SHORT HISTORY OF STANDARD RAILROAD TRACK GAUGE

Track gauge and the rail-wheel interface

Abstract

A review of a historical element of the worldwide industrial revolution riddled with legends and speculation — railroad track gauge. Topics also include North American developments, engineering practices, track gauge safety limits, and rail-wheel considerations.

Frederick R. Smith

fred.smith369@yahoo.com

February 8, 2021, rev. December 22, 2023

Foreword

This paper is a practical overview of rail transportation's critical element, the track gauge. It uses the predominant word "gauge" to describe rail-to-rail spacing. Some authors use "gage" for the dimension, while "gauge" may be a tool to measure it. The word "railway" in the title suggests the historical lineage. The British, giving birth to the rail guideway system we know today, use the word railway in most literature and engineering documents. In North America, some corporations use the word railway in their formal name. Railroad, the more contemporary phrase in North America, follows as the operative word throughout this paper.

Brief Overview

The predominant or standard railroad track gauge throughout the world is 4'8½". There is no critical reason for the chosen standard gauge. Instead, it occurred as an evolutionary process. It began in Britain. Before the Industrial Revolution, the predominant guideway dimension corresponded to the wheel spacing of horse-drawn carts. Historical documents show one of two reasons for the original choice of standard gauge:

1. Early railroad builders chose the measurement of chariot wheel grooves in ancient Rome. The ruts or grooves cut into the pathways by the chariots and carts supply a gauge measurement of less than 4'9". The builders of the precursor to a practical railroad used a guided beam system and chose the predominant gauge - 4'8". With the introduction of a flanged wheel and rail system, engineers increased the gauge ½" to reduce friction.
2. The "L" shape of the first iron rail supplied a guide for the wheels of the horse-drawn carriages. That changed in 1789 with the introduction of the cast iron edge-rail and the flanged wheel. Measured from the outside of the rails, 5'0" constituted the predominant gauge. Later, railroads made wheels with flanges on the inside. The same 5'0" gauge then became 4'8½" due to the deduction of 1¾" width from each rail.

The rail/wheel interface played a significant historical role in developing the standard track gauge. For North American common carrier railroads, the nominal gauge contact point between the rail and wheel is ⅝" below the top of the rail. This point of reference is the same for wheel gauge. The difference between rail and wheel gauge of $^{15}/_{16}$" induces a steering effect for wheels moving through curves.

The standard track gauge can deviate from its as-built dimension. That is because of allowances for materials manufacturing and construction tolerances. Railroads have maintenance standards that address the expected wear of track materials. This wear causes wide gauge. Common carrier railroads in the United States must follow Title 49 of the Federal Code of Regulations. There, Part 213 (Track Safety Standards) includes thresholds for track safety. It sets the most significant allowable gauge deviation. With the gauge deviating wider a significant amount than designed, the track loses its ability to keep a wheel assembly (truck) in line with the track. This increases wheel climb dynamics. It also increases rail rollover forces. Or, in extreme cases, it lets wheels drop between the rails.

Track Gauge Worldwide

The more noticeable railroad engineering attribute is the distance between the rails (track gauge). Measured at right angles to the rails at a point ⅝" below the tread (top) of the railhead, the predominant gauge throughout the world is 4'8½" (1435 mm).[1] With 61% of the track miles in the world constructed at 4'8½" gauge, literature, and railroad engineering manuals call this *standard gauge*. Various gauges less than standard, known as *narrow gauge*, account for 17% of the total. Gauges greater than standard, known as *broad gauge*, account for 22%.[2]

History and physics show no fundamental reason for the chosen worldwide standard gauge. Yet, from a track-train dynamic perspective, *broad gauge* can be a positive aspect. That includes the centrifugal forces that trains place on the track structure. It also supplies passenger comfort when negotiating curves.

Many gauges existed in the United States up to the Civil War. If the North lost, the 5'0" predominant gauge in the South during the war could now be today's standard. In 1886, railroads in the South converted about 11,500 miles of track. Because the Pennsylvania Railroad had several "break in gauge" connections with the southern railroads, engineers chose that carrier's then 4'9" gauge.[3] That coordinated feat of gauge change in the South occurred in 36 hours.

During the construction of the Erie Railroad in New York State in 1835, the engineers opted for an extensive network of track with a gauge of 6'0". For 15 years, the Erie ran on a 6'0" gauge track, gradually expanding westward. However, it became clear that few new railroads in the region were adopting the Erie's gauge width. After meeting challenges with the interchange and transshipment of freight and railcars with other lines, Erie finished converting all its lines to the prevailing 4'8½" gauge network in 1880.

A noteworthy broad-gauge example was the Great Western Railway. In 1833-34, Isambard Kingdom Brunell (1806-1859) chose a 7'0¼" gauge for the Great Western in London, England. This dimension offered stability for speeds up to sixty miles per hour.[4] The Philadelphia, Pennsylvania trolley system has a gauge of 5'2¼". The Bay Area Rapid Transit (California) has a gauge of 5'6". These are examples of broad gauge used today in the United States. Broad gauge is more expensive to build due to the added track material and more real estate. A narrow gauge track consumes less space. Its smaller footprint limits the size of railroad rolling stock.

Narrow gauge railroads of various dimensions moved commodities at mines, industries, and quarries. They played a significant role in freight transportation worldwide during the Industrial Revolution. They also ran passenger operations. The 3'0" gauge lines offered an extensive network in North America. This sub-industry never achieved its grand plan. They wanted to connect separate narrow-gauge companies into a transcontinental railroad. Several 3'0" narrow gauge lines continue as tourist operations. Examples include Durango and Silverton in Colorado. Another example is the White Pass and Yukon. It runs in Alaska, British Columbia, and the Yukon Territory.

Origins of Standard Gauge

Books and periodicals claim gauge began with ancient chariot wheel grooves in Rome. They approximated the distance between the rails. Others assert that this ancient relic of grooves resulted from the need to space the chariot wheels around a horse's posterior. The British (supposedly) adopted this figure. Too many engineering considerations suggest this might be an urban legend. Also, the lack of corroborating evidence adds to the doubt. Researchers can find a conflicting and incomplete picture of the reason for the development of the standard gauge. This includes looking through various books and periodicals.

In the mid-1500s, German coal miners built timber guideways to move coal. To supply steering, a vertical pin attached to the bottom of wagons moved along in a slit in the wood guideway. The German miners imported this design into England in the later 1500s. The pin guide system gave way to the "L" shaped wood plank with an integral wood steering flange. As loads increased, thin metal straps attached to the top of the wood planks helped with load distribution. But they soon proved to be insufficient. As a precursor to rail, iron plates appeared in 1767. A flat-shaped section with a flange-like projection on each side of the plates supplied a groove for the wagon wheels.[5]

The standard track gauge originated with wagonways in the early 1800s. The 5'0" dimension was typical for flanged guided wagon wheels in Britain. Builders measured this dimension from the outside (or field side) of the rails. And, by extension, the wagonway planks. William Jessop (1745-1814) designed the Loughborough to Nanponton Railway initially with the flanges to the outside. In 1789, he changed the design, placing the wheel flanges on the inside, resulting in a 4'8½" gauge.

Richard Trevithick (1771-1833) of Cornwall, England, built the first trial steam-powered locomotive in 1803. In Britain, William Hedley (1885-1900) and Timothy Hackworth (1786-1850) made a rudimentary machine to run on rails in 1813. A year later, George Stephenson (1781-1848) built the first practical steam locomotive. Many historical documents show George Stephenson as the founding father of modern railroading.

George Stephenson pioneered railroad work in Britain. He began his engineering career collaborating with a wooden horse-drawn wagonway. It connected the mines with loading facilities. History books show that Stephenson oversaw the construction of the first practical railroad in 1814 with a 4'8½" gauge. That was 25 years after the first use of 4'8½" gauge.

The prevailing point of view to describe the origins of the standard gauge is a straightforward assumption. That is, it approximates the Roman chariot wheel grooves. Early guideway pioneers constructing the precursor to railroads chose this familiar dimension. The alternate is a detailed engineering perspective. Stephenson's work included an engineering adaptation. It transformed the gauge dimension of 5'0" into the precise measurement of 4'8½". Thus, the reader should draw conclusions based on the perspectives presented below. These are the "chariot wheel" or "engineering evolution" perspectives.

From 1811 to 1850, the builders of most tracks constructed in the United States used a method like the British system described above. This "strap rail" continued until after the Civil War (1865). The metal strap attached to the top of the wood beams tended to curl up due to the load from the passing cars. The loose straps, called "snakeheads," would work up into passenger coaches. This caused casualties and deadly derailments. The first construction of the Baltimore and Ohio Railroad in 1827 to guide a horse dawn car was the first use of a 4'8½" gauge in the United States.

The first iron rails with "T" shape in the United States came from Britain in 1831. Robert Livingston Stevens (1787-1856) used this rail on his 4'10" gauge Camden & Amboy Railroad in New Jersey. Also, in 1831, construction of the Allegheny Portage railway began with a 4'9" gauge. At the start of construction, the Portage builders used strap rails. In 1832, they also used imported iron bullhead rail. The Pennsylvania Railroad (PRR) also chose a 4'9" gauge as it began construction in 1846. As noted in the above section *Track Gage Worldwide*, PRR used that gauge until after the Civil War.[6] Of note, strap rail was the only type of guidance system made in the United States up to 1844.[7]

Chariot Wheel Perspective

Contemporary accounts supply a lineage of "gauges." The origin starts from pre-railroad guideway systems in ancient Rome and goes to wagonways in Britain. Archaeological projects in the 1870s discovered the pathway grooves. They were in Pompeii and other locations and were almost 4'9" apart.[8]

The earliest wagonways in Britain used gauges from about 3'10" to 5'0". The tramway builders called the Wellington Way (1763-64) chose the predominant British gauge, 4'8". That approximates the same dimension as the Roman grooves described above.

George Stephenson oversaw the building of his first steam locomotive in 1814 on the Collingwood Railway. In 1825, he designed the Stockton and Darlington Railway. He designed the Liverpool and Manchester Railway in 1830. Stephenson picked the familiar 4'8" gauge and increased it by ½", resulting in 4'8½". Many consider him to have built the first practical railway. Possibly, Stephenson's arrangement reduced friction between the wheel flanges and rails. Yet, the builders had the opportunity to change the wheel gauge. Thus, for this scenario, the precise reason Stephenson changed the gauge to 4'8½" leads to speculation.[9] Moreover, as seen above, the first use of a 4'8½" occurred earlier in 1787.

Engineering Evolution Perspective

In 1776, the first vertical iron rails appeared in Britain. The "L" shape of this iron shape guided the wheels of the horse-drawn carriages. That changed in 1789 when William Jessop introduced the cast iron edge-rail and the flanged wheel. The 5'0" dimension was typical for guided wagon wheels in Britain. They measured it from the outside of the rails. By extension, the wagonway planks were also 5'0".[10]

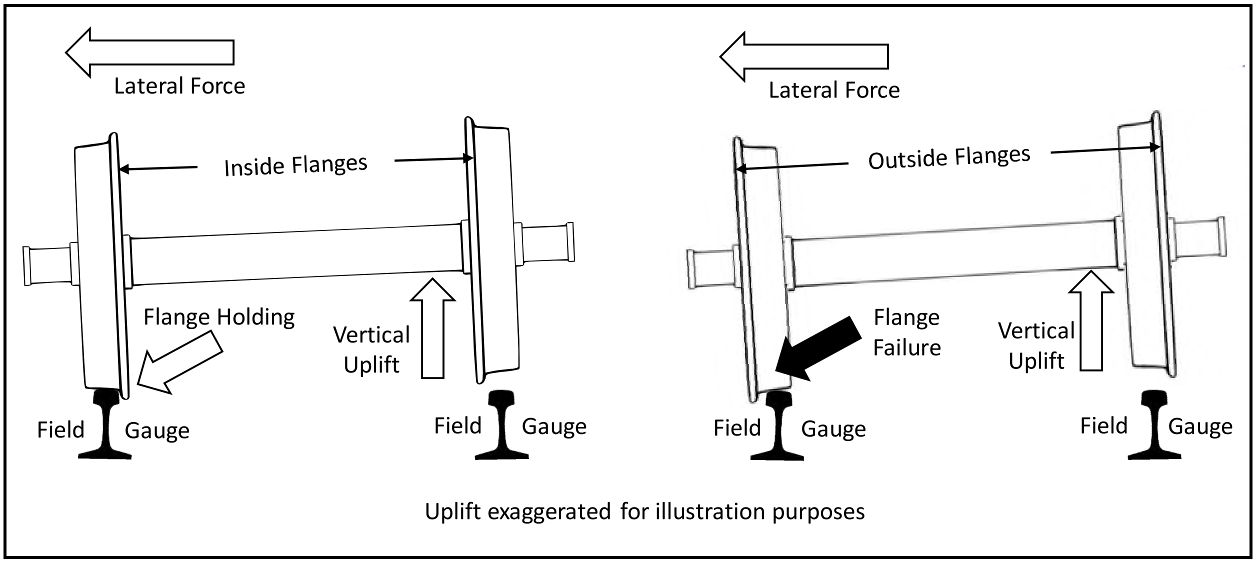

Figure 1

As illustrated in Figure No. 1 above, the dynamic forces imparted by the rolling stock at curves and locations with geometric anomalies can result in wheel lift. The corresponding wheelset is free to slip off the rails with flanges to the outside. Realizing this dynamic, Jessop changed the design and moved the flanges to the inside. That, by happenstance, resulted in the establishment of the 4'8½" gauge.[11 12 13]

Wheel Gauge and Design Profile

An integral element of the track gauge is the wheel gauge dimension. North American common carrier railroads and most rail transit systems measure the track gauge ⅝" below the top of the railhead. This reference point is the same for the wheel gauge – measurement occurs from the inside of the wheel flange ⅝" below the wheel tread. That is the nominal contact interface between the wheel flange and the railhead. See Figure No. 2 below.

In North America, the standard wheel gauge for multiple-wear wheels is 4'7$^{11}/_{16}$". Therefore, the difference between rail and wheel gauge is about $^{15}/_{16}$". In a perfect static setting, this supplies just under ½" free space ($^{13}/_{32}$") between each wheel flange and its corresponding rail.

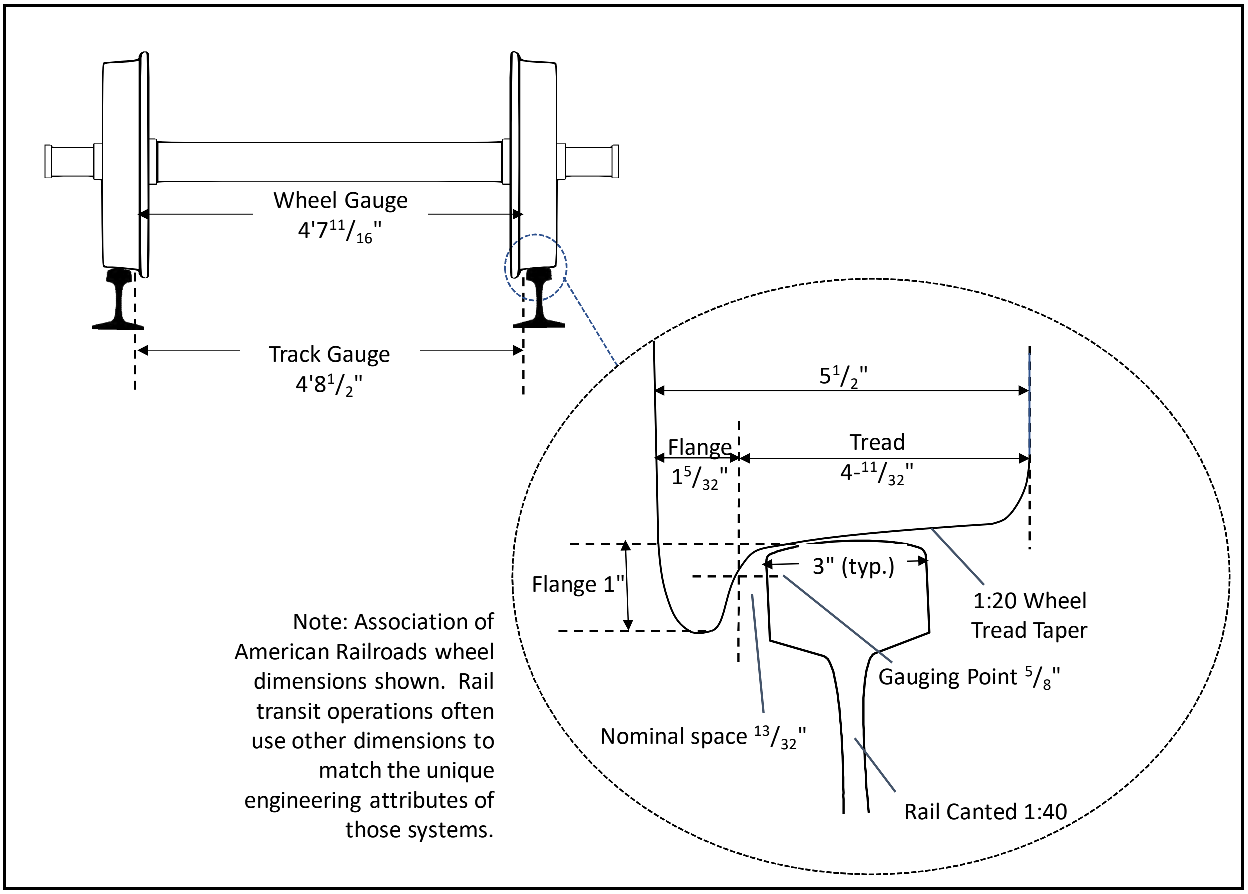

Figure 2

Rolling Radius Differential

The wheel/rail gauge free space described above mitigates the wheel flange and rail friction. That allows the two wheels on each axle to rate at the same rate. This design element accounts for the shorter radius on the inside (low) rail. It also accounts for the longer radius of the outside (high) rail.[14] This layout takes advantage of the wheel tread's conical (or tapered) shape. Otherwise, wheels slip to compensate for the shorter radius of the inside rail vs. the longer radius of the outside rail.

With each wheel fixed to its axle, the centrifugal force of moving trains in curves will cause the flange on the outside rail to be up against this rail. This action causes the flange on the inside rail to pull away from the rail gauge face. Thus, the wheel on the outside rail will be on the longer radius segment of the wheel tread. The inside rail will contact the smaller tread radius. See Figure No. 3 below for an illustration of the attribute known as "rolling radius differential." Supplying steering capabilities can cause "hunting" in a tangent (straight track).[15] The use of 4'8¼" gauge on tangent track has had limited success in reduce hunting.

The normal wheel treads have a 1:20 taper profile to enable steering.[16] Considering this attribute, engineers orient each rail with a lean (canted) toward the track centerline. This focuses the wheel/rail contact patch to the middle of the railhead. A canted tie plate offering a 1:20 cant can provide a matching contact face. Yet, based on past experiences, North American typical practice for wood ties includes a tie plate with a 1:40 cant.[17] Thus, rails have less cant compared to the wheel taper angle.

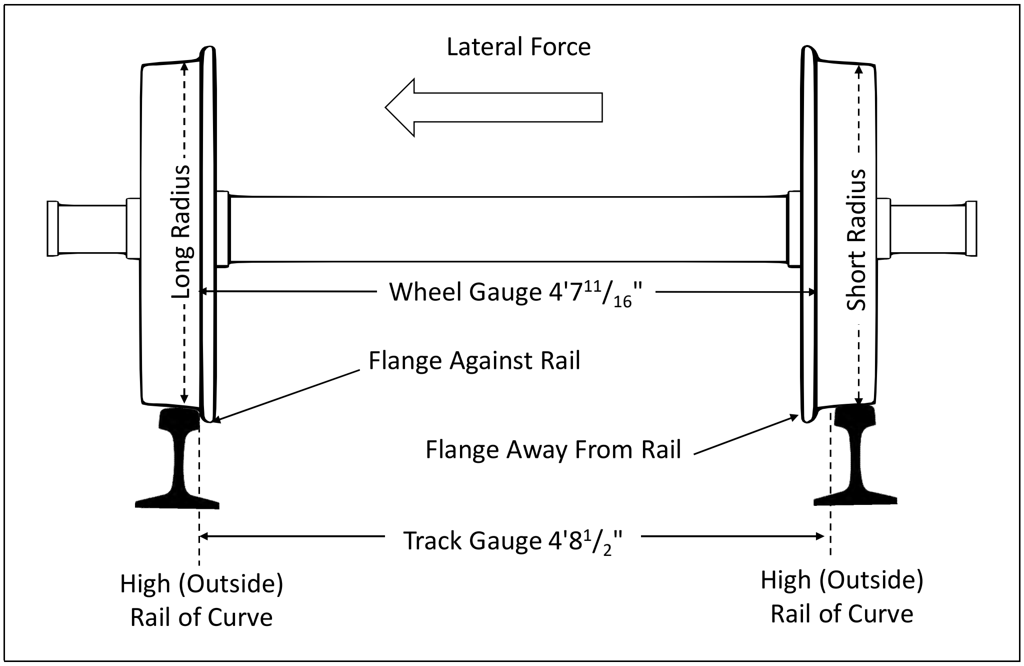

Figure 3

As described above, the wheel-rail interface works well with new wheels and new track. As wheels and rails wear, the system loses its capability to supply a rolling radius differential. Other wheel profiles exist. They are not the standard Association of American Railroads section design. For example, passenger trains running on longer radius curves have a unique wheel profile. This profile incorporates alternating wheel tapers to mitigate wheel hunting. Other profiles, such as those found in certain light rail transit lines, use a cylindrical wheel tread to mitigate truck hunting. It lacks a taper, unlike the conical profile.[18]

Gauge and Track Safety

Construction specifications prescribe an as-built 4'8½" gauge dimension. Gauge thresholds allow for manufacturing and construction imperfections in materials. These standards supply slight allowances due to the above considerations (e.g., plus, or minus ¼"). As tonnage accumulates on a track, the track geometry degrades. The railhead wears, especially the outside rail of curves. This causes a larger-than-designed gauge dimension. Railroads have maintenance standards that address the expected wear of track materials.

The Federal Railroad Administration (FRA) regulates the safety compliance of common carrier railroads in the United States.[19] Title 49 of the Federal Code of Regulations, Part 213 (Track Safety Standards) prescribes safety limits. That includes the largest allowable gauge beyond normal maintenance limits. This applies to various speed groups called Class of Track. Under normal circumstances, 4'10" is the greatest permissible wide gauge.[20] This figure applies to the lowest speed group called "Class 1 track." That designation has a speed limit of ten miles per hour for freight and 15 miles per hour for passenger trains. As the speed increases, the wide gauge threshold decreases for each Class of Track.[21]

Track/Truck Dynamics

With excessive wide gauge, the track loses its ability to keep a truck in line.[22] The wider the gauge, the potential increases for a truck to rotate on its center plate (skew). As a result, the lead flange on the outside rail of a curve will "attack" or "bite" its corresponding rail. This increases wheel climb dynamics and rail turnover forces. It allows the non-biting or "pulled flange" to fall off its corresponding rail. See Figure No. 4 below.

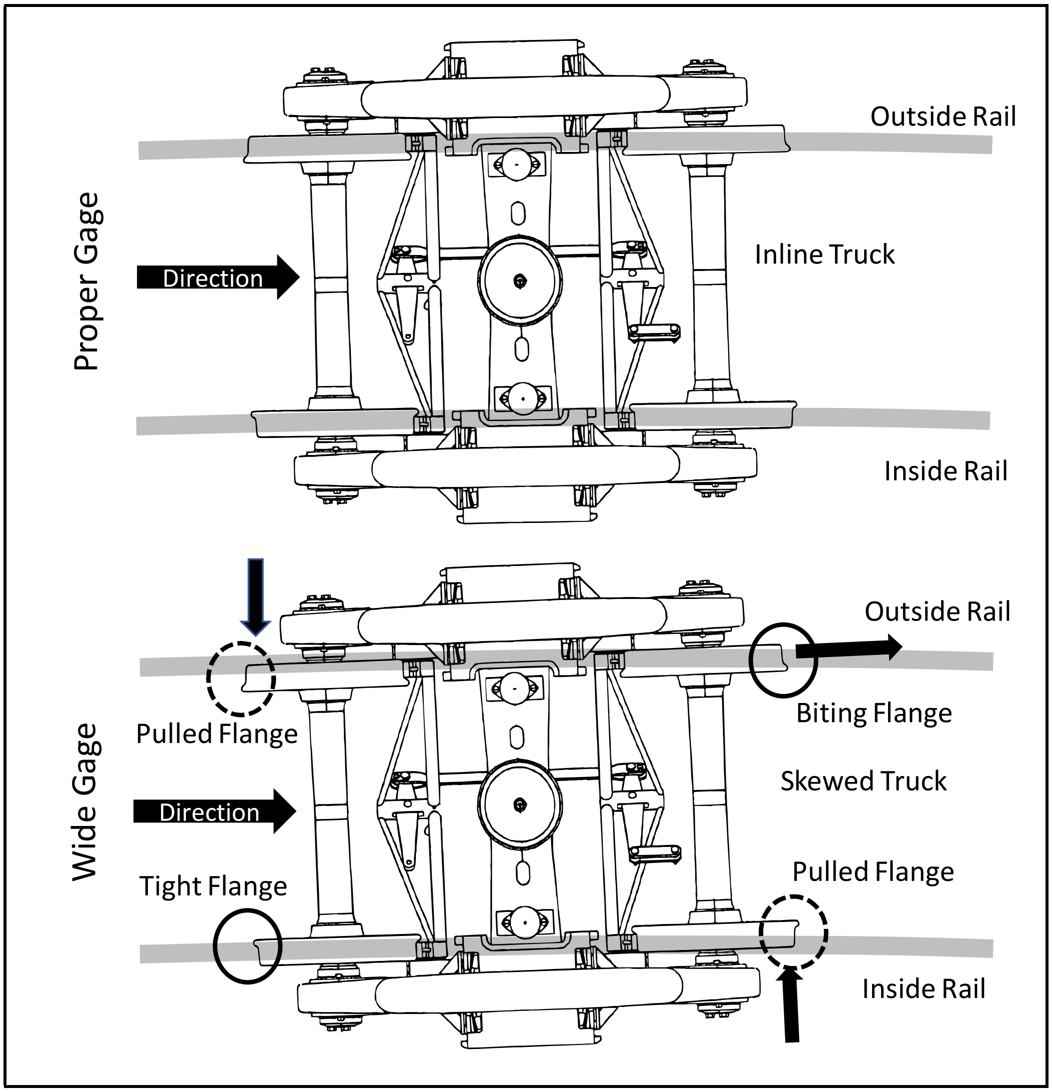

Figure 4

Even with ideal track and equipment, trucks tend to skew. For example, trucks tend to skew with certain equipment configurations and track conditions. For instance, disproportionate rail lubrication can cause skewing. In certain circumstances, trucks can perform like the "wide gauge" scenario illustrated in Figure 4. This is especially true with curves. Such an interaction negates the benefit of the rolling radius differential. It affects the trailing wheelset. Early intermodal equipment with articulated trucks developed this type of dynamic action.

Modern advances, such as radial trucks that position the axles in line with the track curvature, can manage to skew. Rail engineering advancements include modern rail lubrication processes and rail profile grinding. These advancements can help guide or “steer” trucks.

Conclusion

In conclusion, the standard worldwide railroad track gauge is now set at 4'8½". Its evolution reflects a historical journey marked by practical considerations and engineering adaptations. History traces the origins of this standard gauge to ancient chariot wheel grooves in Rome or through the lens of engineering evolution.

The Chariot Wheel Perspective suggests that the gauge lineage began in ancient times. Wagonways in Britain adopted a gauge close to the Roman chariot grooves. George Stephenson is the founding father of modern railroading. He played a crucial role in the adoption of the 4'8½" gauge. He adjusted the gauge to reduce friction and improve efficiency.

The Engineering Evolution Perspective delves into engineering adaptations. This led to the establishment of the 4'8½" gauge. The transition from outside to inside flanges resulted in this gauge. That ensures stable wheelset movement and addresses challenges like wheel climb dynamics.

The chosen standard gauge lacks a critical reason. However, the text emphasizes the global prevalence of the 4'8½" gauge. Sixty-one percent of track miles worldwide adhere to this standard. Other gauges include variations like narrow and broad gauges. The study of various track gauges illustrates their historical significance and practical implications.

Safety regulations, such as those in Title 49 of the Federal Code of Regulations, address the maintenance of gauge. Deviations from the standard gauge, either due to wear or construction imperfections, can lead to safety concerns. These include wheel climb dynamics and rail turnover forces.

The history of the standard railroad track gauge is complex. It involves historical precedents and engineering adaptations. Ancient practices root the 4'8½" gauge. It has endured as a global standard. It has shaped the development of railroads and influenced track safety standards. ■

Notes

[1] Standard measurement points for North American common carrier track.

[2] *Trains Special 2020, Railroad Maps Vol. 2* (Waukesha: Kalmbach Media Company, 2020), pp. 8-9.

[3] Conversion of the southern railroads 5'0" gauge to the Pennsylvania Railroad's 4'9" gauge was shortsighted. Equipment interchanged with some success between railroads with 4'8½" and 4'9" gauge. Therefore, the narrower of the two (standard) would have been a more practical choice as the Pennsylvania Railroad later changed their track to the national standard, 4'8½" gauge.

[4] George W. Hilton, *American Narrow Gauge Railroads* (Stanford: Stanford University Press, 1990), pp. 3-4.

[5] Dr. Arnold D. Kerr, *Roadbed and Rails Fundamentals* (Newark: University of Delaware seminar documentation-handout, 1994), pp. 1-15.

[6] The 4'9" to 4'8½" coexisted with minimal complications because the rolling stock had the same specifications, regardless of these two gauges. PRR changed its gauge from 4'9" to 4'8½" in increments starting in the early 1900s.

[7] Robert C. Reed, *Train Wrecks* (New York: Bonanza Books, 1968), pp. 36-38.

[8] George W. Hilton, *A History of Track Gauge* (Waukesha: Kalmbach Media Company, 2006), https://trn.trains.com/railroads/abcs-of-railroading/2006/05/a-history-of-track-gauge

[9] George W. Hilton, *American Narrow Gauge Railroads* (Stanford: Stanford University Press, 1990), p. 4.

[10] Dr. Arnold D. Kerr, *Roadbed and Rails Fundamentals* (Newark: University of Delaware seminar documentation handout, 1994), p. 1. Here, Dr. Kerr also writes about the stone-paved roads that appeared during the Greek and Roman periods over 2,400 years ago. Initially built by the Etruscans in pre-Roman times, these roads had wheel guiding grooves about 3'0" apart. The Romans adopted this same gauge to connect their provinces. This conflicts with the prevailing 5'0" view.

[11] Ibid. p. 3.

[12] Clement C. Williams, *Design of Railway Location* (New York: John Wiley & Sons, 1924), p. 2. Here Williams supplies an explanation like Dr. Kerr's documentation, Roadbed and Rails Fundamentals. He explains that in 1789 the first cast-iron "edge rails" appeared with an outside measurement of 5'0" and a corresponding inside measurement of 4'8½".

[13] K. F. Antia, *Railway Track* (Bombay: The New Book Company Private Ltd., 1960) p. 4. Here, Antia succinctly matches Dr. Kerr's work with a description of the empirical change from 5'0" to 4'8½". *"The gauge to which the first railway track was laid was 4'8½" and by far the greatest mileage of the railways of the world are laid to this gauge. The reason for this particular dimension lies in the plateways of the eighteenth century. The angle iron plateways mentioned in para 101 were fixed 5'0" apart. When rails were introduced, wheels were provided with flanges on the outside of the rails and the gauge still remained 5'0". Later the flanges of wheels were fixed on the inside of rails and the same 5' gauge then became 4'8½" due to the deduction of the width of 1¾" of each rail. In America and many other countries, the same gauge was adopted."*

[14]As a reference, for most fieldwork, railroad engineering personnel use the *degree of the curve* rather than radius based on a 100'0" chord to measure and calculate horizontal alignment. As such, a 5,729.67' radius theoretically connected to each end of a 100'0" chord equals one (1) degree at the vertex of the same two-length radii. For practical application to conduct field maintenance work, engineers use a derivative of the same calculation as above but use a 62'0" chord (string) held along the gauge side of a

high rail. This arrangement calculates to about one (1) inch of middle ordinate for each degree of curvature.

[15] Truck or wheel hunting is the rapid cyclical side-to-side movement of railroad wheels. This action occurs when a wheel tread slides laterally, causing the flange to strike its corresponding rail. In a rapid-fire motion, the wheel assembly (two wheels and axle) reverses back (kicks), causing the opposing flange to strike its corresponding rail. This action will often repeat in a series of harmonic piston-like motions.

[16] Taper is the expression of an angle as a ratio. The flatter the angle, the larger the number.

[17] For concrete ties, cant is an integral design element to provide rail tilt rather than using canted tie plates.

[18] Rail transit engineers use restraining guard rails. These rails, placed next to the low rail in curves, effectively steer or guide the wheels.

[19] The term *common carrier railroad* describes operations engaged in interstate transportation commerce. The Surface Transportation Board (STB) has economic regulatory oversight of common carrier railroads such as corporate mergers, shipping rates, and the acquisition and abandonment of rail lines. For operations not under the auspices of the STB, but using the same track system, the term *general system railroad* describes such operations for FRA safety regulations. An example of a non-common carrier railroad otherwise regulated by FRA (general system) would be an industry that runs their *plant railroad* trains out from the confines of their facility onto the track of a common carrier railroad. Also see *United States Government, Title 49 Code of Federal Regulations (CFR) Appendix A to Part 209—Statement of Agency Policy Concerning Enforcement of the Federal Railroad Safety Laws* (https://www.ecfr.gov/).

[20] Ibid. at *Part 213.4 Excepted Track.* This provision allows railroads to designate low speed track into a special category. It allows railroads to continue freight only train operations (with limits on hazardous materials cars) up to ten miles per hour on track not in compliance with most but not all the of requirements for Class of Track. Track gauge in the excepted category may not exceed 4'10¼".

[21] Ibid. at *Part 213.53 Gauge*.

[22] A truck is the element of a railcar consisting of an assembly of wheels and axles. Most railcars have a set of two trucks each with four wheels (two axles). Some locomotives, heavy-weight passenger cars, and special duty freight cars have three-axle trucks. Articulated multi-wheel arrangements for heavy-duty cars support the transportation of exceptionally large dimension objects.