁵. Components can be readily fabricated using conventional metalworking processes and have high fracture toughness and moderate strength. Almost all invar specimens measured have temporal dimensional instabilities ranging from 1ppm/yr to as much as 50ppm/yr, depending on thermo-mechanical history. Super invar has a lower CTE than invar and very high dimensional stability^{31,36,37} but can not be used at low temperatures since it undergoes an irreversible phase transition at temperatures below -55°C that substantially changes its dimensions and properties³¹.

Composite materials, by their very nature have the disadvantage of having built in residual stress at the phase boundaries which usually lead to temporal dimensional instability. Graphite-epoxy composites can be designed and fabricated with near zero CTE, low density and high modulus. The properties can be tailored to the application, but are anisotropic due to the fabrication methods which require a preferred fiber direction or at best, a pseudoisotropic layup. The biggest drawback of graphite-epoxy is its moisture sensitivity which effects dimensions, CTE and mechanical properties. An excellent review of moisture and viscoelastic effects on the stability of composites has been given by Wolff¹⁰, and the recent SPIE conference on Design of Optical Instruments³⁸ contained three sessions on composites where much new information was presented on ways to minimize moisture effects.

While silicon carbide (SiC) is not as low expansion a material as those discussed above, its CTE is lower than the borosilicates, and its combined properties, particularly E/ ρ and CTE/D give it a unique combination of properties that should be considered for precision optical instruments. As shown in Table 1, there are several kinds of SiC whose properties are similar, but the fabrication methods and composition can be quite different. The single phase CVD SiC³⁹ is nominally 100% β SiC and since it is made by a deposition method, is well suited to forming faceplates, shells, tubes and other forms that can be replicated from a mandrel. There is normally some locked in stress in CVD SiC as evidenced by a slight springback from the mandrel. The other common form of SiC is a two phase form consisting of 60% to 88% α SiC (the hexagonal form) with the balance pure silicon that is called reaction bonded (RB), reaction sintered or siliconized SiC^{40,41}. This material can be cast into virtually any shape and depending on the processing methods can be formed to net shape. Components such as lightweight mirrors and optical benches have been fabricated from RB SiC, and limited testing has shown the material to be dimensionally stable. Much more testing is required before this developmental material can be qualified for nanospec applications.

5. PROMOTING DIMENSIONAL STABILITY

Now that I have given you some insight into the causes of dimensional instability, you should see that there are many potential pitfalls in the design and fabrication of dimensionally stable components. In order to avoid these pitfalls there is a sequence of actions that can be taken that should lead to stable components.

The first step is to establish a budget for the allowable dimensional change for each component in the system and allocate a tolerance to each element. Structural components usually do not have as tight a tolerance as optical components and all components of each type will not necessarily have the same requirements. Next consider the sources of dimensional change as they relate to the components to be fabricated. For example, if the system is to operate at reduced or elevated temperatures, then both thermal instability and thermal cycling instability are potential types of dimensional instabilities to consider, and the sources for such behavior are changes in internal stress, inhomogeneity and/or anisotropy, and microstructural changes. Consider the fabrication options and how they relate to the dimensional instability sources and the component performance requirements. With all these factors in hand, select the candidate materials and reevaluate the sources and fabrication methods for each material with respect to meeting the budgeted dimensional tolerances in the specified use environment.

In order to make the final materials and fabrication method selection, you need to demonstrate that external stresses will not cause excessive strain in the component; that internal stresses in the component due to fabrication methods or inhomogeneities and/or anisotropies will not change excessively; and that any microstructural changes will not cause excessive strains. While this method may sound complicated, once you understand the basic sources of potential instability and the magnitude of the possible changes for each of the common candidate materials and their respective fabrication methods, the selection process becomes almost intuitive. The difficult part is when you must produce a component that operates in an environment for which properties information is limited, or when designing components to nanotolerances.

A good example of how this process can be implemented is given by Hassler, et al⁴², for the design of a dimensionally stable composite position encoder scale. The stability requirements are defined, the conventional materials are reviewed and consideration is given to composites. Both the CTE and coefficient of moisture expansion (CME), and creep (temporal stability) are taken into account in the design. Fabrication methods were considered, test pieces were fabricated and tests confirmed the efficacy of the final design choices.

The bottom line is that dimensional instability of materials *is* critical in the design of optical instruments and must be considered early in the design process.

6. REFERENCES

1. C.W. Marschall and R.E. Maringer, Dimensional Instability, An Introduction, Pergamon Press, New York (1977)
2. R.A. Paquin, ed., Dimensional Stability, *SPIE Proc.* 1335 (1990)
3. D.A. Vukobratovich and R.A. Paquin, eds., Optomechanics and Dimensional Stability, *SPIE Proc.* 1533 (1991)
4. W.B. Emerson, "Secular length changes of gage blocks during twenty-five years", in Metrology of Gage Blocks, U.S. National Bureau of Standards, Circular 581, p. 71 (1957)
5. Ref. 1, p. 187
6. R. Dawson and D.G. Moffat, "Vibratory Stress Relief: A Fundamental Study of Its Effectiveness", *J. Eng. Matls. and Tech.* 102, p. 169 (April 1980)

7. R.A. Paquin, "Processing Metal Mirrors for Dimensional Stability", Wkshop on Optical Fab. and Test, Tech. Digest, Opt. Soc. of Amer., p. TB-1 (Dec 1981)
8. O. Lindig and W. Pannhorst, "Thermal Expansion and Length Stability of Zerodur in Dependence on Temperature and Time", *Appl. Optics* **24**, p. 3330 (1985)
9. Ref. 1, p. 23
10. E.G. Wolff, "Moisture and viscoelastic effects on the dimensional stability of composites", in Dimensional Stability, R.A. Paquin, ed., *SPIE Proc.* **1335** p. 70 (1990)
11. Ref. 1, p. 13
12. R.A. Paquin, moderator, "Dimensional stability workshop", in Dimensional Stability, R.A. Paquin, ed., *SPIE Proc.* **1335**, p. 217 (1990)
13. A.G. Ingram, M.E. Hoskins, J.H. Sovik, R.E. Maringer and F.C. Holden, "Study of microplastic properties and dimensional stability of materials", U.S. Air Force Materials Laboratory Report AFML-TR-67-232, Part II, Battelle Columbus Laboratories (1968)
14. V.A. Likhachev, "Microstructural strains due to thermal anisotropy", *Sov. Phys. Solid State* **3** p. 1330 (1961)
15. N.N. Davidenkov, V.A. Likhachev and G.A. Malygin, "Investigation of the irreversible thermal change in shape of zinc", *Phys Metals Metallogr. (USSR)* **10** p. 95 (1960)
16. R.S. Barker and J.C. Sutton, "Stress relieving and stress control", Ch. 10 in Aluminum, Vol. III, Fabrication and Finishing, K.R. Van Horn, ed., Amer. Soc. for Metals, Metals Park, OH (1967)
17. W.P. Koster, M. Field, L.J. Fritz, L.R. Gatto and J.R. Kahles, "Surface integrity of machined structural components", U.S. Materials Laboratory Report AFML-TR-70-11, Metcut Research Associates, Inc (1970)
18. M. Field and W.P. Koster, "Surface integrity in conventional machining chip removal processes", Society of Manufacturing Engineers Paper EM 68-516 (1968)
19. C.W. Marschall and R.E. Maringer, "Stress relaxation as a source of dimensional instability", *J. Mater.* **6** p. 374 (1971)
20. J.W. Bender and R.L. Wahl, "Work-induced stress and long-term stability in optically polished silicon", in Optomechanics and Dimensional Stability, D.A. Vukobratovich and R.A. Paquin, eds., *SPIE Proc.* **1533** p. 264 (1991)
21. I.Kh. Lokshin, "Heat treatment to reduce internal stresses in beryllium", *Metal Sci. Heat Treat. (USSR)* p. 426 (May, 1970)
22. H.Y. Hunsicker, "The Metallurgy of heat treatment", Ch. 5 in Aluminum, Vol. I: Properties, Physical Metallurgy and Phase Diagrams, K.R. Van Horn, ed., Amer. Soc. for Metals, Metals Park, OH (1967)
23. K. Parker and H. Shah, "Residual Stresses in Electroless Nickel Plating", *Plating* **58** p. 230 (Mar. 1971)
24. K. Parker, "Effects of Heat Treatment on the Properties of Electroless Nickel Deposits", *Plating and Surface Finishing* p. 71 (Dec. 1981)
25. D.L. Hibbard, "Dimensional stability of electroless nickel coatings", in Dimensional Stability, R.A. Paquin, ed., *SPIE Proc.* **1335** p. 180 (1990)
26. K. Parker and H. Shah, "The Stress of Electroless Nickel Deposits on Beryllium", *J. Electrochem. Soc.* **117** 8 p. 1091 (1970)
27. R.A. Paquin, "Hot Isostatic Pressed Beryllium for Large Optics", *Opt Eng.* **25** 9 p. 1003 (1986)

28. R.A. Paquin, "New technology for beryllium mirror production", in Current Developments in Optical Engineering and Commercial Optics, R.E. Fisher, H.M. Pollicove, W.J. Smith, eds., *SPIE Proc.* **1168** (1989)
29. R.A. Paquin and G.J. Gardopée, "Fabrication of a fast aspheric beryllium mirror", in Reflective and Refractive Optical Materials for Earth and Space Applications, M.J. Riedl, R.R. Hale and T.B. Parsonage, eds., *SPIE Proc.* **1485** p. 39 (1991)
30. S.F. Jacobs, "Unstable Optics", in Dimensional Stability, R.A. Paquin, ed., *SPIE Proc.* **1335** p. 20 (1990)
31. S.F. Jacobs, "Variable invariables - dimensional instability with time and temperature", in Optomechanical Design, P.R. Yoder, ed., *SPIE Proc.* **CR43** (1992)
32. B.P. Dolgin, J. Moacanin and T.P. O'Donnell, "Theoretical limits of dimensional stability for space structures", in Optomechanics and Dimensional Stability, R.A. Paquin and D. Vukobratovich, eds., *SPIE Proc.* **1533** p. 229 (1991)
33. L. Goble, J.R.P. Angel and J.M. Hill, "Steps toward 8m honeycomb mirrors VII. Spin casting an experimental f/1 1.8m honeycomb blank of borosilicate glass", in Large Optics Technology, G.M. Sanger, ed., *SPIE Proc.* **571** p. 92 (1985)
34. R.K. Melugin, J.H. Miller, J.R.P. Angel, P.A.A. Wangsness, R.E. Parks and D.A. Ketelson, "Development of lightweight, glass mirror segments for the Large Deployable Reflector", in Large Optics Technology, G.M. Sanger, ed., *SPIE Proc.* **571** p. 101 (1985)
35. B.S. Lement, B.L. Averbach and M. Cohen, "The dimensional behavior of invar", *Trans. ASM* **43** p. 1072 (1951)
36. S.F. Jacobs, J.N. Bradford and J.W. Berthold, "Ultraprecise Measurement of Thermal Coefficients of Expansion", *Appl. Optics* **9**, 11, p. 2477 (1970)
37. S.R. Patterson, "Dimensional Stability of Superinvar", in Dimensional Stability, R.A. Paquin, ed., *SPIE Proc.* **1335** p. 53 (1990)
38. Design of Optical Instruments, D.M. Aikens, V.L. Genberg, G.C. Krumweide and M.J. Thomas, chairs, *SPIE Tech. Conf.* **1690**, Orlando (1992)
39. M.A. Pickering, R.L. Taylor, J.T. Keeley and G.A. Graves, "Chemically Vapor Deposited Silicon Carbide (SiC) for Optical Applications", in Space Optical Materials and Space Qualification of Optics, R.R. Hale, ed., *SPIE Proc.* **1118** p.2 (1989)
40. R.A. Paquin and M.B. Magida, "Low scatter surfaces on silicon carbide", in Laser Induced Damage in Optical Materials: 1989, H.E. Bennett, A.H. Guenther, L.L. Chase, B.E. Newnam and M.J. Soileau, eds., *NIST Spec. Publ.* **801** p. 256 (1990)
41. M.B. Magida, R.A. Paquin and J.J. Richmond, "Dimensional stability of bare and coated reaction-bonded silicon carbide", in Dimensional Stability, R.A. Paquin, ed., *SPIE Proc.* **1335** p. 60 (1990)
42. W.L. Hassler, Jr. and R.K. Nakayama, "New low-cost dimensionally stable composite linear absolute position encoder scale", in Dimensional Stability, R.A. Paquin, ed., *SPIE Proc.* **1335** p. 80 (1990)