nature reviews physics

https://doi.org/10.1038/s42254-023-00682-0

Perspective

Check for updates

The precision measurement of the *W* boson mass and its impact on physics

Ashutosh V. Kotwal

Abstract

As a mediator of the weak nuclear force, the *W* boson influences many properties of fundamental particles and their interactions. Understanding the *W* boson as accurately as possible, including knowing its mass, has been a priority in particle physics for decades. In the past few years, in a succession of increasing-precision measurements by multiple experiments, a significant tension between the measured and predicted mass has been documented by the CDF Collaboration. Furthermore, smaller differences between different measurements exist. Because the *W* boson mass provides a window on new physics, a comparison between different measurement techniques can inform the path to further investigations. This Perspective article overviews the role of the *W* boson mass in the Standard Model of Particle Physics and its extensions, compares and contrasts its measurement techniques and discusses prospects and future directions.

Sections

Department of Physics, Duke University, Durham, NC, USA. e-mail: ashutosh.kotwal@duke.edu

Introduction

Ever since the discovery of the weak nuclear force in radioactive decays more than 100 years ago, this interaction has continued to be a source of intrigue and new insights into fundamental particles and their interactions at the smallest distances. Among other surprises, the weak nuclear force taught us that nature distinguishes between particles and their mirror images (the so-called parity violation, see ref. 1 and references therein), and that fundamental particle masses are emergent properties arising from their interaction with the constant, non-zero value of the Higgs field in the vacuum[2–8].

The W boson, as one of the carriers of the weak nuclear force (Fig. 1), has a special role in the well-developed theory of fundamental particles referred to as the Standard Model (SM)[9–11]. Despite certain conceptual limitations[12], the SM continues to be extraordinarily successful in predicting the properties of fundamental particles and their interactions as witnessed by the spate of results obtained from experiments at the Large Hadron Collider (LHC)[13,14].

However, in the past few years, the high-precision measurement of the mass of the W boson[15] was found to differ significantly (7σ) from its calculated value in the SM[16–18], becoming the latest surprise in fundamental physics that points towards inadequacies of the SM and opening doors for further revelations. This measurement confirms the trend that almost all previous measurements have also been higher than the SM expectation (Fig. 2) and is the first to achieve the requisite precision to create a sharp inconsistency with the SM.

This Perspective article recaps the role of the W boson mass, M_W, in the SM and its extensions, compares and contrasts M_W measurement techniques and discusses prospects and future directions.

Historical overview

The direct observation of the W (refs. 19,20) and Z (refs. 21,22) bosons, the mediators of the weak interaction, at the CERN S$p\bar{p}$S collider in 1983 was a major milestone in the understanding of fundamental particles and their interactions. Since then, the precise measurement of their masses has been one of the priorities of the subfield of particle physics that focuses on precision measurements. In a few years following the discovery of the W and Z bosons, their masses were measured with a precision of 2-0.5% (refs. 23–27). In the following two decades, the large electron–positron (LEP) collider at CERN[28–36] and the Tevatron proton–antiproton collider at Fermilab[37–46] advanced these measurements. They surpassed 0.1% precision in the W boson mass[47] and achieved 23 parts per million precision in the mass of the Z boson[16]. With the precise measurements of the Z boson mass, the lifetime of the muon and the coupling strength of the electromagnetic interaction, as well as the masses of the top quark and the Higgs boson, within the SM it is possible to calculate the expected mass of the W boson to an accuracy better than 0.01% (ref. 16).

Between 1995 and 2000, the Tevatron collider was upgraded to deliver a factor of 100 more data than its 1992–1995 run. This second run operated between 2000 and 2011. Along with many other physics goals, improving the precision of the W boson mass measurement to test the SM was a priority for the Tevatron experiments CDF and DØ. Using the steadily increasing data sets, both CDF[48–51] and DØ[52–54] produced multiple measurements of the W boson mass with increasing precision, reaching 0.02% in 2012 (ref. 55). In parallel, the LHC started data collection in 2009 and some of these data have been analysed to publish M_W measurements by the ATLAS[56,57] and LHCb[58] experiments, achieving precisions of 0.02% and 0.04%, respectively. The latest result from CDF in 2022 uses the complete Tevatron data set and achieves a precision of almost 0.01% (ref. 15). This measurement reveals a 7σ tension with the SM expectation, prompting a flood of speculations about the implications for new physics.

The measurement of the W boson mass is substantially more complicated than the Z boson mass, even though both particles have clean two-body decay modes containing electrons or muons. Most of these complications arise from the presence of the neutrino in the W boson decay, which is undetectable yet carries away half of the rest-mass energy of the boson. The inference of the latter from the partial information available in the final state is one of the reasons that each iteration of the data analysis has required many years. W bosons also decay to quark–antiquark pairs. This mode has never been used to make a precise measurement of M_W at a hadron collider, because this mode is swamped by non-W background processes and because the measurement of the net four-momentum of the daughter particles is not sufficiently precise or accurate.

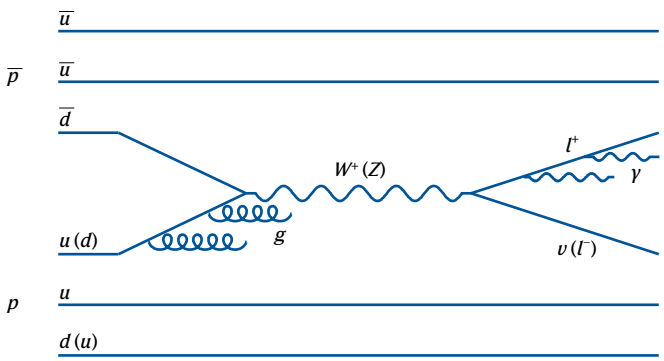

Fig. 1 | Illustration of W (or Z) boson production owing to quark (q)–antiquark ($\bar{q}$) annihilation in $p\bar{p}$ collisions at the Tevatron. The initial-state q ($\bar{q}$) are predominantly intrinsic to the proton p (antiproton $\bar{p}$). Quarks associated with the annihilation to a Z boson are indicated in parentheses. Initial-state gluon (g) radiation and final-state photon (γ) radiation are also shown. The initial-state quantum chromodynamics radiation fragments into a spray of hadrons. At the Large Hadron Collider, the initiating $\bar{q}$ arises from $g \to q\bar{q}$ splitting. Figure reprinted with permission from ref. 51, APS.

Theoretical motivation

The mass of the W boson is a very interesting observable because there is a strong theoretical connection between the masses of the W and Z bosons in the SM. To understand this connection, one can imagine an inaccurate, yet simplest, version of the weak interaction based purely on the SU(2) gauge group. In the SM, the principle of gauge symmetry of the fermionic wave functions is invoked, under which the fermions undergo transformations in an internal space while maintaining the invariance of the theory. The transformation is referred to as a gauge transformation when it is a smooth (but otherwise arbitrary) function of spacetime. The manifold of these symmetry transformations is described by a chosen Lie group, which is referred to as the gauge group.

The simplest gauge group is U(1) whose manifold is a circle; therefore, this gauge transformation changes the phase of the wave function smoothly with respect to spacetime. The principle of gauge symmetry states that the equation of motion of the fermion should be invariant under these arbitrary transformations. This is accomplished by corresponding changes in the induced vector potential by the spacetime derivative of the arbitrary phase function. The theory with this choice of gauge group describes fermions with electromagnetic

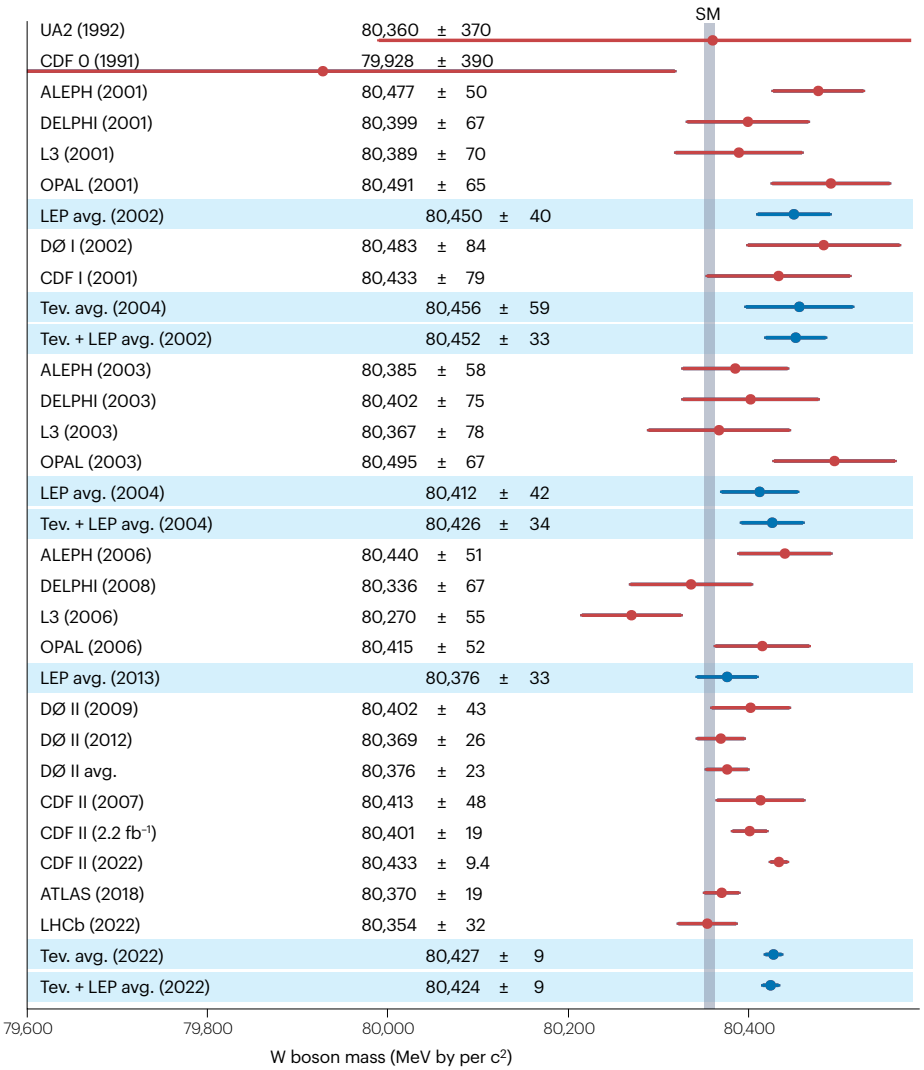

Fig. 2 | *W* boson mass measurements and the Standard Model expectation. The Standard Model (SM) expectation includes the estimated uncertainty (4 MeV) owing to higher-order effects in the perturbation theory, and the uncertainty (4 MeV) induced by the experimental inputs to the calculation, such as the top quark mass. Tev. refers to the CDF and DØ experiments at the Fermilab Tevatron. Experimental measurements and their total uncertainties are shown as red points, and averages of two or more experimental measurements are shown as blue points. ALEPH, DELPHI and L3 measurements at CERN LEP are re-analyses of the full data set (1995–2000), whereas the 2001 and 2003 results of OPAL used the first 60% of the full data set. DØ measurements are published from independent data collected during successive Tevatron running periods. CDF II measurements are published using cumulative data sets so that all available CDF II data are analysed for each publication. For the Tevatron experiments, I and II refer to the 1992–1995 and 2000–2011 running periods, respectively. The CDF II (2.2 fb^{-1}) value is from ref. 15, updated from the value $M_W = 80{,}387 \pm 19$ MeV (refs. 50,51) to include a correction for a technical oversight in the CDF track reconstruction software, as explained in ref. 15. The measurements from the UA1 experiment[23] of $M_W = 82.7 \pm 1.0_{stat} \pm 2.7_{syst}$ ($81.8^{+6.0}_{-5.3}$(stat) $\pm 2.6_{syst}$) GeV in the electron (muon) channel would be plotted to the right of this figure.

interactions. The generators of the Lie group are identified with the vector bosons that mediate the interaction; the U(1) generator is identified with the photon.

When the gauge group is SU(2) (whose manifold is the unit three-sphere S^3), the fermions are described in the internal space by two components that transform into each other under the action of the symmetry. As a gauge transformation, it is again local, meaning that it is a matrix-valued function of spacetime. Therefore, the induced vector potential is also matrix-valued and couples the two components of the fermion wave function to each other. For the weak interaction, the vector bosons thus mediate transformations between the electron and the electron-neutrino, between the up and down quarks and in general between the two components of the weak-SU(2) doublet of fermions.

The infinitesimal SU(2) group transformations have a correspondence with infinitesimal rotations of a sphere. The basis vectors of a two-component fermionic wave function may be visualized as the poles of the sphere. (This analogy captures some of the properties of the SU(2) action, but not all of them. The precise mathematical discussions can be found in refs. 59,60.) This correspondence is in analogy with SU(2) being used to describe rotations of electrons in space, where the components of the spinors of SU(2) describe the spin-up and spin-down states of electrons. Continuing this analogy, the weak interaction transforms a fermion doublet from one component to the other by the emission or absorption of a *W* boson. This description requires the existence of a third transformation that can be visualized as a rotation along the equator of the sphere. Thus, the SU(2) gauge group predicts three gauge bosons, which mathematically serve as the generators of group transformations and physically mediate all possible weak interactions. In this simplest picture, the three types of interactions associated with the weak force must be exactly symmetric because they are related by the rotational symmetry.

The principle of symmetry under gauge transformations (called gauge invariance) predicts that these transformations must be induced by massless gauge bosons in the relativistic quantum theory. This principle has served to beautifully explain the massless photon as the gauge mediator of the electromagnetic interaction and the massless gluons as the mediators of the strong nuclear force when probed at high energies.

The weak interaction, however, appears to be mediated by very massive bosons, which explains why this interaction appears to be

'weak' at low energies. This mystery led to the postulate of the Anderson–Brout–Englert–Higgs–Guralnik–Hagan–Kibble mechanism[2–4,6], in which the Lorentz-invariant Higgs field develops a constant, non-zero value in the vacuum. The interactions of all particles experiencing the weak force with this Higgs condensate impart the emergent property of mass to them as a low-energy manifestation.

In this simplest description of the weak-interaction group, the three gauge bosons will acquire the same mass because of their mutual rotational symmetry – the two W bosons and the third would-be Z boson. From this starting point, one can sequentially incorporate the known principles of nature that modify the original rotational symmetry. The biggest effect is the quantum–mechanical mixing between the third SU(2) generator and another, electromagnetism-like U(1) interaction called hypercharge, yielding the observed Z boson and the photon (γ) as the mixed states[9]. The Higgs field is an SU(2) doublet. The quantum numbers of its condensate are such that one of the mixed states remains massless; this state is identified as the photon[10,11]. In other words, although the condensate (the true vacuum) is charged under $SU(2)_{weak}$ and under $U(1)_{hypercharge}$, the linear combination of these charges that is identified as electric charge is zero for the condensate. Therefore, the vacuum is electrically neutral, gauge invariance is maintained for electromagnetism, electric current is conserved and the photon is massless. However, the weak charge and hypercharge of the Higgs condensate ensure that the W and Z bosons acquire masses whose values are proportional to their respective dimensionless coupling to the condensate and the energy-dimensioned value of the latter. This means that the weak current and the hypercharge current are not separately conserved after the Higgs condensate forms[10,11].

The electromagnetic fine structure coupling α, the Fermi coupling constant G_F (extracted from lifetime of the muon) and the Z boson mass are the most precisely measured observables[16] related to the three parameters of the theory; the SU(2) and U(1) gauge couplings and the value of the Higgs condensate. With the tight constraints on the parameters from these measurements, the theory is rendered predictive for M_W. In other words, the theory predicts M_W when α, G_F and the Z boson mass M_Z are fixed to their measured values[61,62]:

$$M_W^2\left(1-\frac{M_W^2}{M_Z^2}\right)=\frac{\pi\alpha}{\sqrt{2}G_F}(1+\Delta r), \qquad (1)$$

in which Δr denotes the quantum–mechanical radiative corrections (explained in the next paragraph) and classical physics predicts $\Delta r = 0$. The mixing between one of the SU(2) generators and the U(1) generator violates the original rotational symmetry (called custodial symmetry) between the three SU(2) generators; the Z boson mass differs from the W boson mass, with $M_Z > M_W$ as the photon mass $M_\gamma = 0$. M_W is thus a probe of such custodial symmetry-violating effects, which makes its precise measurement of high importance to check for a deviation from the calculated value[63].

The SU(2) × U(1) mixing effect is within the realm of classical physics. In the 1990s, the precision of the M_W measurement became sufficient to discern the small quantum effects denoted by Δr; these are due to vacuum fluctuations wherein particle–antiparticle pairs appear and disappear over very short time intervals, consistent with the Heisenberg uncertainty principle and special relativity. The gauge bosons experience quantum fluctuations into fermion–antifermion pairs[64].

From the perspective of the W boson mass, the most interesting fluctuations involve the top quark and the bottom quark, members of a weak-SU(2) doublet. However, their interaction with the Higgs field is very different, resulting in a large difference in their masses and also a large violation of the custodial symmetry that tries to maintain the similarity of the W and Z boson masses. Hence, quantum fluctuations in the gauge boson propagators involving weak-SU(2) doublet components with different masses induce a calculable modification of M_W for a fixed M_Z; this modification is dominated by fluctuations into top–bottom quarks, which have by far the largest mass difference among all SU(2) doublets. In other words, mass differences between weak-SU(2) doublet fermions represent another violation of the custodial symmetry and therefore induce a shift in M_W through quantum effects[65].

Historically, an estimate of the top quark mass was inferred from such quantum corrections[66]. Since the top quark discovery by the CDF[67] and DØ[68] experiments at the Tevatron in 1995, its mass has been measured directly and included as an input in the calculation[16]. Finally, quantum corrections involving the Higgs boson also induce a shift in the expected value of M_W (ref. 69), and this became a method for inferring the Higgs boson mass before the latter was directly observed.

This history illustrates how the W boson mass, in conjunction with other measurements, has been a harbinger of fundamental principles of nature. Looking towards the future: with increasing precision, M_W can probe additional laws of physics that may follow this pattern of custodial-symmetry violation between the W and Z bosons. Hypotheses include a new, weak force (which might mediate an interaction with dark matter) that mixes with the photon and the Z boson, the existence of additional Higgs-like bosons that interact differently with W versus Z bosons and heavy, weakly-interacting fermion or boson pairs that have a large mass difference akin to the top–bottom mass difference.

A popular hypothesis that accommodates the latter is supersymmetry, in which all SM particles have supersymmetric partners whose intrinsic spin differs by $\hbar/2$ from their respective SM counterparts. (The reduced Planck's constant $\hbar \equiv h/(2\pi)$ is the fundamental unit of angular momentum.) Such heavy particles may also arise as composite states owing to substructure and a new strongly-confining interaction at the TeV energy scale. This new interaction and substructure may provide a dynamical explanation for the formation of the Higgs condensate, and the Higgs boson may be a bound state rather than the fundamental scalar of the SM. Conceivably, this new dynamical sector is connected with dark matter particles, or with the mystery of the excess of matter over antimatter in the Universe, or both. Both supersymmetry and compositeness allow for a new conserved quantum number that stabilizes one or more of the new particles, which, if interacting very weakly with SM particles, can constitute the dark matter.

These are some of the hypotheses being investigated as explanations of the difference between the CDF measurement and the SM expectation of M_W.

Measurement of the W boson mass

The W boson mass has been measured both at proton–(anti)proton colliders and electron–positron colliders. It is healthy for the scientific development of the field to perform measurements under different experimental circumstances and with different techniques.

The largest samples of W and Z bosons are produced at proton–(anti)proton colliders. The bosons are mostly produced singly from the annihilation of a quark and an antiquark, respectively, from the colliding beam particles (Fig. 1). These initial-state particles are strongly interacting, that is, the principle of gauge symmetry with the gauge group SU(3) explains the 'strong' interactions between quarks as being mediated by gluons. Analogous to the electric charge associated with particles that interact electromagnetically, the 'charge'

quantum number associated with the strong interaction is called colour. When a coloured quark and antiquark collide to produce a colourless electroweak boson, they are decelerated, which causes them to radiate gluons. Thus, electroweak gauge boson production is accompanied by multiple radiated gluons, which further split into quark–antiquark pairs. This shower of coloured particles coalesces into a spray of colour-neutral bound states called hadrons, owing to the colour-confining nature of the strong interaction.

The decay products of the W boson provide insufficient information to fully reconstruct the quark–antiquark collision, owing to the invisible momentum carried away by the neutrino. The measurement of the net momentum carried by the accompanying spray of hadrons is therefore necessary to recover as much of the missing information as possible.

The SU(3)-gauge theory of strong interactions called quantum chromodynamics (QCD) has made enormous progress over the past five decades. In one of its great insights, the combined effect of quantum fluctuations involving quarks and gluons is to reduce the effective colour charge as the energy of the scattering process increases[70,71]. This property of asymptotic freedom of QCD makes it possible to calculate high-energy QCD-radiative processes in the perturbation theory. However, the perturbation theory becomes progressively inaccurate when the relevant process energy decreases to the GeV level and the effective QCD coupling is no longer small enough.

Sensitivity to QCD calculations

The inference of the neutrino momentum is aided by the use of theoretical calculations of W boson production and decay, which take into account the effects of the strongly interacting QCD radiation. Over the decades, these calculations have progressed in accuracy by including quantum–mechanical amplitudes at higher order in the perturbation theory.

An important theme in the M_W measurement is the ability to improve the accuracy of these calculations by applying constraints from the data. In this respect, the Tevatron has some advantages over the LHC. First, at the Tevatron, antiprotons already contain antiquarks, whereas antiquarks have to be generated through the strong interaction radiative processes when using proton beams at the LHC[56,57]. Second, the fractional momenta of the (anti)quarks with respect to the (anti)protons at the Tevatron are larger, and their probability distributions are better measured and less affected by QCD radiation, in comparison to the corresponding distributions at the LHC[15]. Third, W boson production at the Tevatron is minimally induced by the heavy bottom, charm and strange quarks, which do contribute substantially at the LHC[56,57] and whose effects are difficult to calculate.

Within the general theme of constraining the QCD modelling of W boson production with measurements of the accompanying QCD radiation, there are differences between the methodologies used by the Tevatron experiments (CDF and DØ) and the LHC experiments (ATLAS and LHCb, with the first measurement from the CMS experiment to be expected).

The Tevatron experiments have consistently used a calculation of vector boson production that includes the effects of multiple soft-gluon radiation and gluon splitting into quark–antiquark pairs, coupled to an exact calculation of a single energetic gluon emission or gluon splitting[72–74]. There has been discussion of this boson production and decay calculation in the context of the highest-precision measurement of the W boson mass by the CDF experiment[15]. (The CDF 2022 measurement uses the latest version of the ResBos program, colloquially referred to as ResBos2, with next-to-next-to-leading logarithm resummation and next-to-leading order accuracy in QCD in the vector boson production and decay.) One of the discussion points is the effect of double gluon radiation or gluon splitting with high transverse momentum (p_{T}). Three facts shed light on this particular point.

First, the phase space of the W bosons selected for this analysis is confined to small p_{T}, in which the 'resummation' of multiple gluon emission and splitting is the most relevant description of the boson production process. The so-called fixed-order calculation of individual emissions is already a minor correction in this phase space. Second, the calculation is tuned on the observed p_{T} spectrum of the Z boson, which constrains the QCD parameters that are largely identical between Z and W boson production. All hadron-collider experiments follow this strategy.

Third, a unique feature of the CDF measurement is the direct application of the W boson p_{T} data as an additional constraint on the corresponding calculation. The perturbative QCD calculation of the W boson p_T spectrum may not be sufficiently accurate even after the calculation has been adjusted so that the calculated and the measured Z boson p_T spectra agree. The residual variation in the calculated p_{T}^{W} spectrum owing to unconstrained degrees of freedom has an impact on the M_W measurement. However, no other experimental analysis has constrained the calculation of the W boson p_{T} spectrum with the corresponding observed distribution in the signal sample, as CDF has.

This constraint would be difficult to apply to the LHC experiments because of the degraded resolution on p_{T}^{W} owing to the large number of additional beam–beam interactions occurring simultaneously (called pile-up events) at the LHC. The Tevatron experiments have the advantage of running with lower pile-up, which permits CDF to measure p_{T}^{W} with adequate resolution to apply this constraint.

In this context, the details of W versus Z boson production merit further discussion. For massless initial-state quarks (a valid approximation at the Tevatron), CDF has used the DYqT calculation[75,76] to confirm that the p_{T}^{W} and p_{T}^{Z} spectra would be identical if the bosons had the same mass. Therefore, the difference between the W and Z bosons in the flavour composition of the initial-state quarks does not impact perturbative-QCD processes. The difference in the boson mass scales, however, does change the phase space available for QCD radiation and impacts the $p_{\mathrm{T}}^{W}/p_{\mathrm{T}}^{Z}$ ratio of the spectra. The variation in this ratio can be parameterized in terms of the renormalization, factorization and resummation scales (in increasing order of sensitivity), which capture the perturbative inaccuracy of the QCD calculation. CDF constrained this parameterization using the measured W boson p_{T} spectrum. In the context of perturbative QCD, the concern that the tuned Z boson production model may not be portable to W bosons has been addressed directly only by the CDF analysis.

In the ATLAS analysis[56,57], an attempt was made to check the theoretical $p_{\mathrm{T}}^{W}/p_{\mathrm{T}}^{Z}$ spectrum ratio against a low-precision measurement made with $\mathcal{O}(1\%)$ of the M_W-measurement data set. Calculations were also compared with the tail of the observed distribution of the component of $\mathbf{p}_{\mathrm{T}}^{W}$ along the charged-lepton direction. This quantity was presented as a kind of proxy for p_{T}^{W} and used to reject certain calculations of the $p_{\mathrm{T}}^{W}/p_{\mathrm{T}}^{Z}$ spectrum ratio.

In the LHCb analysis[58], a measurement of $\mathbf{p}_{\mathrm{T}}^{W}$ was not possible because the detector does not surround the beam collision point. An attempt was made to gain sensitivity to the p_{T}^{W} spectrum by floating an additional nuisance parameter in the fit to the $q^{\mu}/p_{\mathrm{T}}^{\mu}$ spectrum of muons (of charge q^{μ}) from $W^{\pm} \to \mu^{\pm}\nu$ data. This parameter converges to a different value from the one that best fits the p_{T}^{Z} data. If they are

constrained to a common value, the inferred M_W increases by 39 MeV, although the quality of the fit degrades. This situation illustrates why a direct constraint from p_T^W data, as obtained by CDF, is preferable to the use of a proxy or multivariate fits with hidden dependencies that are difficult to control.

At very low boson p_T, the effective strong-interaction coupling becomes large enough that non-perturbative QCD processes dominate, which may differentiate between Z and W bosons. In the limit of massless quarks, this possibility is considered in the CDF analysis by including a mass dependence in the non-perturbative parameters. One parameter is constrained with the CDF data on the Z boson, and a published global fit is used for the second parameter. The global fit includes the measured p_T spectra of low-mass vector bosons[74]. As this global fit was performed about 20 years ago, it would be useful to repeat it with new data on low-mass boson production and with newer computational tools, to help confirm the robustness of the CDF analysis in this respect.

In summary, all hadron-collider analyses of M_W tune parameters of the QCD calculation of boson production to match the observed p_T^Z distribution. Other than CDF, all other analyses assume that the tuned calculation automatically describes W boson production accurately, although additional parameters are known to differentiate between W and Z boson production. What sets the CDF analysis apart is its use of the observed p_T^W distribution to constrain the most important of these parameters, demonstrating its reliability in this respect.

Theoretical issues in non-perturbative QCD. The preceding discussion of electroweak boson production, framed in the context of massless quarks, suggests the following investigation. A fraction of W bosons are produced with a charm quark in the initial state, whose mass of ~1.3 GeV may induce a non-perturbative effect that differentiates W from Z boson production and spoil the portability of the Z boson production model to W bosons. At the Tevatron, only a small percentage of W boson production is induced by charm quarks, whereas this fraction is many times larger at the LHC. The effect of the charm quark mass is expected to be negligible at the Tevatron; a definitive confirmation requires a QCD calculation that includes the charm quark mass. Such a calculation would be a welcome addition in the domain of high-accuracy QCD calculations of electroweak boson production. Similar considerations apply for bottom quark-induced Z boson production at the LHC. As heavy-quark mass effects are a source of systematic uncertainty at the LHC, M_W measurements at the LHC would benefit from this calculation.

Another aspect of non-perturbative QCD is worth exploring. The wave functions of up and down quarks in nucleons are governed by non-perturbative QCD dynamics. As there is a net of two up quarks and one down quark in the proton, Fermi–Dirac statistics suggests a difference between the up-quark and down-quark wave functions and their intrinsic p_T. As the mix of up and down quarks differs between W and Z bosons, the intrinsic p_T difference may lead to small differences in the W and Z boson p_T spectra at small p_T. (We thank D. Soper, J. Collins and T. Han for this suggestion.) This is another example of non-perturbative effects that can spoil the portability of the p_T^Z spectrum parameters to the p_T^W spectrum. The quark-flavour dependence of the intrinsic p_T is not known; its theoretical investigation would benefit all hadron-collider analyses of M_W.

Thus, improvements in the theoretical understanding of heavy quarks and of the flavour-dependent non-perturbative aspects of proton structure would help to port the Z boson model to W bosons at very low p_T.

Momentum calibration of the experiment

One of the most important aspects of the M_W analysis is the calibration of the momentum of the charged lepton (electron or muon) emitted in the W boson decay. Charged-particle momenta are measured using a magnetic tracker, consisting of position-measuring sensors arranged in a concentric cylindrical configuration and immersed in an axial magnetic field. The position measurements along the helical trajectories are used to infer the curvature, the direction and, in conjunction with the magnetic field, the momentum vector. Electrons and photons are measured using electromagnetic calorimeters that convert the energy of the particle into a shower of soft photons and electron–positron pairs in the nuclear electric field of a dense material. As this cloud of charges traverses instrumented active material, it creates scintillation light or ionization, which is converted to a proportional electrical signal as a measure of the incident electron/photon energy.

The momentum calibration philosophies followed by the hadron-collider experiments fall into two categories: one followed by the ATLAS, DØ and LHCb experiments and the other followed by the CDF experiment.

The ATLAS and DØ experiments make no attempt to calibrate the magnetic tracker from first principles. Instead, they calibrate solely on the Z boson mass, that is, they adjust the detector response functions until the Z boson mass measured using the adjusted lepton momenta agrees with the previously measured precise value from the LEP collider. At first glance, this procedure seems sound as $M_Z = 91.19$ GeV is similar to $M_W \approx 80.4$ GeV and the calibration has to be ported over a fraction of the relevant momentum range. However, the difference between M_W and M_Z is 11 GeV; when the M_W measurement reaches the precision level of 10^{-4}, there is no a priori proof that porting the M_Z-based calibration to M_W is fundamentally understood at the 10^{-3} level. In more than three decades of M_W measurements at hadron colliders, the sole reliance on M_Z for calibration has never been formally proven to be sufficiently accurate.

The M_W analysis from the complete Tevatron Run 2 data set by the DØ experiment is an instance in which the M_Z-based calibration strategy did not ultimately succeed. The DØ experiment has not analysed the latter 40% of Run 2 data, apparently owing to tracker degradation. (Private communication from the DØ Collaboration.) The larger implication is that calibration on M_Z alone is insufficient for the requisite accuracy and robustness, owing to a dependence on tracker information. This operational experience and its relevance for M_W measurement would be useful to document.

Conscious of potential weaknesses of a solely M_Z-based calibration, all three CDF publications from Run 2 data emphasized a more robust calibration and analysis strategy from the beginning. In contrast with the other analyses, the CDF procedure for charged-lepton momentum calibration relies on fundamental principles of operation of the sensor technologies. The first step is the alignment and calibration of the drift chamber[77], which yields the calibrated track momentum of all charged particles. CDF has developed a special technique using cosmic rays[78] to pin down the positions of the 30,240 sense wires in the drift chamber to a precision of ~1 μm, eliminating many biases in the measurement of track parameters[79]. The procedure was refined over three iterations in the past two decades, each iteration requiring between 1 and 2 years.

Cosmic rays provide the most direct and transparent method of aligning the position-measuring sensors. As these muons rain down on the detectors from the atmosphere, they traverse the tracking sensors diametrically across their cylindrical geometry. This trajectory builds in a powerful redundancy because the same muon is measured twice,

on opposite sides of the beam axis. In the CDF procedure, a single helical fit is performed to this double-sided trajectory[78]. This fit is immune to many of the deformations of the tracker that bias collider tracks; the latter are prone to bias because they propagate outwards from the beam axis. Owing to this intrinsic rigidity, the double-sided fitted trajectory of the cosmic ray provides the most robust reference with respect to which the sensors can be aligned, thus eliminating the deformation-induced bias[79]. Given the large size of the cosmic-ray sample at CDF, different sensors overlap sufficiently that they are all anchored with respect to each other with a precision of ~1 μm.

The CDF procedure emphasizes the importance of control; the observable (observables) chosen for calibrating any parameter (parameters) should: first, be a strong function of the parameter (parameters) of interest, and second, be as independent as possible of other variables. The latter is a crucial consideration because it minimizes the chance of being misled in the parameter inference. Both requirements are strongly satisfied by the CDF cosmic-ray alignment procedure – it isolates the degrees of freedom of the sensor positions and is almost independent of other variables. It is then possible for CDF to analytically identify the remaining modes of displacement that cannot be measured by the cosmic-ray method; these are pinned down using complementary techniques[15]. These features distinguish the CDF tracker alignment from the procedure published by, for example, the LHCb experiment[58].

The next step in the CDF procedure is the calibration of the magnetic field and its spatial non-uniformity using the reconstructed invariant mass of the J/ψ particle, whose mass is precisely known, in its dimuonic decay channel. This large sample enables a fine-grained spatial scan of the drift chamber. These muons also cover a wide range in track curvature, the quantity directly measured by the drift chamber, and enable the calibration of momentum-dependent effects such as the ionization energy loss in the passive material[15,49,51].

Additionally, CDF uses a large sample of Υ particles decaying to dimuons to demonstrate the consistency and robustness of the tracker calibration. First, Υs with a higher (and again precisely known) mass than the J/ψ cover the phase space closer to W bosons. Second, the muon tracks used for the $\Upsilon \to \mu^+\mu^-$ mass calculation can be reconstructed with two variants of the track reconstruction algorithm. The consistency of the measured Υ mass between these variants is another important demonstration of the robustness of the momentum calibration[15,49,51].

Instead of a purely empirical calibration of the electromagnetic calorimeter, CDF performs a detailed first-principles simulation of electron and photon detection[80] based on the knowledge of the calorimeter geometry and electromagnetic showering calculations encoded in the GEANT program[81,82]. Next, CDF uses the electrons emanating from W boson decay, with their calibrated momenta from the tracker, as an in situ test beam to measure the calorimeter thickness, nonlinear energy response owing to scintillator ageing, spatial and temporal variations in response and the amount of passive radiative material upstream of the tracker. Almost all tuned parameters are causally related to the principles governing electron and photon interactions with various materials[15,49,51].

The ultimate confirmation of these calibration procedures is the independent measurement of the Z boson mass by CDF in both electron and muon channels, using the same 'blinded' procedure as used for the M_W measurement and demonstrating consistency with the LEP value. Among all hadron-collider measurements of the W boson mass, the three CDF measurements published from the Tevatron Run 2 data are unique in pursuing these high-precision calibrations and cross-checks from first principles[15,49,51]. All other hadron-collider measurements have been based solely on rescaling various observables based on the reconstructed Z boson mass.

The LHCb experiment uses a mix of $J/\psi \to \mu^+\mu^-$, $\Upsilon \to \mu^+\mu^-$ and $Z \to \mu^+\mu^-$ data to understand the tracker. Misalignments are parameterized by corrections to the track curvature and the parameters are fitted using a proxy for the reconstructed Z boson mass in the data. The momentum scale and resolution parameters are fit simultaneously to the J/ψ, Υ and Z boson data; however, the mutual consistency of the momentum calibration from independent data sets was not demonstrated[58].

For the M_W measurement, the CDF strategy benefits from unique features of its drift chamber geometry and construction, features that are lacking in the trackers of the other hadron-collider experiments. First, the CDF drift chamber is a single gaseous volume with embedded sense wires. Groups of 12 wires (cells) are mounted on precision-installed cards at each end of the cylindrical chamber, substantially reducing the number of alignment degrees of freedom[77]. In comparison, the LHC experiments use silicon sensor-based trackers, which consist of many tens of thousands of small modules, leading to an explosion in the alignment degrees of freedom and a corresponding increase in tracker complexity for fundamental understanding.

The second advantage conferred by the CDF drift chamber is that, as a charged particle traverses 96 sense wires radially, there are no gaps or discontinuities in the instrumented volume. All particles must pass through 96 drift regions in which their positions are recorded, without exception. In particular, for incremental changes in the curvature or direction of the particle, the relevant active regions change continuously and smoothly, with a handful of transitions between adjacent drift cells. This is the reason the curvature response function for this drift chamber is analytic, that is, the measured track curvature is a smooth function of the true curvature and can be expressed as a polynomial function. Pinning down the few relevant polynomial coefficients is at the heart of the CDF tracker calibration.

In comparison, silicon trackers provide a factor of ten fewer (but substantially more precise) measurements, with many more boundaries between instrumented regions. It is more difficult to guarantee an analytic curvature response with few measurements along the track and a highly tiled detector geometry, which is inevitable when covering a large area with numerous small planar sensors.

Silicon trackers provide considerably higher performance than drift chambers in terms of precision, rate capability, granularity, radiation hardness and the ability to operate in high-flux environments. They are also uniquely suited to be placed close to the beam interaction point and measure the track impact parameters with high precision. It is the geometrical complexity and segmentation of silicon trackers that the curvature measurement is particularly sensitive to, because curvature is related to the second radial derivative of position, whereas impact parameter is related to the first derivative. In addition, for the M_W measurement, it is the track curvature that is crucial for the accurate inference of the transverse momentum, whereas the impact parameter is irrelevant because the daughter leptons are prompt.

This discussion motivates further studies of tracker geometries and an exploration of geometrical deformations that lead to a non-analytic curvature response at small curvature. It behooves the experiments to conduct such studies to fully understand and exploit magnetic trackers for precision measurements. It would be natural for CDF to lead these studies as they are most advanced in the systematic understanding of tracker-based momentum calibration.

In the spirit of making the most reliable measurement possible, it is crucial that the M_W analysis is based on fundamental principles, which are transparent and are in the public domain. The strategy of CDF for calibrating the charged-lepton momentum exemplifies this philosophy.

Event selection

The upside of W boson production at hadron colliders is that the production rate is substantially higher than that at electron–positron colliders (such as LEP II); the complication is that this rate is less than one-millionth of the total hadron–hadron interaction rate. Real-time filtering to record events of interest while rejecting 99.99% of the uninteresting collisions is a crucial aspect of the experimental techniques at hadron colliders. In the offline analysis of the recorded data, further requirements are placed on the event information to suppress high-rate backgrounds from QCD-induced non-W events that mimic high-p_T electrons/muons along with a p_T imbalance owing to resolution, faking a neutrino signature.

It is important to minimize the selection bias on W-boson events while also minimizing the mis-identification backgrounds. Both effects alter the shape of the kinematic distributions from which M_W is inferred. The larger the bias, the more accurately it must be estimated from control samples of data or high-quality simulation.

Among the hadron-collider experiments, CDF has the highest charged-lepton trigger efficiency and the smallest mis-identification backgrounds; for W boson events selected with the charged-lepton $p_T^\ell > 30$ GeV for the analysis, the trigger efficiency is essentially 100% with no dependence on p_T^ℓ. As shown in Supplementary Fig. 40 of ref. 15, the inferred M_W is stable with respect to the p_T^ℓ requirement; no correction for trigger inefficiency is needed here. This is a substantial simplification compared with the p_T-dependent trigger inefficiencies in the ATLAS, DØ and LHCb experiments, which must be estimated and corrected for to prevent biasing the inferred M_W. The dependence of the selection bias on event topology is also the smallest in the CDF analysis, in which the simplest selection requirements are sufficient to achieve charged-lepton mis-identification backgrounds <0.4%.

Both the amount of QCD radiation accompanying the W bosons and the number of pile-up events are much smaller at the Tevatron than at the LHC. Small pile-up implies that the resolution on the inferred p_T of the neutrino p_T^ν is much better at the Tevatron. Both circumstances force the W boson selection to be $p_T^W < 30$ GeV in the ATLAS analysis, which is suboptimal compared with the $p_T^W < 15$ GeV requirement in the CDF and DØ analyses. The former admits more mis-identification background, more QCD radiation, more corrections for efficiency-related biases and considerable smearing of the kinematic distributions sensitive to M_W, all of which increase the experimental and theoretical complications associated with LHC measurements of M_W.

One concludes that the CDF experiment requires the smallest corrections for trigger-induced and selection-induced biases among the hadron-collider experiments. This is an intrinsic advantage in favour of CDF in terms of robustness of the analysis.

Hadronic recoil measurement

The $\mathbf{p}_T^\nu$ vector is inferred by measuring the $\mathbf{p}_T$ vector of all detectable particles and imposing $\mathbf{p}_T$-balance assuming a single neutrino in the event. In addition to the charged lepton (leptons) from $W(Z)$ boson decay, the initial-state QCD radiation which fragments into a spray of hadrons must also be measured (Fig. 1). The vector sum $\mathbf{u}$ of the $\mathbf{p}_T$ vectors of these hadrons is referred to as the hadronic recoil vector, as it represents the measurement of $-\mathbf{p}_T^{W(Z)}$ using the hadrons. The response and resolution functions that characterize the $\mathbf{u}$ measurement are more complicated than those for electrons and muons; fortunately, the required accuracy is in parts per thousand rather than tens of parts per million for the charged leptons.

At a hadron collider, most of the energy of the beam particles is transferred to collision fragments that propagate at small angles along the beam pipe and cannot be detected. Hence, energy and momentum conservation cannot be used to infer the longitudinal momentum of the neutrino; one can work with its $\mathbf{p}_T$ vector only.

Three kinematic quantities are sensitive to M_W; the distributions of p_T of the charged lepton and the neutrino, and the transverse mass m_T, which is akin to the invariant mass but uses the $\mathbf{p}_T$ vectors of leptons. The maximum value of the p_T of leptons (the approximate end point of distributions) is linear in M_W, but is also linear in boson p_T. The latter is incorporated from an accurate calculation, as discussed before, and measured as $\mathbf{u}$. Both inputs are important because neither is sufficient. The special property of the m_T distribution is that its end point does not depend on the boost of the boson, while still being linear in M_W. This linear dependence on M_W is used to infer the value of M_W from the shape of these distributions.

The M_W extraction from the p_T^ν distribution is by far the most sensitive to the imperfections of $\mathbf{u}$. Contrary to popular misconception, the use of the charged-lepton p_T^ℓ distribution is far from immune to $\mathbf{u}$. The m_T-based extraction has yet another set of correlated dependencies on the various inputs. One of the strengths of the Tevatron analyses is that the benefits of the m_T variable are fully exploited, because the small pile-up maintains good resolution on $\mathbf{u}$. In fact, as shown in the CDF and DØ analyses, the $\mathbf{u}$-resolution is good enough to permit a reasonably precise extraction even from the p_T^ν distribution[15,54].

The $\mathbf{p}_T^\ell$ vector has the same role in all three kinematic distributions; they differ only in the usage of $\mathbf{u}$. Hence, the consistency of the M_W values inferred from these three distributions is the most powerful cross-check on the validity of the $\mathbf{u}$ measurement and the associated uncertainties. The covariances of these M_W values with respect to the response and resolution functions of $\mathbf{u}$ are used to quantify their mutual consistency. These consistency checks have been published by CDF[15,49,51]. An additional check to examine the consistency of M_W measurements in bins of pile-up would be useful for CDF to investigate.

Unfortunately, the $\mathbf{u}$ resolution is quite degraded at the LHC owing to the larger pile-up, to the extent that the m_T-based and p_T^ν-based extractions are almost powerless in providing a validation of the $\mathbf{u}$ measurement. Yet, the p_T^ℓ-based extraction, which is essentially the sole method of M_W inference at the LHC, is sensitive to a mis-measurement of $\mathbf{u}$ through the event-selection criteria. As such, the LHC measurements of M_W lack the ability to demonstrate their robustness with respect to the hadronic recoil measurement in the same way that CDF has demonstrated. The alternatives pursued in the ATLAS analysis are not as powerful[56,57]. For example, the W-like reconstruction of the Z boson mass (Fig. 16 of ref. 56) is inconclusive at best.

This weakness can be rectified by the analysis of the low pile-up data recorded at the LHC for the M_W measurement. In the CDF analysis, the $m_T : p_T^\ell : p_T^\nu$-based extractions contribute in the proportion 64:26:10 in the combined final answer, whereas in the ATLAS analysis the p_T^ℓ-based extraction provides 95% of the final answer[83]. Even the p_T^ν-based extraction at CDF is more impactful than the m_T-based extraction at ATLAS. These weights illustrate the usefulness of the low pile-up data at the LHC.

On a related note, the component of **u** parallel to the charged-lepton direction is important; it contributes linearly to the p_T of leptons. In the CDF analysis, special attention is devoted to the understanding of this component and its implications, as documented extensively in each of the three CDF publications from the Tevatron Run 2 data. A corresponding exposition in future ATLAS publications would be welcome, as there is currently little discussion of this key issue.

The point to emphasize is that the lower pile-up confers a significant advantage to the Tevatron over the LHC, which CDF has fully exploited to show cross-checks of the hadronic recoil measurement.

Likelihood fitting

The inference of parameters of interest (POI) from data is the domain of statistical analysis. The accurate description of the data is also controlled by many other 'nuisance' parameters; so named because they need to be constrained to extract the POI, but are not the motivation for the analysis. When the POI is M_W, the nuisance parameters are related to the boson production and decay physics, the detector response and resolution functions and the non-W processes contaminating the data sample. The typical method of statistical inference is likelihood maximization, in which the likelihood function depends on the POI(s) and the nuisance parameters.

There is a divergence on the choice whether to float the POI and the nuisance parameters separately or simultaneously. Two schools of thought have emerged. The traditional, conservative and time-tested methodology in precision measurements is to float them separately. The first step is to constrain all nuisance parameters from independent control samples of data, selected to be strongly constraining on the nuisance parameters while being independent of M_W. For each nuisance parameter, a causal and explainable logic is developed, whose reasoning can be debated and improved, leading to scientific learning and progress. Likelihood maximization on these control data with respect to the nuisance parameters provides well-understood constraints on the latter. In the next step, the likelihood with respect to M_W is maximized on the m_T or the lepton p_T distributions of the data, whereas the nuisance parameters are fixed to their respective values determined from the causal analysis of the control data. The covariance of M_W in the nuisance parameter space is estimated by varying the nuisance parameters by their a priori justified uncertainties and noting the induced shift in the fitted value of M_W. The covariance is determined from the model of the data so that it does not suffer from data-driven statistical fluctuations. This methodology, applied by the Tevatron experiments, has steadily advanced our fundamental understanding of the nuisance parameters and their impact on the M_W extraction, as documented in a set of detailed publications[15,27,38,39,41,43,45,46,49,51,54]. It is considered a robust, interpretable and transparent method of data analysis because it clearly demarcates between the POI and the nuisance parameters and maintains their meaning.

The newer school of thought approaches precision analysis as a machine-learning exercise. The a priori constraints on nuisance parameters from control data are sometimes weak. In an attempt to improve these constraints, a multi-parameter likelihood function is constructed that incorporates the nuisance parameters and the POI on an almost equal footing. In this high-dimensional parameter space, the likelihood is maximized on the m_T or p_T^ℓ distributions of the W boson data. In the ATLAS re-analysis[83] of M_W, the number of nuisance parameters is initially in the hundreds.

This method appears sophisticated and increases the accuracy with which the nuisance and the interesting parameters together describe the m_T and p_T^ℓ distributions. However, the interpretation of the inferred M_W and its uncertainties is now opaque as it is mixed with hundreds of other parameters. The causal interpretation of the nuisance parameters may be lost if all parameters including M_W are reduced to mere smoothing knobs to fit a set of kinematic distributions. This raises the concern that M_W is getting tuned to smooth out internal inconsistencies, thereby losing its rigorous interpretation as a measured quantity. Similar concerns arise in the context of 'over-fitting' in the parametric modelling of data[84,85].

These issues worsen as the dimensionality of the nuisance parameter space increases; even if the initial choice of hundreds of parameters is pruned to 218, as in the ATLAS analysis, the post-fit likelihood function is still impossible to visualize and scrutinize. Mathematically, there is no guarantee that a 218-dimensional likelihood function is well behaved in the entire domain bounded by the a priori uncertainties on the nuisance parameters; meaning that there is a stable global minimum without the existence of local minima, under-constrained 'valleys' and saddle points, which can lead to chaotic paths in the numerical minimization. Such issues may lead to a 'random walk' of M_W of $\mathcal{O}(\sqrt{N}\delta)$ amounting to tens of megaelectronvolts (MeV), in which N is the (large) number of nuisance parameters floated simultaneously in the likelihood function and δ is a typical M_W deviation of a few MeV induced by each of them. Can one a priori defend the hypothesis-testing ability of a method which relies on a single POI that is susceptible to a random walk in a huge, 200+-dimensional parameter space?

If the nuisance parameters are not correlated with the POI, there is no need to float them simultaneously with the POI. If they are correlated with the POI, then it is dangerous to float them simultaneously with the POI. In either case, the more robust procedure is to determine nuisance parameters from control samples of data, not the data used to determine the POI.

The purpose of high-dimensional machine-learning models is to provide a reproduction of training data, that is, a method of mimicking data rather than a method of learning fundamental principles governing the data. There is a rapidly growing forum in computer science related to interpretability and explainability of black-box machine-learning models. Explainable models allow human users to comprehend how the model is extracting information from the data and to debate whether the model and its results can be trusted. Explainability is a requirement when a parameter that has a fundamental meaning is inferred from the data. CDF has maintained this philosophy of explainability for all its M_W measurements and will continue to do so in the future.

In the M_W context, it is important to understand the causal source of information in multi-parameter fitting, such as the reduction of the M_W uncertainty owing to the proton structure in the ATLAS analysis from a priori 28 MeV to 8 MeV after the multi-parameter fit[83]. Which aspects of the m_T and p_T^ℓ distributions are constraining which degrees of freedom pertaining to the proton structure? Are these constrained degrees of freedom consistent with other world data describing the proton structure?

It would be interesting to understand the stability of multi-parameter fitting and to compare what one seems to learn from such fitting with what one expects to learn based on first-principles reasoning. The Fisher information is a measure of how much information about any model parameter is carried by the data. Evaluating the Fisher information for the nuisance parameters can help understand whether multi-parameter fitting is resulting in legitimate or spurious constraints.

Table 1 | Comparisons of the *W* boson mass M_W analyses performed by various experiments at hadron colliders

Criterion	CDF	DØ	ATLAS	LHCb
Sensitivity to proton structure	Low	Low	High	High
Sensitivity to heavy quark effects	Low	Low	High	High
Electron channel analysed	Yes	Yes	Yes	No
Muon channel analysed	Yes	No	Yes	Yes
Check of p_T^Z model with data	Yes	Yes	Yes	Yes
Check of p_T^W/p_T^Z ratio with data	Strong	No	Weak	No
Emphasis on tracker alignment	Yes	No	No	No
Momentum calibration using J/ψ and Υ	Yes	No	No	Partial
Independent M_Z measurement and confirmation	Yes	No	No	No
Event selection bias corrections	Small	Large	Large	Large
Recoil validation with m_T, p_T^ℓ and p_T^ν fits	Strong	Modest	Weak	No
M_W-independent nuisance parameter constraints	Yes	Yes	No	No
Dimensionality of M_W-inference likelihood	1	1	218	8
Explainability of M_W inference procedure	High	High	Low	Low

Using simulated data in which the ground truth is known by design, the likelihood-fitting approaches used by the Tevatron and LHC experiments could be compared. Among other insights, such a study could help understand why re-fitting the same data published in 2018 (refs. 56,57) by ATLAS with identical calibrations caused a 16 MeV shift[83] in the extracted M_W — the only change was that nuisance parameters were floated during M_W extraction. Similarly, M_W changes by 39 MeV in the LHCb analysis when an additional parameter ascribed to the p_T^W spectrum is floated simultaneously[58]. Could these shifts be a symptom of fitting instability in the black-box approach? The spectre of jumping between local minima in the loss function haunts the use of multi-parameter fitting.

Comparisons of M_W analyses at hadron colliders

The preceding sections discussed key aspects of M_W analyses and comparisons between how different experiments addressed them. Table 1 summarizes these comparisons.

Experience of M_W measurements

The prerequisite for higher precision measurements of M_W is the increase of the data set size. It is also a learning experience because reducing the statistical error is only useful if the uncertainties owing to nuisance parameters (systematic uncertainties) are also reduced, and this requires a deeper and more informed analysis. In other words, achieving precision is as much dependent on climbing a long learning curve, as on having more data.

The progress on both fronts is notable in Fig. 2. Starting with the UA1 and UA2 experiments at the CERN S$p\bar{p}$S collider, hadron colliders have reduced the uncertainties from all sources by two orders of magnitude. During these decades, a better understanding of measurement procedures and nuisance parameters has occasionally shifted the value of a new measurement by more than the estimated uncertainty on the previous measurement. For example, the CDF measurement from the first quarter of the Tevatron Run 2 data set appears to be shifted low by 33 MeV compared with their high-precision measurement from the full data set[15]. As their analysis techniques improved substantially together with a quadrupling of the data, it would be useful for CDF to investigate whether there is a causal explanation for the shift, which would contribute to scientific learning. In the same vein, the shift of 16 MeV in the ATLAS measurement owing solely to an updated likelihood estimator and the shift of 39 MeV in the LHCb measurement owing to an additional floated parameter ascribed to the p_T^W spectrum are worth understanding at a deeper level.

Measurements of M_W have also been performed at the LEP collider at CERN using two different methods. In the first method, M_W is inferred from the rise of the WW pair-production cross-section when the electron–positron beam energies are scanned across the WW pair-mass threshold[32–35]. These measurements were limited by the data set size, but this threshold-scan technique is a robust and accurate method of measuring M_W. It could be repeated with extreme precision (~0.5 MeV) at a future electron–positron collider designed to run at much higher luminosity[86]. The threshold-scan measurement of M_W is a prime motivator for such a future collider.

The second method of measuring M_W at LEP used the reconstructed four-momenta of the final-state hadrons and charged leptons emanating from the decay of the pair-produced W bosons. When both W bosons decay hadronically, strong interactions between the decay products from the two W bosons create a bias in the inferred M_W, which is corrected on the basis of hadronization models[32–35]. Methods of attributing the momenta of measured particles to the respective W bosons evolved, as did the hadronization models[28–31]. Between 2003 and 2006–2008, all four LEP experiments updated their measurements by ~75 MeV or ≥1σ each, as seen in Fig. 2 and refs. 36,87,88. In preparation for a new electron–positron collider, revisiting the origin of these rather substantial shifts would be educational.

The lesson learnt is that understanding and control of nuisance parameters is crucial in all M_W analyses, both at lepton and hadron colliders.

Prospects

The experiments running at the LHC are technologically advanced and can be calibrated with careful analysis of in situ control samples of data. They have already collected two orders of magnitude larger data samples than the Tevatron, and another order of magnitude increase is expected from future running. However, these samples will suffer from a large rate of pile-up events because they have been or will be collected with a high beam luminosity, which enables the large signal sample rate. Ironically, the requirements of the M_W measurement from the perspective of precision and robustness are better served by a lower beam energy to simplify the W boson production dynamics and a lower beam luminosity to reduce pile-up. Although the technical and scientific cost to the rest of the LHC physics programme may outweigh the benefit of a lower beam energy, a short-term reduction in the beam luminosity is an attractive option. Data collected at low luminosity will reduce the number of additional proton–proton collisions that complicate the inference of the neutrino momentum. Among other benefits, low-luminosity data will provide the built-in consistency check of M_W extracted from the m_T, p_T^ℓ and p_T^ν distributions, which is only possible when the p_T^ν resolution is not degraded by pile-up.

In the longer term, an electron–positron collider that serves as a Higgs and electroweak factory will be the ultimate precision

machine for the measurement of W, Z and Higgs boson properties[89–92]. The beam-energy scan at the WW pair-production threshold is the most robust method of confirming the difference between the measured M_W and its SM-calculated value[86]. This difference is most striking in the recent CDF measurement, but Fig. 2 shows the historical trend for this difference to be positive.

The M_W measurement at an electron–positron future collider complements an extensive programme of precision measurements of Z bosons, which can improve upon the precision achieved at LEP by a factor of 20–100 (ref. 92). Similarly, certain properties of the Higgs boson can be measured with precision ranging from 1% to 0.2%, exceeding the capability of the LHC in these respects[92]. As with M_W, the motivation for making extremely precise measurements of such observables is that they can be calculated with commensurately high precision in the SM. A pattern of deviations between measured and calculated values may emerge that guides us to develop a deeper theory that subsumes and supersedes the SM.

Weak mixing angle

In the third section, we discussed how the SM predicts the value of M_W when other key parameters are fixed by measurements. We mentioned the weak mixing angle that describes the superposition of two generators of the $SU(2)_{weak}$ and $U(1)_{hypercharge}$ gauge groups, respectively. The SM also predicts the weak mixing angle when the aforementioned parameters are fixed[93]; thus, a precise measurement of the weak mixing angle provides a stringent test of the SM, along the same lines as and independent of M_W.

The two most precise measurements of the weak mixing angle have been performed in the 1990s by the LEP experiments and by the SLD experiment at the SLAC SLC accelerator, respectively[94]. The measurements were based on different properties of Z bosons; the SLD observable depends on leptonic couplings only, whereas the LEP observable involves b-quark couplings. The measurements differ by 3.2σ. The SLD measurement relies on beam polarization and the LEP measurement relies on identification of b-quarks produced from the Z boson decay.

This discrepancy has not been resolved; nevertheless, the two measurements have been averaged. Additional measurements have been made in atomic parity violation, electron and neutrino scattering and at the Tevatron and LHC (reviewed elsewhere[16]). Analysis of the available LHC data may yield a precision similar to that of the SLD measurement. New experiments[95–97] including those at a future electron–positron collider[89–92] plan to improve upon the precision of this parameter.

In the meantime, the status of this discrepancy in the weak mixing angle is similar to the ~3σ difference between the CDF and ATLAS measurements of M_W. Per precedence, the scientific approach is to combine the M_W measurements using established statistical methods while continuing investigations. This approach has been used when even larger differences, such as the 5.5σ between measurements of the fine structure constant α using cesium and rubidium atoms, have been found[16]. An alternate, biased approach has been presented in ref. 98. Unless scientific concerns on any published result are documented, it is not good scientific practice to isolate any measurement as an outlier and remove it from a world average. The Particle Data Group (PDG) has maintained a time-tested procedure for combining measurements that are not in perfect agreement – the PDG review is replete with examples[16].

It is interesting to note that the SLD measurement of the weak mixing angle and the CDF measurement of M_W each point towards the same new physics, qualitatively and quantitatively. This concurrence (Fig. 3) has been pointed out in ref. 99, in which it was shown that both measurements favour a similar violation of the WZ custodial symmetry that we discussed in the third section as an approximate feature of the SM.

Muon magnetic moment

Analogous to M_W and $\sin^2\theta_{eff}$, the magnetic moments of the electron and the muon are precisely measurable and their values are predicted by the SM. The magnetic moment is proportional to the intrinsic spin, the charge/mass ratio and the g-factor, in which $g = 2$ in Dirac's theory of the electron. In the quantum field theory, this prediction is modified by quantum fluctuations, which are calculable as a perturbation series with the electromagnetic fine structure constant α as the expansion parameter. A measurement of $(g - 2)$ is a stringent test of the SM and has been pursued for both electrons and muons for many decades, along with increasing accuracy of its calculations.

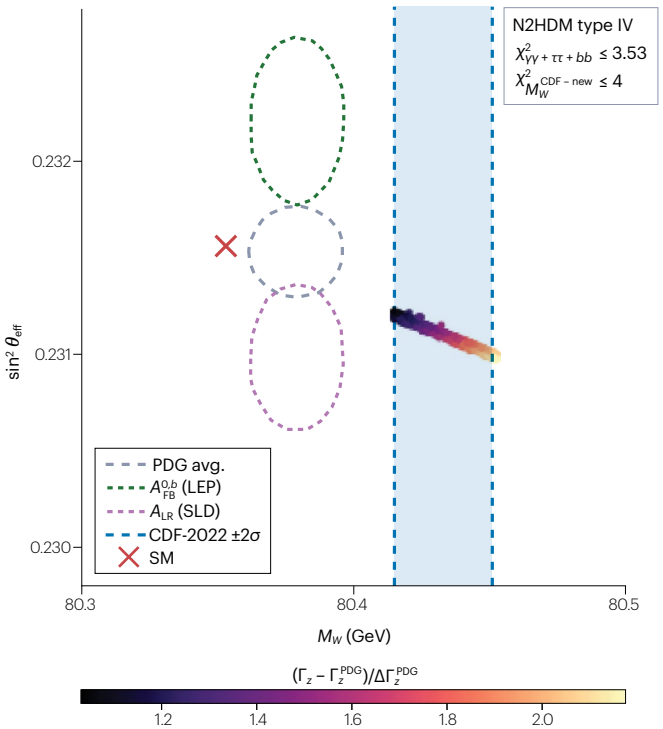

Fig. 3 | Illustration of the two most precise measurements of the weak mixing angle θ_{eff} (shown conventionally as $\sin^2\theta_{eff}$ on the vertical axis) from the large electron–positron (LEP) experiments and the SLD experiment, respectively[94], and their average[16]. The CDF experiment's measurement of the W boson mass is shown on the horizontal axis. The Standard Model (SM) expectation is shown as the orange cross. The SLD experiment's measurement of $\sin^2\theta_{eff}$ and the CDF measurement of M_W concur on a set of new-physics scenarios shown as the line of coloured points. Although a non-supersymmetric framework with additional Higgs bosons (N2HDM) is used in ref. 99, the common feature between these and other models is the violation of WZ custodial symmetry mentioned in the third section. A_{LR} refers to the measurement of the weak mixing angle derived from the cross section asymmetry between left-handed and right-handed polarizations of the electron beam, as performed by the SLD experiment. Both techniques apply to electron–positron annihilation at the Z-boson pole. PDG refers to the Particle Data Group. Figure reprinted from ref. 99, Springer Nature Limited. $A^{0,b}{}_{FB}$ refers to the measurement of the weak mixing angle derived from the forward–backward asymmetry of b-quark production, as performed by the LEP experiments.

In the past few years, a difference between the measured and calculated values of muon $(g-2)$ has been reported[100–103]. The most recent measurement from Fermilab[100] improves the precision of the experimental world average by a factor of two. Simultaneously, the SM calculation of this quantity is being revisited[104] owing to improved inputs from QCD calculations on the lattice[105,106] and from a new measurement of the $e^+e^- \to \pi^+\pi^-$ cross section near threshold[107]. This information changes the hadronic vacuum polarization and the resulting update to α would bring the SM value of the muon $(g-2)$ closer to the measured value.

The effect of low-energy hadronic quantum fluctuations on α propagates to M_W and to the muon $(g-2)$ in opposite directions[108]. In other words, if the discrepancy between the measured and calculated values was to reduce for the muon $(g-2)$ owing to an adjustment of α, the discrepancy must increase for all M_W measurements and vice versa. There is no way to reconcile both discrepancies simultaneously by adjusting this hadronic quantum correction on α.

The interesting conclusion is that the combination of these two measurements makes a striking case for new physics. Whether the new physics manifests in the magnetic moment of the muon, or the W boson mass, or both, these two measurements are hinting at a future beyond the Higgs boson.

Summary

Developed over a century, the SM is one of the crowning achievements of science. It is built on a remarkable set of axioms of local quantum field theory, special relativity and group theory and has been both strongly guided by and predictive of a vast set of experimental observations. The theoretical structure of the SM and past consistency with experimental data together led to the prediction of the Higgs boson, which was spectacularly confirmed by its observation at the LHC in 2012.

Nevertheless, the SM is not expected to be the most fundamental theory of matter and forces. First, the Higgs sector of the SM is not as axiomatic as the rest of the theory; the fermionic interactions and the self-interactions of the Higgs field are parametrized phenomenologically without a fully dynamical construction. Second, the parameter describing the Higgs boson mass receives large quantum corrections that are very sensitive to unknown physics at arbitrarily high-energy scales. The small observed value of the Higgs boson mass raises the concern of 'naturalness' of the Higgs sector, because it requires an extremely tuned set of conditions at higher energies. Such 'fine-tuning' is prevented in other sectors of the SM by its symmetry structure[109] and can be alleviated for the Higgs mass parameter as well by extending the field content and symmetries of the theory. The fine-tuning associated with the Higgs mass increases quadratically with the energy scale of new physics, which creates an expectation that the new physics could be around the corner.

Another set of deficiencies in the SM relates to crucial facts of cosmological significance — the observed excess of matter over antimatter in the universe, neutrino mass, dark matter, dark energy and cosmic inflation, all of which are beyond the scope of the SM. Finally, the force of gravity cannot be accommodated into the quantum field-theoretic framework of the SM.

For these reasons, physics beyond the SM is expected to show up. A priori there is no definitive prediction for the energy scale at which new physics may be revealed. Nevertheless, these considerations justify current and future particle physics experiments that are designed to observe phenomena not explainable by the SM.

We discussed the theoretical reasoning for the W boson mass to be a sensitive probe of certain aspects of new physics, which explains why its measurement has been pursued with increasing precision (progressing by two orders of magnitude) over the past three decades. Finally, we have an M_W measurement that shows a significant tension; it is the largest deviation observed from a prediction of the SM. The most exciting possibility is that the new physics that this deviation points towards will reveal itself as new particles in the near future.

Published online: 31 January 2024

References

1. Wu, C. S., Ambler, E., Hayward, R. W., Hoppes, D. D. & Hudson, R. P. Experimental test of parity conservation in beta decay. *Phys. Rev.* **105**, 1413–1415 (1957).
2. Anderson, P. W. Plasmons, gauge invariance, and mass. *Phys. Rev.* **130**, 439 (1963).
3. Englert, F. & Brout, R. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.* **13**, 321 (1964).
4. Higgs, P. W. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.* **13**, 508–509 (1964).
5. Higgs, P. W. Broken symmetries, massless particles and gauge fields. *Phys. Lett.* **12**, 132–133 (1964).
6. Guralnik, G. S., Hagen, C. R. & Kibble, T. W. B. Global conservation laws and massless particles. *Phys. Rev. Lett.* **13**, 585 (1964).
7. Aad, G. et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B* **716**, 1 (2012).
8. Chatrchyan, S. et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B* **716**, 30 (2012).
9. Glashow, S. Partial-symmetries of weak interactions. *Nucl. Phys.* **22**, 579 (1961).
10. Salam, A. & Ward, J. C. Electromagnetic and weak interactions. *Phys. Lett.* **13**, 168 (1964).
11. Weinberg, S. A model of leptons. *Phys. Rev. Lett.* **19**, 1264 (1967).
12. Salam, G. P., Wang, L. T. & Zanderighi, G. The Higgs boson turns ten. *Nature* **607**, 41–47 (2022).
13. The ATLAS Collaboration. A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery. *Nature* **607**, 52–59 (2022).
14. The CMS Collaboration. A portrait of the Higgs boson by the CMS experiment ten years after the discovery. *Nature* **607**, 60–68 (2022).
15. Aaltonen, T. et al. High-precision measurement of the W boson mass with the CDF II detector. *Science* **376**, 170–176 (2022).
16. Workman, R. L. et al. Review of particle physics. *Prog. Theor. Exp. Phys.* **2022**, 083C01 (2022).
17. Awramik, M., Czakon, M., Freitas, A. & Weiglein, G. Precise prediction for the *W*-boson mass in the standard model. *Phys. Rev. D* **69**, 053006 (2004).
18. Erler, J. & Schott, M. Electroweak precision tests of the Standard Model after the discovery of the Higgs boson. *Prog. Part. Nucl. Phys.* **106**, 68 (2019).
19. Arnison, G. et al. Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV. *Phys. Lett. B* **122**, 103 (1983).
20. Banner, M. et al. Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN $\bar{p}p$ collider. *Phys. Lett. B* **122**, 476 (1983).
21. Arnison, G. et al. Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider. *Phys. Lett. B* **126**, 398 (1983).
22. Bagnaia, P. et al. Evidence for $Z^0 \to e^+e^-$ at the CERN $\bar{p}p$ collider. *Phys. Lett. B* **129**, 130 (1983).
23. Albajar, C. et al. Studies of intermediate vector boson production and decay in UA1 at the CERN proton–antiproton collider. *Z. Phys. C* **44**, 15 (1989).
24. Ansari, R. et al. Measurement of the Standard Model parameters from a study of *W* and *Z* bosons. *Phys. Lett. B* **186**, 440 (1987).
25. Alitti, J. et al. An improved determination of the ratio of *W* and *Z* masses at the CERN $\bar{p}p$ collider. *Phys. Lett. B* **276**, 354 (1992).
26. Abe, F. et al. Measurement of the *W*-boson mass. *Phys. Rev. Lett.* **65**, 2243 (1990).
27. Abe, F. et al. Measurement of the *W*-boson mass in 1.8-TeV $\bar{p}p$ collisions. *Phys. Rev. D* **43**, 2070 (1991).
28. Barate, R. et al. Measurement of the *W* mass and width in e^+e^- collisions at 189 GeV. *Eur. Phys. J. C* **17**, 241 (2000).
29. Abreu, P. et al. Measurement of the mass and width of the *W* boson in e^+e^- collisions at $\sqrt{s} = 189$ GeV. *Phys. Lett. B* **511**, 159 (2001).
30. Acciarri, M. et al. Measurement of mass and width of the *W* boson at LEP. *Phys. Lett. B* **454**, 386 (1999).
31. Abbiendi, G. et al. Measurement of the mass and width of the *W* boson in e^+e^- collisions at 189 GeV. *Phys. Lett. B* **507**, 29 (2001).
32. Schael, S. et al. Measurement of the *W* boson mass and width in e^+e^- collisions at LEP. *Eur. Phys. J. C* **47**, 309 (2006).
33. Abdallah, J. et al. Measurement of the mass and width of the *W* boson in e^+e^- collisions at $\sqrt{s} = 161 - 209$ GeV. *Eur. Phys. J. C* **55**, 1 (2008).

34. Achard, P. et al. Measurement of the mass and the width of the *W* boson at LEP. *Eur. Phys. J. C* **45**, 569 (2006).
35. Abbiendi, G. et al. Measurement of the mass and width of the *W* boson. *Eur. Phys. J. C* **45**, 307 (2006).
36. The ALEPH Collaboration, The DELPHI Collaboration, The L3 Collaboration, The OPAL Collaboration. The LEP Electroweak Working Group. Electroweak measurements in electron–positron collisions at *W*-boson-pair energies at LEP. *Phys. Rept.* **532**, 118 (2013).
37. Abe, F. et al. Measurement of the *W* boson mass. *Phys. Rev. Lett.* **75**, 11 (1995).
38. Abe, F. et al. Measurement of the *W* boson mass. *Phys. Rev. D* **52**, 4784 (1995).
39. Affolder, T. et al. Measurement of the *W* boson mass with the Collider Detector at Fermilab. *Phys. Rev. D* **64**, 052001 (2001).
40. Abachi, S. et al. Measurement of the *W* boson mass. *Phys. Rev. Lett.* **77**, 3309 (1996).
41. Abbott, B. et al. Determination of the mass of the *W* boson using the DØ detector at the Fermilab Tevatron. *Phys. Rev. D* **58**, 012002 (1998).
42. Abbott, B. et al. A measurement of the *W* boson mass at the Fermilab $p\bar{p}$ Collider. *Phys. Rev. Lett.* **80**, 3008 (1998).
43. Abbott, B. et al. Measurement of the *W* boson mass. *Phys. Rev. D* **58**, 092003 (1998).
44. Abbott, B. et al. Measurement of the *W* Boson Mass Using Electrons at Large Rapidities. *Phys. Rev. Lett.* **84**, 222 (2000).
45. Abbott, B. et al. Measurement of the *W* boson mass using large rapidity electrons. *Phys. Rev. D* **62**, 092006 (2000).
46. Abazov, V. M. et al. Improved *W* boson mass measurement with the DØ detector. *Phys. Rev. D* **66**, 012001 (2002).
47. Abazov, V. M. et al. Combination of CDF and D0 results on the *W* boson mass and width. *Phys. Rev. D* **70**, 092008 (2004).
48. Aaltonen, T. et al. First measurement of the *W*-boson mass in Run II of the Tevatron. *Phys. Rev. Lett.* **99**, 151801 (2007).
49. Aaltonen, T. et al. First Run II measurement of the *W* boson mass at the Fermilab Tevatron. *Phys. Rev. D* **77**, 112001 (2008).
50. Aaltonen, T. et al. Precise measurement of the *W*-boson mass with the CDF II Detector. *Phys. Rev. Lett.* **108**, 151803 (2012).
51. Aaltonen, T. et al. Precise measurement of the *W*-boson mass with the Collider Detector at Fermilab. *Phys. Rev. D* **89**, 072003 (2014).
52. Abazov, V. M. et al. Measurement of the *W* boson mass. *Phys. Rev. Lett.* **103**, 141801 (2009).
53. Abazov, V. M. et al. Measurement of the *W* boson mass with the D0 detector. *Phys. Rev. Lett.* **108**, 151804 (2012).
54. Abazov, V. M. et al. Measurement of the *W* boson mass with the D0 detector. *Phys. Rev. D* **89**, 012005 (2014).
55. Aaltonen, T. et al. Combination of CDF and D0 *W*-boson mass measurements. *Phys. Rev. D* **88**, 052018 (2013).
56. Aaboud, M. et al. Measurement of the *W*-boson mass in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector. *Eur. Phys. J. C* **78**, 110 (2018).
57. Aaboud, M. et al. Erratum to: measurement of the *W*-boson mass in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector. *Eur. Phys. J. C* **78**, 898 (2018).
58. Aaij, R. et al. Measurement of the *W* boson mass. *J. High Energ. Phys.* **2022**, 36 (2022).
59. Georgi, H. *Lie Algebras in Particle Physics: From Isospin to Unified Theories* 1st edn (CRC Press, 2000).
60. Greiner, W. & Müller, B. *Quantum Mechanics: Symmetries* 2nd edn (Springer Berlin, 1994).
61. Marciano, W. J. & Sirlin, A. Radiative corrections to neutrino-induced neutral-current phenomena in the $SU(2)_L \times U(1)$ theory. *Phys. Rev. D* **22**, 2695 (1980).
62. Marciano, W. J. & Sirlin, A. Erratum: testing the standard model by precise determinations of $W^{\pm}$ and *Z* masses. *Phys. Rev. D* **31**, 213 (1985).
63. Ross, D. A. & Veltman, M. J. G. Neutral currents and the Higgs mechanism. *Nucl. Phys. B* **95**, 135 (1975).
64. 't Hooft, G. & Veltman, M. Regularization and renormalization of gauge fields. *Nucl. Phys. B* **44**, 189–213 (1972).
65. Veltman, M. J. G. Limit on mass differences in the Weinberg model. *Nucl. Phys. B* **123**, 89–99 (1977).
66. Erler, J. & Langacker, P. Implications of high precision experiments and the CDF top quark candidates. *Phys. Rev. D* **52**, 441 (1995).
67. Abe, F. et al. Observation of top quark production in $\bar{p}p$ collisions with the Collider Detector at Fermilab. *Phys. Rev. Lett.* **74**, 2626 (1995).
68. Abachi, S. et al. Observation of the top quark. *Phys. Rev. Lett.* **74**, 2632 (1995).
69. Veltman, M. J. G. The screening theorem and the Higgs system. *Acta Phys. Polon. B* **25**, 1627–1636 (1994).
70. Gross, D. J. & Wilczek, F. Ultraviolet behavior of non-Abelian gauge theories. *Phys. Rev. Lett.* **30**, 1343–1346 (1973).
71. Politzer, H. D. Reliable perturbative results for strong interactions? *Phys. Rev. Lett.* **30**, 1346–1349 (1973).
72. Balázs, C. & Yuan, C.-P. Soft gluon effects on lepton pairs at hadron colliders. *Phys. Rev. D* **56**, 5558 (1997).
73. Ladinsky, G. A. & Yuan, C.-P. Nonperturbative regime in QCD resummation for gauge boson production at hadron colliders. *Phys. Rev. D* **50**, R4239 (1994).
74. Landry, F., Brock, R., Nadolsky, P. M. & Yuan, C.-P. Fermilab Tevatron Run-1 *Z* boson data and the Collins–Soper–Sterman resummation formalism. *Phys. Rev. D* **67**, 073016 (2003).
75. Bozzi, G., Catani, S., Ferrera, G., de Florian, D. & Grazzini, M. Transverse-momentum resummation: a perturbative study of *Z* production at the Tevatron. *Nucl. Phys. B* **815**, 174–197 (2009).
76. Bozzi, G., Catani, S., Ferrera, G., de Florian, D. & Grazzini, M. Production of Drell–Yan lepton pairs in hadron collisions: transverse-momentum resummation at next-to-next-to-leading logarithmic accuracy. *Phys. Lett. B* **696**, 207–213 (2011).
77. Affolder, T. et al. CDF Central Outer Tracker. *Nucl. Instrum. Methods Phys. Res. A* **526**, 249 (2004).
78. Kotwal, A. V., Gerberich, H. K. & Hays, C. Identification of cosmic rays using drift chamber hit timing. *Nucl. Instrum. Methods Phys. Res. A* **506**, 110 (2003).
79. Kotwal, A. V. & Hays, C. Drift chamber alignment using cosmic rays. *Nucl. Instrum. Methods Phys. Res. A* **762**, 85 (2014).
80. Kotwal, A. V. & Hays, C. Electromagnetic shower properties in a lead-scintillator sampling calorimeter. *Nucl. Instrum. Methods Phys. Res. A* **729**, 25 (2013).
81. Brun, R. et al. *GEANT: Detector Description and Simulation Tool*. Report No. W5013 (CERN Program Library, 1993).
82. Agostinelli, S. et al. geant4 — a simulation toolkit. *Nucl. Instrum. Methods Phys. Res. A* **506**, 250 (2003).
83. ATLAS Collaboration. *Improved W boson Mass Measurement using 7 TeV Proton–Proton Collisions with the ATLAS Detector*. Report No. ATLAS-CONF-2023-004 (CERN Document Server, 2023).
84. Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference* 2nd edn (Springer, 2002).
85. Dyson, F. A meeting with Enrico Fermi. *Nature* **427**, 297 (2004).
86. Azzurri, P. The *W* mass and width measurement challenge at FCC-ee. *Eur. Phys. J. Plus* **136**, 1203 (2021).
87. The LEP Collaborations: ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, the LEP Electroweak Working Group, the SLD Heavy Flavour & Electroweak Working Group. A combination of preliminary electroweak measurements and constraints on the Standard Model. Preprint at https://doi.org/10.48550/arXiv.hep-ex/0112021 (2002).
88. The LEP Collaborations: ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, the LEP Electroweak Working Group, the SLD Electroweak & Heavy Flavour Groups. A combination of preliminary electroweak measurements and constraints on the Standard Model. Preprint at https://doi.org/10.48550/arXiv.hep-ex/0312023 (2004).
89. Aryshev, A. et al. The International Linear Collider: report to Snowmass 2021. Preprint at https://doi.org/10.48550/arXiv.2203.07622 (2022).
90. Cheng, H. et al. The physics potential of the CEPC. Prepared for the US Snowmass community planning exercise (Snowmass 2021). Preprint at https://doi.org/10.48550/arXiv.hep-ex/0112021 (2022).
91. de Blas, J. et al. The CLIC potential for new physics. Preprint at https://doi.org/10.48550/arXiv.hep-ex/0312023 (2019).
92. Blondel, A. & Janot, P. FCC-ee overview: new opportunities create new challenges. *Eur. Phys. J. Plus* **137**, 92 (2022).
93. Sirlin, A. Radiative corrections in the $SU(2)_L \times U(1)$ theory: a simple renormalization framework. *Phys. Rev. D* **22**, 971 (1980).
94. The ALEPH Collaboration, the DELPHI Collaboration, the L3 Collaboration, the OPAL Collaboration, the SLD Collaboration, the LEP Electroweak Working Group & The SLD Electroweak and Heavy Flavour Groups. Precision electroweak measurements on the *Z* resonance. *Phys. Rept.* **427**, 257–454 (2006).
95. Benesch, J. et al. The MOLLER experiment: an ultra-precise measurement of the weak mixing angle using Møller scattering. Preprint at https://doi.org/10.48550/arXiv.1411.4088 (2014).
96. Arrington, J. et al. The Solenoidal Large Intensity Device (SoLID) for JLab 12 GeV. Preprint at https://doi.org/10.48550/arXiv.2209.13357 (2023).
97. Becker, D. et al. The P2 experiment. *Eur. Phys. J. A* **54**, 208 (2018).
98. Amoroso, S. et al. Compatibility and combination of world *W*-boson mass measurements. Preprint at https://doi.org/10.48550/arXiv.1411.4088 (2023).
99. Biekotter, T., Heinemeyer, S. & Weiglein, G. Excesses in the low-mass Higgs-boson search and the *W*-boson mass measurement. *Eur. Phys. J. C* **83**, 450 (2023).
100. Aguillard, D. P. et al. Measurement of the positive muon anomalous magnetic moment to 0.20 ppm. *Phys. Rev. Lett.* **131**, 161802 (2023).
101. Abi, B. et al. Measurement of the positive muon anomalous magnetic moment to 0.46 ppm. *Phys. Rev. Lett.* **126**, 141801 (2021).
102. Albahri, T. et al. Measurement of the anomalous precession frequency of the muon in the Fermilab Muon $g-2$ Experiment. *Phys. Rev. D* **103**, 072002 (2021).
103. Bennett, G. W. et al. Final report of the E821 muon anomalous magnetic moment measurement at BNL. *Phys. Rev. D* **73**, 072003 (2006).
104. Castelvecchi, D. Dreams of new physics fade with latest muon magnetism result. *Nature* **620**, 473–474 (2023).
105. Borsanyi, S. et al. Leading hadronic contribution to the muon magnetic moment from lattice QCD. *Nature* **593**, 51–55 (2021).
106. Lehner, C. High-precision lattice QCD calculations of the muon anomalous magnetic moment. *Nat. Rev. Phys.* **4**, 14–15 (2022).
107. Ignatov, F. V. et al. Measurement of the $e^+e^- \rightarrow \pi^+\pi^-$ cross section from threshold to 1.2 GeV with the CMD-3 detector. Preprint at https://doi.org/10.48550/arXiv.2302.08834 (2023).
108. Athron, P. et al. Hadronic uncertainties versus new physics for the *W* boson mass and muon $g-2$ anomalies. *Nat. Commun.* **14**, 659 (2023).
109. 't Hooft, G. Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking. *NATO Sci. Ser. B* **59**, 135–157 (1980).

Acknowledgements
The author thanks K. Agashe, S. Gupta and R. Plesser for helpful discussions.

Author contributions
The author contributed to all aspects of the article.

Competing interests
The author is also the corresponding author of 'Aaltonen, T. et al. (CDF Collaboration), *Science* 376, 170–176 (2022)', a member of the CDF Collaboration and the ATLAS Collaboration and a past member of the D0 Collaboration and the E665 Collaboration.

Additional information
Peer review information *Nature Reviews Physics* thanks Claudia-Elisabeth Wulz, Jens Erler and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.