

Rolling friction

Warren D. Smith, August 2022. warren.wds@gmail.com

ABSTRACT. We theoretically examine energy losses experienced during rolling, then compare versus experiments (which we criticize), and suggest applications such as building better pendulum clocks, and draw conclusions such as offering new better advice about wheel-engineering. As far as I know this is the best treatment available. In particular we resolve the >180-year-old Coulomb-Dupuit controversy. by explaining their disagreement both with each other, and versus (better than their) experiments.

1. Introduction

The **wheel** often is claimed to be one of the greatest human inventions (believed to date to 3100-4500 BC in Mesopotamia). So it would be nice to understand it.

The Problem (and Notation). A wheel (several different kinds will be considered) with diameter D , rolls along a "road" at speed V ; the two are pressed together by weight W . Ideally, the wheel and road each would be rigid and rolling would engender zero friction. In reality, both are elastic and there is some frictional "drag" force F , which must be countered to maintain the rolling. The problem is to predict F as a function of D , W , what kinds of wheel and road we have, and what materials they are made from.

We'll consider three types of wheel: (1) solid cylinder, (2) thin-walled cylindrical pipe, (3) solid ball. We will not treat spoked wheels per se, but they are something of a hybrid of (1) and (2). Also, the weight W could be applied (a) at the wheel hub, or (b) just because the entire wheel weighs W , or (c) via another road (which weighs W) riding on top of the wheel, in which case perhaps a better word than "wheel" would have been "roller" or "bearing." One of the advantages of our simple "dimensional analysis" approach is that we do not actually care which – all three a,b,c shall yield the same laws, up to multiplicative constant factors which we usually have no intention of evaluating exactly. Our goal is merely to find the correct form of the laws in the limit of something (most usually the elastic strain) being small; we leave the precise values of the constants to be determined by experiment. It's often simplest and most symmetrical if the "road" is identical to the wheel (i.e. we want to understand the "drag" for two identical wheels rolling against each other, pressed together by weight W), but changing that to a flat road again will only alter the friction laws by multiplicative constants. Indeed, by reflection symmetry a wheel rolling on an ideal perfectly-rigid flat road ought to be mathematically *identical* to two cloned wheels rolling against each other.

To measure how "**elastic**" and "**hard**" the wheel and road are, we'll use their [Young's moduli](#) Y and their "plastic deformation stress threshold" aka "yield strength" P . There are other elastic constants too, e.g. bulk modulus, Poisson ratio, etc, but we mostly shall not examine them and again generically do not expect them to alter matters by more than a constant factor versus just basing everything on Young's modulus. For example, Hertz 1882 thinks he found the exact leading term for the displacement distribution for two contacting almost-rigid elastic balls, and his formulas as recounted by Timoshenko & Goodier are functions only of the "effective elastic modulus" $\tilde{Y}=Y/(1-$

σ^2) where σ is the material's [Poisson ratio](#). For most real isotropic materials $0 < \sigma < 1/2$, in which case this causes an error in our pseudo-Y bounded by a factor in $[3/4, 1]$. But it must be admitted that it is theoretically allowed for isotropic Poisson ratios σ to be anywhere in $[-1, 1/2]$, and artificial mathematical constructions (which can be instantiated in triply-periodic "3D printed" artificial "materials") do approach -1 arbitrarily closely. (Also *anisotropic* materials can have Poisson ratios below -1 or above $1/2$ and as great as 1.99, in the right special directions.) A hypothetical hard isotropic material with $\sigma = -0.99999$ would presumably, according to Hertz, make a better ball bearing, but this is not an issue in practice since for practical purposes no such material exists.

However, there is one thing I shall insist on. Our wheels must be made of an **isotropic** substance, e.g. amorphous, or polycrystal with grain size smaller than D and any other length scale that matters, so that they are **rotationally symmetric** and hence during constant-speed operation we assume/expect a **steady state** to analyse. If the wheel were a single crystal, or made of highly anisotropic materials such as woods or aligned polymers, then the answers would depend on orientations, the wheel could highly asymmetrically be happy to squash in one direction but resist squashing in the perpendicular direction, there could be anisotropic resonant excitations, etc – and I have no intention of entering that analysis-hell.

We shall first analyse rolling friction, then discuss how to minimize it, applications, and comparison versus historical experiments. Despite the simplicity of our arguments and the age of this topic, apparently-new conclusions will be reached; and I do not think anybody has yet had the full breadth of understanding, nor in such a simple way, as we shall explain here.

2. Derivation of rolling-friction laws

There are several **reasons** for rolling friction, each of which can be predominant: (1) "trench digging," (2) adhesion, (3) dust particles, cobblestones, and noncircularity, (4) sliding, (5) elastic hysteresis losses (of various kinds).

We have ordered these simplest-first. We shall examine and propose laws for them. The last is the most interesting, and I would claim the first and last are the most important. All our arguments shall be based on simple geometry, the fact $\text{work} = \text{energy} = \text{force} \times \text{distance}$, and elementary "dimensional analysis." No solving of partial differential equations or exact evaluation of difficult integrals will be needed.

Trench digging. It is easy to ride your bicycle along a hard paved road. But when the road switches to unpaved soft sand, your pedalling effort greatly increases! Evidently, the predominant cause of rolling friction is the softness of the sand. If your bike is stationary, your weight W presses down on that sand, causing your wheel (assumed rigid) to sink into it to some depth Z , "digging a trench" Z deep and some distance S long. From geometry (because the bottom of a circular wheel $x^2 + y^2 = D^2/4$ is locally a quadratic curve $y + D/2 \approx x^2/D + x^4/D^3 + \dots$), we see that S , if Z is small in comparison to D , has order $(DZ)^{1/2}$. So after you ride distance L , you've "dug a trench" Z deep using energy of order WZL/S . We'll assume a constant fraction of that energy is lost to heat. Hence $F = \text{energy} \cdot \kappa / L$ where κ is a dimensionless constant, so

$$F_{\text{tren}} = \kappa W D^{-1/2} Z^{1/2}.$$

Adhesion. If you roll a *cylinder* across a tabletop, the friction is tiny. Now coat the tabletop with sticky flypaper glue and try that again – it stops rolling almost immediately! Obviously, by far the predominant cause of this is the sticky glue. It is known that liquid surfaces (between two liquids, or liquid-gas, or liquid-solid) are characterized by a "surface energy density" also called by the stupider, but unfortunately more common, name "surface tension." I.e. the energy of such a surface equals its area times a constant. That constant depends on the two materials and can be of either sign. If the wheel-road-contact surface obeys the same sort of energy law, and if we suppose that whenever that contact is created/destroyed a constant fraction of its energy is lost (i.e. converted to heat), then for a rolling cylindrical wheel of width Δ , we predict

$$F_{\text{adh.cyl.}} = \kappa \Delta$$

where κ is a constant having dimensions energy/area=force/distance.

For a rolling *ball*, the calculation is more interesting. A rigid sphere would contact a rigid plane road at a single point. However, that single point (zero area) would support the nonzero weight W of the ball, yielding infinite pressure. Real balls are not rigid enough to withstand infinite pressure. In fact, the ball deforms to get a contact surface of nonzero area ("Hertzian contact"; two identical balls would by symmetry have an exactly-flat circular contact patch). If this area is A , then the pressure is of order W/A , and from local-quadratic geometry we see that the contact patch is pushed flat by a vertical displacement of order A/D . That causes strain of order A/D^2 throughout the ball's volume of order D^3 , or strain of order $A^{1/2}/D$ throughout a volume of order $A^{3/2}$, the latter being the one that matters in the sense it stores much more elastic energy. [Under linear elastic theory, elastic energy is proportional to $\int (\text{strain})^2 dx dy dz$.] If strains are small enough for linear elasticity theory to apply, the integrated elastic energy is of order $YA^{5/2}D^{-2}$, which must be of the same order as WA/D , hence A is proportional to $(WD/Y)^{2/3}$ and hence the width Δ of the contact patch, which has order \sqrt{A} , must be asymptotically proportional to $(WD/Y)^{1/3}$. Hence

$$F_{\text{adh.ball}} = \kappa \cdot (WD/Y)^{1/3}$$

where κ again is a constant having dimensions energy/area.

Dust particle. Assume a rigid wheel on a rigid rail encounters a tiny (width ϵ) rigid dust particle. This happens when their centers are horizontally separated by distance $\approx [D\epsilon]^{1/2}$. In order to rise above the dust particle, the weight W needs to be lifted by ϵ , costing energy. Indeed, if the wheel initially is stationary, then from local quadratic geometry we see the horizontal force F needed to initiate motion is

$$F_{\text{dust}} = (\epsilon/D)^{1/2}W.$$

This calculation was already made by T.Tredgold in 1825. Tredgold mentioned experiments by Palmer in which a slight covering of dust on a tramway increased rolling friction by nearly 20%. However, that calculation ignored the possibility that energy is returned upon descending on the other side of the dust particle, and tells us the *maximum* force needed during the wheel's adventures, not the *average*. One way to overcome those objections would be to claim/assume that: (i) a constant fraction κ of the energy is somehow lost to heat (perhaps by moving other dust

particles?) and not recovered by the descent on the other side of our dust particle: (ii) the dust is magically distributed at just the right density so that as soon as we stop descending from one dust particle, we begin surmounting the next. That all sounded rather dubious, but if so then we would deduce for the distance-averaged force

$$F_{\text{dust avg}} = \kappa W/D.$$

Cobblestones. Here is a more-realistic redo of Tredgold not subject to the preceding objections. I point out that riding a bicycle on cobblestones seems to require substantially more power than on a flat road; and if you bicycle faster than a certain threshold speed $V_{\text{thresh}} = (Dg/2)^{1/2}$, which for today's bicycles is about 2 meters/sec \approx 6 km/hour (which is slower than anybody bicycles) then your wheel flies off the end of the cobble into open air, thus *not* usefully recovering energy via a controlled descent on the other side of the cobble. Assume cobbles have flat tops and a standard length L , and let ϵ now denote the typical intercobble height discrepancy. If $V < 2^{-1/2}(g/\epsilon)^{1/2}(L - [D\epsilon])^{1/2}$ then the bicycle wheel will fall onto a new (lower) cobble *without* overflying it and still with enough room to begin the next rise. If $\epsilon = 1\text{mm}$ and $L = 30\text{cm}$ and $D = 622\text{mm}$ (the former two are not at all unreasonable for modern paving [stones](#), the latter is a standard bicycle wheel diameter) then the right hand side is 19 meters/sec \approx 69 km/hour, considerably faster than you are likely to bicycle. (As of 2020 the bicycling world hour record is 55089 meters by Victor Campenaerts.)

Then the cobblestone-redo of Tredgold finds that F averaged over travel distance is

$$F_{\text{cobblestones}} = \kappa W \epsilon / L,$$

(where κ is a dimensionless positive constant) provided $(Dg/2)^{1/2} < V < 2^{-1/2}(g/\epsilon)^{1/2}(L - [D\epsilon])^{1/2}$.

If instead of a circular wheel rolling on cobblestones we consider a slightly **noncircular wheel** (e.g. regular N -gon) rolling on a straight line, a more elegant version of essentially the same theory will happen. One interesting feature of this theory is its behavior in the limit of high rolling speed V . The wheel will continually bounce off of and back onto the line. These bounces have vertical speed proportional to N^2V , hence vertical kinetic energy component of order N^4V^2 , perhaps times W , and duration between bounces proportional to N^2V/W during which time the wheel travels horizontal distance proportional to N^2V^2/W . So if each bounce loses a constant fraction of its vertical kinetic energy (Newton "coefficient of restitution" model of bouncing) we expect energy loss per distance rolled to be

$$F_{\text{noncirc.bounce}} = \kappa N^2 W^e.$$

where $e=1$ or 2 , and where $\kappa > 0$ is a constant for any particular wheel which if $e=1$ is dimensionless. These cobblestone and noncircularity friction mechanisms, since "dynamic" rather than "steady state," really go outside the scope of this paper and do not, strictly speaking, concern mathematical "rolling" at all, but rather "bouncing" in which departures of the wheel from the road are logically essential.

Sliding while rolling. Although a rigid wheel on a rigid flat road would not slide at all, if the wheel rolls within a *groove* then the sides of the groove will rub against the wheel. Furthermore, even with

a flat road, with imperfect rigidity the flexing of the wheel and road actually can cause some sliding between them. The latter was proposed as a cause of rolling friction by O.Reynolds in 1875, and supposedly he verified that this effect exists by experiments showing "creep" of india rubber coatings. Call this effect "**Reynolds creep**."

Elastic hysteresis loss (and some more Trench digging). Although each of the preceding causes of rolling friction *can* be predominant in contrived situations (with the plausible exception of Reynolds creep), in most machine-design situations where we can intentionally design to reduce friction, it is trivial to make them **negligible**. For example, if we use hard materials for the road and keep stresses below levels that cause plastic deformation, then no "digging" of permanent "trenches" happens. (Nevertheless whenever loads W are heavy enough to make "plastic deformation" important, then "trench digging" effectively will happen; we'll discuss that more soon.) If we place pieces of metal foil on a railroad track, then roll a train over them, they do not even get "picked up," demonstrating adhesion is negligibly tiny. By polishing our wheel and road arbitrarily smooth and perfectly shaped, "cobblestone" and "noncircularity" friction could be reduced arbitrarily small. We simply do not roll wheels in narrow grooves so they can slide, because we are not idiots. Because rolling friction seems largely unaffected by lubricating films (and indeed to the extent lube has an effect, it experimentally almost always *increases* rolling friction) it is obvious Reynolds creep is negligible.

In those "engineered for low friction" situations, the main cause of rolling friction is elastic hysteresis losses. That is: consider a lump of material. Stress and hence strain it, thus storing some amount of elastic energy. Release the stress and it bounces back. Each such cycle loses some energy – it is not returned 100% efficiently.

Let H be the fraction of energy lost. The **simplest hypothesis about H** would be that it is a constant, $0 < H < 1$, that depends only upon the material, but not on the strain (at least for strains small enough for linear elasticity theory to be approximately valid) and also not on the frequency of the cycling (at least for all sufficiently-low frequencies). And furthermore, for "hard, elastically efficient" materials, i.e. ones an engineer would select, $H \ll 1$. In this simple model $H = 2\pi \tan(\theta_{lag})$ where θ_{lag} is the phase-angle lag between stress and strain in sinusoidal vibrations. For this reason $(2\pi)^{-1}H$ sometimes is called the "loss tangent."

This "simplest hypothesis" about H can be correct – but also can be incorrect! We'll now survey some phenomena which cast light on this.

1.Bouncing. When two hard objects bounce off each other, a fraction $H = 1 - C^2$ of their original kinetic energy is lost to heat where C with $0 < C < 1$ is called the "**coefficient of restitution**." This model was known and experimentally tested already by Isaac Newton, who stated in his [Principia](#) that he'd found $C \approx 5/9 \approx 0.56$ for tight balls of wool, $C \approx 15/16 \approx 0.94$ for glass balls, and "almost 1" for steel balls. It predicts an exponentially decreasing sequence of bounce-heights for a ball off a hard floor, with height ratio \sqrt{C} and time-of-flight ratio C . The total time T_∞ to complete the infinite number of bounces equals the first fall time $T_1 = (2h/g)^{1/2}$ for a ball dropped height h in gravitational acceleration g , *times* $(1+C)/(1-C)$ if bouncing is falsely modeled as instantaneous and if there is no air resistance. The neglect of air resistance means that really, this only yields a *lower bound* $C \geq (T_\infty - T_1)/(T_\infty + T_1)$. For example, in a [video](#) by [Liquidmetal.com](#), hard steel balls are shown bouncing in

air on stainless steel, titanium, and bulk metallic glass plates with initial fall time about 0.35 seconds ($h \approx 60$ cm, first impact speed $v \approx 340$ cm/sec). They stop bouncing after about 20, 25, and 81 seconds respectively, from which I compute $C \geq 0.9656$, $C \geq 0.9724$, and $C \geq 0.9914$. Reasons C is so large for Liquidmetal.com's bulk metallic glass product (an [alloy](#) of Zr,Cu,Ni,Nb,Al) presumably include its high yield stress 1.2GPa and yield strain 1.6% and high Vickers-indentor-hardness score 460 (versus stainless steel 0.22GPa, 0.1%, 152 and titanium 0.24GPa, 0.2%, 145) and the fact that since it is amorphous, it has neither crystal grain boundaries nor crystal "slip planes" to serve as "weak spots" for plastic deformation.

Newton's model agrees with our "simplest hypothesis" if C is *constant*, i.e. independent of impact speed v . I have seen many undergraduate experiments on this, using, e.g. 1 inch diameter balls and 1 meter drop heights, which unfortunately ignore air drag and therefore most of whose conclusions about alleged slight decrease of C as a function of v are unjustified. Experimentally C indeed appears roughly constant for all small-enough impact speeds. For example Lifshitz & Kolsky 1964 used electrical timing to measure C for impacts of a metal pendulum bob against a massive metal block (each impact closed an electrical circuit, allowing very precise timing). Their fig.1 p.38 shows $C \approx 0.94$ is constant for steel balls ($\frac{1}{2}$ -inch diam.) impacting a stationary 23kg mild-steel block at speeds $6 \leq v \leq 30$ cm/sec, provided they pre-polish all surfaces to a smooth mirror finish. (With less-smooth surfaces, they measure smaller C 's with much greater variation.) But there is a clear "corner" in their $C(v)$ curve at $v = 30$ cm/sec, with plastic deformation ("yield") thereafter causing C to decrease, e.g. $C \approx 0.65$ at 140 cm/sec. Lifshitz & Kolsky claim that for a $\frac{1}{4}$ -inch (6.35mm) diameter steel ball hitting a steel block at speed 46 cm/sec, stresses > 1.4 GPa happen in the contact region (i.e. 5-7 times the plastic yield stress threshold for typical steels, hence well outside the regime of applicability of linear elasticity theory), with impact duration < 30 microseconds. C can be near zero for very plastic materials such as soft clay. The Leeb "rebound hardness" tester measures C for a tungsten carbide ball hitting any thick smooth surface. For two hard polished steel balls (diam=22.3mm) bouncing slowly and symmetrically off each other in a commercial "Newton cradle" apparatus, Cross 2018 found using 300 frame/sec video analysis, and after correcting for drag, that $C = 0.9976 \pm 0.0002$, and confirmed that amplitude indeed decayed quite-precisely-exponentially with time throughout a 35:1 energy range (each ball having speed 47-277 mm/sec), i.e. confirming the constancy of C throughout this range. For 2.5cm diameter balls made of various metals, C appears to be constant provided $v \leq 8$ cm/sec in measurements by Baca et al 2016, but again can decrease substantially at faster impact speeds. Materials with high plastic-yield strength tend to have high C . For example Baca et al found $C = 0.27-0.45$ for annealed copper at speeds between 105 and 3 cm/sec, but if we adulterate the 100% copper with about 5% tin and 0.2% phosphorus to get "phosphor bronze," which is similar in density but much harder and with $10 \times$ the yield strength, then $C = 0.67-0.99$ under the same conditions. For most balls dropped ≤ 1 meter, $0.25 < C < 0.99$.

2. Resonator Q. It is commonly observed that low-strain vibrating mechanical resonators, such as bells or quartz crystal oscillators, exhibit slow amplitude decay that is an exponential function of time. Indeed, this is customarily described by "Q", where the oscillator loses a fraction $H = 2\pi/Q$ of its stored energy per cycle. Bells have $Q \approx 1000$ while high quality quartz oscillators in vacuum can have $Q > 10^5$. This is entirely compatible with our simple hypothesis.

3. Thermodynamics. Next, consider an ideal gas in a cylinder. Adiabatically compress it by a factor $(1+\epsilon)$ using a piston. (Or, if $\epsilon < 0$ this is a rarefaction, not compression.) That alters the pressure p by a factor of $(1+\epsilon)^\gamma$ where $\gamma > 1$ is the [ratio](#) of the specific heat at constant pressure, divided by specific

heat at constant volume, for that kind of gas. By evaluating the integral we find that performing this compression costs mechanical energy $(\gamma-1)^{-1}[(1+\epsilon)^{\gamma-1}-1]+(1+\epsilon)^{-1}-1$ in units where the gas originally contains thermal energy 1. In the limit $\epsilon \rightarrow 0$ of small strains this is $(\gamma/2)\epsilon^2 - (1/6)(5-\gamma)\gamma\epsilon^3 + O(\epsilon^4)$.

γ : For an ideal monoatomic gas (He, Ne, Ar) $\gamma=5/3 \approx 1.67$, for a diatomic gas $\gamma=7/5=1.40$ (CO, N₂, O₂), for triatomic gases $\gamma=4/3 \approx 1.33$ (N₂O, H₂O). For real polyatomic gases such as CO₂, CH₄, SF₆, C₂H₆, benzene, and hexane smaller values $1.06 \leq \gamma \leq 1.33$ arise; and for non-ideal gases near to becoming liquids or solids, γ greater than 1.67 can arise, such as $\gamma=1.76$ for Ar at 93°K.

If we now allow our compressed gas to cool back to its original temperature T, then in view of the ideal gas law equation of state $pV=nkT$ (where p =pressure, V =volume, n =#gas molecules, k =Boltzmann constant) the pressure falls to only $(1+\epsilon)$ times its original value. Suppose we now attempt to recover as much as possible of our mechanical energy by isothermally re-expanding the gas by a volume-factor of $(1+\epsilon)$, thus restoring the original geometry. This recovers energy $(1+\epsilon)^{-1} + \ln(1+\epsilon) - 1$. In the $\epsilon \rightarrow 0$ limit this is $\epsilon^2/2 - (2/3)\epsilon^3 + (3/4)\epsilon^4 - \dots$

Hence in the limit $\epsilon \rightarrow 0$ of small strains, the fraction of supplied energy that is lost is $H = [(\gamma-1)/\gamma] [1+\epsilon/3 + \dots]$ This is independent of the frequency of the compression cycles for all frequencies low enough to allow cooling; albeit at higher frequencies H should decrease. But keep in mind that for *large* strains H becomes *nonconstant*, and increases, ultimately approaching 1 from below.

For a "solid" wheel made of closed-cell *foam*, I expect this gas-compression/heating effect will be the predominant cause of elastic hysteresis losses (because the solid gets distorted far less than the gas). And indeed, at least for sound between about 100 and 1000-2000 Hz, most materials that are commonly used today (or considered for future use) for sound dampening in home construction, including silicone rubber open-cell foams, hemp, fibreboard, "Linacoustic RC" duct liner, polymerizing spray-foam insulation, and mineral wool, all exhibit acoustic transmission losses that rise roughly linearly with frequency, i.e. are entirely compatible with our "simplest hypothesis"; and all these materials indeed are gas/solid mixtures.

But even for gas-free solids and liquids, the same sort of effect occurs (just with smaller coefficients): solids have greater specific heats under constant-pressure than constant-volume conditions ($C_p > C_v$); therefore a solid heats when compressed; then if it cools to the original temperature, some amount of the mechanically-supplied energy is lost to heat and not recovered by decompression. For small strains, the lost energy-fraction is $H = (C_p - C_v)/C_v$. It is known from thermodynamics that $C_p - C_v = 9Tp^{-1}\alpha^2B$ where T =temperature, ρ =density, α is the [coefficient](#) of thermal linear-expansion (hence for volume-expansion use 3α), and B is the [bulk modulus](#). So if you want to diminish this effect, use a low- α material such as diamond, fused quartz, or invar. However experiment shows this effect usually is only a minor contributor ($\approx 10\%$) to elastic losses, even in fairly high- α metals such as brass. This effect generally will require less time to operate in *inhomogeneous* materials, e.g. containing grains of different compositions, because of rapid heat conduction from hotter to colder grains.

4.Seismology. If a sound wave is transmitted a long distance, its amplitude is observed to fall exponentially with distance even though in an ideal elastic medium there would be no such loss of

energy. In earth mantle rocks, a constant factor loss per acoustic cycle is typically observed at all frequencies from 100 to 1000 Hz; and in rocks tested in the lab, at all frequencies from 1 to 10^6 Hz – just as expected from our "simplest hypothesis" and the above "Q" picture. I.e. the energy fraction absorbed per unit length is proportional to frequency. It is likely that most of the acoustic loss in rocks in labs at atmospheric pressure is due to friction across cracks. But in rocks ≥ 30 km deep at pressures ≥ 10 kbar, that will be unimportant because the cracks will close. Chemical relaxation due to e.g. molecular bond rearrangements at crystal *grain boundaries* is suspected to be the predominant cause of acoustic loss in earth's mantle. That view seems highly supported by the fact that polycrystalline rocks experimentally have ≥ 10 times more acoustic loss than mineral single crystals. This also leads to the conclusion that if you want more efficient wheels (at least under *light* loads W), then if they are polycrystalline materials you should use *larger grain sizes*. Using N times larger grains will cause N times smaller H -values until the grains become so large that single-crystal bulk effects dominate the picture. Beyond that point, further grain-enlargement will cease to help.

But for *heavy* loads W where wheels enter the plastic deformation regime, it might (paradoxically) be more efficient, at least in some materials, to use small grain sizes, because grain boundaries can serve as a barrier to crystal dislocation plastic deformation and thereby actually increase yield strength. The sound intensities in seismology usually are too tiny for plastic deformation to be relevant.

5.High frequencies. However, our "simplest model" is *not* what happens for high-frequency sound traveling through air, water, or earth consisting of solid particles mixed with water and/or gas. The energy fraction absorbed per unit length due to viscosity and thermal conduction in a classical ideal monatomic gas (or even a solid, cf. Landau & Lifshitz 1970 EQ 35.3) is proportional to *squared* frequency; but in dry air at 20°C effects due to molecular relaxation of N_2 and O_2 molecules dominate that at frequencies below 1000 Hz for N_2 and below 60000 Hz for O_2 (albeit high humidity in the air can greatly diminish absorption below about 500 Hz). In ocean water, the main effect absorbing sound below 10 kHz is thought to be chemical relaxation of boric acid in the water, and between 10-500 kHz magnesium sulfate. Above 500 kHz – or in fresh water, which lacks those two chemicals – viscosity dominates, yielding absorption again proportional to *squared* frequency. Those effects for us would correspond to *nonconstant* H increasing linearly proportionally to frequency. However, obviously that linear growth cannot continue forever, since by definition $H < 1$. E.g. certainly once we reach sound-frequencies with wavelengths comparable or smaller than gas-molecule mean free paths, we expect $H \approx 1$. In air at sea level the mean free path is about 60 nm, corresponding to frequency $f \approx 5$ GHz.

6.Plastic deformation. In metals, small strains cause only small energy losses because "bounce back" restores most of the energy. However, strains larger than the "plastic deformation threshold," aka "yield strain," cause permanent deformations that do not "bounce back." Yield strains are 3% for nylon 66, 1-2% for PTFE, 1.6% for liquidmetal.com's bulk metallic glass, but 0.1-0.4% for most metals in common use.

Therefore we expect H to remain constant for small strains, but to be much larger, approaching 1, for strains large enough to cause plastic deformation. Based on the ideal gas example we might expect H rising for large strains proportional to the $(\gamma-1)$ -power of the strain, where for the gases discussed this exponent ranges between 0.06 and 0.76; *but* always obeying $0 < H \leq 1$. However, if so

one might expect considerably greater such powers for solids and liquids than for ideal gases, because for solids and liquids very strong interatomic repulsion forces will come into play at, say, 50% compression, whereas "ideal" gases have no interatomic forces at all, no matter how great the compression. (And indeed gases near to becoming liquids or solids experimentally have greater γ .) The wheels of heavily loaded railroad coal cars indeed operate at such high stresses that plastic deformation continually occurs near their rims, albeit over time this averages out rotationally so that they stay round. (It is fortunate for them that ductile metals enjoy a substantial strain range in which they "plastically deform" rather than "fracture.") H can change over time too, e.g. both rails and coal car wheels "work harden" over time.

We naively expect (based on all examples so far, and **7** the theory of low-Reynolds-number flows of viscous liquids) that H 's frequency dependence will become *linear* at all high-enough frequencies, albeit *constant* at all low-enough frequencies.

The **simplest function** of frequency f with those properties is $H(f)=H_0+(cf)^2/(1+bf)$ where H_0 , b , and c are positive constants (b and c have the dimension of time, while H_0 is dimensionless). This function obeys $H(0)=H_0$ and $\lim_{f \rightarrow \infty} H(f)/f=c^2/b>0$, and $H(f)$ is monotone increasing and concave-U.

Now that we have some understanding of what elastic hysteresis is and where it comes from, we are ready to...

Derive the rolling friction laws expected from elastic hysteresis. We'll mostly assume linear elasticity theory, but in some cases instead resort to a simple ad hoc model of plastic deformation. In all cases the hysteretic loss per cycle will be a fraction $H(f,S)$ of the stored elastic energy, where we regard H as a function of cycle-frequency f and maximum strain S , and H ought to be *constant* for all small-enough S and f ; but considerably larger for large S , perhaps growing for intermediate S proportionally to S^e for some positive real exponent e , and growing proportional to f for intermediate frequencies f ; but those growths cannot continue forever because $0<H<1$.

At the high- S end, where plastic deformation occurs, the situation instead should be described by our above "trench digging" model with whatever "sinking depth" Z happens empirically as a function of W (we'll also be able to predict Z theoretically); and similarly at high frequencies f , corresponding to high rolling speeds, our sinking depth Z would be an empirical function of both the rolling speed and the weight W . Obviously "harder" materials, as measured by P or by the Vickers indenter and similar tests, will be the ones with the least Z and hence greatest wheel-efficiency in the plastic deformation regime.

Solid cylinder. Assume either that the width Δ of the cylindrical wheel is large in comparison to its diameter D , or that the wheels ride on rails of the same width, so that we can treat this as a 2- and not 3-dimensional problem. The wheel-circle resting on the road is elastically distorted so that we get a contact region of nonzero width L . The pressure on it then is of order $W/(L\Delta)$ instead of ∞ , and from local-quadratic geometry we see that the contact patch is pushed flat by a vertical displacement of order L^2/D . That causes strain of order L^2D^{-2} throughout the cylinder's volume of order $D^2\Delta$, and a strain of order L/D throughout a smaller volume of order $L^2\Delta$. Either (if these strains are small enough for linear elasticity theory to apply) causes elastic energy of order $YL^4D^{-2}\Delta$, which must be of the same order as WL^2/D , hence L is proportional to $W^{1/2}Y^{-1/2}D^{1/2}\Delta^{-1/2}$.

Hence what I'll call the **naive total stored elastic energy** has order $W^2Y^{-1}\Delta^{-1}$. We assume that when the wheel rolls distance L that constitutes one hysteretic cycle, losing fraction H of its stored energy. Hence

$$F_{\text{solid.cyl.naive hyst.}} = \kappa W^{3/2}Y^{-1/2}D^{-1/2}\Delta^{-1/2} \cdot H(f,S)$$

where κ is some dimensionless positive constant and where we regard H as a function of cycle-frequency f and maximum strain S . (Actually, we could just combine κ and H into a single new "H.") For us that frequency is $f=V/L$ so that f has order $VW^{-1/2}Y^{1/2}D^{-1/2}\Delta^{1/2}$. And the maximum strain S has order L/D , that is $W^{1/2}Y^{-1/2}D^{-1/2}\Delta^{-1/2}$. As we've said H ought to be constant for all small-enough S and f ; but considerably larger, perhaps proportional to S^e for some positive real exponent e , for large S ; and growing proportional to f for high frequencies f ; but always with $0 < H \leq 1$.

The alert reader might now ask why we used the word "naive" above. The answer is that the elastic energy at the smallest length scale L , and at the largest D , and also at every intermediate length scale, all have the *same* values, and those energies *sum* (or more precisely, integrate, but at our level of precision where we are ignoring multiplicative constants, that distinction is irrelevant). This scale-independence is a special feature of 2 space-dimensions. In all other dimensions either the largest or smallest length scale dominates the strain-energy. Since there are $\log(D/L)$ such length scales, our naive stored-energy estimate was a factor of order $\log(D/L)$, or equivalently $\log([D/L]^2)$, too small. So now (the reader enquires) shouldn't we correct our F formula by multiplying it by this log? The answer is **no** – our "naive" F formula was correct! The reason is that the shortest length scale is the one for which the elastic cycling frequency V/L applies. At X times larger length scales, elastic cycling frequencies are X times lower. Hence, as far as hysteretic energy-loss rates are concerned, the shortest length scale really does dominate all the others.

The above analysis was focused on a "light load" regime where linear elasticity theory applied (or at least, came close to applying). If our wheel instead operates in a "heavy load" regime where plastic deformation is important, then we expect (at least, at slow rolling speeds) the wheel to keep plastically deforming until the width L of the flat spot grows large enough that the pressure $W/(L\Delta)$ shrinks below the plastic-yield stress P of the wheel and road materials. Deformation then stops. If so, then L will have order $W/(P\Delta)$ and from local-quadratic geometry the vertical displacement Z will be of order $L^2/D=W^2D^{-1}P^{-2}\Delta^{-2}$ and therefore our prior trench digging [law](#) shows

$$F_{\text{solid.cyl.plastic deform}} = \kappa W^2D^{-1}P^{-1}\Delta^{-1}$$

where κ now denotes a (new) dimensionless positive constant. At fast rolling speeds where there is not enough time for full plastic deformation, we expect smaller F .

Ball. If we redo essentially the same argument but now based on Hertzian contact for a ball rather than cylinder, then we already [saw](#) that the width Δ of the contact patch under linear elasticity theory must be asymptotically proportional to $W^{1/3}D^{1/3}Y^{-1/3}$ with vertical displacement of order Δ^2/D , hence the stored elastic energy must have order $W\Delta^2/D$, that is $W^{5/3}D^{-1/3}Y^{-2/3}$. We assume that when the wheel rolls distance Δ that constitutes one hysteretic cycle, losing fraction H of its stored energy. Hence

$$F_{\text{ball.hyst.}} = \kappa W^{4/3} D^{-2/3} Y^{-1/3} \cdot H(f, S).$$

Here κ is some dimensionless positive constant and we regard H as a function of cycle-frequency f and maximum strain S . (Or again, we could just combine κ and H into a single new "H.") For the allegedly exact value of κ based on Hertz's allegedly-asymptotically-exact elastic displacement solution for contacting balls, see Tabor 1955, but Greenwood, Minshall, Tabor 1961 concluded this "exact" κ disagreed by factor 2-3 versus experiment, probably due to our "simple H model" *not* involving the same H locally everywhere, then attempted to suggest some corrections. Here that frequency is $f=V/\Delta$, while the maximum strain S has order Δ/D , that is $S \approx W^{1/3} Y^{2/3} D^{-2/3}$. Again H ought to be constant for all small-enough S and f ; but considerably larger, perhaps proportional to S^e for some positive real exponent e for large S ; and growing proportional to f for high frequencies f ; but always with $0 < H \leq 1$.

Again the above analysis was focused on a "light load" regime. If our ball bearing instead operates in a "heavy load" regime where plastic deformation is important, then we expect (at least, at slow rolling speeds) the ball to keep plastically deforming until the width Δ of the flat spot grows large enough that the pressure $W\Delta^{-2}$ shrinks below the plastic-yield stress P of the wheel and road materials. Deformation then stops. If so, then Δ will have order $(W/P)^{1/2}$ and from local-quadratic geometry the vertical displacement Z will be of order $\Delta^2/D = WD^{-1}P^{-1}$ and therefore our prior [trench digging](#) considerations show

$$F_{\text{ball plastic deform}} = \kappa W^{3/2} D^{-1} P^{-1/2}$$

where κ now denotes a (new) dimensionless positive constant. At fast rolling speeds where there is not enough time for full plastic deformation, we expect smaller F .

Exact treatment of "Hertzian contact"? Hertz 1882 claimed to find the exact infinitesimal displacement field for two quadratic surfaces (e.g. elastic ellipsoids, and in particular balls) pressed together. Consequently it ought to be possible to do better than our "dimensional analysis" by working out all the constants *exactly* (in the limit of small displacements) in our elastic hysteresis "simple model" of rolling resistance for balls and cylinders. *However*, Hertz admitted that important parts of his argument were left unproven. Later authors re-presenting Hertz's work, such as the books on elasticity by A.E.H.Love (pp.190-200 of the 1906 edition), L.Landau & E.Lifshitz (pp.26-37 of 1970 edition) and S.Timoshenko & J.Goodier (pp.362-390 of the 1951 edition; this is the best treatment I saw) then simply ignored those deficiencies and pretended the matter was settled. My personal stance is

- i. In 140 years, it is unclear to me whether a full proof was ever presented, and while I consider it plausible and beautiful, I remain *not* completely satisfied.
- ii. Given the present very incomplete experimental/theoretical understanding of elastic hysteresis losses in materials, even if we accept the exactness of Hertz it remains dubious that it can provide any benefit as far as quantitative understanding of rolling resistance is concerned, as compared with our much simpler dimensional-analysis inexact treatment. And indeed Tabor 1955 claimed to work out the exact constant for ball rolling resistance losses based on Hertz, but then Greenwood, Minshall, Tabor 1961 concluded this "exact" formula disagreed by factor 2-3 versus experiment.

- iii. Attempts by Hertz and others to apply that contact theory to modeling impacts often *clearly* are unjustified, e.g. [since](#) a 6.35mm-diam steel ball impacting a 26kg steel block at 46 cm/sec experiences some stresses well into the plastic deformation, *not* linear elasticity theory, regime; and since the 30μsec impact duration is shorter than the vibration-period of the anvil, demonstrating Hertz's "slow change" approximation is ridiculous.

Hollow cylindrical "rolling thin-walled pipe" wheel. We model the pipe wall using slender-beam elasticity theory. Recall that in that theory, the elastic energy stored in a beam equals $(\frac{1}{2})YM\int(\text{curvature})^2 ds$ where Y is the Young's modulus of the beam-material, M is the beam-cross-section's "moment of inertia" about the "x axis," meaning more precisely $M=\int\int y^2 dx dy$ where we integrate over the fixed 2-dimensional beam cross-section, the bending is in the y-direction, and the beam's central surface (which corresponds to the x-axis of our beam cross-section) is assumed bent according to some curvature(s) function, with $s=\text{arclength}$.

Let the pipe-wall thickness be ζ . It is convenient to let $\tilde{D}=D-\zeta$ because the beam's *central surface* is a cylinder of the slightly-reduced diameter \tilde{D} . Now one might imagine, similarly to the ball and solid cylinder, that the bottom of the wheel again will flatten to a width-L contact region ($L \times \Delta$ rectangular region), and that that distortion is responsible for the hysteresis that yields the rolling friction. That imagining turns out to be correct for small W, but wrong if W is above a certain critical threshold value. While it is true that we get such a flattening, its strain energy, and the vertical distortion it causes, both are small in comparison to the larger strain energy from the entire wheel being squashed vertically into some noncircular shape having height $D-\delta$ rather than D. (We assume there is another "road" riding on top of the wheel, with weight W, providing the load; the wheel itself is weightless.)

It actually is possible to solve this problem exactly. The squashed wheel must assume a "rectangle plus two semicircles" shape having the same perimeter of its beam central curve $\pi\tilde{D}$ as the original circle. (Because: symmetry combined with the concave-u nature of the squaring function forces all elements of the slender beam to deform in an identical manner, which is possible only if the curves are circular arcs and straight line segments.) So instead of its original curvature= $2/\tilde{D}$, this new shape has two curvatures:

- 0 along length 2L (namely L on top and L on bottom of the rectangle), and
- $2/(\tilde{D}-2L/\pi)=2/(\tilde{D}-\delta)$ along two semicircles of combined length $\pi\tilde{D}-2L=(\tilde{D}-\delta)\pi$.

The height of our circle, originally \tilde{D} , then becomes $\tilde{D}-2L/\pi=\tilde{D}-\delta$. Evidently an equivalent way to express this height reduction is to say $\delta=2L/\pi$. Here L is the horizontal sidelength of the rectangle, $0 \leq L \leq \pi\tilde{D}/2$. When $L=0$ we have the original circle. And when $L=\pi\tilde{D}/2$ it has been squashed flat (which would require infinite elastic energy in the slender-beam approximation, hence will not happen for any finite W, but does happen in the limit $W \rightarrow \infty$).

The gravitational energy loss then equals $W\delta$. This must equal the stored elastic energy:

$$W\delta = (\pi/3) (Y\zeta^3\Delta) \{ \tilde{D}^{-2} \delta + [(\tilde{D}-\delta)^{-1} - \tilde{D}^{-1}]^2 (\tilde{D}-\delta) \}.$$

The right hand side equals zero if $\delta=0$, and increases monotonically to ∞ as δ increases to \tilde{D} . Therefore, if

$$W > W_{\text{crit}} \equiv (\pi/3) (Y\zeta^3\bar{D}^{-2}\Delta)$$

then there must exist a unique $\bar{\delta}$ with $0 < \bar{\delta} < \infty$ solving the equation. Indeed, that solution has this unexpectedly simple exact expression:

$$\bar{\delta} = \bar{D} - (\pi/3)W^{-1}\bar{D}^{-1}Y\zeta^3\Delta = (1 - W_{\text{crit}}/W)\bar{D}.$$

But if $0 \leq W \leq W_{\text{crit}}$ then the unique $\bar{\delta} \geq 0$ solving the equation instead is $\bar{\delta} = 0$. The latter does *not* mean that the wheel remains perfectly rigid for all W with $0 \leq W \leq W_{\text{crit}}$. Rather, it means that slender beam theory was inapplicable, because the deformation of the wheel is *local*, meaning occurring on a length scale smaller than the beam thickness (i.e. pipe wall thickness) ζ . In that case our prior "solid cylinder" analysis applies.

The fact that we are dealing with a *hollow* cylinder evidently only is relevant if $W > W_{\text{crit}}$, and only then can slender beam theory be claimed applicable. And then we obtain F in our usual manner by assuming fraction H of the stored elastic energy $W\bar{\delta}$ is lost per "cycle," where now a cycle means half a wheel-revolution. Hence

$$F_{\text{pipe hyst.}} = \kappa [\bar{D}W - (\pi/3)\bar{D}^{-1}Y\zeta^3\Delta] H(f, S)$$

Here κ is some positive constant and we regard H as a function of cycle-frequency f and maximum strain S , and again we could combine κ and H into a single new "H" if desired. That frequency is $f = 2V/(\pi D)$, while the maximum strain S has order $\zeta \max\{\bar{D}^{-1}, (\bar{D} - \zeta)^{-1} - \bar{D}^{-1}\}$. As usual H ought to be constant for all small-enough S and f ; but considerably larger, perhaps proportional to S^e for some positive real exponent e , for large S .

3. Applications

For most engineering purposes, I expect the elastic hysteresis F formulas will be the ones that matter.

Transporting a fixed weight W a fixed travel distance. Friction losses will be minimized by choosing the largest possible diameter D for the wheels and making them and the rails out of the hardest (largest Y and P) possible stuff. Solid cylinder wheels are better than either hollow cylinders or rolling balls of the same diameter in the sense that latter two store more elastic energy. Also, balls have greater maximum strain.

What if we can choose between one wheel with twice the diameter, versus 4 normal wheels (each carrying a quarter of the load; all wheels solid cylinders with same widths)? The latter is predicted by $F_{\text{solid.cyl.naive hyst.}}$ to yield a factor $\sqrt{2}$ less friction and maximal strain, while $F_{\text{solid.cyl.plastic deform}}$ predicts the latter will halve the friction.

What if we can choose between one wheel with twice the diameter, versus 2 normal wheels (each carrying half the load)? With either $F_{\text{solid.cyl.naive hyst.}}$ or $F_{\text{solid.cyl.plastic deform}}$ both choices are predicted to yield the *same* friction and same maximal strain.

A hollow cylinder roller can, paradoxically, have *less* maximum strain than a solid cylinder roller-bearing of equal diameter. E.g. consider a load $W=2W_{\text{crit}}$, which causes our hollow cylinder wheel to lose half its beam-center height ($\delta=\bar{D}/2$) and thus causes the bending |stress| and |strain| each to be *uniform* over the entire circular beam. This is a huge distortion storing huge elastic energy $W\bar{D}/2$. Meanwhile the solid cylinder only loses height of order ζ , a far smaller height-distortion, therefore storing much smaller total elastic energy, *but* with a constant factor greater *maximum* strain. This mathematically could cause greater rolling friction F for the solid cylinder than the hollow one (at same loading W and same diameter D) provided the solid cylinder's maximum stress exceeded the plastic yield-stress P , while the hollow cylinder's did not, and provided that material's H was small enough for strains below plastic deformation threshold. Indeed F could in this way be reduced by an *arbitrarily large factor!*

However this paradox can only be enjoyed at just the right range of loads W (starting a bit above the threshold needed to cause plastic deformation of the solid wheel), with the hollow cylinder's wall-thickness ζ chosen just right, and with a material with small-enough H . If the wheel width Δ is made wide enough to keep all stresses below P and prevent plastic deformation, then the solid cylinder will always have smaller F than the hollow one at same loading W (assuming the wheel's own weight is negligible compared to W).

Maximizing "Q" for a wheel oscillatorily rolling along a vertically-curved track of fixed travel-length. If the weight W of the solid cylinder wheel is assumed proportional to $D^2\Delta$, then Q will be proportional to $D^{5/2}Y^{-1/2}$, so it helps to increase D and Y . Increasing Δ with unchanged diameter D will not change Q . However, the maximum strain will grow proportionally to $D^{1/2}$ and once it nears the yield-strain P/Y the benefits of increasing D likely will cease. Incidentally, a rigid track shaped like an upside-down [cycloid](#), except expanded by $D/2$ (i.e. the track consists of points at distance= $D/2$ from the cycloid) in theory with a rigid diameter- D roller would yield a perfect simple harmonic oscillator.

Maximizing "Q" for a pendulum of fixed weight, rod-length, and swing-angle. The frictional *torque*, as opposed to force, is got by multiplying the force by $D/2$ where D is the wheel diameter. Imagine a pendulum whose upper end is supported *not* by a sliding-contact "journal bearing," but rather by two wheels rolling back and forth by some fixed angle on some sort of track. Then, to minimize the rolling-friction torque $DF_{\text{solid.cyl.naive hyst.}}/2$ and hence maximize Q , we want with solid-cylinder wheels to *minimize* D . Meanwhile for $F_{\text{solid.cyl.plastic deform}}$ the torque is independent of D so we again might as well minimize D . [Swinging at fixed angle thus leads to exactly the opposite recommendation as traveling fixed distance!] So it is best to make the "wheels" actually be sharp *corners* of, say, a prism, with radius of curvature approaching *zero*; and to make both the prism and its track from the hardest possible stuff, e.g. diamond.

More precisely, though, the corner optimally should have very small, but *not* zero, radius of curvature, specifically we want to keep the radius of curvature just large enough to prevent plastic deformation. So for example for a 20kg pendulum supported by a triangular-prism edge made of hardened steel 20cm long rolling on a track also of hardened steel, rounding the prism "corner" to have radius of curvature $\approx 500\mu\text{m}$ will keep the maximum strain below the plastic deformation threshold $P \approx 0.002$. By using [liquidmetal.com](#)'s bulk metallic glass instead of steel, we can shrink the radius of curvature to $30\mu\text{m}$, comparable to a typical human hair.

This swinging kind of pendulum only has fixed period in the limit of *small* swing angles, unlike the cycloidal track whose period remains constant regardless of amplitude. However, far greater Q should be achievable for the swinging pendulum. Also, because this kind of swinging-rod pendulum is readily "temperature compensated" to make its period independent of temperature despite thermal expansion, while the cycloidally-rolling-cylinder is not so easily thermally-compensated, in practice I expect better timekeeping with the rod.

4. Comparisons versus experiment

Famed physicist Ch.A. de **Coulomb** published the first(?) experiments on rolling friction in 1785, testing lignum vitae and elm cylinders rolling on smooth oak rails, and applying the load W by hanging a cord over the cylinder with equal weights on each end. Because wood is *anisotropic*, and the cord-hanging method created confusion because of also having to worry about energy possibly wasted straightening and bending the cord, and since excess cord on one side creates additional pull, this was approximately the stupidest possible experimental design!

His lignum vitae cylinders were 6-in. and 2-in. diameter tested with W=100, 500, and 1000 lb. The elm cylinders were 12-in. and 6-in. diam. and tested with W=1000 lb only. Rolling friction was taken as the force F, which if applied at the axis would just overcome static friction to permit continuous slow motion. Coulomb applied pulling force F tangentially with a flexible cord.

W(lb)	D=2in LigVit	D=6in LigVit	D=6in Elm	D=12in Elm
100	F=1.6	F=0.6		
500	F=9.4	F=3		
1000	F=18	F=6	F=10	F=5

Coulomb observed that his data above was compatible with the hypothesis $F_{\text{Coulomb}} = \kappa W/D$ to within $\pm 6\%$, where $\kappa = 0.0356 \text{ inch} = 0.904 \text{ mm}$ for Lignum Vitae and $\kappa = 0.06 \text{ inch} = 1.52 \text{ mm}$ for Elm. "Coulomb's law of rolling resistance" then unfortunately was mindlessly repeated by handbooks continuing to the present day, despite often being contradicted by later experimenters.

Next into the breach was Arene Jules E.J. **Dupuit** during 1837-1842. Dupuit became Chief Engineer of Roads and Bridges in France. Dupuit instead found the law $F_{\text{Dupuit}} = \kappa W D^{-1/2}$ after a much larger number of more varied tests than Coulomb (diameters D=10-100cm, more materials, more loads W) using loaded wheel-axle pairs on an inclined plane. Dupuit initially suggested F was independent of rolling speed, but later he and others found some speed dependence. Dupuit's exponent $-1/2$ agrees with my theoretical $F_{\text{solid.cyl.hyst}}$ law. However, Coulomb's D^{-1} dependence agrees with my my theoretical $F_{\text{solid.cyl.plastic deform}}$ law! **Thus our two solid cylinder laws explain the 180-year old Coulomb-vs-Dupuit controversy.**

Both Coulomb's and Dupuit's W^1 dependencies are incorrect for smooth surfaces, but they agree with my theoretical F_{cobble} law.

Dupuit found $\kappa = 21\text{-}36 \mu\text{m}^{1/2}$ for wagon wheels rolling on 4 different kinds of real road surface, the best being dry [ballasted](#) "very smooth," and the worst "wet earth"; and $\kappa = 15.8 \mu\text{m}^{1/2}$ for wood wheels on wood, and $\kappa = 12.4 \mu\text{m}^{1/2}$ for iron wheels rolling on wood.

J.Poiree 1852-1853 in tests on iron rails & wheels ($D=0.9-1.6$ meter, $W=3000$ kg) found $\kappa=0.85-0.96\mu\text{m}^{1/2}$. O.Reynolds 1875-1876 tested a smooth iron wheel with $D=14.5$ cm and $W=6.4$ kg rolling on various surfaces, finding the best surface was iron with $\kappa=0.10-0.22\mu\text{m}^{1/2}$, and the worst rubber with $\kappa=1.3\mu\text{m}^{1/2}$. Note the factor 4-10 discrepancy between Reynolds' and Poiree's iron-iron Dupuit- κ values. **This discrepancy would be resolved if we agreed that the Dupuit law's dependence on W^1 was incorrect because its exponent 1 should be replaced by a number between 1.2 and 1.4.**

Whittemore & Petrenko 1921 measured rolling friction for steel ball bearings with diameter $D=1, 1.25$ and 1.5 inch and also $D=1.25$ inch diam ($\Delta\approx 5$ inch thick) steel roller bearings (solid cylinders), all rolling on steel, with loads from 0 up to 50000 lb. They confirmed my theoretic claim that rollers both can safely carry more load W (in their case $10\times$ more) and reduce friction at the same loading (in their case by factors 1.2-2.2) versus balls of the same diameter.

Whittemore & Petrenko 1921's roller experiment also confirmed the falsity of Dupuit law's exponent 1 on W , finding instead the exponent 1.1 both for loads 250-2500 lb applied between flat plates, and for loads 5000-25000 lb applied between cylinders (inner cylinder 20inch diam). That 1.1 seems in considerable disagreement with my $F_{\text{solid.cyl.hyst.}}$ law's $3/2$, although it would agree better in the large- W limit (for W still within the realm of applicability of linear elasticity theory) with my $F_{\text{pipe hyst.}}$ law's W -exponent 1. For loads 20000-50000 lb now instead applied between cylinders (inner cylinder 20inch diam), Whittemore & Petrenko' data yields the exponent 1.9-2.1, in agreement with my theoretical $F_{\text{solid.cyl.plastic}}$ law's W^2 dependence.

Whittemore & Petrenko 1921's ball experiments suggested an exponent between 1.166 and 1.503 on W , as compared with my $F_{\text{ball.hyst.}}$ prediction $4/3$ and my $F_{\text{ball.plast.}}$ prediction $3/2$. Whittemore & Petrenko's experiments suggest an exponent between -0.71 and -1.15 on D , versus my $F_{\text{ball.hyst.}}$ prediction $-2/3$ and my $F_{\text{ball.plast.}}$ prediction -1. Unfortunately W&P's ball experiments are subject to confusion because their balls ran in grooves ("races") rather than on flat plates, and also since the speed was *not* being altered as a function of load in the right way to really compare with my theoretical $F_{\text{ball.hyst.}}$ exponent on W .

Experiments seemingly having higher quality than any of the above were by Hersey & Downes 1969 (but describing experiments during 1924-1925!) who tested spoked monolithic chilled-cast-iron (machined surfaces, five wheel diameters $D=8-24$ inches; they provide photograph; the smallest size only was a non-spoked disk) wheels on 90ft of steel rails on concrete in a specially constructed lab, using wheel-pairs with disc-weights threaded onto 42inch axles (all one rigid assembly). Because the rims of their spoked wheels were fairly thick and their loads fairly small, I would guess that the fact their wheels were spoked made little difference versus if they all had been fully-solid disks. They used automated electrical timing. Loads W per wheel were 100 to 1800 lb, speeds V from 0 to 7 mph. Static/starting tests also included. (They estimate rolling friction of harder, commercially finished wheels might well be as low as half that for their machined surfaces.) In the event of future tests on wheels by deceleration they recommend that runs be made in both directions to eliminate any doubt as to the grade or level of the track. The exponent of D (which in Dupuit's law was $-1/2$) in Hersey & Downes' experiments ranged from -0.38 to -0.66 in table 4 (mean -0.48), or -0.42 to -0.60 in table 8 (mean -0.51). The exponent of W (which in Dupuit's law was 1,

and in my $F_{\text{solid cyl.hyst}}$ law is $3/2$) was 0.67, 0.73, 0.78, 0.81, and 0.87 at speed $V=0$ in Hersey & Downes' table 3, while at constant rolling speed either $V=3.6\text{mph}$ or $V=1.4\text{mph}$, the exponent of W based on their table 5 appeared near 1. **But** at rolling speed V *proportional* to $W^{1/2}$, which is what theoretically matters for comparing to the $F_{\text{solid cyl.hyst}}$ law, the exponent of W appeared based on their table 5 to be between 1.1 and 1.7.

So based on Hersey & Downes' experiments, which to my knowledge were the best ever performed, conclude that **my $F_{\text{solid cyl.hyst}}$ proposed law is confirmed**. This confirmation is not tremendously impressive because of the rather large gap between 1.1 and 1.7, but that interval contains W 's theoretical exponent 1.5 and does not contain Dupuit & Coulomb's 1, and also does not contain 2 as in our $F_{\text{solid cyl.plastic deform}}$ law. The confirmation is more impressive as far as the exponent on D (theoretically $-1/2$) is concerned.

"*Static*" rolling friction did *not* appear to obey my $F_{\text{solid cyl.hyst}}$ proposed law (perhaps due to trench-digging and "creep," or manufacturing imperfections in roundness or rotation asymmetry of the initial weight distribution?). In any case "static" tests seem to me a rather stupid idea.

Sasaki & Okino 1962 found that for a ball rolling at **high speed** at light loads, friction ultimately grew almost proportionally to speed, in a monotonic and concave- U manner, and also highly influenced by the surface roughness of the balls. That also confirms our theoretical [expectations](#).

Hersey in a review of experimental work concludes: "**Qualitative summary**. There is general agreement that dry rolling friction is reduced by increasing the hardness and improving the surface finish of the bodies in contact, and increasing the roller diameter. The effects of load and speed variation are differently reported by the several investigators. Lubricants tend to *increase* rolling friction." That all agrees with our theory.

I do not know whether anybody ever observed my "hollow cylinder [paradox](#)."

Hersey also found a report (which I have not seen) by Noonan & Strange 1934 testing rolling friction of various solid cylindrical rollers on thick flat metal plates under loads W ranging between 500 and 5000 psi (3.4-34 MPa) in terms of projected area. Hersey claims his review's table 1 summarizes Noonan & Strange's data. The dependence of F on W in these experiments behaved proportionally to W^e where $e=1.50$ for steel roller on [monel](#), $e=1.55$ for Tobin bronze on steel, $e=0.90$ for steel on steel, $e=0.52$ for monel on "hy-ten-sl" bronze, and $e=0.26$ for monel on monel. Obviously the first two e 's support my $F_{\text{solid cyl.hyst}}$ proposed law's $e=3/2$ and agree with each other, while the remaining three do not support $e=3/2$ and also all disagree with each other. I have absolutely no idea what was going on (adhesion? Unquantified rolling-speed effects?).

Unfortunately it is evident that **no experimenter has ever done a good job** for comparing reality versus our theoretical predictions (at least among those I surveyed so far) – in the sense that we've pointed out obvious ways to greatly improve their experiments for that purpose. Modern technology such as computers might help too. (You might have hoped that after 5000-6000 years of work on wheels that would have happened. If so, you would have been a naive pollyanna.)

Hall 1996 described his **attempt to build the world's most accurate pendulum clock**. His pendulum consisted of a 6.4 kg cylindrical bob (75 mm diam, about 190 mm high) on a rod 9.6 mm

in diameter and about 1 meter long (about 0.8 kg) – both made of "invar" alloy pre-aged for 15 years. The bob was mounted on a 38.5mm-high slug of 304 stainless steel whose purpose was to zero the pendulum's temperature coefficient. The pendulum swung with horizontal amplitude about 1 cm in a 2 nanobar "high vacuum" chamber (air molecule mean free path \approx 33 meters) firmly mounted on concrete, maintained at constant $30\pm 0.05^\circ\text{C}$ temperature by a thermostat heating system, with each swing sensed by an LED-photodiode light-beam interruptor. Occasional "kicks" were provided by a current-controlled magnetic drive coil.

What matters for us was the method Hall used to suspend his pendulum: an agate prism with inverted equilateral-triangle cross-section, attached to the top of his rod, riding on a plane surface also made of agate. That is exactly the method I [recommended](#) (radius \rightarrow 0+ cylindrical "wheel") to maximize "Q". And in fact, Hall obtained **Q=120000** in a 2 nanobar vacuum, 120 times his Q=10000 in ordinary 1 bar air. As far as I know, this Q far exceeds any other pendulum clock ever constructed.

The commercial "[Shortt clocks](#)," about 100 of which were sold between 1922 and 1956 and which during about 1921-1945 were considered the most accurate clocks in the world, also were based on an invar pendulum swinging in a vacuum chamber and "kicked" occasionally. These kicks were not directly supplied by an electromagnet, but rather by dropping a tiny weight. It is important to perform such kicks at the *bottom* of the pendulum swing, and instantly, since then (and only then) they do not affect its period.

However, Shortt foolishly employed a flexing-leaf-spring suspension, and hence only achieved Q=110000 in vacuum, only 4.2 times his Q=25000 at atmospheric pressure. (If Shortt had been aware that the main cause of rolling friction is elastic hysteresis, then it would have been obvious to him that those hysteresis losses would necessarily be far greater in a leaf spring than in an optimal rigid roller, since the latter in the limit of perfect rigidity would have no such losses at all.) Short only used a weak vacuum (30 millibar) after discovering that lower pressures would be pointless because below 30 millibar, the losses from his leaf spring suspension dominate those from air drag. Despite Shortt's poor Q, testing by Boucheron 1985 of the Shortt#41 pendulum (as modified to now employ optical sensing) versus a Cesium atomic clock showed Shortt achieved accuracy of ± 1 second per 12 years, i.e. 3×10^{-9} , and probably by more careful temperature regulation (Shortt did none) that could be improved. Shortt clocks were originally thought to have only 1 second/year accuracy, but part of the reason for that misperception was caused by measuring them versus astronomical observations. In fact Shortt's pendulum if used in Boucheron's manner is more accurate than the constancy of Earth's rotation, and it readily detects the gravitational presence of the Moon and Sun (which cause cyclic "time" variations of +140 and -140 microseconds with about 2 such cycles per day), which, in fact, limit its accuracy. The first such "detection of the moon" was by A.Loomis in 1931 by comparing Shortt versus quartz clocks.

Historically, Shortt-style clocks were supplanted by quartz oscillators, and those in turn by atomic clocks, which nowadays transmit time information over the GPS system (which can be used to continually re-synchronize quartz). One commercial "double oven" temperature-regulated quartz oscillator as of year 2022 claims to offer short term frequency stability of 10^{-10} , but its frequency changes relatively due to "aging" by $\pm 15\times 10^{-9}$ in the first year and $\pm 300\times 10^{-9}$ in 10 years. Thus over the long term, Shortt's pendulums from the early 1920s, if used in Boucheron's manner, appear superior to year-2022 quartz! And I believe pendulums could be improved well beyond

Shortt, see below, albeit still could not compete with atomic clocks.

Despite his greater Q , Hall apparently failed to obtain better accuracy than Shortt/Boucheron! A big reason for that failure, pointed out by Byran Mumford, was that Hall mis-designed his kicker. Shortt's kicker, based on a small weight falling a fixed distance, tended to maintain constant angular amplitude for his pendulum-swing, regardless of slight changes in gravity due to e.g. the Moon. Hall measured at-bottom swing speed and kicked whenever it got too small, thus *not* maintaining constant angular amplitude, but rather constant at-bottom swing speed. (The two are inequivalent if gravity is nonconstant.) So if gravity got smaller, causing his clock to slow, Hall misinterpreted that as "too small angular amplitude" (even though, actually, let us say, the angular amplitude was exactly at the design value), and therefore increased amplitude to compensate, which in turn made his clock even slower! So Hall actually mis-controlled his clock, making Moon-caused errors worse, not better!

It is commonly asserted that a pendulum clock measures, not time t , but rather the time-integral of the square root of the local gravitational acceleration $\int g^{1/2} dt$, and therefore g -variations set the ultimate limit on pendulum clock true-time accuracy. And Shortt's design tried to approach that limit.

But that assertion (no matter how many times "great authorities" reiterate it!) is incorrect: it is possible in principle to **build a pendulum clock immune to variation of g** . To explain how: measure *both* the bottom-speed v , and the angular amplitude θ , of your swing. My point is that the *ratio* v/θ is proportional (at leading order, in the limit of small θ) to \sqrt{g} . Apply kicks designed to keep not θ , and not v , but rather $[1+\theta^2/16]\theta/v$, constant. (For example: if this quantity is too small, then perform a kick.)

This will precisely compensate for g -variations by means of amplitude variations, *but* only to leading order in both effects. Hence there will still be imperfectness, but it is demoted to being proportional to a higher power of θ in the small-amplitude limit $\theta \rightarrow 0$. Furthermore, that blemish may be overcome (at least in the limits $Q \rightarrow \infty$ of a low-loss infinitely-rigid pendulum, with slow g -variation, $|g'| \rightarrow 0$) as follows. The ratio $v^2/(1-\cos\theta)$ *exactly* equals $2g$. And then the period of the pendulum *exactly* equals $2\pi(Lg/J)^{-1/2}/\text{AGM}(1,\cos(\theta/2))$ where L is the distance between the pendulum's hub and center of mass, J is its moment of inertia about the hub per unit mass, and $\text{AGM}(a,b)$ denotes the Lagrange-Gauss "[Arithmo-geometric mean](#)" function. Then design your kicking strategy to keep *that* constant.

Admittedly, that is a bit easier to say than to do. For example, if our pendulum-swing angular amplitude is $\theta=0.01$ (i.e. $\pm 1\text{cm}$ swing of 1m rod) then changing 1cm to 1.0002cm would compensate for a 5 parts in 10^9 gravity increase due to the Moon. So this would require measuring θ and v to at least 4 significant figures of accuracy each swing, then computing $([1-\cos\theta]J/L)^{1/2}v/\text{AGM}(1,\cos(\theta/2))$, then kicking if it is too small (or, for even finer control, different *size* kicks could be delivered depending on *how much* too small it was...). Boucheron's way to measure θ is to mount a mirror on the pendulum rod, reflect a laser beam off it, and observe the reflected beam (thus measuring 2θ) with a telescope. A 650nm (red) laser-pointer diode costs \$5, and 405nm (blue-violet) laser-pointer diodes also are available at about 10 times that price, and 375nm (soft UV) diodes are also available. At beam diameter 1.5mm , those should enable angular accuracy $(3-5)\times 10^{-4}$ respectively for 2θ , and hence half that for θ ; with 3cm beam diameter we

could get 20 times better angular accuracy. Hall's way to measure v was to interrupt a light beam near its focus with a thin object attached to the pendulum, and time the interruption. Probably a better way is to time how long it takes Boucheron-style reflected beams to change angle by some specified amount. That could be criticized as "cheating" since we are timing it with some other kind of clock. But I claim we aren't really cheating, since it would suffice merely to measure the *ratio* of the time for θ to change by a small amount near swing bottom, versus the time for a full-amplitude swing. (And a different way, which definitely is not "cheating," is to attach a small magnet to the pendulum and then the voltage it induces round a small stationary wire loop is proportional to flyby speed.)

How much better pendulums can we build? First, Hall's choice of "agate" was not optimal. Agate, which is SiO_2 , has Young's modulus $Y=66\text{-}75$ GPa. If we replaced the agate with diamond ($Y\approx 1215$ GPa) then since $F_{\text{solid cyl. hyst.}}$ predicts friction proportional to $Y^{-1/2}$, we expect this would improve over Hall by a factor of $(1215/70)^{1/2}\approx 4.2$. And *single crystal* diamond is likely to improve over polycrystalline by perhaps a factor of 30 due to elimination of grain-boundary loss-mechanisms. Further improvement by a factor of about 5 would likely be obtained by cooling to 30°K . Similarly, if you are going to use quartz, I suspect single-crystal quartz would drastically outperform polycrystalline agate; and intrinsic mechanical losses of quartz dramatically decrease below 10°K .

Second, based on their performances in 1 bar air, Hall's pendulum evidently was aerodynamically 2.5 times worse than Shortt's. And making the bob from a denser material, e.g. lead (density= 11.3 kg/liter, CTE= $29\text{ppm}/^\circ\text{K}$) or tungsten (Elmet sells 99.95% pure tungsten plate up to 1.5 inch thick; density= 19.3 kg/liter, CTE= $4.3\text{ppm}/^\circ\text{K}$, $Y=410$ GPa) or tungsten "heavy alloys" (density= $17\text{-}18.5$ kg/liter, CTE= $4.4\text{-}5.4$ ppm/ $^\circ\text{K}$, $Y\geq 275$ GPa, excellent stability, machinable with standard tools) as well as simply increasing bob mass, both would reduce drag/weight ratio, by, e.g, a factor of 2.5.

Third, Hall's plot of Q versus vacuum pressure p featured Q behaving proportionally to $p^{-0.15}$ for all $p\leq 10^{-5}$ bars, with no end in sight. The question is how hard a vacuum you are willing and able to sustain. I suspect a $p=10^{-11}$ bar vacuum would be feasible. That by extrapolation of Hall's graph would improve his Q by a factor of 1.8. Commercial vacuum tubes in old radios, and cathode ray tubes in old televisions, contained 10^{-10} to 10^{-12} bar vacuums (achieved using roughing- plus turbo-pumps while baking at $375\text{-}475^\circ\text{C}$ for up to 1 hour, followed by sealing with introduction of barium "getters"). Even lower pressures ought to be attainable by cryogenic operation. The LIGO gravitational wave detector has a multi-kilometer-long ultra-high-vacuum ambient-temperature chamber at 10^{-12} bar. The ATLAS detector at CERN/LHC features a 10^{-14} to 10^{-13} bar vacuum. The atmosphere of the moon is thought to be a 4×10^{-16} bar vacuum in day and 3×10^{-15} at night. The [best](#) vacuums available require rather painful techniques such as extreme cleanliness; never touch anything with human hands; wash things with approved solvents; polish then electropolish metal surfaces then bake whole system during first pump-out; use only approved materials (many metals and most plastics forbidden); no screws and small pits on metal, only smooth surfaces; short fat connections to multiple stages of vacuum pumps; cryogenic cooling during operation, etc.

So I believe that $Q>10^7$ (conceivably even 10^8 with heroic effort) should be attainable, if anybody wanted to, by thus-improving the Shortt and Hall pendulums. Then accuracy could be improved by

our new gravity-compensation kicking mechanism (never used before in any clock). Also Hall & Shortt's use of the low thermal-expansion (0.6 to 1.2 parts per million per °K) alloy "[invar](#)" was subject to the problem that invar has poor dimensional stability, tending to "grow" gradually after manufacture (Hall ameliorated that problem by using 15 year old invar and Boucheron by waiting 60 years after Shortt manufactured his clock) and to "creep" under stress. As Hall knew, better, albeit more expensive, rod materials are Corning's "[ULE glass](#)" and Schott's "[Zerodur](#)" glass-ceramic, both with considerably lower CTE, as low as $\pm 7 \times 10^{-9}$ per °K between 20 and 300°C for the best grades. ULE glass has been measured to shrink by 2-5 parts in 10^{12} per day for many years, at a gradually decreasing rate. A perfect crystal (invar and glasses are not) would neither grow nor shrink with aging. Note that it is better to rely more on ultralow-CTE materials and less on than "temperature compensation" schemes using two different-nonzero-CTE materials, because temperature compensation only works if the temperature always is *spatially uniform*. Also, fused silica has *zero* CTE (minimum length as a function of temperature T) at just the right temperature – which unfortunately appears to vary as much as 60°C between specimens, but a typical value is $T = -80^\circ\text{C} = 193^\circ\text{K}$.

5. References

Renee N. Baca & 6 others: [A Novel Experimental Method for Measuring Coefficients of Restitution](#), SAND2016-5693 June 2016.

Pierre H. Boucheron: [Just How Good Was the Shortt Clock?](#), Bulletin of Natl Assoc of Watch and Clock collectors Inc 235,27,2 (1985) 165-173.

Rod Cross: Multiple collisions of two steel balls in a Newton's cradle, European J. Physics 39,2 (2018) 025001, 16pp.

James A. Greenwood, H.Minshall, D.Tabor: [Hysteresis Losses in Rolling and Sliding Friction](#), Proceedings of The Royal Society A Mathematical Physical and Engineering Sciences 259,1299 (Jan.1961) 480-507.

E.T.Hall: The Littlemore Clock, Horological science (June 1996).

Mayo D. Hersey: Rolling Friction, I - Historical Introduction, J. of Lubrication Tech. 91,2 (Apr 1969) 260-263.

Mayo D. Hersey & M.S.Downes: Rolling Friction, II - Cast-Iron Car Wheels, J. of Lubrication Tech. 91,2 (Apr 1969) 264-268.

Mayo D. Hersey: Rolling Friction, III - Review of Later Investigations, J. of Lubrication Tech. 91,2 (Apr 1969) 269-275.

Heinrich R. Hertz: Über die Berührung fester elastischer Körper und Über die Härte, 1882; English translation *On the contact of rigid elastic solids and on hardness*, pp.163-183 of HH's [Miscellaneous papers](#), Macmillan, New York & London 1896.

L.D.Landau & E.M.Lifshitz: Theory of elasticity, Pergamon 1970.

J.M.Lifshitz & H.Kolsky: Some experiments on anelastic rebound, J. of the Mechanics and Physics of Solids 12,1 (Feb. 1964) 35-43.

N.G.Noonan & W.H.Strange: Tests on Rollers, Technical Memorandum No. 399, U.S. Department of the Interior, Bureau of Reclamation, Sept. 1934, 20 pp. (Also N.G.Noonan, MS thesis, University of Colorado, Boulder, Colo., 1951, 45 pp. Unfortunately I have seen neither.)

Tokio Sasaki & Norio Okino: [Rolling Friction at High Speed](#), Bulletin of Japan Society of Mechanical Engineers 5,18 (1962) 360-373.

David Tabor: Elastic work involved in rolling a sphere on another surface, British J. Applied Physics 6,3 (1955) 79-81.

S.Timoshenko & J.Goodier: [Theory of Elasticity](#), McGraw, 2nd ed, 1951.

H.L.Whittemore & S.N.Petrenko: [Friction and carrying capacity of ball and roller bearings](#), USA Natl Bureau of Standards report #201, October 1921, 38pp.

[Return to main page](#)