The Standard Model as a Certified Attribute Grammar

Feynman rules as compiler phases on an $[8, 4, 4]$ record alphabet

David Elliman
Neuro-Symbolic Ltd
`dave@neusym.ai`
https://neusym.ai

3 July 2026

Abstract

We formalise the interaction vertices of the Standard Model (SM) as a typed rewriting grammar over a sixteen-letter matter alphabet, and we certify the construction by machine enumeration. The alphabet is the codeword set of the extended Hamming / Reed–Muller code $[8, 4, 4] = \mathrm{RM}(1, 3)$, whose sixteen words we identify with the sixteen Weyl fermions of one Standard Model generation (equivalently, one **16** of SO(10)). Interactions are typed productions $w \to w' +$ response: two *read* channels (γ, Z), one chirality-typed *flip* ($W^{\pm}$), one colour rotation (g), one left–right *bridge* (Yukawa), a $\Delta L = 2$ Majorana portal, and the gauge/scalar self-productions. Conservation laws act as a static type system; parity violation enters as a typing rule rather than a dynamical add-on. Two certificates are supplied by an accompanying exact-arithmetic program: (i) the grammar generates the SM tree-level vertex set exactly—no missing and no spurious classes—with "sterility" of the right-handed neutrino emerging as a derived typing fact (the unique letter read by neither neutral channel); and (ii) the alphabet is anomaly-free—the $[\mathrm{U}(1)^3]$, $[\mathrm{grav}^2\,\mathrm{U}(1)]$, $[\mathrm{SU}(2)^2\,\mathrm{U}(1)]$ and $[\mathrm{SU}(3)^2\,\mathrm{U}(1)]$ sums vanish identically over the sixteen letters. The claim this certifies is therefore not that the labels resemble the Standard Model, but that *the record alphabet is capable of supporting a consistent chiral gauge theory, and the Standard Model sits naturally inside that alphabet.* We then locate the grammar in the formal-language hierarchy: the tree-level planar fragment is context-free (planar Wick contractions are the Dyck language); multiple simultaneous conservation laws are multicounter constraints; non-planar contractions realise crossed serial dependencies, the same structure by which natural language exceeds context-freeness; and loop diagrams require graph rewriting—the tree/loop boundary of quantum field theory coincides with the tree/graph boundary of rewriting theory. As a corollary, 't Hooft's planar large-N limit is precisely the context-free fragment of the grammar, so the $1/N$ expansion is an expansion in grammatical complexity. We list the structural falsifiers the grammar stakes (no right-handed charged currents; no fourth sequential generation; no tree-level $\gamma\gamma\gamma$; a Majorana-only route to leptonic CP violation) and point to companion preprints where the underlying substrate programme registers quantitative predictions.

1 Introduction

Feynman's diagrams [20] are usually presented as bookkeeping for perturbation theory. Viewed from computer science they are something more specific: *derivations in a rewriting system.* Particles are terminal symbols, vertices are productions, a diagram is a parse structure, and the S-matrix element is the semantics attached to the parse. This paper takes the compiler analogy literally and asks the two questions a language designer would ask of any proposed grammar:

1. **Completeness/soundness.** Does the grammar generate exactly the intended language—here, the tree-level vertex set of the Standard Model—with no missing and no spurious productions? Is the type system sound?
2. **Classification.** Where does the grammar sit in the Chomsky hierarchy [8, 9], and what do the classification boundaries correspond to physically?

The compiler structure we defend has four layers:

Primitive grammar	the productions of Table 1: two reads (γ, Z), the left-typed W rewrite, colour rotation, the Yukawa bridge, the Majorana portal
Type rules	charge, colour, chirality, lepton/baryon number, generation validity
Soundness certificate	anomaly cancellation (Thm. 2) — a global property of the rule set, not a per-derivation check
Composite parse trees	loop-induced processes: $H \to \gamma\gamma$, $gg \to H$, light-by-light
Compiler output	a *response action* whose perturbative expansion is the Feynman series; running the program is perturbation theory

Both questions turn out to have sharp answers, and the second is, to our knowledge, new in the form given here. The completeness question is settled by exhaustive enumeration over a finite alphabet (Sec. 4), with anomaly cancellation [1, 6, 7] playing the role of a type-system soundness certificate. The classification question (Sec. 6) yields a dictionary: context-freeness corresponds to tree-level planarity; the failure modes of context-freeness correspond, one by one, to named physical features (multiple conservation laws, non-planarity, loops); and 't Hooft's planar limit [34] is identified as the context-free fragment.

The alphabet we use is not arbitrary. It is the codeword set of the extended Hamming code $[8, 4, 4]$, equivalently the first-order Reed–Muller code $\mathrm{RM}(1, 3)$ [28], whose sixteen codewords we place in bijection with the sixteen left-handed Weyl fermions of one SM generation—one spinor **16** of SO(10) [4, 21, 23]. This identification arises in a broader research programme in which the code cell is a physical substrate and particle species are its stabilised records [13, 16, 17]; that programme registers quantitative, pre-timestamped predictions elsewhere [14, 15, 18]. The present paper is deliberately self-contained: every result below is a statement about the SM vertex algebra and about formal languages, and can be checked without adopting the substrate interpretation. Readers who wish to treat the $[8, 4, 4]$ alphabet as merely a convenient labelling lose none of Secs. 3–6.

Relation to prior art. The idea that physical processes are morphisms composed diagrammatically is classical in category-theoretic physics: Penrose's tensor diagrams [29], the "Rosetta stone" of Baez and Stay [5], and the ZX-calculus of Coecke and Duncan [12], which gives a *complete* graph-rewriting calculus for qubit quantum mechanics. Our contribution is complementary and more specific: a *typed string grammar* for the SM's tree-level matter–gauge–scalar vertex algebra, certified complete by enumeration, together with an explicit placement of the full diagram calculus in the formal-language hierarchy. The type-system framing is Knuth's: conservation laws enter exactly as the attributes of an attribute grammar [26], which is also how production compilers implement the non-context-free parts of real programming languages.

2 The lexer: a sixteen-letter matter alphabet

Definition 1 (Matter alphabet)**.** *The matter alphabet* Σ_m *is the set of sixteen codewords of* $\mathrm{RM}(1, 3)$*, identified with the Weyl fermions of one SM generation in the left-handed basis:*

$$\begin{aligned} Q = (u, d)_L\ (\mathbf{3}, \mathbf{2})_{1/3}, \quad & u^c\ (\bar{\mathbf{3}}, \mathbf{1})_{-4/3}, \quad & d^c\ (\bar{\mathbf{3}}, \mathbf{1})_{2/3}, \\ L = (\nu, e)_L\ (\mathbf{1}, \mathbf{2})_{-1}, \quad & e^c\ (\mathbf{1}, \mathbf{1})_{2}, \quad & \nu^c\ (\mathbf{1}, \mathbf{1})_{0}, \end{aligned}$$

Class	Production	Typing condition
EM read	$w \to w + \gamma$	$Q(w) \neq 0$; w unchanged (read, not rewrite)
Weak-NC read	$w \to w + Z$	$g_Z(w) = T_3(w) - Q(w)\sin^2\theta_W \neq 0$; w unchanged
CC flip	$w \to \sigma_{\text{doublet}}(w) + W^{\pm}$	**left-handed only**; within-doublet partner; colour preserved; flavour change enters only through the CKM/PMNS dressing of this one production
Colour rotation	$q_c \to q_{c'} + g$	coloured letters only; $c \neq c'$; flavour, chirality unchanged
Yukawa bridge	$w_L \leftrightarrow w_R + H$	same flavour, charge, colour; chirality toggled (includes the Dirac neutrino bridge $\nu_L \leftrightarrow \nu_R$)
Majorana portal	$\nu^c\,\nu^c \to \varnothing$ (+ conj.)	$\Delta L = 2$; gauge-singlet letter only; distinct from the $\Delta L = 0$ bridge
Scalar–gauge	$H \to W^+W^-$, $H \to ZZ$ (+ seagulls)	the mass-giving productions
Scalar self	$H \to HH$, $HH \to HH$	the quartic of the scalar potential
Gauge self	W^+W^-Z, $W^+W^-\gamma$, ggg, $gggg$, $WWWW$, $WWZZ$, ...	non-abelian channels only

Untypable at tree level: $\gamma\gamma\gamma$, $\gamma\gamma\gamma\gamma$; lepton → coloured lepton; $w_R \to \cdot + W$; fourth-generation letters.

Table 1: The production rules. Loop-induced effective vertices ($H\gamma\gamma$, ggH, light-by-light) are deliberately absent: they are composite parse structures, not primitives (Sec. 6).

with hypercharge normalised by $Q = T_3 + Y/2$*. Counting colour and weak components:* $6 + 3 + 3 + 2 + 1 + 1 = 16$*.*

That these sixteen states fill one spinor of SO(10) is the classic grand-unification observation [21, 23]; the pedagogical mapping is given in Baez and Huerta [4]. The substrate programme adds a reason for the *code*: $[8, 4, 4]$ is forced as the unique self-dual record floor by an enumeration over check families [13], and the count $\binom{16+1}{2} = 136$, $136 + 1 = 137$ underlies its bare-coupling identification [17]. None of that is needed below.

Definition 2 (Response alphabet)**.** *The response alphabet is* $\Sigma_r = \{\gamma,\ Z,\ W^{\pm},\ g^{a=1..8},\ H\}$, *carrying the attribute triple (electric charge, colour label, channel kind), with kinds* $\{$*read*, *flip*, *colour*, *bridge*$\}$.

3 The grammar and its type system

Definition 3 (Typed production)**.** *A* legal vertex *is a production*

$$w \;\longrightarrow\; w' \;+\; r, \qquad w, w' \in \Sigma_m,\ r \in \Sigma_r,$$

admitted by the type system below, together with the gauge/scalar self-productions of Table 1. A tree-level Feynman diagram is a derivation; crossing symmetry corresponds to reading productions in any causal orientation.

Definition 4 (Type system)**.** *Attributes: electric charge Q, colour label with triality, weak isospin T_3, chirality $\chi \in \{L, R\}$, lepton and baryon number, generation index $G \in \{1, 2, 3\}$. Predicates: (i) charge conservation at every production; (ii) colour closure (triality 0 for any asymptotic sentence); (iii)* ***chirality typing****: the* flip *kind acts only on $\chi = L$ letters; (iv) L*

and B conserved by all productions except the Majorana portal ($\Delta L = 2$); (v) G is a lexer-level index: three copies of Σ_m, with the algebraic fourth corner excluded from the alphabet, and no *production changes G except the CKM/PMNS-dressed flip.*

Three design points deserve emphasis. First, electromagnetism and the weak neutral current are *reads*: the letter is not rewritten; its attributes are read by a response channel. The two neutral reads differ only in their weight functions (Q versus g_Z). Second, parity violation is not dynamical input but a *typing rule*: the flip kind is defined on the left-handed sub-alphabet, so a right-handed charged current is not suppressed—it is untypable. Third, the two neutrino mass operations are grammatically distinct: the Dirac bridge is the same Yukawa production every flavour has, while the $\Delta L = 2$ portal is a production available *only* to the gauge-singlet letter; their composition is the seesaw.

4 Certificate I: the grammar compiles exactly the SM

All claims in this section are verified by an accompanying exact-arithmetic program[1] that enumerates all $16 \times 16 \times |\Sigma_r|$ candidate productions against the type system.

Theorem 1 (Tree-level completeness)**.** *The typed grammar of Sec. 3 generates the Standard Model tree-level vertex set exactly: every SM tree vertex class is produced (both neutral reads; the left-only charged-current flip with precisely the within-doublet pairs; colour rotation for all and only coloured letters; Yukawa bridges for all four flavours including the Dirac neutrino bridge; the $\Delta L = 2$ portal; HVV, HHH; non-abelian gauge self-couplings), and no spurious class survives (no γ self-coupling, no right-handed flip, no lepton–gluon production, no fourth-generation letter).*

Observation 1 (Sterility is a derived typing fact)**.** *The enumeration yields exactly one letter invisible to both neutral reads: ν_R, the unique w with $Q(w) = 0$ and $g_Z(w) = 0$. Its complete grammatical role is therefore the Dirac bridge and the $\Delta L = 2$ portal. "Sterile neutrino" is not a label added to the theory; it is the type system's unique blind spot, and the seesaw is the only sentence structure available to it.*

Observation 1 is the grammar's most consequential output: it *forces* the neutrino-mass sector to factor through the portal, which is what elevates the Majorana route to leptonic CP violation (Sec. 7).

5 Certificate II: the type system is sound (anomaly freedom)

A per-production type check cannot see the deepest consistency condition of a chiral gauge theory: cancellation of triangle anomalies, a *global* sum over the alphabet [1, 6, 7]. In compiler terms this is a soundness proof for the type system: it certifies that the typing rules, applied to this particular alphabet, define a consistent theory rather than a lookalike.

Theorem 2 (Anomaly certificate)**.** *Over the sixteen letters, in exact rational arithmetic,*

$$\sum \mathrm{U}(1)^3 = 0, \qquad \sum \mathrm{grav}^2\,\mathrm{U}(1) = 0, \qquad \sum \mathrm{SU}(2)^2\,\mathrm{U}(1) = 0, \qquad \sum \mathrm{SU}(3)^2\,\mathrm{U}(1) = 0.$$

Anomaly cancellation is of course a known property of one SM generation [7]; the point here is its *role*: it is the certificate that distinguishes the $[8, 4, 4]$ letter assignment from an arbitrary sixteen-species chiral set, for which these sums generically fail. Indeed the constraints are strong enough to nearly fix the hypercharges outright [22]. The alphabet passes with no adjustable freedom, and—in the substrate reading—had no freedom to adjust.

---

[1] `quantum_grammar_compiler_certification.py`, in the public archive [19]; twelve independent assertions, exit status tied to their conjunction.

6 Where the grammar sits in the Chomsky hierarchy

We now classify. Throughout, "diagram" means the derivation structure itself, not a linearisation of it.

Proposition 1 (The context-free skeleton)**.** *The production set of Table 1 is a finite set of single-nonterminal rewrites; tree-level diagrams are derivation trees; and* planar *Wick contraction structure is the Dyck language of balanced brackets, the archetypal context-free language [10]. Hence the tree-level planar fragment of the vertex calculus is context-free.*

Proposition 2 (Multiple charges are multicounter)**.** *A single additive conservation law is a one-counter constraint and stays within context-free power. The* simultaneous *conservation of several independent charges (Q, colour, L, B) is an intersection of context-free constraints; context-free languages are not closed under intersection [24], and the resulting class is that of multicounter / vector-addition systems. The type system of Sec. 3 implements exactly this layer—which is why it is an* attribute *grammar [26] rather than a context-free one, precisely as in production compilers (the syntax of C is context-free; "declared before use" is not).*

Proposition 3 (Non-planarity is mild context-sensitivity)**.** *A non-planar contraction contains two matched pairs in crossing configuration $a\,b\,a'\,b'$: a crossed serial dependency. Crossed serial dependencies are the classical mechanism by which natural language was shown to exceed context-freeness [32] and are the defining capability of the mildly context-sensitive formalisms (TAG, MCFG) [25, 31]. The non-planar sector of the vertex calculus therefore sits in the mildly context-sensitive class: quantum field theory and Swiss German fail context-freeness by the same structural mechanism.*

Proposition 4 (Loops require graph rewriting)**.** *A derivation in any string grammar is a tree. A diagram containing a loop is a graph with a cycle and is therefore not the derivation structure of any string grammar; the appropriate formalism is typed graph rewriting [30], equivalently composition in a compact/traced monoidal category [5, 12]. Thus the tree/loop boundary of quantum field theory coincides with the tree/graph boundary of rewriting theory: renormalisation lives exactly where string grammars end.*

Proposition 5 (The planar large-N limit is the context-free fragment)**.** *In 't Hooft's double-line representation the $1/N$ expansion grades diagrams by genus [34]. Genus-zero (planar) gluon contractions are non-crossing perfect matchings, in canonical bijection with Dyck words (Catalan structures) [27, 33]; each unit of genus inserts crossings, i.e. crossed serial dependencies (Prop. 3). Hence the $N \to \infty$ limit retains exactly the context-free fragment of the grammar, and the $1/N$ expansion is an expansion in grammatical complexity.*

We flag a pun to prevent its misuse: Prop. 5 says QCD is "asymptotically context-free at large N"—a statement about grammar, logically unrelated to asymptotic freedom of the running coupling.

The dictionary assembled by Props. 1–5:

Physical feature	Formal-language cost
finite vertex set; tree diagrams	context-free (derivation trees)
planar contraction structure	Dyck / Catalan (CF)
one additive charge	one-counter ($\subset$ CF)
several independent charges	multicounter / vector addition (not CF)
non-planar contractions	crossed dependencies: mildly context-sensitive
loops	graph rewriting (beyond string grammars)
confinement (singlet closure)	global semantic predicate (attribute level)

7 Insights and falsifiable consequences

The grammar is tree-level bookkeeping; it does not by itself compute amplitudes. Its value is structural, and structure can be falsified. We separate (a) consistency re-derivations, (b) structural exclusions the grammar stakes, and (c) quantitative predictions registered in the companion substrate programme.

(a) Consistency re-derivations

The neutrino couples to Z and not to γ (the 1973 neutral-current signature) as a two-line typing consequence; anomaly freedom holds identically over the alphabet (Thm. 2); the number of light active neutrino species is the number of $g_Z \neq 0$ neutral letters per generation, one, consistent with the LEP invisible width $N_\nu = 2.9840 \pm 0.0082$ [2].

(b) Structural exclusions (kill conditions for the grammar)

1. **No right-handed charged currents.** The flip kind is typed on $\chi = L$. Any confirmed W_R or right-handed charged-current contamination falsifies the typing, not a parameter choice.

2. **No fourth sequential generation.** G is a two-bit lexer index with the fourth corner excluded from the alphabet; a fourth generation at any mass falsifies the lexer.

3. **No tree-level $\gamma\gamma\gamma$ (or 4γ).** The abelian read channel carries no self-record; observation of a tree-strength photon self-coupling falsifies the read/kind split.

4. **The ν_R is portal-only.** By Obs. 1 the only available neutrino-mass sentences factor through the Dirac bridge and the $\Delta L = 2$ portal. In the companion programme this forces leptonic CP violation into the Majorana sector with a Dirac-CP null: a high-significance nonzero long-baseline Dirac phase ($J_\ell \neq 0$) falsifies that branch [15].

(c) Registered quantitative predictions (companion preprints)

The substrate programme that motivates the $[8, 4, 4]$ alphabet registers pre-timestamped, kill-ruled predictions that are out of scope here but inherit the alphabet: Newton's constant from the proton mass ($G = 6.674311 \times 10^{-11}\,\mathrm{m}^3\,\mathrm{kg}^{-1}\,\mathrm{s}^{-2}$, kill at 3σ once $\sigma_G \leq 15\,\mathrm{ppm}$) [14]; a primordial tensor null ($r_{\mathrm{linear}} = 0$, kill at $r \geq 10^{-3}$) [18]; and the Majorana-only leptonic CP programme ($m_{\beta\beta} \approx 1.8$–$3.1\,\mathrm{meV}$, $\Sigma m_\nu \approx 60\,\mathrm{meV}$, normal ordering; killed by inverted ordering, by $m_{\beta\beta} \gtrsim 10\,\mathrm{meV}$, or by a Dirac phase) [15]. The grammar adds to these a cheap, purely structural test surface: items (b)(1)–(b)(3) require no new theory runs, only experiments that would also surprise the SM—the point being that the *grammar* stakes them as typing facts rather than as accidents of parameter choice.

(d) Two structural insights

First, the tree/loop boundary of field theory is the tree/graph boundary of rewriting (Prop. 4); the technology of renormalisation is, in this reading, the price of leaving string grammars. Second, the $1/N$ expansion is an expansion in grammatical complexity (Prop. 5); the long-observed simplicity of planar theories acquires a formal-language explanation: planar QCD is the largest fragment of the theory that a pushdown automaton can parse.

(e) A diagnostic that existing tools do not provide

Existing pipelines of the FeynRules/MadGraph class [3, 11] *verify* legality against a model supplied to them; they neither explain why the allowed vertex set has the shape it does, nor classify the *absences*. The grammar does both, and the second yields a decision procedure for beyond-Standard-Model searches. Every process not yet observed falls into exactly one of two classes:

- **Absent with a parse.** A composite derivation exists in the grammar; the process is suppressed (loop order, mixing, phase space) but reachable. More energy, luminosity or sensitivity is the rational response.
- **Absent without a parse.** The process is untypable: no derivation exists at any order. No amount of energy helps; observing it requires *new grammar*—either a new production class, or an extension of the alphabet. And alphabet extensions are priced by the soundness certificate: any proposed new letter must arrive in an anomaly-complete set (Thm. 2 applied to the extended alphabet), or the type system itself fails.

Proposed process	Parse status	Class
$0\nu\beta\beta$	$\Delta L = 2$ portal parse exists	sensitivity
$H \to \gamma\gamma$, $gg \to H$	composite (loop) parse exists	luminosity
tree-level FCNC ($t \to cZ$)	no tree parse; loop parse exists	luminosity (GIM-suppressed)
proton decay	no parse: no ΔB production	**new grammar** (a ΔB portal)
right-handed currents (W_R)	untypable: chirality typing	**new grammar** (retype the flip)
fourth generation	no letter	**new alphabet** (anomaly-complete row)
millicharged fermion	no letter; breaks the anomaly ledger	**new alphabet** (must complete)

The operational content of "compiler" is exactly this distinction: a type error is categorically different from a long runtime. The contrast in the first and fourth rows is instructive—neutrinoless double-beta decay and proton decay are often grouped as "rare searches," but the grammar separates them absolutely: $0\nu\beta\beta$ has a parse and is a sensitivity question; proton decay has none and is a constitutional question.

8 Discussion

Why an attribute grammar and not a category? The monoidal / string-diagram formalisation of processes is well developed [5, 12, 29] and is the right home for the *loop-level* calculus (Prop. 4). The attribute-grammar presentation is chosen for the tree level because it makes two things visible that the categorical packaging hides: the compiler phase structure (context-free syntax versus context-sensitive typing—Prop. 2), and the classification boundaries of Sec. 6, which are statements about *grammars* and disappear once everything is absorbed into morphism composition.

Limitations. (i) The grammar is tree-level; loop-induced effective vertices ($H\gamma\gamma$, ggH, light-by-light) are composite parse structures by design. (ii) Coupling magnitudes, mixing-matrix entries and masses are semantics, not syntax: the grammar fixes which sentences exist, not their probabilities. The division of labour is exactly the compiler split:

the grammar says what can be written;
the response layer says what it measures as.

Couplings, masses, CKM/PMNS entries, loop amplitudes and precision cross-sections all live on the response side. (iii) Confinement enters as a global predicate (asymptotic sentences are colour singlets), not as a production rule. (iv) Gravity is absent from the alphabet; in the substrate programme it is an emergent readout rather than a channel letter, and its predictions are registered separately [14].

Relation to the substrate programme. For this paper the $[8, 4, 4]$ identification may be read as a labelling. In the companion programme it is load-bearing: the code is forced as the unique self-dual record floor; the response/record split that makes "read" versus "rewrite" a physical distinction is proven at operator grade; and the alphabet's Sym^2 count underlies the bare coupling [13, 16, 17]. The grammar then stops being an analogy: the compiler is the physics.

9 Conclusion

The Standard Model's tree-level interaction structure is a certified attribute grammar over a sixteen-letter code alphabet: sixteen codewords, nine production classes, five attribute families, two machine-checked certificates (exact vertex completeness; exact anomaly cancellation), four structural kill conditions, and a clean placement in the formal-language hierarchy in which planarity is context-freeness, non-planarity is mild context-sensitivity, and loops are graphs. A Feynman diagram is a parse tree; conservation laws are a type system; anomaly cancellation is its soundness proof; and the $1/N$ expansion measures how far a sentence strays from what a pushdown automaton can parse. The grammar says what can be written; the response layer says what it measures as.

Reproducibility. The certification program (twelve assertions, exact rational arithmetic, single file, no dependencies beyond NumPy) is available in the public archive [19].

References

[1] Stephen L. Adler. Axial-vector vertex in spinor electrodynamics. *Physical Review*, 177: 2426–2438, 1969. doi: 10.1103/PhysRev.177.2426.

[2] ALEPH, DELPHI, L3, OPAL, SLD Collaborations, LEP Electroweak Working Group, SLD Electroweak and Heavy Flavour Groups. Precision electroweak measurements on the Z resonance. *Physics Reports*, 427(5–6):257–454, 2006. doi: 10.1016/j.physrep.2005.12.006.

[3] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7):79, 2014. doi: 10.1007/JHEP07(2014)079.

[4] John C. Baez and John Huerta. The algebra of grand unified theories. *Bulletin of the American Mathematical Society*, 47(3):483–552, 2010. doi: 10.1090/S0273-0979-10-01294-2.

[5] John C. Baez and Mike Stay. Physics, topology, logic and computation: A Rosetta Stone. In Bob Coecke, editor, *New Structures for Physics*, volume 813 of *Lecture Notes in Physics*, pages 95–172. Springer, 2010. doi: 10.1007/978-3-642-12821-9_2.

[6] J. S. Bell and R. Jackiw. A PCAC puzzle: $\pi^0 \to \gamma\gamma$ in the σ-model. *Il Nuovo Cimento A*, 60:47–61, 1969. doi: 10.1007/BF02823296.

[7] C. Bouchiat, J. Iliopoulos, and Ph. Meyer. An anomaly-free version of Weinberg's model. *Physics Letters B*, 38(7):519–523, 1972. doi: 10.1016/0370-2693(72)90532-1.

[8] Noam Chomsky. Three models for the description of language. *IRE Transactions on Information Theory*, 2(3):113–124, 1956. doi: 10.1109/TIT.1956.1056813.

[9] Noam Chomsky. On certain formal properties of grammars. *Information and Control*, 2(2): 137–167, 1959. doi: 10.1016/S0019-9958(59)90362-6.

[10] Noam Chomsky and Marcel-Paul Schützenberger. The algebraic theory of context-free languages. In P. Braffort and D. Hirschberg, editors, *Computer Programming and Formal Systems*, pages 118–161. North-Holland, Amsterdam, 1963. doi: 10.1016/S0049-237X(08)72023-8.

[11] Neil D. Christensen and Claude Duhr. FeynRules – Feynman rules made easy. *Computer Physics Communications*, 180(9):1614–1641, 2009. doi: 10.1016/j.cpc.2009.02.018.

[12] Bob Coecke and Ross Duncan. Interacting quantum observables: categorical algebra and diagrammatics. *New Journal of Physics*, 13:043016, 2011. doi: 10.1088/1367-2630/13/4/043016.

[13] David Elliman. Foundations and methodology for a finite-QEC substrate: Code, crystallisation, ledgers, and audit protocol. `https://doi.org/10.5281/zenodo.21111257`, 2026.

[14] David Elliman. Newton's constant from the proton mass: A locked zero-parameter prediction. `https://doi.org/10.5281/zenodo.21140769`, 2026.

[15] David Elliman. Leptonic CP is Majorana-only: A registered prediction addendum. `https://doi.org/10.5281/zenodo.21100148`, 2026.

[16] David Elliman. From counts to observables: A measurement discipline for discrete record-based physics. `https://doi.org/10.5281/zenodo.21132677`, 2026.

[17] David Elliman. A finite-QEC substrate program for particle physics and cosmology: Current canon and audit methodology. `https://doi.org/10.5281/zenodo.21111287`, 2026.

[18] David Elliman. A primordial-tensor null from boundary printing: Pre-registering $r_{\text{linear}} = 0$ and a scalar-induced floor. `https://doi.org/10.5281/zenodo.21062225`, 2026.

[19] David Elliman. It-from-bit public executable scripts. `https://github.com/dgedge/itfrombit`, 2026. Public script archive; certification program `quantum_grammar_compiler_certification.py`.

[20] R. P. Feynman. Space-time approach to quantum electrodynamics. *Physical Review*, 76: 769–789, 1949. doi: 10.1103/PhysRev.76.769.

[21] Harald Fritzsch and Peter Minkowski. Unified interactions of leptons and hadrons. *Annals of Physics*, 93(1–2):193–266, 1975. doi: 10.1016/0003-4916(75)90211-0.

[22] C. Q. Geng and R. E. Marshak. Uniqueness of quark and lepton representations in the standard model from the anomalies viewpoint. *Physical Review D*, 39:693–696, 1989. doi: 10.1103/PhysRevD.39.693.

[23] Howard Georgi. The state of the art—gauge theories. In C. E. Carlson, editor, *Particles and Fields—1974*, volume 23 of *AIP Conference Proceedings*, pages 575–582. American Institute of Physics, 1975. doi: 10.1063/1.2947450.

[24] John E. Hopcroft and Jeffrey D. Ullman. *Introduction to Automata Theory, Languages, and Computation.* Addison-Wesley, Reading, MA, 1979.

[25] Aravind K. Joshi. Tree adjoining grammars: How much context-sensitivity is required to provide reasonable structural descriptions? In D. R. Dowty, L. Karttunen, and A. M. Zwicky, editors, *Natural Language Parsing*, pages 206–250. Cambridge University Press, 1985. doi: 10.1017/CBO9780511597855.007.

[26] Donald E. Knuth. Semantics of context-free languages. *Mathematical Systems Theory*, 2(2): 127–145, 1968. doi: 10.1007/BF01692511. Correction: ibid. 5 (1971) 95–96.

[27] Sergei K. Lando and Alexander K. Zvonkin. *Graphs on Surfaces and Their Applications*, volume 141 of *Encyclopaedia of Mathematical Sciences.* Springer, Berlin, 2004. doi: 10.1007/978-3-540-38361-1.

[28] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes.* North-Holland, Amsterdam, 1977.

[29] Roger Penrose. Applications of negative dimensional tensors. In D. J. A. Welsh, editor, *Combinatorial Mathematics and its Applications*, pages 221–244. Academic Press, London, 1971.

[30] Grzegorz Rozenberg, editor. *Handbook of Graph Grammars and Computing by Graph Transformation, Vol. 1: Foundations.* World Scientific, Singapore, 1997. doi: 10.1142/3303.

[31] Hiroyuki Seki, Takashi Matsumura, Mamoru Fujii, and Tadao Kasami. On multiple context-free grammars. *Theoretical Computer Science*, 88(2):191–229, 1991. doi: 10.1016/0304-3975(91)90374-B.

[32] Stuart M. Shieber. Evidence against the context-freeness of natural language. *Linguistics and Philosophy*, 8:333–343, 1985. doi: 10.1007/BF00630917.

[33] Richard P. Stanley. *Catalan Numbers.* Cambridge University Press, 2015. doi: 10.1017/CBO9781139871495.

[34] Gerard 't Hooft. A planar diagram theory for strong interactions. *Nuclear Physics B*, 72(3): 461–473, 1974. doi: 10.1016/0550-3213(74)90154-0.

A The certification checks

The accompanying program asserts: (G1a) both neutral read channels exist; (G1b) ν is Z-active and γ-blind; (G1c) ν_R is the unique doubly-blind letter; (G2) gluon couples to all and only coloured letters; (G3) all charged-current vertices are left-handed and exactly the doublet pairs; (G4) the scalar sector comprises the four Yukawa bridges, HVV, $HHH/HHHH$, and the distinct $\Delta L = 2$ portal; (G5) non-abelian self-couplings present, abelian absent; (G6) loop-induced vertices absent from the tree grammar; (G7) three generations with flavour change confined to the dressed flip; (C) the four anomaly sums vanish exactly; (S) the generated class set equals the SM tree set. All pass; the program's exit status is the conjunction.

B Anomaly arithmetic over the alphabet

With multiplicities $n_c \times n_w$ and $Q = T_3 + Y/2$:

letter	mult.	Y
Q	6	1/3
u^c	3	−4/3
d^c	3	2/3
L	2	−1
e^c	1	2
ν^c	1	0

$$\begin{aligned}
\textstyle\sum Y^3 &= 6 \cdot \tfrac{1}{27} - 3 \cdot \tfrac{64}{27} + 3 \cdot \tfrac{8}{27} - 2 + 8 + 0 = 0,\\
\textstyle\sum Y &= 2 - 4 + 2 - 2 + 2 + 0 = 0,\\
\textstyle\sum_{\text{doublets}} n_c Y &= 3 \cdot \tfrac{1}{3} + 1 \cdot (-1) = 0,\\
\textstyle\sum_{\text{triplets}} n_w Y &= 2 \cdot \tfrac{1}{3} - \tfrac{4}{3} + \tfrac{2}{3} = 0.
\end{aligned}$$