Semiring Parsing the S-Matrix

Packed forests, recursion as dynamic programming, and typed pruning in perturbative field theory

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

5 July 2026

Abstract

A companion paper certified the Standard Model's tree-level vertex algebra as a typed grammar: sixteen lexical codewords, productions for the reads, flips, and bridges, conservation laws as a type system, and exact anomaly freedom as the global consistency certificate [1]. This sequel asks what the grammar's *evaluation theory* says about amplitudes. The central observation is structural: at the combinatorial (skeleton) layer, amplitude assembly is *inside evaluation over a packed parse forest* in the sense of semiring parsing [2]. The Boolean semiring is the legality front end, the counting semiring counts Feynman skeletons, and the complex-weight version of the same inside recurrence, with couplings as production weights, computes the amplitude skeleton. The Born rule is then the reading law converting inside scores to record probabilities. Three consequences follow. (1) The amplitudes community's recursion revolution — Berends–Giele currents [3] and BCFW [4] — is dynamic programming on the shared forest: off-shell currents are memoized subderivations, and parsing theory *predicts* the polynomial-versus-factorial separation the field found empirically. (2) The planar/tree fragment is context-free-class: inside evaluation is $O(n^3)$-polynomial over Catalan-many parses, and the $1/N$ expansion [5] acquires a complexity reading. (3) Typed pruning terminates illegal subderivations before expansion, where the path integral generates and cancels by interference; we keep the algorithmic claim strictly separate from any ontological one. Two worked examples run end to end ($e^+e^- \to \mu^+\mu^-$ as forest semantics with visible interference; proton decay as a syntax error whose repair is priced), and every quantitative statement is an assertion of an accompanying self-checking evaluator whose measured scaling (chart work $n^{3.00}$; forest growth $e^{1.32}$ per leg against $\ln 4 = 1.386$) is reproduced on commodity hardware.

1 Introduction and scope fence

The predecessor paper [1] established that the Standard Model's tree-level interaction content can be written as a typed grammar in the strict sense: a lexicon (the sixteen Weyl fermions of one generation, realised as the codewords of an $[8, 4, 4]$ code, plus the response alphabet $\{\gamma, Z, W^\pm, g, H\}$), typed productions (reads, the left-handed flip, the Yukawa bridge, the $\Delta L = 2$ portal), a type system (charge, colour triality, chirality, B, L, generation), and a global consistency certificate (all four gauge anomaly sums vanish exactly). A Feynman diagram is a derivation; an amplitude is a compiled sentence.

That paper deliberately stopped at *recognition*: which sentences are legal. This paper is about *evaluation*: what the parse is worth. Its single load-bearing idea is imported from computational linguistics [2]: once a grammar's derivations are organised as a *packed forest* — the polynomial-size chart whose nodes are shared subderivations — then legality, counting, best derivation, and weighted sums are all the *same* inside algorithm run over different semirings. We claim that perturbative amplitude assembly has exactly this structure at the skeleton layer.

The scope fence (stated now, and again in Section 2). The semiring claim lives at the combinatorial layer: the assembly of an amplitude from vertex factors and propagator *symbols*. Loop *integrals* — the analytic evaluation of the weights attached to a skeleton — are the weight-evaluation step, not the parse, and nothing in this paper claims parsing complexity controls them. Where we say "polynomial" we mean polynomial in the number of shared skeleton evaluations, the precise sense in which tree-level recursion is polynomial [3]. Claims are fragment-scoped (planar/tree); general non-planar complexity is left open.

2 Parse-forest semantics

Fix the typed grammar G of [1]. A *sentence* is an assignment of external letters (asymptotic states). A *derivation* is a tree of productions consuming the sentence; the set of derivations, quotiented by shared subderivations, is the packed forest $F(s)$. Semiring parsing evaluates

$$\mathrm{Inside}(s) = \bigoplus_{d \in F(s)} \bigotimes_{p \in d} w(p), \tag{1}$$

with $(\oplus, \otimes)$ the semiring operations and w the production weights, computed on the chart in time polynomial in the chart size [2, 6]. The dictionary:

semiring	$w(p)$	physics reading
Boolean	`true`	legality (selection rules)
counting	1	number of Feynman skeletons
complex	coupling × propagator symbol	amplitude skeleton

Claim 1 (Forest semantics)**.** *At tree level, and more generally at fixed perturbative order before loop integration, the skeleton amplitude for a legal sentence is the complex-weight inside score of its packed forest; meaning attaches to the* forest, *not to any single tree. Interference between derivations is the difference between the coherent inside score and the incoherent per-tree sum. In real electroweak observables, the forward–backward asymmetry at the Z pole is one familiar place where such coherent channel and vector/axial interference becomes visible; Example A, Section 7, keeps only the simpler γ/Z forest to show the mechanism.*

The reading law is the companion instrument theory's D2a [7]: inside scores are response objects; squaring against the record basis (the Born rule) converts them to record probabilities. In that framework, stable records are the classical readout endpoints of the quantum response [7, 8]; the forest lives entirely on the response side.

3 The Berends–Giele/BCFW bridge

The amplitudes community discovered, empirically and repeatedly, that factorial diagram enumeration is the wrong algorithm. The Berends–Giele recursion [3] computes n-gluon trees in polynomial time by building *off-shell currents*; BCFW [4] reduces on-shell amplitudes to smaller on-shell amplitudes; Parke–Taylor [9] showed the summed object can be radically simpler than any diagram-by-diagram accounting.

Observation 1 (The bridge)**.** *These recursions are inside evaluation on the shared forest. An off-shell current* $J(1, \ldots, k)$ is *the chart cell for the span* $\{1, \ldots, k\}$ *— a memoized subderivation* *value; the Berends–Giele relation is the chart's recurrence* (1)*; and the polynomial-versus-factorial separation is the textbook packed-forest phenomenon: Catalan-many trees share* $O(n^2)$ *spans evaluated in* $O(n^3)$ *work [6]. Parsing theory predicts the recursion revolution; the field found it by ingenuity.*

The accompanying evaluator measures exactly this on the cubic-vertex toy: chart work fitted at $n^{3.00}$ while the forest it evaluates grows at $e^{1.32}$ per leg (the exact asymptotic is $\ln 4 = 1.386$, the Catalan growth rate), with counts Catalan-*exact* to $n = 16$ against both the closed form and brute-force enumeration (Appendix A).

4 The context-free fragment, sharpened

The predecessor paper graded the grammar by fragments; here the grading becomes complexity statements. The *planar/tree* fragment — derivations whose skeletons are planar at fixed cyclic order — is context-free-class: its forest is the bracketing chart, and inside evaluation is $O(n^3)$ [6]. The 't Hooft $1/N$ expansion [5] is then a complexity filtration: leading N selects precisely the fragment whose evaluation shares maximally, and non-planar corrections are, in grammar terms, context-sensitive insertions whose sharing structure degrades in a controlled way. We state this as a two-sided scope: polynomial evaluation is *claimed* for the CF fragment (and demonstrated); the complexity of the full non-planar grammar is *left open*, and nothing below depends on it.

5 Typed pruning versus generate-and-cancel

Two evaluation strategies coexist in field theory. In a Lagrangian or path-integral representation one often sums over a larger space and lets symmetries, projectors, Ward identities, and cancellations enforce the physical sector. The grammar *prunes* earlier: a subderivation that violates a typing rule (charge, chirality, B, L) is terminated before expansion. These are compatible descriptions of the same selection rules, but they are not the same *algorithm*, and the difference is the entire content of Sections 3–4.

We keep two claims strictly separate. The *evaluation claim* (algorithmic, safe): syntax-directed evaluation with typed pruning is the efficient normal form of skeleton assembly, and the community's recursions already instantiate it. The *ontological claim* (interpretation-fenced, never load-bearing here): in the substrate programme that hosts the grammar [10], service rules act as per-event type checks, so "nature prunes" is a live reading of why exact conservation laws are exact. Readers who reject the second claim lose nothing of Sections 2–4.

6 Unitarity as admissible weighting

Not every weighting of a legal grammar is physical. In semiring terms, unitarity is a constraint on the weight assignment and regularisation, not on the forest alone: the complex weights must be such that the squared reading of (1) over any complete record basis sums to one. Cutting rules are the forest restatement — the imaginary part of an inside score is a sum over ways of splitting a derivation into two on-shell subderivations, i.e. over chart cells — and crossing is the statement that one packed forest serves several sentences. We record this as the boundary of the paper: the constraint is stated, its systematic exploitation (which weightings on which fragments) is future work.

7 Two end-to-end examples

Example A — a legal sentence: $e^+e^- \to \mu^+\mu^-$**.** Front end: all letters in the lexicon; two derivations found, the γ-read and Z-read productions composed head to head in the s-

channel. Attributes check (charge, colour trivial, chirality read-blind with distinct g_Z weights, generation preserved). Middle end: the packed forest is $\{\gamma_s, Z_s\}$ — a forest, not a tree; the amplitude is the coherent inside score, and the evaluator exhibits the interference term explicitly ($|A_\gamma + A_Z|^2 - |A_\gamma|^2 - |A_Z|^2 = +0.079$ at its toy kinematic point): the forward–backward asymmetry at the Z pole is parse interference made observable. Back end: the certificate LEGAL(tree, order e^2) with both derivations; the skeleton $\sum_{V\in\{\gamma,Z\}}[\bar{v}\,\Gamma_V\,u]\,P_V(s)\,[\bar{u}\,\Gamma_V\,v]$; and (for the certifier tool of the next paper in this series) a model-file stanza per vertex. For contrast, $e^+e^- \to e^+e^-$ parses to four derivations ($\{\gamma, Z\} \times \{s, t\}$): the count ratio 2 is a forest cardinality statement the evaluator checks exactly.

Example B — a syntax error: $p \to e^+\pi^0$. The grammar is written for elementary fields, so the hadronic sentence is first expanded to its quark-level content. After that resolution the front-end verdict is sharp: all letters are valid, but *no parse exists at any order* in the certified grammar, because every production conserves B and the process needs $\Delta B = -1$. This is a syntax error of class *new grammar* — not merely a rare amplitude and not merely a suppressed one. Back end: the falsifier report prices the repair. Legalising the sentence requires a $\Delta(B-L) = 0$ baryon violating production (for example a leptoquark-class letter or a dimension-six-style bridge); the grammar's anomaly-completeness filter then *forces* the partner content of any such letter, so the repair predicts other vertices — the repair is priced, not free. The diagnostic's teeth show in the contrast row: neutrinoless double-beta decay is superficially similar ("rare B/L process") but *has* a parse — the $\Delta L = 2$ Majorana portal is already a production — so it is a sensitivity class, not new grammar. The compiler separates what the phrase "rare search" conflates. (A third class — $\mu \to e\gamma$ as a portal-dressed loop — completes the five-way classification and is exercised in the certifier paper.)

8 Concluding remark

The grammar paper ended with a certified alphabet; this paper ends with a certified algorithm. What remains is engineering with teeth: the certifier that emits community-standard model files from a grammar certificate, and the repair engine that turns anomalies into priced grammar extensions with forced partner content — both specified, both gated on reproduction benchmarks, neither claimed here.

Reproducibility. Every quantitative claim is asserted by the self-checking script `python_code/c1_forest_inside_evaluator_gate.py`. It uses only the Python standard library and exits 0 only if all five checks pass. The certified compiler underlying the grammar is `python_code/quantum_grammar_compiler_certification.py` (12/12) [1].

A The certification program

G1: Boolean-semiring chart agrees with brute-force parse existence ($n \leq 10$). G2: counting-semiring chart equals the Catalan number C_{n-1} exactly ($n \leq 16$) and brute force where affordable. G3: measured scaling — chart work $n^{3.00}$ (fitted), forest growth $e^{1.32}$ per leg against $\ln 4 = 1.386$; the polynomial/exponential separation asserted, not assumed. G4: Example A's forests carry exactly 2 and 4 derivations; the coherent/incoherent gap is $+0.079$ at the toy point. G5: Example B is untypable, one priced portal legalises it, and $0\nu\beta\beta$ already parses.

References

[1] David Elliman. The grammar of quantum physics: A typed vertex algebra for the standard model. Zenodo preprint, 2026.

[2] Joshua Goodman. Semiring parsing. *Computational Linguistics*, 25(4):573–605, 1999.

[3] Frits A. Berends and Walter T. Giele. Recursive calculations for processes with n gluons. *Nuclear Physics B*, 306:759–808, 1988. doi: 10.1016/0550-3213(88)90442-7.

[4] Ruth Britto, Freddy Cachazo, Bo Feng, and Edward Witten. Direct proof of the tree-level scattering amplitude recursion relation in Yang-Mills theory. *Physical Review Letters*, 94: 181602, 2005. doi: 10.1103/PhysRevLett.94.181602.

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

[6] Daniel H. Younger. Recognition and parsing of context-free languages in time n^3. *Information and Control*, 10:189–208, 1967. doi: 10.1016/S0019-9958(67)80007-X.

[7] David Elliman. Records and responses: Dressing-blind theorems for monitored quantum instruments. Zenodo preprint, 2026.

[8] Wojciech H. Zurek. Decoherence, einselection, and the quantum origins of the classical. *Reviews of Modern Physics*, 75:715–775, 2003. doi: 10.1103/RevModPhys.75.715.

[9] Stephen J. Parke and Tomasz R. Taylor. Amplitude for n-gluon scattering. *Physical Review Letters*, 56:2459–2460, 1986. doi: 10.1103/PhysRevLett.56.2459.

[10] David Elliman. An executable record grammar for quantum correlations. Zenodo preprint, 2026.