Published for SISSA by Springer

Received: *July 4, 2023*
Revised: *October 6, 2023*
Accepted: *October 30, 2023*
Published: *November 22, 2023*

Statistics of Feynman amplitudes in ϕ^4-theory

Paul-Hermann Balduf

Department of Combinatorics and Optimization, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada

E-mail: paul.balduf@uwaterloo.ca

Abstract: The amplitude of subdivergence-free logarithmically divergent Feynman graphs in ϕ^4-theory in 4 spacetime dimensions is given by a single number, the Feynman period. We numerically compute the periods of 1.3 million completed graphs, this represents more than 33 million graphs contributing to the beta function. Our data set includes all primitive graphs up to 13 loops, and non-complete samples up to 18 loops, with an accuracy of ca. 4 significant digits.

We implement all known symmetries of the period in a new computer program and count them up to 14 loops. Combining the symmetries, we discover relations between periods that had been overlooked earlier. All expected symmetries are respected by the numerical values of periods.

We examine the distribution of the numerically computed Feynman periods. We confirm the leading asymptotic growth of the average period with growing loop order, up to a factor of 2. At high loop order, a limiting distribution is reached for the amplitudes near the mean. A small class of graphs, most notably the zigzags, grows significantly faster than the mean and causes the limiting distribution to have divergent moments even when normalized to unit mean. We examine the relation between the period and various properties of the underlying graphs. We confirm the strong correlation with the Hepp bound, the Martin invariant, and the number of 6-edge cuts. We find that, on average, the amplitude of planar graphs is significantly larger than that of non-planar graphs, irrespective of $O(N)$ symmetry.

We estimate the primitive contribution to the 18-loop beta function of the $O(N)$-symmetric theory. We find that primitive graphs constitute a large part of the beta function in MS for $L \to \infty$ loops. The relative contribution of planar graphs increases with growing N and decreases with growing loop order L.

Keywords: $1/N$ Expansion, Large-Order Behaviour of Perturbation Theory, Renormalons, Renormalization Group

ArXiv ePrint: 2305.13506

https://doi.org/10.1007/JHEP11(2023)160

Contents

JHEP11(2023)160

1 Introduction

1.1 Motivation

One of the most widely used frameworks to compute observables in quantum field theory is perturbation theory: an approximation to the sought-after function is constructed as a power series in a coupling parameter, where the series coefficients are given by Feynman integrals, which in turn can be indexed by Feynman graphs. The first coefficients of the perturbation series often give a decent approximate result, comparable with current experimental precision, but there are various reasons to believe that the perturbation series is ultimately divergent. This does not necessarily invalidate perturbation theory, because it might be possible to infer non-perturbative finite results from a divergent perturbation series via resummation. More broadly, the properties of divergent series can be examined within the framework of *resurgence* theory, but such a study needs as input the asymptotic growth of series coefficients at high order. Promising steps in this direction have been taken in the last years, see for example [1–6], however, they are often restricted to cases where extra information, such as a differential equation or an exact solution of the involved Feynman integrals, are available to efficiently generate high-order series coefficients.

The number of non-isomorphic Feynman graphs per loop order grows factorially with the loop order, see section 2.2. Considering only a structurally simple subclass is not a reliable approximation for the amplitude of "typical" Feynman graphs of the theory at high loop order. The goal of the present work is to provide such a broad picture, by numerically computing as many graphs as possible, albeit only to limited precision. This allows us to not only examine the average of those amplitudes, but also their distribution and various correlations, which is vital for estimating the reliability of non-complete samples or extrapolations.

Concretely, we examine all subdivergence-free Feynman integrals contributing to the 4-point function (the *periods*, see section 1.2) in $O(N)$-symmetric ϕ^4-theory up to, and including, 13 loops, and non-complete samples of such graphs up to 18 loops. Details are explained in section 1.3.

1.2 The period

The renormalized perturbative 4-point function (amplitude) of ϕ^4-theory is given by a sum over all vertex graphs of ϕ^4-theory, that is, graphs which are 4-regular except for having

JHEP11(2023)160

exactly 4 external edges,

$$G_{\mathcal{R}}^{(4)} = \sum_{G \text{ is } \phi^4 \text{ vertex graph}} \frac{4!}{|\mathrm{Aut}(G)|} \mathcal{F}_{\mathcal{R}}(G), \tag{1.1}$$

Here $\mathcal{F}_{\mathcal{R}}(G)$ denotes the renormalized Feynman integral associated to the graph G. In full generality, a renormalized Feynman integral is a function of masses, external momenta, coupling constants, the renormalization scheme and renormalization scale, and the dimension and signature (Euclidean or Minkowski) of spacetime.

In the present work, we consider Euclidean ϕ^4-theory with interaction $\frac{\lambda}{4!}\phi^4$ in $D = 4-2\epsilon$ spacetime dimensions. The vertex graphs of this theory have $|E_G| = 2L_G$ internal edges and $|V_G| = L_G + 1$ vertices, where L_G is the loop number. Their superficial degree of convergence is $\omega_G = |E_G| - L_G\frac{D}{2} = L_G \cdot \epsilon$, making them superficially log-divergent. The unrenormalized Feynman integral can be expressed as a parametric integral according to

$$\mathcal{F}(G) = \frac{(-\lambda)^{|V_G|}}{(4\pi)^{|L_G|\frac{D}{2}}}\Gamma(\omega_G) \prod_{e\in E_G} \int_0^\infty \mathrm{d}a_e\, \delta\left(1 - \sum_{e=1}^{|E_G|} a_e\right) \frac{\psi_G^{\omega_G - \frac{D}{2}}}{\phi_G^{\omega_G}}. \tag{1.2}$$

See [7, 8] for details and derivations. In eq. (1.2), the structure of the graph is encoded in the first and second Symanzik polynomials ψ_G and ϕ_G [9]. These polynomials are homogeneous functions of the Schwinger parameters $\{a_e\}$, and ϕ_G is additionally a function of masses and momenta. Using dimensional regularization, $D = 4-2\epsilon$, the divergence of the integral eq. (1.2) takes the form of poles in ϵ.

Furthermore, we restrict ourselves primitive graphs, this means that no proper subgraph of G has zero or negative degree of convergence in the limit $\epsilon \to 0$. In that case, the only ultraviolet divergence in eq. (1.2) is the gamma function $\Gamma(\omega_G) = \frac{1}{L\epsilon} + \mathcal{O}(\epsilon^0)$, and any potential divergence of the integral itself is of kinematic or infrared nature.

Let p be a suitable linear combination of external momenta, then $s \propto p^2$ defines an overall *scale*, and all masses and momenta can be expressed relative to s by scale-free ratios, called *angles*. For dimensional reasons, we also need a reference scale s_0, which is arbitrary but fixed. Expanding eq. (1.2) in ϵ, we then find the general structure of the integral of a primitive log-divergent graph:

$$\mathcal{F}(G) = \Lambda\left(\frac{\mathcal{P}(G)}{L_G}\frac{1}{\epsilon} - \mathcal{P}(G)\ln\left(\frac{s}{s_0}\right) + C_G(\{\theta\}) + \mathcal{O}(\epsilon)\right). \tag{1.3}$$

Here, $\Lambda = (4\pi)^{L_G(-2)}(-\lambda)^{|V_G|}$, and C_G is a finite quantity which might depend on the angles $\{\theta\}$, but not on the scale s. $\mathcal{P}(G)$ is the *period*

$$\mathcal{P}(G) = \left(\prod_{e\in E_G}\int_0^\infty \mathrm{d}a_e\right)\, \delta\left(1 - \sum_{e\in E_G} a_e\right) \frac{1}{\psi_G^2}. \tag{1.4}$$

The period is free of UV-divergences and it does not depend on the kinematics of the graph. For a more rigorous discussion of the convergence of integrals and the precise dependence on angle variables, as well as alternative integral representations for the period,

JHEP11(2023)160

see [10–12]. The fact that the same coefficient describes the $\ln(s)$-dependence and the ϵ^{-1}-pole is expected from the renormalization group [13]. In ϕ^4-theory, the period is known analytically for the infinite family of zigzag-graphs (eq. (4.11)), and hundreds of other graphs [11, 14–19].

For a primitive graph, renormalization merely amounts to subtraction of the pole term and — depending on the scheme — potentially finite terms of the unrenormalized amplitude. Crucially, renormalization can not alter the s-dependent part of the primitive amplitude because that would require an s-dependent, hence non-local, counterterm. Therefore, the renormalized integral of a primitive graph eq. (1.3) has the form

$$\mathcal{F}_{\mathcal{R}}(G) = \Lambda \left(-\mathcal{P}(G) \ln\left(\frac{s}{s_0}\right) + \text{finite terms depending on } \theta, \text{ but not on } s + \mathcal{O}(\epsilon) \right). \tag{1.5}$$

In particular, the period encodes the s-dependence at s_0 of eq. (1.5),

$$\frac{\partial}{\partial \ln(s)} \mathcal{F}_{\mathcal{R}}(G)\Big|_{s=s_0} = -\Lambda \cdot \mathcal{P}(G). \tag{1.6}$$

The central physical motivation to consider the period is its direct contribution to the beta function via eq. (1.6), as will be discussed in section 5. However, the period is also interesting from a conceptual perspective because it is structurally simpler than Feynman integrals with subdivergences, and is being studied with respect to the algebraic structures or the number content of Feynman integrals. Thirdly, a period in the pure mathematics sense is any number which is defined by a rational integral with rational boundaries, and these periods constitute a class of numbers between algebraic and complex numbers. Consequently, periods have received considerable interest in the past two decades, an incomplete list of relevant publications is [11, 20–32].

1.3 Samples

The periods contributing to the 4-point function of ϕ^4-theory are invariant under completion, that is, joining the 4 external edges of the graph G to a new vertex, see section 2.1. The resulting graph is a primitive (subdivergence-free) vacuum graph of ϕ^4-theory, called a *completion* of G. Owing to the completion symmetry of the period, the present paper is largely organized in terms of completions, not in terms of the underlying vertex graphs called *decompletions*.

We generated all completions and decompletions up to $L = 16$ loops, counts will be discussed in section 2.

For numerical integration, we use the algorithm recently developed by Borinsky [33]. Up to including $L = 13$ loops, we computed the periods of all primitive completions — even of those which are known to have equal periods, see section 3. For higher loop order, we drew random samples, generated by multiple runs of the `genrang` procedure supplied with `nauty` version 2.7 [34], and corrected for non-uniform sampling according to appendix A. The concrete numbers of graphs (prior to correction) are reported in table 1, where the non-complete samples are indicated by appending "s" to the loop number. To assess the

JHEP11(2023)160

L	computed completions	proportion	relative accuracy	computed decompl.	planar decompl.	inferred compl.	inferred decompl.
1	1	1	4/4	1	1	1	1
3	1	1	3/3	1	1	1	1
4	1	1	3/3	1	1	1	1
5	2	1	6/8	3	2	2	3
6	5	1	5/6	10	5	5	10
7	14	1	7/9	44	19	14	44
8	49	1	9/15	248	58	49	248
9	227	1	24/33	1688	235	227	1688
10	1354	1	68/90	13094	880	1354	13094
11	9722	1	36/90	114016	3623	9722	114016
12	81305	1	56/90	1081529	14596	81305	1081529
13	755643	1	217/289	11048898	60172	755643	11048898
13s	10000	$1.4 \cdot 10^{-2}$	143/284	148389	849	26329	364065
13$\star$	10000	$1.4 \cdot 10^{-2}$	130/284	146084	843	32476	412477
14s	216190	$2.9 \cdot 10^{-2}$	246/289	3436145	7314	489993	7690996
14$\star$	109999	$1.4 \cdot 10^{-2}$	217/240	1734993	3649	302704	4677617
15s	101207	$1.2 \cdot 10^{-3}$	220/288	1713395	1283	258802	4320337
15$\star$	67048	$8.1 \cdot 10^{-4}$	212/260	1129937	879	211718	3475930
16s	10196	$1.1 \cdot 10^{-5}$	217/274	182990	39	23329	414073
17s	5178	$4.6 \cdot 10^{-7}$	172/262	98180	7	11644	217002
18s	912	$6.2 \cdot 10^{-9}$	108/221	18206	0	1743	34442

Table 1. Overview of the graphs computed in this work. Loop number of decompletion, with s indicating non-complete samples and $\star$ indicating a sample drawn from the complete set of graphs. For "s", the numbers refer to the samples prior to correction (appendix A). The columns describe the following: number of completed primitives computed; proportion computed relative to total number (table 2); average/maximum relative accuracy, in ppm, average taken over computed completions; decompletions (= non-isomorphic 4-valent Feynnman graphs) of the computed completions; planar decompletions; completions which can be inferred from the computed graphs by symmetries, including the computed ones; decompletions which can be inferred from the computed graphs. The latter is the total number of non-isomorphic graphs whose period is known numerically.

JHEP11(2023)160

reliability of the sampling, we generated, using `genrang`, a random sample of 10000 13-loop graphs, processed it analogously to the other samples and compared results to the full data set of all 13-loop graphs.

For 13, 14 and 15 loops, we drew uniform random samples out of the set of *all* completions using `Mathematica` version 13.1, they are denoted 13⋆, 14⋆, and 15⋆ in table 1. Those samples serve as a check for potential bias in the random graph generation algorithm of `genrang`. The overlap between the samples 14s and 14⋆, and between 15s and 15⋆, is very small.

In total, we computed the period of more than 1.3 million different completions, they correspond to more than 16.5 million different decompletions (= non-isomorphic Feynman graphs of the 4-point function eq. (1.1)). The periods of over 33 million different decompletions can be inferred from this data by exploiting known symmetries of the period, see the last colum of table 1.

We express the accuracy of our results as the standard deviation, relative to the absolute value. For $L \leq 12$ loops, all numerical periods have ≤ 90 ppm relative accuracy, that is, slightly more than 4 significant digits. For all other graphs, we reach ≤ 290 ppm accuracy, see column 4 in table 1.

1.4 Content and results

A large-scale statistical analysis necessarily involves many tables and plots to show the various coefficients, distributions, and correlations. To make this work more accessible to the reader, the present section is a survey of all central outcomes. References point to the concrete subsections and tables in the main text which support our claims empirically.

Section 2 concerns the number of graphs, frequency of subclasses, growth rate, and symmetry factors.

- Section 2.1 is a review of completions, decompletions, and symmetry factors.
- A well-understood model in graph theory, which contains the completions of ϕ^4-theory as a subset, is the class of random 4-regular simple graphs $\mathcal{G}_{n,4}$ (theorem 1).
- If multiedges are excluded, most of the vertex-type graphs of ϕ^4-theory are primitive, in accordance with the prediction from $\mathcal{G}_{n,4}$ (section 2.2).
- The proportion of planar graphs at high loop order is vanishingly small (table 2).
- With growing loop order, automorphisms and symmetry factors become increasingly irrelevant (section 2.3), at 15 loops, the average symmetry factor is 0.95, more than 90% of all completions have symmetry factor unity and more than 99% of all decompletions are non-isomorphic (table 3). This trend is expected from $\mathcal{G}_{n,4}$. The "asymptotic" regime starts rather abruptly at $L_{\text{crit}} \approx 9$ loops (figures 4(a) and 7).
- In terms of symmetry factors, planar graphs behave similarly to the full data set (table 4), and there is no significant correlation between the symmetry factor of a completion and how many of its decompletions are planar (section 2.4).

In section 3, we discuss the symmetries of the period (not to be confused with the symmetry factors of section 2), and how they are used for the numerical integration.

JHEP11(2023)160

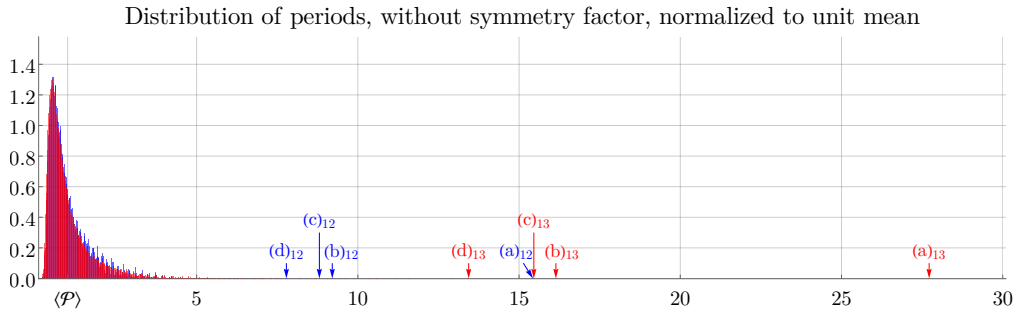


Figure 1. Distribution of periods for 12 loops (blue) and 13 loops (red), normalized with respect to their mean $\langle \mathcal{P} \rangle$. Labels $(a)_L, (b)_L, \ldots$ refer to the largest, 2nd largest, ... period of the L-loop sample, see section 4.3. $(a)_L$ is the L-loop zigzag graph. Note that the "continuous" part around $\langle \mathcal{P} \rangle$ (shown in figure 2(b)) is almost identical for the two distributions, whereas the "outliers" of the 13-loop sample are larger than those of the 12-loop sample by a constant factor ≈ 2.

- In section 3.1, we review the known symmetries and the Hepp bound, and comment on our programs to compute each of them.
- The number of symmetric graphs is reported in table 5. The bulk of the symmetries can be interpreted as Twist and/or Fourier split. Pure Fourier identities are very rare.
- Our result agrees with earlier publications at lower loop order, but we find slightly more graphs related by symmetries at higher loop order than previously assumed. This is because we use more combinations of the known symmetries (section 3.3).
- Still, not all cases where the Hepp bound agrees can be explained by known symmetries. These cases are rare compared to the total number of graphs (table 5).
- The tropical Feynman quadrature algorithm [33] works as anticipated. The process has been fully automated in order to handle large sample sizes (section 3.4).
- Numerical results agree whenever we expect the graphs to be related by a period symmetry. We use this to improve the accuracy of the numerical results.

Section 4 is about the empirical distribution of periods as observed from our samples.

- For $L \leq 13$ loops, we determine the mean of the period to 6 significant figures with (table 7) and without (table 6) symmetry factor. For $14 \leq L \leq 18$ loops, the uncertainty of the mean is, by far, dominated by the sampling uncertainty (section 4.1). The mean grows by a factor of $\frac{3}{2}$ per loop order (figure 2(a), section 5.4).
- The average period of decompletions is similar to that of completions. The average period of planar graphs is significantly larger than that of all graphs (section 4.2).
- The distribution of periods consists of two qualitatively different regions: a continuous distribution around the mean, and a small set of very large outliers (figure 1).
- Even when the distribution is normalized to unit mean, the variance and higher moments grow monotonically with loop order and are likely infinite for the limiting distribution (figure 7). Even large non-complete samples of periods have significant sampling uncertainty, which can be estimated as discussed in appendices B and C.

JHEP11(2023)160

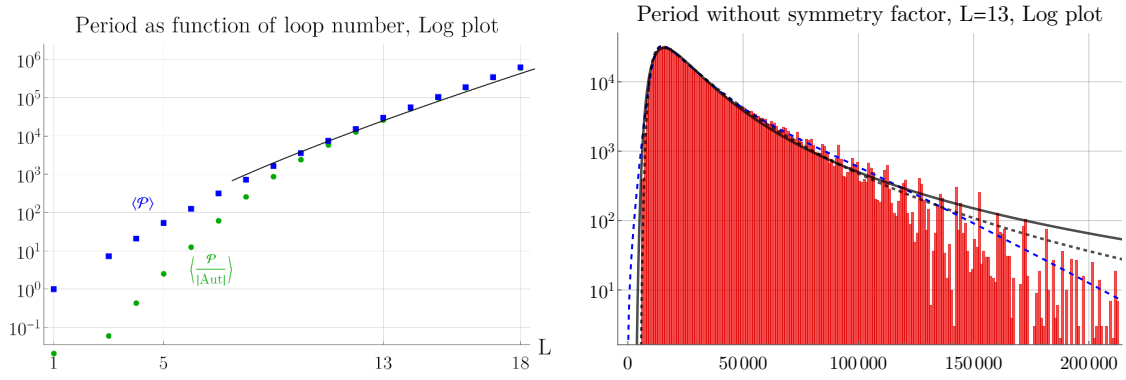


Figure 2. (a) Mean period of completed primitive graphs, where L is the loop number of decompletions. Blue squares: without symmetry factor (eq. (4.1)), green circles: with symmetry factor (eq. (4.5)). Black line: supposed asymptotics eq. (5.12). (b) Histogram of periods, without symmetry factor, absolute count with bin size 1000. The horizontal plot range corresponds to $[0, 7]$ in figure 1. Black line: quadratic model function eq. (4.16), Black dashed line: quartic model eq. (4.17). Blue long dashed line: empirical model eq. (4.13). Note that the quartic model accurately captures the falloff on the left, while the empirical model is superior for the decay on the right.

- The divergence of moments is caused by the outliers which grow faster than the average period and lie farther off the mean the higher the loop order. For each loop number, we find the largest graph to be the $(1, 2)$-circulant ("zigzag") of that loop number. The next-largest graphs can be described as "almost zigzags", where larger distortion from the zigzag shape results in smaller periods. (section 4.3.)
- Besides the $(1, 2)$-circulant being the largest period, the other circulants don't show remarkable features, see table 10. For $L \leq 11$ loops, the smallest period is a circulant.
- The continuous part of the distribution, roughly the periods between 0.5 and 4 times the mean, can be described well by an empirical distribution function (section 4.4). In this regime, the probability for a period $x \cdot \langle \mathcal{P} \rangle$ falls off like $e^{-1.2x}$. The parameters of the distribution function change only slowly with loop order and likely assume finite limiting values as $L \to \infty$.
- The inverse square root of the period has an almost symmetric distribution, which can be well approximated by a normal distribution (section 4.5) This gives rise to a second class of functions to model the distribution of periods. Either of the two models is reliable in a certain region, see figure 2(b).
- The logarithm of the period behaves much better in terms of outliers, it is possible, but not certain, that its limiting distribution has finite moments. Conversely, the logarithm is affected by symmetry factors much more than the period itself (section 4.6).

In section 5, we examine the consequences of our data for the beta function of a $O(N)$-symmetric ϕ^4-theory.

JHEP11(2023)160

- In section 5.1, we review the computation of the beta function and the conjecture that primitive graphs dominate the large-order asymptotics of the beta function in MS.
- Coefficients of the primitive beta function for $L \leq 18$ are listed in table 14. They agree with previous 7-loop analytic results and with an 11-loop extrapolation based on the Hepp bound.
- Regarding the asymptotics, our results are ambivalent: at $L \approx 18$, the primitive coefficients reach, and even exceed, the supposed leading order asymptotics. This supports the conjecture that the primitive coefficients dominate the beta function in MS, but the numerical values come out significantly too large to indicate convergence to the limit (section 5.2). One possible interpretation is that the leading asymptotics is missing a factor of two. An alternative interpretation is that the asymptotic regime will only be reached at $L \gg 18$ loops.
- We argue, and find empirically, that the primitive contribution to the $O(N)$-symmetric beta function always exceeds the leading asymptotic value for any fixed L if N is chosen large enough. This implies that our data, even at $L = 18$, is not yet close to the asymptotics if N is significantly larger than unity (section 5.3).
- The relative contribution of planar graphs to the beta function grows with growing N, but becomes relevant only at $N \gg 10$. Extrapolating the planar graphs to all graphs gives a too large result even at $N = 1$ since planar graphs have above average periods (table 9).
- The supposed asymptotic growth of the coefficients of the beta function implies a leading asymptotic growth of the mean period. Our data is consistent with the expected leading growth rate $\left(\frac{2}{3}\right)^L L^{\frac{5}{2}}$. Again, introducing a factor of two into the absolute value of the asymptotics significantly improves the convergence, see section 5.4.

Finally, in section 6, we examine the statistical correlations between the period and various properties of the graph.

- The period of a graph is only weakly correlated with its symmetry factor or the number of non-isomorphic decompletions (section 6.1). However, there is a strong, almost linear correlation between the period and the number of *planar* decompletions of a graph, which gets more pronounced with growing loop order (figure 22).
- The diameter or the girth of a graph turn out to be not useful for the loop orders under consideration since they only assume very few distinct values (theorem 1 and section 6.3).
- There is a significant correlation between the mean distance between two vertices in a graph, and its period (section 6.2). This correlation is consistent across loop orders *without* any normalization, that is, graphs with similar mean distance tend to have similar periods (in absolute value), even if their loop order differs (figure 23).
- The period is correlated with the number of triangles, and the pattern is regular once the period is scaled to its leading asymptotic growth (figure 24). We find an empirical lower bound for the periods in our samples (eq. (6.1)). The period is only

JHEP11(2023)160

weakly correlated with the number of 4-cycles, and anticorrelated with the number of 5-cycles.

- The correlation between the period and the number of 6-edge cuts is so strong that it allows to predict the period with an average error of less than 15% (section 6.4).
- The Hepp bound grows faster with loop order than the period, its higher moments are divergent similarly to the ones of the period, but its overall distribution is more concentrated around the mean (table 16).
- We confirm the known strong correlation between period and Hepp bound, and determine new fit coefficients (section 6.5). This allows to predict, using a 5-parameter fit function for each loop order, the value of every period in our samples with a maximum error below 5%, and an average error below 1% (table 18).
- The residues of the prediction by the Hepp bound fall into distinct strips, all graphs within a given strip share the same number of 6-edge cuts (figure 28). This will be used in a future publication to predict the period.
- The period is strongly correlated with the Martin invariant (section 6.6), but not quite as strong as with the Hepp bound.

1.5 Discussion

All in all, our data supports the conjecture that primitive graphs represent a large contribution to the asymptotic beta function in MS at high loop order, see section 5.2. In particular, we confirm the expected growth per loop order both for the average period and for the coefficients of the beta function. Interestingly, this growth ratio quickly approaches its asymptotic value starting from a loop order $L \approx 9$ (see figure 20(b)), which is very similar to the critical loop order observed for the purely graph-theoretic properties, such as symmetry factors (section 2.3).

However, we can not clearly confirm the convergence of primitive graphs towards the supposed asymptotics of the MS beta function in absolute value. At $L \approx 18$, our data for primitive graphs is larger than the supposed asymptotics by a factor ≈ 2. If one introduces a correction factor 2 into the asymptotics, it becomes possible to extract a finite subleading correction coefficient to from our data for loop orders $L \geq 10$, see figure 20(a). This is no proof for the behavior $L \to \infty$, but at least it suggests that with a factor of 2, the asymptotics becomes a useful description for $10 \leq L \leq 18$. Interestingly, in that scenario the transition to the "asymptotic" regime occurs at the same loop order $L_{\text{crit}} \approx 9$ for the period values as it does for the pure graph-theoretical properties such as symmetry factors (section 2.3). In any case, our findings show that the asymptotic approximation is inaccurate below $L_{\text{crit}} \approx 9$ loops.

Another remarkable outcome concerns the role of planar graphs. Firstly, for large loop orders, the proportion of planar graphs is very small (theorem 1) and their contribution to the beta function is negligible. If we introduce $O(N)$ symmetry then also non-planar graphs contribute to the leading order in N (section 5.3). Nevertheless, for very large values of N, the contribution of planar graphs to the beta function becomes significant (figure 19(b)). This effect is noticeable, but relatively small, with $N \gg 1000$ required

JHEP11(2023)160

already at $L \approx 15$. This shows that for an $O(N)$ symmetric theory, one should not expect to obtain a good approximation to the beta function by leaving out non-planar graphs, at least for reasonably small values of N.

Secondly, planar graphs on average have considerably higher periods than non-planar ones (table 9). Consequently, planar graphs can not simply be "extrapolated" without introducing significant error. In view of both effects, great care is required when planar graphs are used as an approximation to the beta function in ϕ^4-theory.

We stress that this question is different from a situation where, by some other mechanism, the theory contains *only* planar graphs to begin with. In that case, our data is compatible with the assumption that such a planar beta function is a convergent power series in the coupling, at least for fixed small N.

The present work impressively demonstrates the power of the tropical Feynman quadrature algorithm [33]. We can conveniently evaluate the periods up to $L \approx 15$ loops. The limiting factors for higher loop orders are the non-parallelized preprocessing (several hours per graph) and the memory requirement (increasing by a factor of 4 per loop order).

Most likely, all higher moments of the normalized period distribution diverge in the limit $L \to \infty$. This has sobering consequences for the reliability of samples. We have seen that even a naive sample of far more than 10^5 completions, corresponding to over 1 million non-isomorphic 4-valent graphs, still barely produces 3 significant digits for the beta function (compare 14s and 14$\star$ in table 6 or the experiments in appendix C). The zigzag graphs, which are the only analytically known infinite class of ϕ^4-periods, are a very poor proxy for a "typical" period, as they are by far the largest graphs of a given loop order (section 4.3).

Regardless of the concrete numerical values for ϕ^4 periods, some qualitative conclusions from the present work can likely be extended to other classes of Feynman amplitudes, for example that the distribution parameters change monotonically and predictably with growing loop order (eq. (4.6)) or that "special" classes of graphs tend to be a poor proxy for the "generic" case.

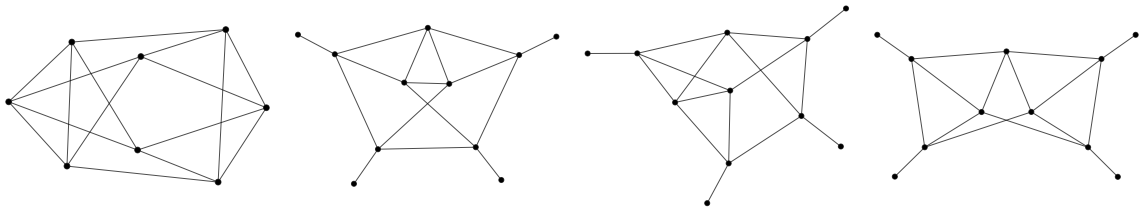
A different pathway to higher loop orders is to use the correlations between the period and various graph properties (section 6) to predict amplitudes without explicit numerical integrations. Such an extrapolation, based on the Hepp bound, had been given in [36] up to $L = 11$ loops and we found it confirmed (table 14). At higher loop orders, the number of graphs grows factorially and the accuracy of predicting individual periods becomes largely irrelevant as long as the prediction is correct on average. It seems well possible that an approximation of all 83 million completions at 15 loops from those correlations produces a more accurate estimate of the beta function than numerically computing 10^5 of them as in the present work. A more systematic study of that approach is currently in preparation.

2 Number of graphs and automorphisms

2.1 Graph automorphisms and completions

We consider Feynman graphs of ϕ^4-theory, this means the graphs contain one type of particle (all edges are of the same type) and a 4-point elementary interaction (all internal

JHEP11(2023)160



(a) (b) (c) (d)

Figure 3. (a) a 6-loop completion, and its 3 decompletions (b), (c), (d). External edges are highlighted by adding a 1-valent vertex. Even if (a) has 8 vertices, each of the 8 possible decompletions is isomorphic to either (b) or (c) or (d). Note that (a) and (b) are not planar, but (c) and (d) are.

vertices are 4-valent). We restrict ourselves to graphs which have either 4, or zero, external edges. The *loops* of a Feynman graph are a basis of cycle space, not to be confused with the term loop in graph theory. A connected Feynman graph G with 4 external edges and L_G loops has

$$|V_G| = L_G + 1 \tag{2.1}$$

vertices. As always for Feynman graphs in physics, by "graph" we mean an isomorphism class of labeled graphs which do not need to be simple. Two graphs are called *isomorphic*, and are being identified, if they differ only by a different labeling of their internal edges or vertices. An *automorphism* is a permutation of labels which transforms the labeled graph into itself, $\mathrm{Aut}(G)$ is the group of automorphisms of a graph G. In the perturbation series (eq. (1.1)), the graph is weighted with a *symmetry factor* $|\mathrm{Aut}(G)|^{-1}$.

If a graph G has 4 external edges, we can introduce one new vertex v and join the 4 external edges of G to v. This way, we obtain a 4-regular graph (that is, every vertex is 4-valent, not just the internal ones) without external edges. In physics terminology, this new graph $G \cup v$ is a vacuum graph of ϕ^4-theory. We call $G \cup v$ the *completion* of G, or conversely, we call G a *decompletion* of $G \cup v$. In the following, *decompletion* will always mean "graph of ϕ^4-theoy with 4 external edges". Technically, the completion has 3 more loops than its decompletion, however, we define that the *loop number of a completion* coincides with the number of loops of its decompletion. By eq. (2.1), all decompletions of a given completion have the same loop number. As an example, figure 3 shows a 6-loop completion (i.e. the decompletions have 6 loops), and its three non-isomorphic decompletions.

The symmetry factors of completions and decompletions are related by the identity

$$(L_G + 2)\,\frac{1}{|\mathrm{Aut}(G)|} = \sum_{\substack{g \text{ decompletion} \\ \text{of } G}} \frac{1}{|\mathrm{Aut}(g)|}, \tag{2.2}$$

where the automorphisms do not have fixed vertices. Note that this convention is different from the usual physics one where external vertices are demanded to be fixed by automorphisms. In our convention, a decompletion contributes to a 4-point amplitude (eq. (1.1))

JHEP11(2023)160

with all 4! different orientations of external edges (unless two external edges are incident to the same vertex, but, except for the 1-loop graph, our data set does not include such graphs as they are not primitive, see below), whereas in the usual physics convention, only the non-isomorphic orientations contribute.

We use the following notation to identify graphs: all graphs are transformed to the "canonical form for multigraphs" as given by `nauty` version 2.7 [34]. Note that this is different from `nauty`'s canonical form for simple graphs. The so-obtained graph has vertices numbered from 0 to $|V_G| - 1$; we transform these labels to single-digit integers and letters, $0, 1, \ldots, 8, 9, a, b, c, \ldots$. We then express the edges of the graph by the *Nickel index* [37], that is, a list of vertex adjacencies according to

$$/\text{vertices adjacent to } 0/\text{vertices adjacent to } 1/\text{vertices adjacent to } 2/\ldots/. \tag{2.3}$$

We leave out any adjacency that has been specified earlier, consequently, our Nickel index contains $|E_G|$ numbers, not $2 \cdot |E_G|$. For example, a 1-loop multiedge has Nickel index /11// and the graph (a) in figure 3 has /4567/2367/357/46/57/6///.

In ϕ^4-theory in 4 dimensions, a graph is superficially divergent if and only if it has 4 or fewer external edges. Consequently, a graph is *primitive* (free of subdivergences) if and only if it is internally 6-edge connected, that is, no non-trivial subgraph is connected to the remainder by fewer than 6 edges. A decompletion is primitive if and only if its completion is primitive [11]. For a fixed loop number L, we define the following counts of primitive graphs in ϕ^4-theory:

$$\begin{aligned} N_L^{(D)} &:= \text{Number of primitive decompletions with } L \text{ loops} \\ N_L^{(C)} &:= \text{Number of primitive completions whose decompletions have } L \text{ loops.} \end{aligned} \tag{2.4}$$

The completion of a given decompletion is unique by construction. A given completion generally has multiple non-isomorphic decompletions, therefore $N_L^{(D)} \geq N_L^{(C)}$. By eq. (2.1), a completion with L_G loops has $L_G + 2$ vertices and removing any of them potentially produces a decompletion, therefore, $N_L^{(D)} \leq (L_G + 2) N_L^{(C)}$. The number of decompletions is lower than this upper bound if some of the decompletions are isomorphic, as happens in the example figure 3.

2.2 Asymptotics of the numbers of graphs

We generated, using `nauty`'s `geng` routine [34], all 2-vertex connected 4-regular simple graphs of a fixed loop order $L \leq 16$. In physics terminology, this amounts to those 1-particle irreducible completed graphs of ϕ^4-theory which have neither multiedges nor tadpoles as subgraphs, and which can not be disconnected by removing a single vertex. The count of these graphs is reported in table 2 as "1PI compl.". For $L \leq 8$, the number coincides with the number of 4-regular connected simple graphs [38, A006820], for $L \geq 9$, they differ as not every connected graph is 2-vertex connected.

In the next step, we removed from the above sets those graphs which are not 4-edge connected. From a physics perspective, this step amounts to removing all remaining

JHEP11(2023)160

L	1PI completions	4-edge conn. completions	$N_L^{(C)}$	pla. com.	3-vtx irr. compl.	$N_L^{(D)}$	planar decompl.
3	1	1	1	0	1	1	1
4	1	1	1	1	1	1	1
5	2	2	2	0	1	3	2
6	6	6	5	1	4	10	5
7	16	16	14	1	11	44	19
8	59	58	49	2	41	248	58
9	264	262	227	2	190	1688	235
10	1542	1532	1354	9	1182	13094	880
11	10768	10726	9722	11	8687	114016	3623
12	88126	87915	81305	37	74204	1081529	14596
13	805281	804108	755643	79	700242	11048898	60172
14	8036251	8028714	7635677	249	7160643	120451435	246573
15	86214189	86159508	82698184	671	78270390	1393614379	1015339
16	985816860	985373593	952538165			17041643034	4169613

Table 2. Counts of 4-regular-graphs. The columns describe the following: Loop order of the decompletion; number of... 2-vertex connected 1PI completed *simple* graphs (i.e. no multiedges, no tadpoles); 4-edge connected simple completions; primitive completions; planar completions; 4-vertex-connected (i.e. 3-vertex irreducible) completions; primitive decompletions; primitive planar decompletions.

propagator corrections (recall that multiedges and tadpoles had been excluded before). The resulting number is "4-e compl." in table 2.

Lastly, we removed the graphs which are not internally 6-edge connected, that is, we remove all subgraphs which are quantum corrections to the 4-valent vertex. The remaining graphs are the primitive completed graphs (eq. (2.4)), their number is "$N_L^{(C)}$" in table 2. Automatically, those graphs are at least 3-vertex connected.

Table 2 additionally shows the number of 4-vertex-connected primitive completions, and of planar graphs. Those numbers will be relevant for the symmetries of the period to be discussed in section 3. Our results for $N_L^{(C)}$ and the 4-vertex connected 1PI graphs agree with the numbers listed, for $L \leq 14$, in [11]. The number of planar completions is consistent (i.e. lower or equal than) the number of planar 4-valent (but not necessarily primitive) graphs generated by `genreg` [39] and reported on the website [40].

For a qualitative understanding of the results of the present work, we recall some facts from the theory of large random graphs. Let $\mathcal{G}_{n,d}$ be the probability space of d-regular simple graphs with n labeled vertices, where all graph occurs with equal probability [41–43]. The graphs in $\mathcal{G}_{n,4}$ are not necessarily 1PI, or even connected, but they are still useful for our setting in the limit $n \to \infty$ due to the following properties:

Theorem 1. *Consider a random 4-regular simple graph $G \in \mathcal{G}_{n,4}$. Then the following statements hold in the limit where G becomes infinitely large, $n \to \infty$:*

JHEP11(2023)160

1. *(Wormald [42, 44]). G is almost surely 4-edge connected, and the probability of being not 4-edge connected scales like n^{-2}.*
2. *(Mentioned in [41] as known). If F is a graph with more edges than vertices, then G almost surely does not contain a subgraph isomorphic to F.*
3. *(Bollobás, Wormald [45, 46]). Let X_k be the number of cycles of length $k \geq 3$ in G. Then the X_k are independent Poisson distributed random variables with means $\frac{3^k}{2k}$.*
4. *(Bender & Canfield, Bollobás [45, 47, 48]). The number of unlabeled 4-regular simple graphs with $n = L - 1$ vertices grows like*

$$\frac{1}{4!(L+2)} \frac{\Gamma(L+3)e^{-\frac{15}{4}}}{\sqrt{2}\pi} \left(\frac{2}{3}\right)^{L+3} \left(36 + \mathcal{O}\left(\frac{1}{L}\right)\right).$$

5. *(Bonichon et al. [49]). The set of unlabeled planar graphs with n vertices, which contains as a subset the 4-regular planar graphs with n vertices, grows asymptotically slower than 31^n.*

Many more results of this type can be found in [41].

Theorem 1 helps to interpret the numbers reported in table 2 for larger loop order:

1. Most of the 1PI simple completions are 4-edge connected, in accordance with theorem 1 (1).
2. Most of the 1PI simple completions are primitive. Having excluded multiedges and tadpoles, this can be understood from theorem 1 (2): each possibly divergent subgraph has asymptotically zero probability of being present, and by theorem 1 (4), the number of such subgraphs is relatively small since subgraphs have lower L than the original graph. See also the explicit results in [50, 51].
3. For the completions, and even more strikingly for the decompletions in table 2, the proportion of planar graphs is small at higher loop order. This is plausible from theorem 1 (5), namely that the number of planar graphs grows only exponentially, not factorially.

2.3 Asymptotics of symmetry factors

Symmetry factors, or equivalently graph automorphisms, play a crucial role in the QFT perturbation series eq. (1.1) at low loop orders. We define the sum of all symmetry factors of a loop order, taking into account the 4! different orientations of a decompletion, by

$$N_L^{(\mathrm{Aut})} := \sum_{\substack{g \text{ decompletion} \\ L \text{ loops}}} \frac{4!}{|\mathrm{Aut}\, g|} = \sum_{\substack{G \text{ completion} \\ L \text{ loops}}} \frac{4!\,(L_G+2)}{|\mathrm{Aut}\, G|}. \tag{2.5}$$

Here, the sums are over non-isomorphic graphs and the automorphisms do not have any fixed vertices. Recall that the loop order of a completion is defined to be the number of loops of its decompletions and that, by eq. (2.1), a L-loop completion has $n = L+2$ vertices.

JHEP11(2023)160

The great advantage of $N_L^{(\mathrm{Aut})}$ over $N_L^{(C)}$ and $N_L^{(D)}$ (eq. (2.4)) is that $N_L^{(\mathrm{Aut})}$ can be obtained directly from a manipulation of the various generating functions of graphs [50, 51], without explicitly generating the graphs themselves. Consequently, the values $N_L^{(\mathrm{Aut})}$ and their asymptotic behavior, including subleading corrections, can be calculated to almost arbitrary order.

Theorem 2. *The following statements hold asymptotically in the limit $L \to \infty$ for completions, where L is the loop number of the decompletion:*

1. *(Bollobás, McKay & Wormald [48, 52]). A graph $G \in \mathcal{G}_{n,4}$ almost surely has no non-trivial automorphism, $|\mathrm{Aut}(G)| \to 1$.*
2. *(Borinsky [51]). The sum of symmetry factors of 1PI primitive decompletions, eq. (2.5), grows like $\bar{N}_L^{(\mathrm{Aut})} \left(1 + \mathcal{O}\left(L^{-2}\right)\right)$, where*

$$\bar{N}_L^{(\mathrm{Aut})} := \frac{\Gamma(L+3)e^{-\frac{15}{4}}}{\sqrt{2}\,\pi} \left(\frac{2}{3}\right)^{L+3} \left(36 - \frac{3}{2}\frac{243}{2}\frac{1}{L+2}\right).$$

We used the methods of [50, 51] to determine $N_L^{(\mathrm{Aut})}$ (eq. (2.5)) for $L \leq 18$, see table 3. For loop orders $L \geq 9$, $N_L^{(\mathrm{Aut})}$ differs from its asymptotics $\bar{N}_L^{(\mathrm{Aut})}$ (theorem 2) by no more than 1%.

When symmetry factors become trivial, the sum of symmetry factors $N_L^{(\mathrm{Aut})}$ approaches the sum of graphs $N_L^{(C)}$ (eq. (2.4)), as reflected by the coincidence of their leading asymptotics theorem 1 (4) and theorem 2 (2). As $N_L^{(\mathrm{Aut})}$ is much easier to compute, this leads to the question: how good is the approximation of $N_L^{(C)}$ by $N_L^{(\mathrm{Aut})}$ at finite loop order? The corresponding ratio, by eq. (2.5), can equivalently be interpreted as the average symmetry factor,

$$\frac{N_L^{(\mathrm{Aut})}}{4!(L+2)N_L^{(C)}} = \frac{1}{N_L^{(C)}} \sum_{\substack{G \text{ completion} \\ L \text{ loops}}} \frac{1}{|\mathrm{Aut}(G)|} = \left\langle \frac{1}{|\mathrm{Aut}|} \right\rangle. \tag{2.6}$$

From theorem 2 (1), this quantity is expected to converge to unity.

We used `nauty` [34] to compute the automorphism groups of all 1.04 billion primitive completions up to $L = 16$ loops. We verified that the sum of symmetry factors for each loop order coincides, according to eq. (2.5), with the value of $N_L^{(\mathrm{Aut})}$ determined earlier. The ratio eq. (2.6) is reported in table 3, we see that it indeed converges to unity, but considerably slower than the ratio between $N_L^{(\mathrm{Aut})}$ and its asymptotics $\bar{N}_L^{(\mathrm{Aut})}$, the relative difference being $\geq 5\%$ as high as $L \leq 15$.

For future reference, it will also be useful to know the averages

$$\langle |\mathrm{Aut}| \rangle := \frac{1}{N_L^{(C)}} \sum_{\substack{G \text{ completion} \\ L \text{ loops}}} |\mathrm{Aut}(G)|\,, \qquad \langle \ln |\mathrm{Aut}| \rangle := \frac{1}{N_L^{(C)}} \sum_{\substack{G \text{ completion} \\ L \text{ loops}}} \ln |\mathrm{Aut}(G)|\,.$$

Note that, since x^{-1} and $\ln x$ are non-linear functions of x, the relationship between the three averages of symmetry factors is non-trivial, as can be seen in table 3.

JHEP11(2023)160

L	$N_L^{(\mathrm{Aut})}$	$\frac{N_L^{(\mathrm{Aut})}}{\bar{N}_L^{(\mathrm{Aut})}}$	$\left\langle \frac{1}{\lvert\mathrm{Aut}\rvert} \right\rangle$	$\langle \lvert\mathrm{Aut}\rvert \rangle$	$\langle \ln \lvert\mathrm{Aut}\rvert \rangle$	$\frac{\#(\lvert\mathrm{Aut}\rvert=1)}{N_L^{(C)}}$	$\left\langle D^{\mathrm{rel}} \right\rangle$
5	$\frac{31}{2}$	1.494	0.04613095	31.000000	3.25512917	0	0.214286
6	$\frac{529}{6}$	1.201	0.09184027	240.00000	3.29312666	0	0.250000
7	$\frac{2277}{4}$	1.085	0.18824405	15.000000	2.16260435	0	0.349206
8	$\frac{16281}{4}$	1.031	0.34610969	17.551020	1.52008078	0.081633	0.506123
9	$\frac{254633}{8}$	1.006	0.53112276	4.0088106	0.89829287	0.281938	0.676011
10	$\frac{2157349}{8}$	0.995	0.69154313	3.3589365	0.56087794	0.497784	0.805884
11	$\frac{39327755}{16}$	0.990	0.81034314	1.7121991	0.31139120	0.670438	0.902125
12	$\frac{383531565}{16}$	0.990	0.87745449	1.4215362	0.19321715	0.780272	0.950152
13	$\frac{7968073183}{32}$	0.991	0.91534365	1.2316028	0.12832106	0.843603	0.974790
14	$\frac{87837901499}{32}$	0.992	0.93616672	1.1601109	0.09456207	0.879872	0.985926
15	$\frac{6145866829559}{192}$	0.994	0.94869319	1.1209582	0.07484757	0.902197	0.991283
16	$\frac{75602802752227}{192}$	0.995	0.95690877	1.0982303	0.06226253	0.917166	0.993932
17	$\frac{1957218888255863}{384}$	0.997					
18	$\frac{26597510458983139}{384}$	0.998					

Table 3. The columns describe the following: loop order; sum of symmetry factors; ratio with the asymptotics theorem 2, including the $\frac{1}{L+2}$ correction; average symmetry factor; average size of the automorphism group; average logarithm of the size of the automorphism group; proportion of graphs with symmetry factor unity; average relative number of non-isomorphic decompletions.

Apart from the average size of automorphism groups, it is also interesting to know how many graphs have a trivial automorphism group, or symmetry factor unity. The number is reported in table 3, normalized to the total number of graphs. No graphs have trivial symmetry factor for $L \leq 7$, but around half of the graphs have trivial symmetry factor for $L = 10$, and more than 90% for $L \geq 15$.

As discussed in section 2.1, a completion can have between 1 and $(L+2)$ decompletions. Similarly to eq. (2.6), the average number of decompletions per completion can equivalently be expressed as a ratio of the counts of decompletions and completions,

$$\left\langle D^{\mathrm{rel}} \right\rangle := \left\langle \frac{\#(\text{decompletions of } G)}{L_G + 2} \right\rangle = \frac{N_L^{(D)}}{(L+2)N_L^{(C)}}. \tag{2.7}$$

For growing loop number, this ratio is expected to approach unity from below, expressing that for large loop orders, a completion with $L+2$ vertices on average gives rise to almost $(L+2)$ non-isomorphic decompletions, or equivalently, almost all $(L+2)$ decompletions are non-isomorphic.

JHEP11(2023)160

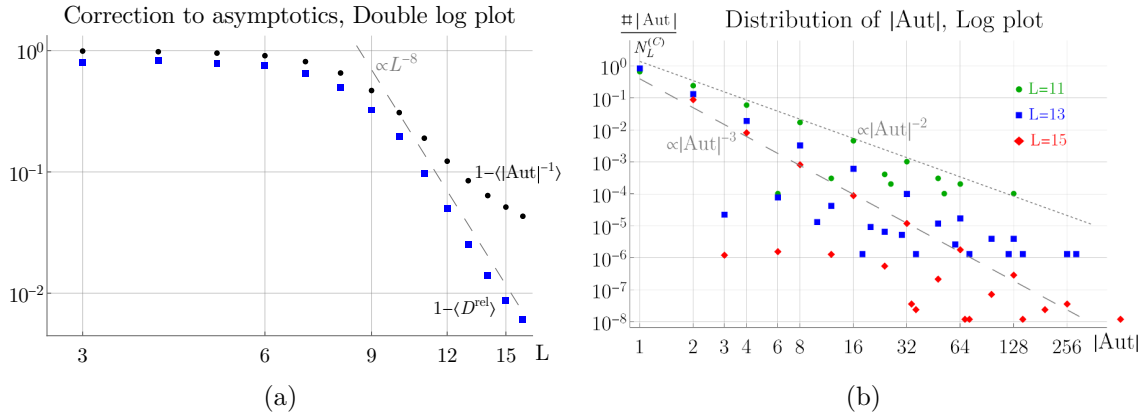


Figure 4. (a) Deviations of $\langle D^{\mathrm{rel}}\rangle$ and $\left\langle \frac{1}{|\mathrm{Aut}|}\right\rangle$ from their asymptotic value unity. Note the abrupt change in slope at $L_{\mathrm{crit}} \approx 9$ loops. (b) Distribution of the size of the automorphism group. The relative prevalence of large groups decays polynomially with growing |Aut|, and the decay becomes steeper as the loop number increases, from $\sim |\mathrm{Aut}|^{-2}$ at $L = 11$ to $\sim |\mathrm{Aut}|^{-3}$ at $L = 15$.

The data in table 3 shows that the convergence of $\left\langle D^{\mathrm{rel}}\right\rangle$ (eq. (2.7)) is significantly faster than the convergence of the average symmetry factor eq. (2.6). For both ratios, we examined the deviation from their asymptotic value unity, see figure 4(a). One might expect a $\frac{1}{L}$ correction, therefore we choose a double logarithmic plot. Interestingly, we do not find such a $\frac{1}{L}$ scaling. Instead, the correction is large below some "critical" loop number $L_{\mathrm{crit}} \approx 9$, and falls off sharply for $L > L_{\mathrm{crit}}$, and the decay is much faster than $\frac{1}{L}$, perhaps even exponential.

Having examined various averages, we also look at the concrete distribution of symmetry factors. In figure 4(b), we show the count of completed graphs for each size of the automorphism group, i.e. the distribution of $|\mathrm{Aut}(G)|$. The larger automorphism groups are less frequent than the smaller ones, the prevalence falls off roughly following a power law, where the power is smaller (the decay steeper) for higher loop orders. Nevertheless, in absolute numbers, there are more graphs with large symmetry group at higher loop order. For example, there are 24 graphs with $|\mathrm{Aut}(G)| = 128$ at 15 loops, but only three at 13 loops.

An analogous behavior is observed for the distribution of D^{rel}, see figure 5(a), but all slopes are inverted since $D^{\mathrm{rel}} \leq 1$ and the limit is $\left|D^{\mathrm{rel}}\right| \to 1$. Remarkably, the distribution shows a local maximum at $D^{\mathrm{rel}} = \frac{1}{2}$, apart from the global maximum at $D^{\mathrm{rel}} = 1$. Moreover, the overall tendency is linear in the log plot figure 5(a), hinting at an exponential decay instead of the polynomial one in figure 4(b). A possible explanation for this is that |Aut| is not bounded, i.e. the slopes in figure 4(b) would be different if we were to normalize |Aut| to the unit interval for each L.

Finally, the relation between $|\mathrm{Aut}(G)|$ and $D^{\mathrm{rel}}(G)$ is shown in figure 5(b). We expect that if a graph has large automorphism group, then many of its decompletions will be isomorphic, so $D^{\mathrm{rel}}(G)$ will be small. This is indeed what happens on average, again the functional dependence is roughly exponential with a notable deviation at $D^{\mathrm{rel}} = \frac{1}{2}$ and significant fluctuations.

JHEP11(2023)160

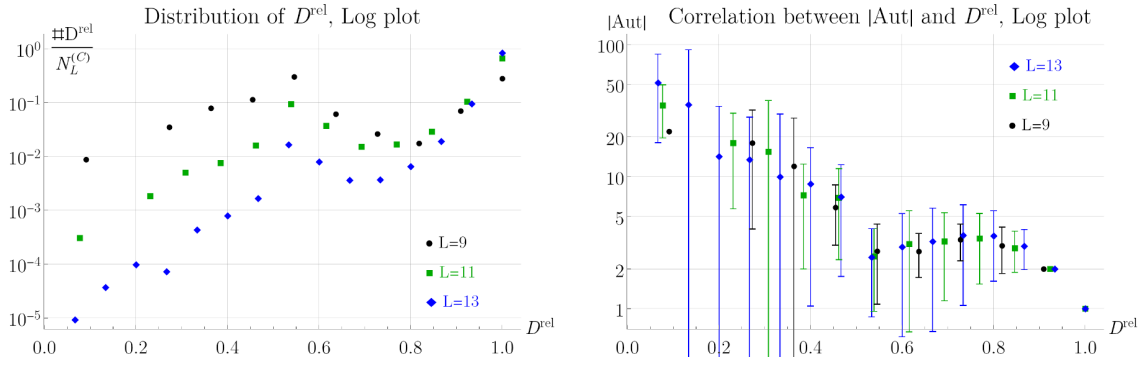


Figure 5. (a) Number of graphs for a given D^{rel}, scaled to the total number of graphs of that loop order. Most graphs have $D^{\text{rel}}(G) = 1$ as expected, but note the second local maximum at $D^{\text{rel}} = \frac{1}{2}$.

(b) Correlation between $D^{\text{rel}}(G)$ and $|\text{Aut}(G)|$. Error bars correspond to the standard deviation. The relation is approximately exponential (linear in the log plot), with notable fluctuations.

L	$\langle \lvert\text{Aut}\rvert \rangle_{\text{dec}}$	$\langle \ln \lvert\text{Aut}\rvert \rangle_{\text{dec}}$	$\left\langle \frac{1}{\lvert\text{Aut}\rvert} \right\rangle_{\text{dec}}$	$\left\langle D^{\text{rel}}_{\text{pla}} \right\rangle$	$\langle \lvert\text{Aut}\rvert \rangle_{\text{pla}}$	$\langle \ln \lvert\text{Aut}\rvert \rangle_{\text{pla}}$	$\left\langle \frac{1}{\lvert\text{Aut}\rvert} \right\rangle_{\text{pla}}$
5	10.000000	1.98354752	0.2152778	0.14285714	9.0000000	1.73286795	0.2812500
6	17.300000	1.43861918	0.3673611	0.12500000	3.0000000	0.91286964	0.4833333
7	3.6818181	0.86643398	0.5390625	0.15079365	4.4210526	0.91203576	0.5476974
8	2.4112903	0.55107975	0.6838458	0.11836735	1.9310345	0.39437684	0.7640086
9	1.8329383	0.35800954	0.7856715	0.09411294	1.9148936	0.34142627	0.7981161
10	1.4272949	0.22310430	0.8581177	0.05416051	1.3034091	0.16068412	0.8964489
11	1.2842496	0.15583254	0.8982601	0.02866615	1.3066520	0.13870596	0.9116431
12	1.1953716	0.11434124	0.9234890	0.01282297	1.1508633	0.08456242	0.9435972
13	1.1473675	0.08968771	0.9390163	0.00530868	1.1300771	0.07254937	0.9512887
14	1.1172909	0.07338441	0.9495301	0.00201826	1.0919890	0.05557029	0.9620677
15	1.0972817	0.06198054	0.9570355	0.00072221	1.0823823	0.04984923	0.9658697
16	1.0827565	0.05342383	0.9627510	0.00024319	1.0692237	0.04316497	0.9702493

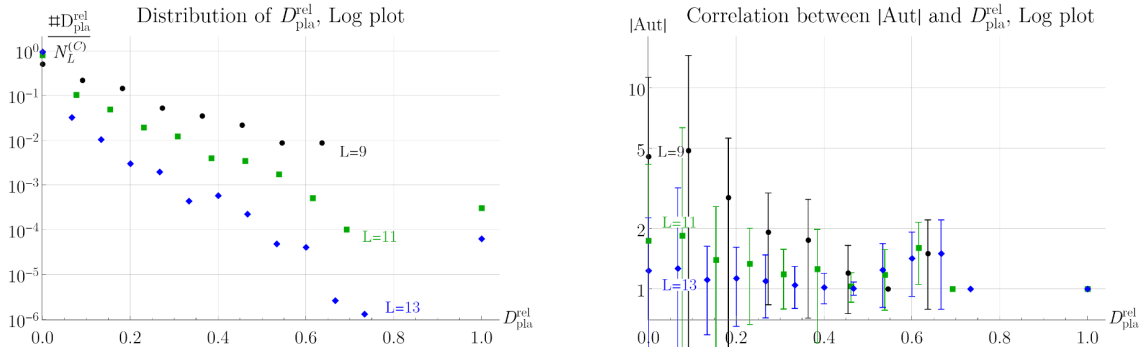
Table 4. Averages of automorphisms of decompletions according to eq. (2.8). The columns describe the following: loop order; 3 columns referring to all decompletions; 4 columns referring to planar decompletions.

2.4 Planar graphs

While the set of primitive 1PI Feynman graphs can be reasonably well approximated by the set of random 4-regular graphs, the mathematical literature on the set of *planar* 4-regular graphs is sparse.

Planar 2-connected graphs are known to have exponentially large automorphism groups in the asymptotic limit [53]. Conversely, our empirical data in table 4 shows that for ϕ^4-theory, the average symmetry factor of planar decompletions is not significantly different from the one of the full set of decompletions; in fact, it is slightly closer to unity than in the non-planar case. The set of all 2-connected planar graphs is therefore not a good model for our case since planar decompletions of ϕ^4-theory constitute only a very small subset.

JHEP11(2023)160



(a) (b)

Figure 6. Plot analogous to figure 5, but for *planar* decompletions (eq. (2.9)). Note that the slope of (a) is reversed compared to figure 5(a), and that the correlation with |Aut| shown in (b) is much weaker than in figure 5(b).

The number of decompletions per completion depends on the particular graph, consequenly, the averages taken over completions, reported in table 3, do not necessarily coincide with the averages taken over decompletions. With $\langle\cdots\rangle_{\text{dec}}$, we indicate that an average is taken over decompletions, and with $\langle\cdots\rangle_{\text{pla}}$ over planar decompletions, as in

$$\langle|\text{Aut}|\rangle_{\text{dec}} := \frac{1}{N_L^{(D)}} \sum_{\substack{\text{decompletions } g \\ L \text{ loops}}} |\text{Aut}(g)|\,, \tag{2.8}$$

$$\langle|\text{Aut}|\rangle_{\text{pla}} := \frac{1}{\#\text{planar decompletions}} \sum_{\substack{\text{planar dec. } g \\ L \text{ loops}}} |\text{Aut}(g)|\,.$$

Recall that even if a completion is non-planar, some or all of its decompletions can be planar, as happens for the example in figure 3. Averages over decompletions are reported in table 4, note that their concrete numerical values differ from table 3 in particular for low loop orders, but the qualitative behavior for large L is similar to that of the averages over completions.

In analogy to eq. (2.7), we define the relative count of planar decompletions,

$$D^{\text{rel}}_{\text{pla}}(G) := \frac{\#(\text{planar decompletions of } G)}{L_G + 2} \leq D^{\text{rel}}(G). \tag{2.9}$$

This quantity converges to zero as $L \to \infty$. As seen in figure 6(a), the distribution indeed is concentrated near $D^{\text{rel}}_{\text{pla}} = 0$. Note that $D^{\text{rel}}_{\text{pla}}$ does not show a local maximum at $\frac{1}{2}$ as does D^{rel} (figure 5(a)). Moreover, the correlation between $D^{\text{rel}}_{\text{pla}}$ and |Aut|, plotted in figure 6(b), is almost negligible, again in contrast to the behavior of D^{rel} in figure 5(b).

3 Period symmetries and numerical integration

3.1 Period symmetries

By a *symmetry of the period*, we mean an operation $G_1 \mapsto G_2$ such that $\mathcal{P}(G_1) = \mathcal{P}(G_2)$ and both G_1 and G_2 are primitive decompletions of ϕ^4-theory and not isomorphic to each

JHEP11(2023)160

other. Several symmetries of the period are known, we briefly list them below. A more precise description and illustrations can be found in the literature cited or in [54].

Completion [11]. If two decompletions G_1 and G_2 have the same completion, then $\mathcal{P}(G_1) = \mathcal{P}(G_2)$. In view of this identity, we define $\mathcal{P}(G)$, for a completion G, to mean the period of any, and hence all, of its decompletions.

Product [11]. Consider two completions G_1 and G_2, each of which has at least one triangle $t_1 \in G_1$, $t_2 \in G_2$. Removing the edges of the triangles produces three 2-valent vertices in each graph. Merge those vertices pairwise to obtain a graph $G_1 \times G_2$, called the product, which has $\mathcal{P}(G_1 \times G_2) = \mathcal{P}(G_1) \cdot \mathcal{P}(G_2)$. Conversely, every completed graph which has a 3-vertex cut is a product of two smaller graphs.

Fourier [55]. If the decompletion G_2 is isomorphic to the planar dual $\tilde{G}_1$ of a decompletion G_1, then $\mathcal{P}(G_2) = \mathcal{P}(G_1)$. This symmetry expresses a Fourier transform of the Feynman integral.

Extended Fourier [11]. In full generality, planar duality operates on integrands non-unit edge weights, where a negative weight corresponds to a factor in the numerator, see e.g. [31, 56]. Fourier symmetry holds even in this more general setting, where the planar dual $\tilde{G}_1$ of a decompletion might not be a ϕ^4 decompletion itself. One may still construct the completion $\tilde{G}_1 \cup \{w\}$, and pick a vertex $v \neq w$ to construct the decompletion G_2 of this graph. Even if G_2 still is not a decompletion of ϕ^4 theory, one might now be able to reach a valid decompletion $\tilde{G}_2$ by computing the planar dual of G_2, and iterating this procedure.

Schnetz Twist [11]. Consider a 4-vertex cut of a completion $G_1 = G_A \cup \{v_1, v_2, v_3, v_4\} \cup G_B$. Pick two of the vertices $\{v_1, v_2\}$, and, in G_B, change those edges going to v_1 to go to v_2, and those originally going to v_2 to go to v_1. Repeat with the remaining two vertices $\{v_3, v_4\}$. One might need to introduce integral numerators (= edges with negative weight) to make the resulting graph 4-regular, see e.g. [31]. If the resulting graph is 4-regular without such negative weights, it is a ϕ^4 period which coincides with $\mathcal{P}(G_1)$.

Fourier split [29]. Consider a 4-vertex cut of a completion $G_1 = G_A \cup \{v_1, v_2, v_3, v_4\} \cup G_B$. Remove v_4 and split the remaining graph at the three vertices. Introduce three new edges $\{e_1, e_2, e_3\}$ to G_B, forming a triangle at the three cut vertices such that e_i is not incident to v_i. If the resulting graph G'_B is planar, take its dual. The triangle becomes a star $\{e'_1, e'_2, e'_3\}$. Remove the central vertex of the star, and let v'_i be the unique remaining vertex incident to e'_i. Join these three vertices pairwise to the three cut vertices of G_A, where v'_i is joined to v_i. If the resulting graph is a decompletion of a 4-regular graph G_2 then $\mathcal{P}(G_1) = \mathcal{P}(G_2)$. The Fourier split can be promoted to an "extended" version similarly to the ordinary Fourier symmetry.

All the above symmetries are known, and all except the "extended" versions had been implemented in previous work. Our own contribution is threefold:

JHEP11(2023)160

1. We wrote a new `C++` implementation of all known symmetries, using `nauty` [34] to check for graph isomorphisms and the `boost graph library` [57] for planar duals. Being independent of all previous implementations, this serves as a check for the previously obtained counts of symmetries.

2. We include a version of the "extended" Fourier symmetry. Concretely, we compute the planar dual $\tilde{G}_1$ of a decompletion, and if $\tilde{G}_1$ has exactly one vertex v of valence greater than 4, we complete $\tilde{G}_1$ and remove v, reaching a new graph $\tilde{G}_2$. We then compute its dual G_2. If G_2 is a decompletion of ϕ^4 theory, we are done. If G_2 again has one vertex of too high degree, we iterate this procedure until either we either reach a decompletion of ϕ^4 theory, or we are back to a graph that we already know, or we have visited more than 1000 intermediate graphs without finding a valid ϕ^4 decompletion (the latter is rare and happened only for $L \geq 14$ loops). We use a similar iteration for the Fourier split symmetry, with Fourier splits instead of planar duals for the intermediate non-ϕ^4 decompletions.

   Our implementation is not the most general conceivable extension of the period symmetries, for three reasons: (1) we do not allow for intermediate graphs to have more than one vertex of valence > 4, nor to have non-trivial edge weights, (2) we do not consider Twist symmetries for these intermediate non-ϕ^4 graphs, and (3) we do not consider the possibility of having a Fourier-like symmetry for non-planar graphs, where intermediate objects are matroids.

3. We compute the symmetries for higher loop numbers than previous authors.

Note that the computational effort for the individual symmetries is vastly different: completions and Fourier transforms are very fast. Extended Fourier, Product, and Twist take considerably longer because they require trying out many possible operations. The extended Fourier split is slower yet because it typically requires to construct thousands of transformation candidates for a single input graph, each of which involves non-trivial operations such as planarity tests or finding and merging particular vertices. Our algorithm heavily relies on caching of the graphs that have already been visited, this entails that every intermediate step needs to transform the resulting graph to `nauty`'s standard form in order to compare with the cached graphs, and it limits the performance that can be gained by parallelization as the caches need to be synchronized.

3.2 Invariants

Over the last years, there has been effort to find quantities which share the symmetries of the period, but are easier to compute or more directly related to properties of the underlying graphs. Notable examples are the c_2 invariant [29, 58] and the permanent [26, 28].

The *Hepp bound* [54] is defined as the tropical approximation of the period integral eq. (1.4). In that approximation, the nested integrals can be solved analytically and the computation of the Hepp bound reduces to a purely combinatoric operation on the subgraphs of the graph in question. The Hepp bound $\mathcal{H}(G)$ is a rational number. We wrote another `C++` program to compute the Hepp bound approximately (to 64-bit floating point

JHEP11(2023)160

precision) using the flag formula [54]:

$$\mathcal{H}(G) = \sum_{\substack{1\text{ PI }\gamma\subset G\\ L_\gamma = L_G - 1}} \frac{\mathcal{H}(\gamma)}{\omega(\gamma)} = \sum_{\text{hyperplane }\gamma\subset G} \frac{\mathcal{H}(\gamma)}{\omega(\gamma)}. \tag{3.1}$$

Here, $\omega(\gamma)$ is the superficial degree of divergence, a subgraph is 1 PI (bridgeless in graph theory language) if it is 2-edge connected, and a *hyperplane* is an induced subgraph $\gamma \subset G$ such that both γ and the subgraph γ' induced from the vertices of $(G \setminus \gamma)$ are connected. Note that even if eq. (3.1) is a sum over graphs of lower loop number, these subgraphs are not superficially log divergent, therefore the subgraphs do not coincide with the graphs in our period samples at lower loop number.

The linear coefficient of the Martin polynomial (eq. (5.3)) of the graph has been identified as an invariant only recently [59]. This *Martin invariant* is a sequence of invariants which are obtained from the graph by duplicating edges. This sequence is a generalization of earlier period invariants. It is conjectured [59] that the full sequences of Martin invariants of two graphs concide if and only if the periods coincide. In the present work, we only compute the first Martin invariant $M^{[1]}$ of the sequence ("first Martin" in table 5). We verified for all graphs in our samples that it coincides whenever two graphs are symmetric. Conversely, the first Martin invariant alone is not sufficient to indicate the opposite direction, there are many pairs of graphs where the first Martin invariant coincides but the periods do not.

3.3 Counts of symmetries

We computed the symmetries as described for all completions with $L \leq 14$ loops. For $L = 15$, we computed Schnetz Twists and extended Fourier symmetries, but not the Fourier splits. Due to the product- and completion symmetries, the most interesting counts are the numbers of symmetries for 3-vertex irreducible completions, reported in table 5. Quite often, one relation between two graphs can be obtained by more than one symmetry, this is why the counts of the individual symmetries in table 5 do not add up to the total number of independent periods.

Within a symmetry orbit, not every pair of graphs is directly related via a known symmetry, but some only via intermediate graphs.

For comparison, we list the number of distinct Hepp bounds (eq. (3.1)), and the number of Hepp bounds which coincide in cases where the period is not known to be symmetric. The numbers of Hepp bounds in table 5 are not entirely reliable for high loop orders because we computed the Hepp bounds numericlaly to floating-point precision and counted them as equal if their relative difference was smaller than 10^{-14}. Our result for $L \leq 11$ coincides with [54] and we expect the results for $L \in \{12, 13\}$ to be at least very close to the true numbers.

We have verified in every single case that those graphs which are expected to be symmetric have the same Hepp bound (within rounding errors). Further, in many cases we numerically computed the periods of graphs which are known to be symmetric (see table 1). In these cases, we compared the estimated accuracy of numerical results to the

JHEP11(2023)160

L	Twists	F	Twists & Fourier	Fourier Split	only Split	independ. periods	Hepp	dupl. Hepp	first Martin
5	0	0	0	0	0	1	1	0	1
6	0	0	0	0	0	4	4	0	4
7	1	2	2	2	0	9	9	0	9
8	9	3	10	9	0	31	29	2	25
9	48	14	55	53	1	134	129	5	100
10	336	21	350	334	13	819	776	42	409
11	2387	43	2420	2276	70	6197	6030	158	1622
12	18680	60	18728	17040	280	55196	54552	618	5799
13	155547	90	155630	138164	1077	543535	541196	2246	19278
14	1386809	117	1386919	1209177	4581	5769143			61757
15	13153096	184	13153272			65117118			192113

Table 5. Prevalence of symmetries for 3-vertex irreducible completions. "One identity" means a reduction of the number of independent graphs by one. The columns describe the following: loop order of decompletion; total number of... Schnetz Twist; extended Fourier identities; Twist and extended Fourier combined; extended Fourier splits; identities which can only be explained as (extended) Split, not as combinations of non-Split symmetries; upper bound of independent 3-vertex irreducible periods, exploiting known symmetries as described in the text; numerically distinct Hepp bounds of irreducible graphs at machine precision; unique Hepp bounds coinciding for sets of graphs not related by known symmetries; distinct Martin invariants $M^{[1]}$.

observed differences for symmetric graphs. We found that the difference are approximately normal distributed with a standard deviation close to the expected uncertainty. This indicates that the expected uncertainty is a good approximation of the deviation between the numerical result and the true period. Moreover, we take these findings as evidence that our implementation of the symmetries does not label any graphs as symmetric which really are not.

The number of independent periods has been reported in the literature for $L \leq 11$ loops. We find our results to be consistent with (i.e. smaller or equal than) [11] when excluding the Fourier split, and with [29] when including Fourier split. In fact, we find agreement if we do not use the *extended* version of Fourier symmetry and Fourier split.

Since we take into account an "extended" version of the Fourier symmetry, we find more graphs related by Fourier identities than previously authors. The number of Fourier splits grows even more by including the extended Fourier symmetry, so that the number of Fourier splits which can not be explained by other identities ("only Split" in table 5) comes out larger than in [29]. Moreover, the number of coinciding Hepp bound without underlying known symmetry, reported as "dupl. Hepp" in table 5, is slightly lower than the numbers given in [54, table 3] for $L \leq 10$. We note in passing that the circulant graphs (of which the zigzag eq. (4.11) is the most prominent one) for $6 \leq L \leq 18$ do not have any non-trivial period symmetries.

Our numbers of table 5 extend the previously known results both in terms of higher loop orders, and in terms of systematically including extended Fourier transforms. The

JHEP11(2023)160

overall picture is that Fourier symmetries are rare (in accordance with planar graphs being rare, see table 2), that the symmetries obtained by Twist show a large overlap with the ones obtained by Fourier split, and that there are graphs where the Hepp bound coincides even if they are not known to be symmetric.

3.4 Numerical integration

For the numerical integration in this work, we have used the reference implementation of the tropical Monte Carlo quadrature for Feynman integrals developed recently by Borinsky [33].[1] We slightly optimized this program for the present case of massless log-divergent graphs in $D = 4$ spacetime dimensions, by removing computation steps which are only required for more general Feynman integrals. Further, we added an automatic adjustment of the number of sample points to meet the target precision. Note that the original algorithm has recently been generalized to a much larger class of amplitudes beyond periods [35].[2]

The scale of the project requires that many intermediate steps be automated. In order to handle large tables of graphs or results, most of our programs directly operate on bzip2-compressed text files, where graphs are described by their Nickel index (eq. (2.3)). We found this format to be comparable in size to binary formats such as the native graph files of `nauty` [34], but having the advantage that files can be inspected or edited in an ordinary text editor after decompression.

Conceptually, the workflow looks as follows:

1. We use `nauty` to generate a list of completed graphs to be integrated, and filter it for 1PI primitive ones as described in sections 1.3 and 2.2.
2. We split this list across many computers, generate a decompletion of each graph, and integrate it with the tropical Feynman quadrature algorithm where the number of sampling points is adjusted automatically to reach a relative target precision.
3. We collect the results, re-do missing graphs, sort, and check if symmetries are fulfilled as described in section 3.3.
4. We exploit the known symmetries to improve the numerical accuracy.
5. We compute symmetry factors (section 2.1), $O(N)$-factors (section 5) and various other properties (section 6), and write the results to one file per loop order.

All steps are implemented as `C++` programs, using `nauty` version 2.7 [34] and the `boost graph library` version 1.82 [57]. The first item of the workflow has been discussed in section 2. Points 3 and 4 use pre-computed tables of symmetries (section 3.1). In step 4, the program combines all graphs of the same loop order which are known to be symmetric (and whose numerical results pass step 3) by computing an average weighted by the individual uncertainties, see appendix D. This program is capable of utilizing duplicate computations of the same graph without wasting previous results. For example, we can gradually increase

[1] The implementation used in the present work is available from https://github.com/michibo/tropical-feynman-quadrature/.

[2] Available from https://github.com/michibo/feyntrop.

JHEP11(2023)160

the accuracy of a selected set of graphs by integrating them multiple times, and the resulting improvement will consistently propagate through the whole data set of that loop order.

Apart from checking the symmetries, we compared our numerical results to the (analytical) data in [17] and found agreement, where the differences have a standard deviation that matches our estimate.

For the non-complete samples at $L \geq 14$, we furthermore used the symmetries to infer the periods of a large number of additional graphs which were not computed in the original samples. The numbers are reported in table 1, but we do not use those graphs for the statistical evaluation in the subsequent chapters since they would systematically distort the random samples in favor of graphs with larger symmetry orbits.

The numerical integrations ran on multiple machines with significantly different performance, most notably on servers of the institute of Mathematics and a cluster of the institute of Physics at Humboldt-Universität zu Berlin. A smaller batch was done on the servers of the Faculty of Mathematics at University of Waterloo. The total time spent on integration was well above 2 million CPU-core hours. The other tasks such as generating graphs, constructing period symmetries, or merging the results, were considerably faster, but still had run times of several days to a few weeks for the largest data sets.

4 Distribution of periods

4.1 Mean and moments

Unless specified otherwise, we define a mean to extend over completions of a given loop order, normalized to the number of completions from eq. (2.4),

$$\langle \mathcal{P} \rangle := \frac{1}{N_L^{(C)}} \sum_{\substack{\text{completions } G \\ L \text{ loops}}} \mathcal{P}(G). \tag{4.1}$$

If our sample does not include all completions, the mean is computed by summing the graphs in the sample, and using N, the number of graphs in the sample, instead of $N_L^{(C)}$.

The samples generated by `genrang` are not distributed uniformly, but proportionally to their symmetry factor. This has been corrected for, as described in appendix A. The influence of the non-uniform distribution on the sample mean $\langle \mathcal{P} \rangle$ (eq. (4.1)) is smaller than one might expect. Firstly, at $L \geq 14$ most graphs have trivial symmetry factor (section 3.3), and secondly, there is no strong correlation between the symmetry factor and the period (section 6.1). For example, ignoring the effect and plainly summing over all graphs, the sample $14s$ gives rise to $\langle \mathcal{P} \rangle = (553.2 \pm 1.0) \cdot 10^2$, which is 4.5% smaller than the true value (table 6). Adding the duplicated periods, for $14s$, the sample size increases from $N = 216189$ (the true sample size reported in table 1) to $N_{\text{corr}} = 230884$, but the relevant factor $\frac{1}{\sqrt{N}}$ for estimating the uncertainty decreases by only 3%.

The mean of the period grows quickly with the loop order, see table 6. In order to compare the distributions at different loop orders, we scale the results with respect to

JHEP11(2023)160

the mean,

$$p(G) = \frac{1}{\langle \mathcal{P} \rangle} \mathcal{P}(G). \tag{4.2}$$

By construction, $\langle p \rangle = 1$, and deviations from the mean can be characterized by the central moments

$$C_j := \left\langle (p-1)^j \right\rangle, \qquad c_k := C_k \cdot C_2^{-\frac{k}{2}}, \quad k \geq 2. \tag{4.3}$$

The standardized moments c_3 and c_4 are known as skewness and kurtosis, the latter is a measure for how "heavy-tailed" the distribution is. All moments can be converted back to the distribution prior to the scaling eq. (4.2) using the value $\langle \mathcal{P} \rangle$. In particular, the standard deviation of the un-scaled sample is

$$\sigma = \langle \mathcal{P} \rangle \cdot \sqrt{C_2}. \tag{4.4}$$

An alternative measure for deviations from the mean are quantile boundaries. In table 6, we give the boundaries $[1\%, 99\%]$, that is, the minimum and maximum of a subset that includes all periods except for the smallest and the largest 1%.

There are two different sources of uncertainty for our results. Firstly, the uncertainty of the numerical integration, stated in table 1, give rise to an uncertainty Δ_{num}. Secondly, for the non-complete samples, there is a sampling uncertainty Δ_{samp} which expresses that the empirical mean of a sample can deviate from the mean of the underlying population. Details are discussed in appendix D. We find that for all non-complete samples, even if they contain 100000s of graphs, the statistical uncertainty by far dominates the numerical one, as evidenced in table 6.

For the perturbation series in physics (eq. (1.1)), we need the sum over decompletions, weighted by their symmetry factors. Owing to eq. (2.2), this sum can equivalently be written as a sum over completed graphs, where we omit the trivial factor $4!(L_G + 2)$ in order to facilitate comparisons with eq. (4.1):

$$\left\langle \frac{\mathcal{P}}{|\text{Aut}|} \right\rangle := \frac{1}{N_L^{(C)}} \sum_{\substack{\text{completions } G \\ L \text{ loops}}} \frac{\mathcal{P}(G)}{|\text{Aut}(G)|} = \frac{1}{(L+2)N_L^{(C)}} \sum_{\substack{\text{decompletions } g \\ L \text{ loops}}} \frac{\mathcal{P}(g)}{|\text{Aut}(g)|}. \tag{4.5}$$

For the "s"-samples, the correction eq. (A.3) cancels the symmetry factors from eq. (4.5) as should be expected: a sum of $\frac{\mathcal{P}}{\text{Aut}}$ is equivalent to a sum of $\mathcal{P}$, but distributed proportionally to $\frac{1}{\text{Aut}}$.

Analogously to eq. (4.2), we normalize the distribution of symmetry-factor-weighted periods with respect to the mean eq. (4.5) and examine the quantiles and moments of the resulting distribution. The results are reported in table 7. The numerical- and sampling accuracy behave completely analogously to table 6, we have omitted them in table 7 and instead included the median, i.e. the 50% quantile. Note the very small 1% quantile at $L = 6$ and $L = 8$ in table 7 caused by graphs with very large symmetry factors. For $L \geq 10$, those graphs lie outside the 1% quantile boundary. The minimum symmetry-factor weighted period at 10 loops is 1.982 and at 12 loops it is 4.483.

JHEP11(2023)160

L	$\langle\mathcal{P}\rangle$	Δ_{num}	Δ_{samp}	$[1\%\ ;\ \ 99\%]$	C_2	c_3	c_4
5	53.8013 ± 0.0003	$3 \cdot 10^{-4}$	0	$52.02; 55.58$	0.001	0.000	1.000
6	125.8743 ± 0.0003	$3 \cdot 10^{-4}$	0	$71.51; 168.3$	0.072	−0.399	1.917
7	315.875 ± 0.0007	$7 \cdot 10^{-4}$	0	$183.0; 527.7$	0.089	0.572	2.613
8	727.559 ± 0.001	$1 \cdot 10^{-3}$	0	$311.0; 1716$	0.147	0.982	4.374
9	1654.858 ± 0.003	$3 \cdot 10^{-3}$	0	$671.0; 3620$	0.185	1.402	6.940
10	3581.796 ± 0.003	$3 \cdot 10^{-3}$	0	$1346; 8757$	0.241	1.629	8.917
11	7516.003 ± 0.003	$3 \cdot 10^{-3}$	0	$(2.617; 21.17) \cdot 10^3$	0.311	1.838	10.46
12	15235.118 ± 0.003	$3 \cdot 10^{-3}$	0	$(4.916; 50.27) \cdot 10^3$	0.402	2.124	12.25
13	30023.445 ± 0.007	$7 \cdot 10^{-3}$	0	$(8.992; 111.8) \cdot 10^3$	0.511	2.512	15.12
13s	29963 ± 298	$7 \cdot 10^{-2}$	$3 \cdot 10^2$	$(8.947; 104.1) \cdot 10^3$	0.508	2.345	11.14
13⋆	30091 ± 217	$6 \cdot 10^{-2}$	$3 \cdot 10^2$	$(9.056; 112.6) \cdot 10^3$	0.520	2.390	11.04
14s	$(578.1 \pm 1.0) \cdot 10^2$	$4 \cdot 10^{-2}$	$1 \cdot 10^2$	$(1.605; 24.44) \cdot 10^4$	0.637	2.908	16.30
14⋆	$(577.9 \pm 1.4) \cdot 10^2$	$5 \cdot 10^{-2}$	$2 \cdot 10^2$	$(1.613; 24.44) \cdot 10^4$	0.635	2.880	15.58
15s	$(1080.3 \pm 2.9) \cdot 10^2$	$9 \cdot 10^{-2}$	$3 \cdot 10^2$	$(2.847; 50.08) \cdot 10^4$	0.774	3.660	25.73
15⋆	$(1085.8 \pm 3.8) \cdot 10^2$	$2 \cdot 10^{-1}$	$4 \cdot 10^2$	$(2.854; 50.22) \cdot 10^4$	0.790	3.560	24.23
16s	$(196.8 \pm 1.8) \cdot 10^3$	$6 \cdot 10^{-1}$	$2 \cdot 10^3$	$(4.997; 94.29) \cdot 10^4$	0.853	3.742	25.02
17s	$(35.8 \pm 4.9) \cdot 10^4$	$2 \cdot 10^0$	$5 \cdot 10^3$	$(8.327; 181.0) \cdot 10^4$	1.002	4.238	30.54
18s	$(65.4 \pm 2.3) \cdot 10^4$	$4 \cdot 10^0$	$3 \cdot 10^4$	$(1.588; 35.37) \cdot 10^5$	1.122	4.373	30.15

Table 6. Mean of the period (eq. (4.1)) with combined numerical and sampling uncertainty. The exact values [18] for $L \in \{5, 6, 7\}$ are, to 10 digits, $\{53.80156133,\ 125.8743179,\ 315.8747617\}$. The columns describe the following: numerical uncertainty; sampling uncertainty; 1% and 99% quantiles; second central moment of the normalized distribution; skewness of the normalized distribution; kurtosis of normalized distribution.

The data in tables 6 and 7 shows that the average period grows significantly with loop order. Moreover, the normalized variance C_2 of $\langle\mathcal{P}\rangle$ grows by more than one order of magnitude between $L = 6$ and $L = 18s$. For both $\langle\mathcal{P}\rangle$ and $\left\langle \frac{\mathcal{P}}{|\text{Aut}|} \right\rangle$, skewness and kurtosis grow at a similar rate. Recall that these central moments refer to the normalized distribution eq. (4.2), the observed growth is not caused by the growth of $\langle\mathcal{P}\rangle$. Similarly, due to the normalization eq. (4.3), the growth of c_j is not explained by a growing standard deviation, but conversely, we find that the higher moments grow *faster* than what should be expected from the growth of C_2. This leads us to the conclusion that in the limit $L \to \infty$, the higher moments $C_{j\geq 2}$ are very likely infinite.

A similar growth, but not shown in the tables, is observed for the non-central moments $\langle p^j \rangle$ and the cumulants $\langle e^{jp} \rangle$. To examine the growth of the central moments C_j more systematically, we plot them as a function of L in figure 7. Apparently, they can be

JHEP11(2023)160

L	$\left\langle \frac{\mathcal{P}}{\lvert\mathrm{Aut}\rvert} \right\rangle$	50%	[1% ; 99%]	C_2	c_3	c_4
5	2.52702 ± 0.00002	1.0837	1.084; 3.970	0.3262	0.000	1.000
6	12.56811 ± 0.00005	10.521	0.062; 33.06	0.7782	0.945	2.674
7	61.3302 ± 0.00015	31.845	3.538; 190.4	0.9729	0.966	2.413
8	257.7329 ± 0.0005	211.48	0.972; 966.8	0.7060	1.180	4.164
9	871.8986 ± 0.0017	690.62	27.52; 2631	0.5265	0.857	3.045
10	2422.493 ± 0.003	2181.7	76.30; 7351	0.4158	0.936	3.951
11	5880.015 ± 0.003	5119.3	609.8; 17770	0.3699	1.362	5.750
12	12834.033 ± 0.003	10701	$(1.865; 41.94) \cdot 10^3$	0.4019	1.905	8.713
13	26320.414 ± 0.007	21098	$(5.250; 95.65) \cdot 10^3$	0.4718	2.439	12.69
13s	$(264.9 \pm 1.9) \cdot 10^2$	20618	$(8.241; 95.93) \cdot 10^3$	0.4860	2.501	12.92
13⋆	$(263.2 \pm 1.9) \cdot 10^2$	20952	$(5.228; 97.27) \cdot 10^3$	0.4838	2.513	13.03
14s	$(518.0 \pm 0.9) \cdot 10^2$	$3.911 \cdot 10^4$	$(15.03; 212.7) \cdot 10^3$	0.5975	2.959	17.13
14⋆	$(517.3 \pm 1.2) \cdot 10^2$	$4.015 \cdot 10^4$	$(1.191; 20.69) \cdot 10^4$	0.5622	2.904	16.71
15s	$(981.7 \pm 2.7) \cdot 10^2$	$7.252 \cdot 10^4$	$(2.688; 44.13) \cdot 10^4$	0.7186	3.642	26.07
15⋆	$(985.7 \pm 3.2) \cdot 10^2$	$7.412 \cdot 10^4$	$(24.00; 441.9) \cdot 10^3$	0.6943	3.677	28.30
16s	$(181.8 \pm 1.7) \cdot 10^3$	$1.307 \cdot 10^5$	$(4.756; 86.66) \cdot 10^4$	0.8212	3.877	27.55
17s	$(334.9 \pm 4.6) \cdot 10^3$	$2.365 \cdot 10^5$	$(8.060; 171.0) \cdot 10^4$	0.9593	4.257	31.08
18s	$(60.3 \pm 2.1) \cdot 10^4$	$4.156 \cdot 10^5$	$(1.531; 29.59) \cdot 10^5$	1.047	4.524	33.48

Table 7. Analogous data as table 6, but for the periods weighted with symmetry factors. "50%" denotes the median. The results do not include the overall factor $4!(L+2)$. The exact values [18] for $L \in \{5, 6, 7\}$ are, to 10 digits, $\{2.527040439,\ 12.56808867,\ 61.33012120\}$.

approximated by a power law

$$C_j \approx e^{a_j} \cdot \left(\frac{L}{8}\right)^{b_j} . \tag{4.6}$$

The offset 8 is motivated by the approximate intersection observed in figure 7. Note that this value is close to $L_{\mathrm{crit}} \approx 9$ of figure 4(a). The empirical values of a_j, b_j are shown in table 8. The parameter a_j is approximately constant for growing j, while b_j increases approximately linearly.

The presence of symmetry factors has a notable influence on the growth parameters, see table 8, but they do not change the overall picture that all higher moments seem to diverge as $L \to \infty$. The crucial property is *not* that the C_j grow with growing j, for a fixed L, but conversely, that every $C_{j>1}$, for fixed j, tends to infinity as L grows.

The overall very large standard deviations (and higher moments) imply that the sampling uncertainty (see appendix D) is high even for large samples. We discuss a concrete example for our $L = 13$ data set in appendix C.

JHEP11(2023)160

		C_2	C_3	C_4	C_5	C_6
$\mathcal{P}$	a_j	−2.00	−3.05	−2.50	−2.59	−2.24
	b_j	2.69	6.07	7.91	12.9	17.8
$\frac{\mathcal{P}}{\lvert\mathrm{Aut}\rvert}$	a_j	−1.54	−2.81	−2.47	−2.18	−2.15
	b_j	1.87	5.51	7.65	9.21	12.5

Table 8. Growth parameters of the central moments C_j according to eq. (4.6), as shown in figure 7.

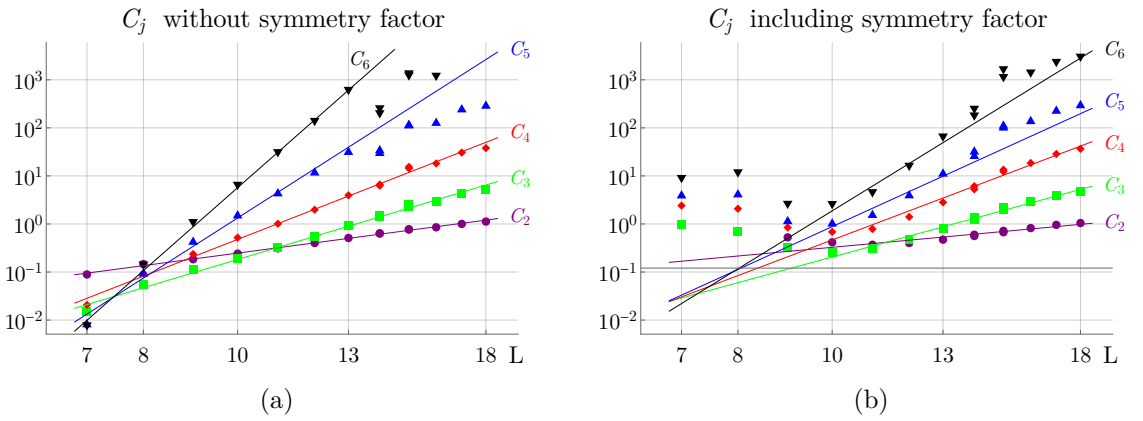


Figure 7. Dependence of the central moments C_j on the loop order L. The lines are eq. (4.6) with coefficients in table 8. For $L \geq 14$, our samples are incomplete and their observed higher moments C_j are increasingly unreliable for growing j. (a) For periods without symmetry factor, the growth is well described by the fit functions. (b) Including the symmetry factor results in a significant deviation for $L \leq L_{\text{crit}} \approx 9$.

4.2 Mean of decompletions and planar graphs

So far, we have normalized all means with respect to the number of completions. Alternatively, we can consider the mean of decompletions. Using eqs. (2.2), (2.7) and (4.5), we define

$$\left\langle \frac{\mathcal{P}}{|\mathrm{Aut}|} \right\rangle_{\text{dec}} := \frac{1}{N_L^{(D)}} \sum_{\substack{\text{decompletions } g \\ L \text{ loops}}} \frac{\mathcal{P}(g)}{|\mathrm{Aut}(g)|} = \frac{1}{\langle D^{\text{rel}} \rangle} \cdot \left\langle \frac{\mathcal{P}}{|\mathrm{Aut}|} \right\rangle \geq \left\langle \frac{\mathcal{P}}{|\mathrm{Aut}|} \right\rangle . \tag{4.7}$$

The correction factor $\left\langle D^{\text{rel}} \right\rangle$ converges to unity as $L \to \infty$. Note that the numbers in table 3 for $L \geq 14$ refer to all completions, not just to our samples. Analogously, let

$$\langle \mathcal{P} \rangle_{\text{dec}} := \frac{1}{N_L^{(D)}} \sum_{\substack{\text{decompletions } g \\ L \text{ loops}}} \mathcal{P}(g). \tag{4.8}$$

In this case, we can not leverage eq. (2.2) and, unlike eq. (4.7), the sum in eq. (4.8) can not be rewritten in terms of completions. Nevertheless, we expect $\langle \mathcal{P} \rangle_{\text{dec}} \to \langle \mathcal{P} \rangle$ as automorphisms become increasingly unimportant in the limit $N \to \infty$, see section 2.4.

JHEP11(2023)160

L	$\langle\mathcal{P}\rangle_{\text{dec}}$	$\left\langle\frac{\mathcal{P}}{\lvert\text{Aut}\rvert}\right\rangle_{\text{dec}}$	$\langle\mathcal{P}\rangle_{\text{pla}}$	$\left\langle\frac{\mathcal{P}}{\lvert\text{Aut}\rvert}\right\rangle_{\text{pla}}$	F_{pla}	$F_{\text{pla}}^{(\text{Aut})}$
5	53.206837	11.792786	53.801340	15.521770	21.29	6.142
6	130.07186	50.272453	146.39450	68.971371	11.65	5.488
7	318.74446	175.62729	355.88168	196.84586	5.803	3.210
8	744.30225	509.23037	901.14361	684.50231	3.496	2.656
9	1664.3771	1289.7695	2225.0521	1763.3185	2.552	2.022
10	3591.9116	3006.0073	5490.1060	4853.5088	2.266	2.004
11	7486.4986	6517.9587	13373.880	12028.176	2.274	2.046
12	15143.497	13507.354	32494.348	30213.802	2.532	2.354
13	29822.743	27001.114	78580.447	73666.453	2.986	2.799
$13s$	$2.9801 \cdot 10^4$	$2.7094 \cdot 10^4$	$8.1908 \cdot 10^4$	$7.8462 \cdot 10^4$	3.022	2.825
$13\star$	$2.9904 \cdot 10^4$	$2.7022 \cdot 10^4$	$7.9377 \cdot 10^4$	$7.4925 \cdot 10^4$	3.016	2.847
$14s$	$5.7436 \cdot 10^4$	$5.2546 \cdot 10^4$	$1.9039 \cdot 10^5$	$1.8045 \cdot 10^5$	3.627	3.437
$14\star$	$5.7419 \cdot 10^4$	$5.2471 \cdot 10^4$	$1.8696 \cdot 10^5$	$1.7716 \cdot 10^5$	3.614	3.425
$15s$	$1.0740 \cdot 10^5$	$0.9900 \cdot 10^5$	$4.6530 \cdot 10^5$	$4.3709 \cdot 10^5$	4.709	4.423
$15\star$	$1.0773 \cdot 10^5$	$0.9933 \cdot 10^5$	$4.5417 \cdot 10^5$	$4.3269 \cdot 10^5$	4.612	4.394
$16s$	$1.959 \cdot 10^5$	$1.823 \cdot 10^5$	$8.896 \cdot 10^5$	$8.725 \cdot 10^5$	4.865	4.772
$17s$	$3.567 \cdot 10^5$	$3.363 \cdot 10^5$	$2.434 \cdot 10^6$	$2.434 \cdot 10^6$	7.264	7.264

Table 9. The columns describe the following: average period of decompletions eq. (4.8); eq. (4.7); average period of planar decompletions eq. (4.10); eq. (4.9); ratios between average of planar decompletions and average of completions as defined in eqs. (4.9) and (4.10). All uncertainties are similar to the ones in tables 6 and 7. Note that the factors F_{pla} are significantly larger than unity.

Analogous to section 2.4, dividing by the number of *planar* decompletions, we define

$$\left\langle\frac{\mathcal{P}}{\lvert\text{Aut}\rvert}\right\rangle_{\text{pla}} := \frac{1}{\#\text{planar decompletions}} \sum_{\substack{\text{planar dec. } g \\ L \text{ loops}}} \frac{\mathcal{P}(g)}{\lvert\text{Aut}(g)\rvert} =: F_{\text{pla}}^{(\text{Aut})} \cdot \left\langle\frac{\mathcal{P}}{\lvert\text{Aut}\rvert}\right\rangle . \tag{4.9}$$

We do not have any a priori information about how the factor F_{pla} introduced in eq. (4.9) behaves in the limit $L \to \infty$. In particular, F_{pla} does not (only) encode information about automorphisms of the graphs, but about whether or not the numerical value of the period of a planar graph is representative for the period of all graphs. Finally, we define

$$\langle\mathcal{P}\rangle_{\text{pla}} := \frac{1}{\#\text{planar decompletions}} \sum_{\substack{\text{planar dec. } g \\ L \text{ loops}}} \mathcal{P}(g) =: F_{\text{pla}} \cdot \langle\mathcal{P}\rangle . \tag{4.10}$$

In all cases, the non-uniform samples are corrected according to appendix A.

The data in table 9 shows that indeed, the factors F_{pla} and $F_{\text{pla}}^{(\text{Aut})}$ are significantly larger than unity, and grow with the loop order. This means that the period of planar graphs is much larger than the average, and the effect gets even stronger for higher loop orders.

JHEP11(2023)160

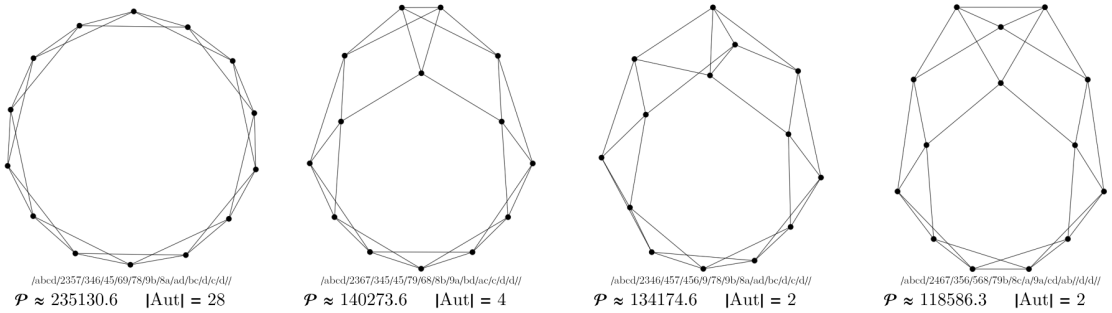


Figure 8. The four 12-loop graphs with the largest period.

4.3 Largest and smallest periods

Having computed all graphs up to 13 loops, we can identify the largest and smallest periods. In all our samples, the graphs with the largest period per loop order are the $(1, 2)$-circulants, known as zigzags Z_L, see graphs (a) in figures 8 and 9. Their period is known to be [14, 15]

$$\mathcal{P}(\ (a)\ ) = \mathcal{P}(Z_L) = 4\frac{(2L-2)!}{L!(L-1)!}\left(1 - \frac{1-(-1)^L}{2^{2L-3}}\right)\zeta(2L-3). \tag{4.11}$$

The next-largest graphs are almost of the zigzag shape, concretely, they are either 3-vertex products of zigzags or they are zigzags with "local" distortions that only affect a few vertices. For $L \geq 8$, the second-largest graphs, (b) in figures 8 and 9, is a 3-vertex product of zigzags $Z_{L-2} \times Z_3$,

$$\mathcal{P}(\ (b)\ ) = \mathcal{P}(Z_{L-2})\mathcal{P}(Z_3) = 24\frac{(2L-6)!}{(L-2)!(L-3)!}\left(1 - \frac{1-(-1)^L}{2^{2L-7}}\right)\zeta(2L-7)\zeta(3).$$

Graphs of type (c) are not 3-vertex reducible, but we find empirically that $(c)_L \approx (a)_{L-1}$ when they are scaled to the mean of the respective loop order, concretely

$$\frac{\frac{\mathcal{P}((c)_{11})}{\langle\mathcal{P}\rangle_{11}}}{\frac{\mathcal{P}((a)_{10})}{\langle\mathcal{P}\rangle_{10}}} = 0.9655, \qquad \frac{\frac{\mathcal{P}((c)_{12})}{\langle\mathcal{P}\rangle_{12}}}{\frac{\mathcal{P}((a)_{11})}{\langle\mathcal{P}\rangle_{11}}} = 0.9852, \qquad \frac{\frac{\mathcal{P}((c)_{13})}{\langle\mathcal{P}\rangle_{13}}}{\frac{\mathcal{P}((a)_{12})}{\langle\mathcal{P}\rangle_{12}}} = 1.0016.$$

The 4^{th} largest graph, type (d), for $L \geq 9$, is again a product

$$\mathcal{P}(\ (d)\ ) = \mathcal{P}(Z_{L-3})\mathcal{P}(Z_4) = 80\frac{(2L-8)!}{(L-3)!(L-4)!}\left(1 - \frac{1+(-1)^L}{2^{2L-9}}\right)\zeta(2L-9)\zeta(5).$$

This pattern can be continued, among the next-larges periods is $\mathcal{P}(Z_{L-4})\mathcal{P}(Z_5)$ and so on. Also, more graphs of "almost-zigzag" shape, such as (c), and their products, start to appear. The period of zigzags grows much faster than the mean period, for example

$$\mathcal{P}(Z_L) \sim \frac{4^L}{4\sqrt{L\pi}}\left(\frac{1}{L} + \mathcal{O}\left(\frac{1}{L^2}\right)\right), \qquad \frac{\mathcal{P}(Z_L)}{\mathcal{P}(\ (b)\ )} \sim \frac{8}{3\zeta(3)} + \mathcal{O}\left(\frac{1}{L}\right). \tag{4.12}$$

JHEP11(2023)160

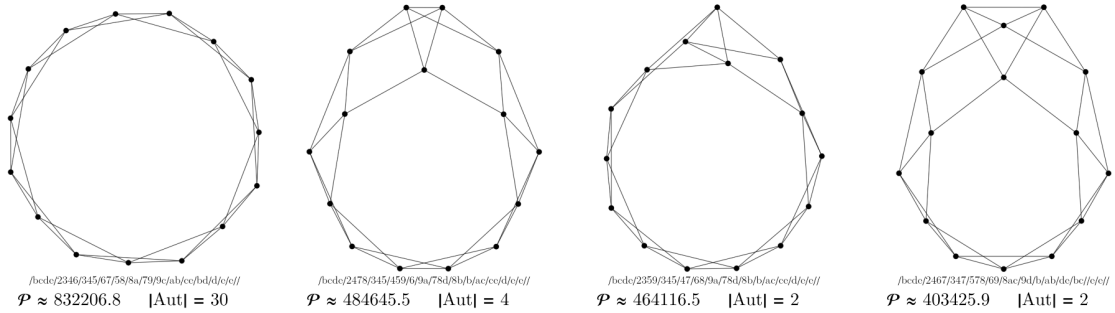


Figure 9. The four 13-loop graphs with the largest period. Note that each of them corresponds to one of the largest 12-loop graphs, shown in figure 8.

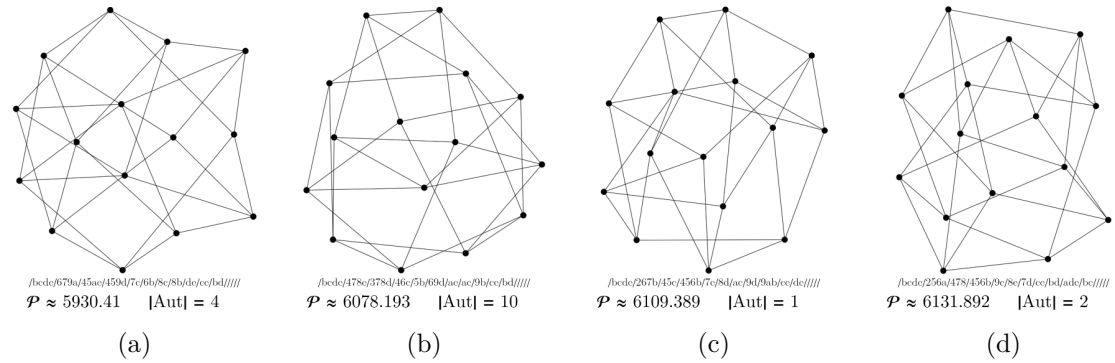


Figure 10. The four 13-loop graphs with the smallest period.

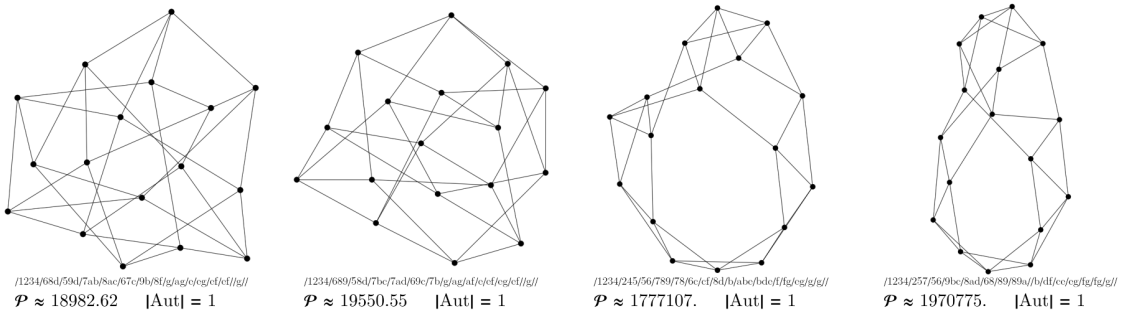
Concretely, if we normalize to unit mean and consider the relative period $x \cdot \langle \mathcal{P} \rangle$, then at $L = 11$ loops, the zigzag is at $\mathcal{P}((a)_{11}) = 8.94 \cdot \langle \mathcal{P} \rangle_{11}$, while at 12 loops it is at $x = 15.4$ and the 13-loop zigzag is at $x = 27.7$, compare figure 1 in the introduction.

At every fixed loop number, there is only one zigzag, one graph of type (b), and so on. That means that for $L \to \infty$, there is a comparatively small set of periods which are arbitrarily much larger than the average, and also arbitrarily far apart (i.e. they differ by asymptotically constant factors such as $\frac{8}{3\zeta(3)}$ in eq. (4.12)). These observations make it plausible that, even if the distribution of periods is normalized to unit mean as in eq. (4.2), still all higher moments diverge as $L \to \infty$. Conversely, we can not expect to find a continuous, monotonically decaying probability density function for the largest periods.

The smallest periods, shown in figure 10, do not show any obvious regular pattern for $L \geq 12$. Conversely, they appear to be particularly "dense" or "irregular". For $L \geq 14$ loops, we can not expect to find the overall smallest and largest periods of the populations in our non-complete samples. Even in the non-complete samples, the smallest graphs appear more dense than the largest ones, see figures 11 and 12.

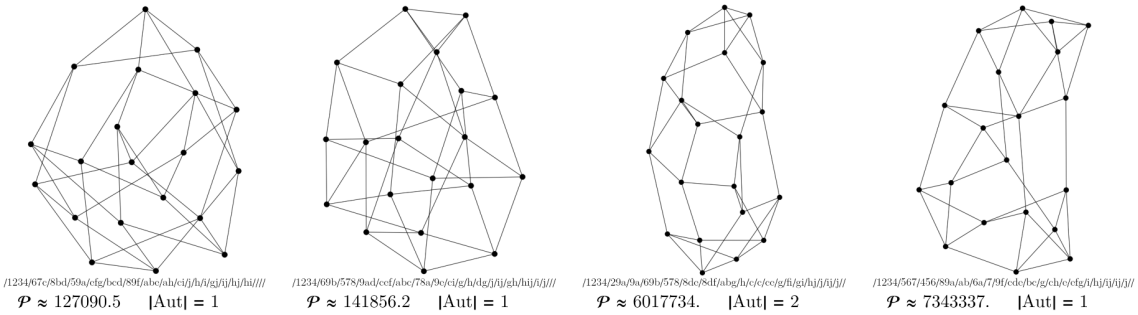
The fact that the $(1, 2)$-circulants constitute the largest periods of each loop order motivates to examine the other (i, j)-circulants as well. To this end, we express their period relative to the average of the corresponding loop order like in eq. (4.2), $\mathcal{P} = x \cdot \langle \mathcal{P} \rangle$. The

JHEP11(2023)160



(a) (b) (c) (d)

Figure 11. The two smallest and the two largest graphs of the 15-loop samples.



(a) (b) (c) (d)

Figure 12. The two smallest and the two largest graphs of the 18-loop samples. Note that the difference in shape is less pronounced than in the previous examples because this sample is very small, compared to the total number of 18-loop graphs, see table 1.

results are given in table 10 in the form (i, j, x). Note that many circulants are isomorphic, the choice of (i, j) is generally not unique.

The data in table 10 shows that even if the zigzag is by far the largest period, many other circulants are smaller than the average. The smallest circulant, for $L \geq 12$, does not coincide with the smallest period of the same loop order. The difference grows with the loop order, such that even in our $L = 18s$ sample, which is tiny compared to the number of 18-loop graphs, we find a graph with smaller period than the smallest 18-loop circulant.

4.4 Probability distribution function near the mean

In the present subsection, we consider the distribution of periods normalized to unit mean as in eq. (4.2). We have established in sections 4.1 and 4.3 that the higher moments of this distribution are most likely infinite in the limit $L \to \infty$. Nevertheless, our data suggests that in the vicinity of the mean, the distribution converges to a non-degenerate limiting distribution, shown in figure 13. Our goal is to describe this distribution near the mean, ignoring the extreme values which cause the moments to diverge.

We have tried fitting several common distributions to the empirical data, but none of them was satisfying. For illustration, the plot figure 13 shows the best fits for a gamma distribution and a log-normal distribution, both of them obviously miss qualitative features of the distribution.

JHEP11(2023)160

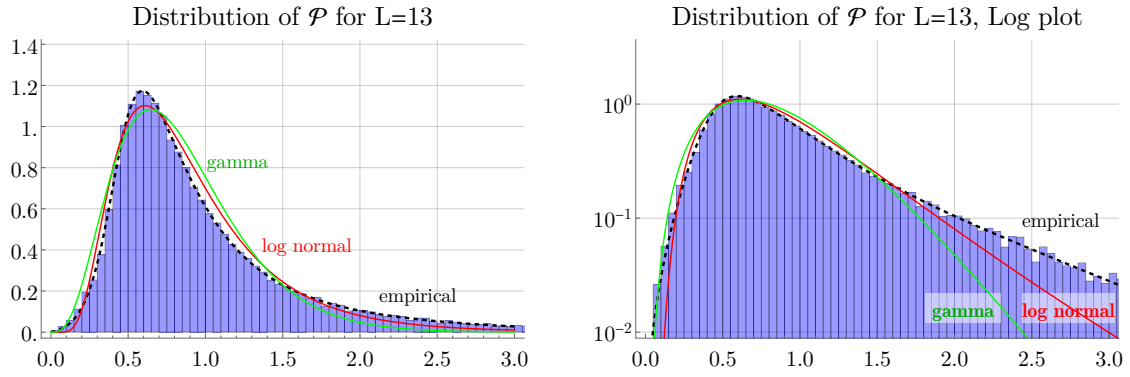
L	Circulants in the form $(i, j, \mathcal{P}/\langle\mathcal{P}\rangle)$
5	$(1,\ 2,\ 1.033)$
6	$(1,3,0.568),\quad (1,2,1.337)$
7	$(1,3,0.634),\quad (1,2,1.671)$
8	$(1,4,0.427),\quad (1,3,0.446),\quad (1,2,2.359)$
9	$(1,3,0.366),\quad (1,2,3.456)$
10	$(2,3,0.284),\quad (1,3,0.365),\quad (1,5,0.425),$ $(1,4,0.661),\quad (3,4,0.676),\quad (1,2,5.430)$
11	$(1,5,0.254),\quad (1,3,0.360),\quad (1,2,8.939)$
12	$(1,4,0.238),\quad (1,3,0.374),\quad (1,6,0.527),\quad (1,2,15.43)$
13	$(1,4,0.219),\quad (1,6,0.228),\quad (1,3,0.413)$ $(1,5,1.128),\quad (3,5,1.130),\quad (1,2,27.71)$
14	$(1,6,0.230),\quad (1,4,0.241),\quad (1,3,0.469)$ $(1,7,0.772),\quad (1,2,51.41)$
15	$(1,4,0.227),\quad (1,5,0.241),\quad (1,3,0.555),\quad (1,2,99.03)$
16	$(1,5,0.195),\quad (1,4,0.216),\quad (2,3,0.261),\quad (1,3,0.681)$ $(1,8,1.306),\quad (1,6,2.778),\quad (1,2,197.1)$
17	$(1,4,0.221),\quad (1,7,0.287),\quad (1,3,0.844),\quad (1,2,395.1)$
18	$(1,8,0.195),\quad (1,4,0.234),\quad (2,5,0.274),\quad (1,5,0.277),\quad (4,5,0.282)$ $(1,6,0.314),\quad (1,3,1.047),\quad (1,9,2.327),\quad (1,2,792.2)$

Table 10. Periods of non-isomorphic circulants. The zigzag $(1,2,x)$ is by far the largest period of each loop order, see also figure 1. For $L \leq 11$, the smallest circulant is also the smallest period of that loop order, but this is false for $L \geq 12$.

We construct an empirical probability distribution function $\rho(x)$ by inspecting the characteristic features of the histogram in a logarithmic plot such as figure 13(b). We find that

1. The log-density grows approximately linearly for small x,
2. The log-density then falls of approximately linearly, but with different slope, at $x \sim 1$, and
3. For $x \gg 1$, the log-density falls of linearly with yet another slope.
4. The density vanishes at $x = 0$ since periods are positive.

JHEP11(2023)160



(a) (b)

Figure 13. Histograms of the symmetry-factor-weighted period, normalized to unit mean, for $L = 13$, with y-axis scaled linearly 13(a) or logarithmically 13(b). The red resp. green lines indicate the best fits for log-normal resp. gamma distributions. Both clearly fail to describe the empirical distribution. The dashed black line is the observed distribution function eq. (4.13).

One possible function with the sought-after properties is

$$\rho(x) = s \cdot x \frac{e^{l\cdot(a-b)} + e^{mx}}{e^{-l\cdot(x-b)} + e^{(m+r)x}}. \tag{4.13}$$

Here, the parameters a and b are defined such that they take values near unity for our data, and the slopes l, m, r are positive. Out of the six parameters in eq. (4.13), only four are independent; the overall scale s is determined by total probability being unity, and one of the remaining parameters $\{a, b, m, l, r\}$ can be eliminated by enforcing unit expectation. See table 11 for fit parameters.

As shown in figure 13 (dashed black line), this density function describes the data well for $L = 13$ loops. Even for small loop orders, the distribution eq. (4.13) is a fairly good approximation to the empirical histograms, see figure 14, but the shape is visibly different from higher loop orders and automorphisms have a large influence. Conversely, for high loop orders $L \geq 15$, our data sets are smaller and the fit parameters are less reliable. Nonetheless, the fitted empirical distribution resembles the histograms acceptably well, see figure 15.

The fit parameters of table 11 are similar in magnitude across all loop orders, but not quite identical. In particular, the left slope l varies significantly, this might indicate that an exponential (linear in the log plot) growth is not an accurate model for the limiting distribution. Conversely, the middle slope $m \approx 2.3$ and the right slope $r \approx 1.2$ are fairly consistent. This implies that the probability for a period to be x-times the mean of the corresponding loop order falls of approximately $\propto \exp(-1.2x)$. Nevertheless, this description is accurate only for moderately large x, not for the outliers (section 4.3). In particular, the distribution function eq. (4.13) has finite moments for all values of its parameters, it therefore arbitrarily much underestimates the probability of large outliers in the limit $L \to \infty$.

JHEP11(2023)160

L	a	b	m	l	r	a	b	m	l	r
11	3.521	1.818	1.636	1.636	1.738	0.741	0.610	2.907	13.21	2.066
12	2.463	1.496	3.166	3.167	1.872	0.677	0.508	2.325	15.16	1.622
13	1.114	0.696	2.362	7.142	1.572	0.656	0.458	2.373	16.19	1.361
13s	0.711	0.493	2.392	15.37	1.357	0.701	0.485	2.576	15.06	1.386
13⋆	1.001	0.613	2.222	8.508	1.407	0.604	0.436	2.374	19.01	1.355
14s	0.684	0.445	2.378	15.826	1.176	0.642	0.426	2.489	16.44	1.235
14⋆	0.883	0.531	2.263	10.11	1.285	0.645	0.428	2.461	16.25	1.254
15s	0.740	0.449	2.256	12.94	1.177	0.589	0.393	2.459	17.89	1.260
15⋆	0.739	0.444	2.313	13.39	1.118	0.590	0.389	2.546	18.63	1.176
16s	0.708	0.414	2.389	14.14	1.041	0.626	0.402	2.648	15.59	1.295
17s	0.749	0.424	2.500	12.91	1.065	0.631	0.387	2.652	15.86	1.169

Table 11. Best fit parameters for the distribution eq. (4.13). The columns describe the following: loop order; 5 columns for periods weighted by symmetry factors; 5 columns for periods without symmetry factors. Observe that the parameters are similar for all loop orders, the distributions have similar shapes. For $L = \{11, 12\}$ with symmetry factor, the parameters m and l are degenerate, this corresponds to the strong distortion visible in figure 14.

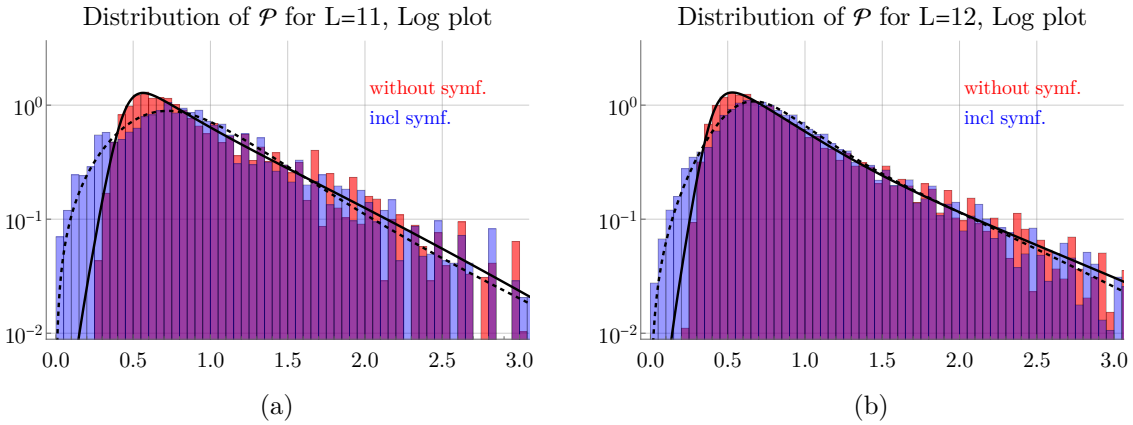


Figure 14. Histograms of the period for smaller loop numbers. The empirical distribution function eq. (4.13) still captures the data, but it is visibly distorted from its shape at higher loop orders.

4.5 Almost-Gaussian model for the period distribution

Consider the distribution of the (normalized) t^{th} power of the period,

$$\frac{1}{\langle \mathcal{P}^t \rangle} \mathcal{P}^t, \qquad t \in \mathbb{R}. \tag{4.14}$$

Here, $t = 1$ reproduces the normalized period examined in section 4.4. Another obvious choice is $t = -1$. The distribution of the inverse period $\mathcal{P}^{-1}$ has an overall wider, more round peak, compared to $\mathcal{P}$ in figure 13, but we skip the details at this point.

A remarkable, but unexpected, phenomenon occurs at $t = -\frac{1}{2}$. If no symmetry factors are taken into account, the quantity $\frac{1}{\sqrt{\mathcal{P}}}$ almost follows a normal distribution, see figure 16

JHEP11(2023)160

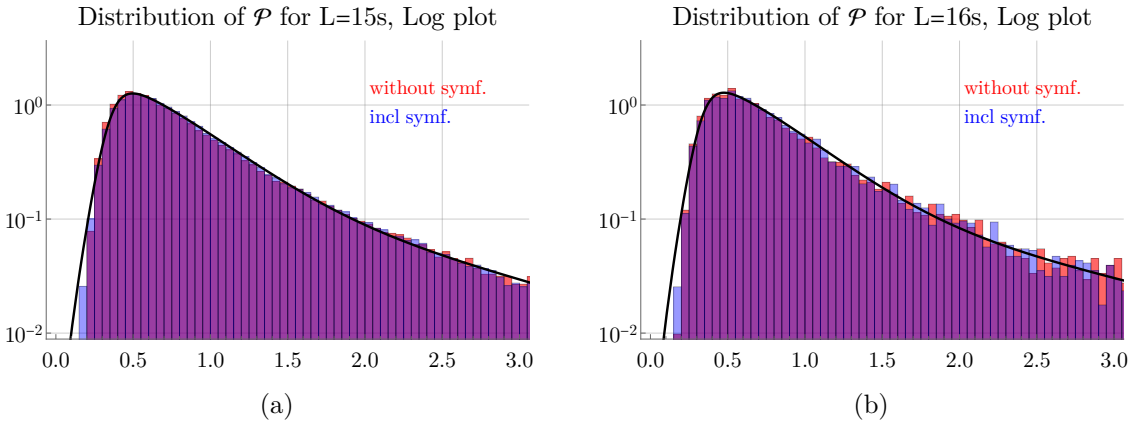


Figure 15. Histograms of the period for higher loop numbers, together with best fit empirical distribution. The shape is very similar to the one for $L = 13$ loops (figure 13), apart from larger fluctuations in the histograms. At 16 loops, the influence of symmetry factors is barely visible.

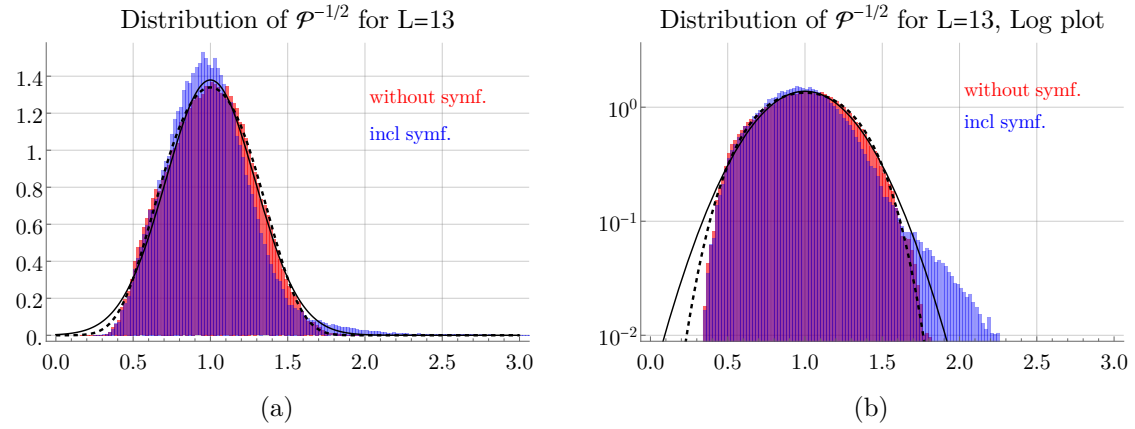


Figure 16. Histogram of $\mathcal{P}^{-\frac{1}{2}}$ for 13 loops. The solid line is the best fit normal distribution to the periods without symmetry factor. A quartic correction eq. (4.15) (dashed line) improves the fit, but still overestimates the presence of large periods (small values of $\mathcal{P}^{-\frac{1}{2}}$).

for $L = 13$. We scanned the parameter range $t \in (-3, +3)$ to confirm that $t = -\frac{1}{2}$ is in fact the only choice where the distribution is close to a Gaussian.

As seen in figure 16, the normal distribution (i.e. a parabola in the log plot) is a good description near the center $x = 1$. An even better fit is obtained with a quartic distribution

$$\rho(x) = \frac{b e^{-\frac{c^4}{b^4}}}{2c^2 K_{\frac{1}{4}}\left(\frac{c^4}{b^4}\right)} \cdot \exp\left(-\frac{(x-1)^2}{2b^2} - \frac{(x-1)^4}{32c^4}\right), \qquad c > 0. \tag{4.15}$$

Here, the first factor is a normalization to unit total probability, $K_{\frac{1}{4}}$ is the modified Bessel function of the second kind, and the parameters b, c are generalizations of the parameter σ of the normal distribution. We remark that the empirical model eq. (4.15) can not be exactly correct since it assigns a (tiny) non-zero probability to negative periods.

JHEP11(2023)160

L	$1\Big/\left\langle\frac{1}{\sqrt{\mathcal{P}}}\right\rangle$	σ	b	c
11	78.860132	0.2621	0.4383	0.1902
12	110.22109	0.2776	0.3788	0.2215
13	151.94024	0.2892	0.3510	0.2554
13s	151.774	0.289	0.351	0.254
13$\star$	151.874	0.293	0.372	0.245
14s	207.131	0.299	0.342	0.287
14$\star$	207.258	0.298	0.340	0.288
15s	279.405	0.306	0.348	0.297
15$\star$	279.543	0.307	0.346	0.303
16s	373.295	0.313	0.346	0.320
17s	497.744	0.311	0.325	0.386
18s	666.063	0.318	0.334	0.385

Table 12. Best fit parameters for the distribution of $\mathcal{P}^{-\frac{1}{2}}$ without symmetry factors. The columns describe the following: loop order; normalization factor; standard deviation of normal distribution; parameters b, c of eq. (4.15).

Best fit parameters, both for a normal distribution and for eq. (4.15), are shown in table 12. Even at high loop orders, the transformed distribution continues to be well approximated by a normal distribution. Our data suggests a diminishing influence of the quartic correction with higher loop order.

The central moments (eq. (4.3)) of the distribution of $\mathcal{P}^{-\frac{1}{2}}$ behave much better than those of $\mathcal{P}$ in tables 6 and 8 and figure 7. Firstly, the odd moments C_3, C_5 are significantly smaller than the even ones C_2, C_4, C_6, which confirms that the distribution is almost symmetric around its mean. Secondly, the moments C_j overall become smaller as j increases, in contrast to the behavior for the non-inverted distribution. Thirdly, the odd moments stay constant with growing loop number L, while the even ones grow only moderately. Our data makes it plausible, but not certain, that all moments of $\mathcal{P}^{-\frac{1}{2}}$ have a finite limit as $L \to \infty$.

The models of the distribution of $\mathcal{P}^{-\frac{1}{2}}$ imply corresponding models for the distribution of $\mathcal{P}$. For a normal distribution, where we introduce a lower cutoff $\delta > 0$ to exclude negative periods, a transformation $t \mapsto t^{-2}$ results in

$$\rho_N(x) = \left(\left\langle\frac{1}{\sqrt{\mathcal{P}(G)}}\right\rangle\right)^2 \cdot \frac{1}{\sqrt{2\pi}\sigma\left(2-\operatorname{erfc}\left(\frac{1-\delta}{\sqrt{2}s}\right)\right)} \frac{e^{-\frac{(\sqrt{x}-1)^2}{2x\sigma^2}}}{x^{\frac{3}{2}}}. \tag{4.16}$$

For the quartic distribution, we find (if no lower limit δ is imposed)

$$\rho_Q(x) = \frac{be^{-\frac{c^4}{b^4}}}{4c^2 K_{\frac{1}{4}}\left(\frac{c^4}{b^4}\right)} \cdot \frac{\exp\left(-\frac{(\sqrt{x}-1)^2}{2xb^2} - \frac{(\sqrt{x}-1)^4}{32x^2c^4}\right)}{x^{\frac{3}{2}}}. \tag{4.17}$$

JHEP11(2023)160

A corresponding distribution centered around the empirical distribution mean can be obtained as usual by scaling,

$$\rho(u) \cdot \left(\left\langle \frac{1}{\sqrt{\mathcal{P}(G)}} \right\rangle \right)^{-2}, \qquad \text{where} \qquad u = x \cdot \left(\left\langle \frac{1}{\sqrt{\mathcal{P}(G)}} \right\rangle \right)^{-2}. \tag{4.18}$$

The distribution functions in eqs. (4.16) and (4.17) constitute alternative empirical models, called "inverse models" hereafter, for the distribution of periods. These models complement the "exponential model" eq. (4.13). As they are continuous, both types of models fail to describe the outliers $\mathcal{P} \gg \langle \mathcal{P} \rangle$ and neither of them is a candidate for an exact limiting distribution as $L \to \infty$. In fact, the presence of a set of discrete outliers (which is a small set, but growing nonetheless) makes it very unlikely that a limiting distribution can be given in closed form at all.

The two different empirical models have quite different features:

- For $x \to \infty$, both eq. (4.16) and eq. (4.17) decay $\sim x^{-\frac{3}{2}}$, consequently, their expectation $\int \mathrm{d}x\, x \cdot \rho(x)$ and all higher moments are infinite. The inverse models correctly reproduce a divergence of higher moments in the limit $L \to \infty$.
- As the higher moments of $\mathcal{P}^{-\frac{1}{2}}$ are finite, the inverse models have a well-defined limit for $L \to \infty$. Moreover, the inverse models can be improved systematically by measuring these moments in samples and constructing a corresponding distribution. Conversely, the exponential model has no obvious way to increase the accuracy, apart from introducing ad-hoc terms and fitting them to a histogram.
- The inverse models are very accurate for the distribution of small values $\mathcal{P} < \langle \mathcal{P} \rangle$, whereas the exponential model is superior for values of $\mathcal{P} > \langle \mathcal{P} \rangle$ Compare the performance for the left resp. right side of the histogram figure 2(b) in the introduction section.

4.6 Logarithm of the period

Although the distribution of periods is well-behaved near the mean (section 4.4), a systematic examination is considerably hindered by the fact that the empirical moments diverge. The presence of very large outliers suggests considering the logarithm $\ln \mathcal{P}(G)$ instead of the period itself. Analogously to section 4.4, we normalize our distribution to unit mean by dividing by $\langle \ln \mathcal{P} \rangle$. The normalization factor is given explicitly in table 13 as it can not be inferred from $\langle \mathcal{P} \rangle$ given in tables 6 and 7. The weighting with symmetry factors can be reconstructed from

$$\left\langle \ln \frac{\mathcal{P}}{|\mathrm{Aut}|} \right\rangle = \langle \ln \mathcal{P} \rangle - \langle \ln |\mathrm{Aut}| \rangle \,,$$

where the second summand is listed in table 3. Furthermore, the standard deviation can be obtained from the mean and C_2 according to eq. (4.4). We computed the central moments (eq. (4.3)) for the distribution of $\ln \mathcal{P}$, both with and without symmetry factors, results are again reported in table 13. The median and quantile boundaries of $\ln \mathcal{P}$ can be reconstructed from tables 6 and 7.

JHEP11(2023)160

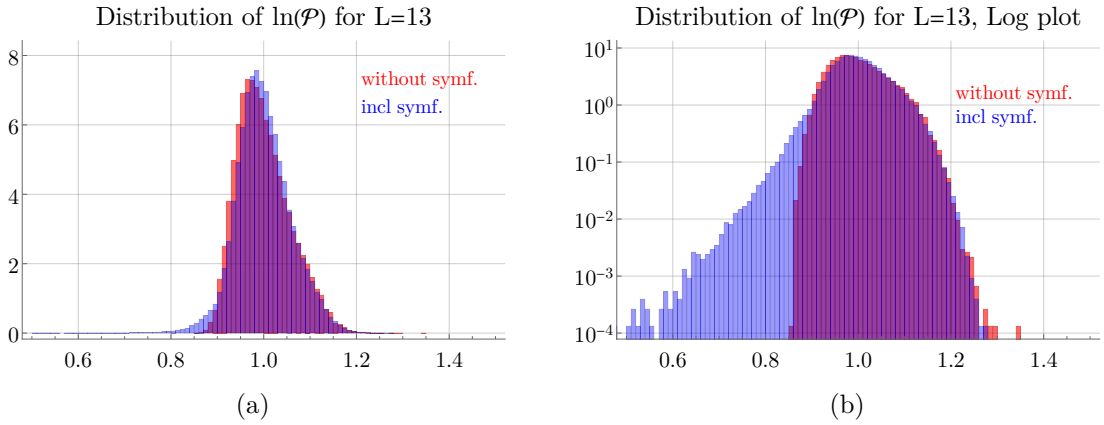


Figure 17. Histograms of the logarithm of the period for 13 loops. Note the different scale of the x-axis compared to previous plots, and the large influence of symmetry factors in (b).

Figure 17 shows the normalized distribution of the logarithm of the period for $L = 13$ loops. Comparing to figure 13, we see that the influence of symmetry factors is much more pronounced for $\ln \mathcal{P}$ than for $\mathcal{P}$. Although, at 13 loops, already 84% of all symmetry factors are unity (see table 3), their presence significantly alters the distribution for small values of $\ln \mathcal{P}$.

The parameters in table 13 indicate that the distribution of $\ln \mathcal{P}$ differs considerably whether the symmetry factor is taken into account or not. In particular, for $L \leq 12$, the skewness for the symmetry-weighted periods is negative, reflecting a tail to the left of the distribution, caused by symmetry factors. For larger loop orders, the difference between the two distributions becomes smaller and the normalized variance seems to converge to a value $C_2 \approx 25 \cdot 20^{-4}$. The skewness and the kurtosis, on the other hand, seem to grow with loop order, albeit less quickly as for the non-logarithmic period in tables 6 and 7.

From the present data, we can not decide whether the higher moments stay finite as $L \to \infty$, but it seems plausible that the limiting distribution at least has finite variance. This has important consequences for sampling since it allows to estimate the uncertainty of the sample mean as outlined in appendix D. Unfortunately, a reliable mean and uncertainty of the distribution of $\ln \mathcal{P}$ does not directly translate to a mean of the distribution of $\mathcal{P}$.

5 Beta function and $O(N)$-dependence

5.1 Background

In the present section, we consider the implications of our data for the beta function of $O(N)$-symmetric massless ϕ^4-theory in 4 spacetime dimensions. The scalar field ϕ is promoted to a N-component vector $(\phi_1, \ldots, \phi_N)$ and the Lagrangian density reads

$$\mathcal{L} = -\frac{1}{2}\left(\partial_\mu \phi_1 \partial^\mu \phi_1 + \partial_\mu \phi_2 \partial^\mu \phi_2 + \ldots + \partial_\mu \phi_N \partial^\mu \phi_N\right) - \frac{\lambda}{4!}\left(\phi_1^2 + \phi_2^2 + \ldots + \phi_N^2\right)^2 . \quad (5.1)$$

See [60] for a review of the background and physical significance of $O(N)$-symmetry.

JHEP11(2023)160

L	$\langle \ln \mathcal{P} \rangle$	$C_2 \cdot 10^4$	$c_3 \cdot 10^2$	c_4	$C_2 \cdot 10^4$	$c_3 \cdot 10^2$	c_4
5	3.9847486	0.693	0	1	7918	0	1
6	4.7935274	39.35	−71.97	2.197	21540	−128.6	3.003
7	5.7115799	26.86	5.173	2.132	1040	−16.03	2.152
8	6.5197021	33.01	2.769	2.476	773.2	−174.2	6.715
9	7.3279257	30.63	15.80	2.574	216.9	−101.0	4.355
10	8.0781075	31.50	24.09	2.529	137.9	−173.4	9.166
11	8.7952253	31.96	35.49	2.564	64.21	−99.85	5.747
12	9.4747560	32.50	46.66	2.718	44.25	−56.40	5.731
13	10.125860	32.55	56.65	2.928	35.06	7.075	3.894
13s	10.124	32.57	57.95	2.914	31.20	60.29	3.026
13⋆	10.126	32.88	58.76	2.927	35.16	9.231	3.945
14s	10.754	32.27	64.77	3.137	30.73	65.89	3.217
14⋆	10.755	32.09	65.03	3.154	31.43	38.57	3.521
15s	11.358	31.07	70.66	3.377	29.66	70.56	3.419
15⋆	11.359	31.56	72.90	3.380	29.66	54.84	3.516
16s	11.918	28.48	74.15	3.503	28.69	74.15	3.503
17s	12.522	28.35	79.30	3.695	27.40	80.16	3.763
18s	13.109	27.07	85.06	3.782	25.73	84.40	3.830

Table 13. Mean and central moments (eq. (4.3)) of the logarithm of the period. The columns describe the following: four columns referring to $\ln \mathcal{P}$; three columns referring to $\frac{\ln \mathcal{P}.}{|\mathrm{Aut}|}$. Note that C_2 stays bounded as L increases, unlike for the non-logarithmic period in table 6.

The beta function encodes the overall scale dependence of a theory. For ϕ^4-theory, it is given by the derivative of the 1PI 4-point vertex Green function $G^{(4)}$with respect to logarithmic energy scale $\ln s$ at the renormalization point,

$$\beta(g) := 2 \cdot \frac{\partial}{\partial \ln s} G^{(4)}\Big|_{s=s_0}.$$

Equivalently, the beta function is the coefficient of the simple pole of the 4-point counterterm in dimensional regularization. The beta function depends on the chosen renormalization scheme, see [61] for a recent discussion. Analytical results for the beta function up to 7 loops in the minimal subtraction scheme can be found in [18, 62].

In the present work, we restrict ourselves to primitive graphs. Their amplitude has the form eq. (1.3), and different renormalization schemes merely amount to adding finite constants. According to eq. (1.6), the dependence on the logarithmic momentum is given, in all renormalization schemes, by the period. We redefine the coupling constant $\lambda \to g = \frac{\lambda}{(4\pi)^2}$ of the Lagrangian in order to absorb the factor Λ in eq. (1.6). We define a series expansion of the contribution of primitive graphs to the beta function according to

$$\beta^{\mathrm{prim}}(N, g) := \sum_{L \geq 1} (-g)^{L+1} \beta_L^{\mathrm{prim}}(N) := 2 \sum_{\substack{G \text{ primitive} \\ \text{compl.}}} (-g)^{L_G+1} \cdot \frac{4!(L+2) \cdot 3T(G,N) \cdot \mathcal{P}(G)}{N(N+2)\,|\mathrm{Aut}(G)|}. \tag{5.2}$$

JHEP11(2023)160

Here, $\mathcal{P}(G)$ is the period eq. (1.4) and $T(G,N)$ is a polynomial in N which encodes the $O(N)$-symmetry of the completion, see eq. (5.4). The automorphism group $\mathrm{Aut}(G)$ has no fixed vertices and the sum is over all non-isomorphic completions, we include the factor 4! to reproduce the usual definition in physics as discussed in section 2.1. As explained in appendix A, we do not need to know the total number of graphs $N_L^{(C)}$, but only $N_L^{(\mathrm{Aut})}$, to estimate the beta function from the "s"-samples. For the "$\star$"-samples, this is irrelevant as we know $N_L^{(C)}$ for $L \leq 16$ (table 2).

The $O(N)$-symmetric 4-valent interaction in eq. (5.1) describes a coupling between two pairs of fields with identical indices. Graphically, every 4-valent vertex is a sum of 3 terms representing the 3 ways to connect its 4 adjacent edges into pairs. The Euclidean vertex Feynman rule is

$$-\lambda \frac{\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}}{3}.$$

The polynomial $T(G,N)$ is the product of vertex Feynman rules, without the factor $(-i\lambda)$, and summed over all indices $i,j,\ldots \in [1,N]$, where the indices of the external edges are set to 1. The resulting polynomial is closely related to the Martin polynomial $M(G,x)$ and the circuit partition polynomial $J(G,x) = r_G(x)$ [63–65] by

$$3^{|V_G|}T(G,x) = J(G,x) = xM(G,x+2). \tag{5.3}$$

The linear coefficient of $M(G,x)$ is the *Martin invariant* [59, 66]. Computation, and examples, of $T(G,N)$ can be found in [67]. We compute $T(G,N)$ recursively and store the intermediate results in a cache, similarly to our computation of the Hepp bound (section 3.3). For example, for a single vertex $T(v,N) = 1$ and if G is the 3-loop multiedge arising from joining all 4 edges of a vertex to a new vertex, we find $T(G,N) = \frac{N(N+2)}{3}$. A 1-loop multiedge G with 4 external edges has $T(G,N) = \frac{N+8}{9}$ and if the 4 external edges are joined to a new vertex, the resulting triangular graph G has $T(G,N) = \frac{16N+10N^2+N^3}{27}$ and the graph $G = (a)$ in figure 3 has

$$T\left(G,N\right) = \frac{3004N + 2740N^2 + 749N^3 + 67N^4 + N^5}{3^8}.$$

In fact, the recursive structure of $T(G,N)$ implies that the result for a completion G is related to that of its decompletion $G \setminus \{v\}$ by

$$T(G,N) = \frac{N(N+2)}{3} T(G \setminus \{v\}, N). \tag{5.4}$$

The order of N appearing in the polynomial $T(G,N)$ is bounded. Concretely, a factor N arises from each closed path (circuit) in the graph, subject to each edge of the graph appearing in exactly one circuit. The maximum number of circuits is reached if they are as short as possible, that is, 3 edges long. A L-loop completion has $2L+4$ edges, therefore, for a completion G with $L \geq 3$ loops

$$\text{Order of } N \text{ of the polynomial } T(G,N) \leq \left\lfloor \frac{|E_G|}{3} \right\rfloor = \left\lfloor \frac{2}{3} L_G + \frac{4}{3} \right\rfloor. \tag{5.5}$$

JHEP11(2023)160

All graphs in our data set respect this inequality. For $L \leq 15$ the data set contains graphs where equality holds and these graphs do not always have planar decompletions.

Taking into account eq. (5.4), the inequality eq. (5.5) implies that terms of higher order in N start to appear only later in the perturbation series, limiting the reliability of asymptotic predictions from our data. For example, terms of order N^8 start to appear only at $L \geq 13$ loops, so we effectively only have the first 6 loop orders, not the first 18, which contribute to the order-N^8-term in β^{prim}. Moreover, only a single graph in our $L = 15$ data sets contributes to the order N^9. This implies that in order to reach high accuracy for large N, one not only needs high loop orders L, but also large samples.

The L-loop coefficient β_L of the beta function in the minimal subtraction (MS) scheme of an $O(N)$-symmetric ϕ^4-theory grows asymptotically as $\beta_L \sim \bar{\beta}_L\left(1+\mathcal{O}\left(L^{-1}\right)\right)$, where [68, 69]

$$\bar{\beta}_L := (L+1)^{3+\frac{N}{2}}\Gamma(L+2)\frac{36\sqrt{3}\cdot 3^{\frac{N}{2}}}{\pi\Gamma\left(2+\frac{N}{2}\right)A^{2N+4}}e^{-\frac{3}{2}-\frac{N+8}{3}\left(\frac{3}{4}+\gamma_{\text{E}}\right)}. \tag{5.6}$$

Here, $A \approx 1.28242713$ is the Glaisher-Kinkelin constant [70] and $\gamma_{\text{E}} \approx 0.57721566$ is the Euler-Mascheroni constant. The coefficient β_L in eq. (5.6) is not the same as the primitive coefficient β_L^{prim} in eq. (5.2), but it is conjectured [69] that their leading asymptotic growth coincides for $L \to \infty$, or equivalently, that primitive graphs constitute the leading contribution to the beta function in minimal subtraction. We will use our data to assess the validity of this conjecture.

5.2 Numerical results

Table 14 contains our results for the primitive contribution to the beta function eq. (5.2). Note that even if the individual periods for $L = 13$ have $\approx 250\,\text{ppm}$ numerical uncertainty (table 1, we reach an expected uncertainty of 0.25 ppm, i.e. more than 6 significant digits, for the beta function, reflecting the large number of graphs in that sample.

Our numerical results for $L \leq 7$ are consistent[3] with the exact results of [36]. Moreover, [36, appendix B] contains an estimate for the primitive beta function up to 11 loops, based on the Hepp bounds of the corresponding graphs, which also is confirmed by our results. For the choice $N = 1$, the $O(N)$-symmetric theory reduces to the ordinary ϕ^4-theory. We have verified for every single graph in our samples that the factor eq. (5.3) satisfies $T(G, 1) = 1$.

If the conjecture is true that eq. (5.6) coincides with the asymptotics of the primitive beta function, then their ratio should converge to unity,

$$\frac{\beta_L^{\text{prim}}}{\bar{\beta}_L} \overset{?}{\sim} 1 + \mathcal{O}\left(L^{-1}\right). \tag{5.7}$$

A plot of this ratio for $L \leq 11$ loops in [36, figure 1] indicated that the ratio is below unity at 11 loops, but growing. We show our data for the choices $N = 0$ and $N = 2$ in figure 18(a), the data for $N = 1$ lies between those points and has been omitted for clarity. We confirm the observation that the ratio eq. (5.7) grows with the loop order. It appears to reach unity around $L \approx 18$, but with a non-vanishing slope. To confirm the limit unity, we computed a second-order Richardson extrapolation and found that it approaches values far

[3] Using the data provided in [17], the coefficient of β^{prim} at $L = 6$ loops and $N = 1$ is 24130.7302..., consistent with our estimate (24130.778 ± 0.080), despite being rounded to 24130 in [36].

JHEP11(2023)160

L	$\beta_L^{\text{prim}}(N=1)$	relative accuracy	$\beta_L^{\text{prim}}(N=1)$ exact	estimate from Hepp	only planar decompl.	$\beta_L(N=1)$ exact
1	$3.000002 \cdot 10^0$	$4.5 \cdot 10^{-6}$	$3 \cdot 10^0$		$3.000 \cdot 10^0$	$3 \cdot 10^0$
3	$1.4424675 \cdot 10^1$	$2.1 \cdot 10^{-6}$	$1.4424683 \cdot 10^1$		$1.442 \cdot 10^1$	$3.25 \cdot 10^1$
4	$1.2443163 \cdot 10^2$	$2.6 \cdot 10^{-6}$	$1.2443133 \cdot 10^2$		$1.244 \cdot 10^2$	$2.72 \cdot 10^2$
5	$1.6981612 \cdot 10^3$	$5.7 \cdot 10^{-6}$	$1.6981712 \cdot 10^3$		$1.490 \cdot 10^3$	$2.85 \cdot 10^3$
6	$2.4130778 \cdot 10^4$	$3.3 \cdot 10^{-6}$	$2.4130730 \cdot 10^4$	$2.41 \cdot 10^4$	$1.655 \cdot 10^4$	$3.48 \cdot 10^4$
7	$3.7092484 \cdot 10^5$	$2.4 \cdot 10^{-6}$	$3.7092457 \cdot 10^5$	$3.71 \cdot 10^5$	$1.795 \cdot 10^5$	$4.75 \cdot 10^5$
8	$6.0618784 \cdot 10^6$	$1.8 \cdot 10^{-6}$		$6.06 \cdot 10^6$	$1.906 \cdot 10^6$	
9	$1.0450228 \cdot 10^8$	$1.9 \cdot 10^{-6}$		$1.05 \cdot 10^8$	$1.989 \cdot 10^7$	
10	$1.8893120 \cdot 10^9$	$8.9 \cdot 10^{-7}$		$1.89 \cdot 10^9$	$2.050 \cdot 10^8$	
11	$3.5671276 \cdot 10^{10}$	$4.0 \cdot 10^{-7}$		$3.57 \cdot 10^{10}$	$2.092 \cdot 10^9$	
12	$7.0121257 \cdot 10^{11}$	$2.1 \cdot 10^{-7}$			$2.117 \cdot 10^{10}$	
13	$1.4319963 \cdot 10^{13}$	$2.5 \cdot 10^{-7}$			$2.128 \cdot 10^{11}$	
13s	$1.43237 \cdot 10^{13}$	$7.1 \cdot 10^{-3}$			$2.308 \cdot 10^{11}$	
13⋆	$1.43445 \cdot 10^{13}$	$7.0 \cdot 10^{-3}$			$2.295 \cdot 10^{11}$	
14s	$3.03718 \cdot 10^{14}$	$2.3 \cdot 10^{-3}$			$2.205 \cdot 10^{12}$	
14⋆	$3.03473 \cdot 10^{14}$	$1.2 \cdot 10^{-3}$			$2.155 \cdot 10^{12}$	
15s	$6.62502 \cdot 10^{15}$	$2.7 \cdot 10^{-3}$			$2.189 \cdot 10^{13}$	
15⋆	$6.65288 \cdot 10^{15}$	$3.3 \cdot 10^{-3}$			$2.107 \cdot 10^{13}$	
16s	$1.49588 \cdot 10^{17}$	$9.0 \cdot 10^{-3}$			$1.498 \cdot 10^{14}$	
17s	$3.535 \cdot 10^{18}$	$1.4 \cdot 10^{-2}$			$1.766 \cdot 10^{15}$	
18s	$8.671 \cdot 10^{19}$	$3.4 \cdot 10^{-2}$				

Table 14. The columns describe the following: primitive contribution to the beta function for $N = 1$; estimated relative uncertainty; analytic value, to 8 significant digits, from [17]; estimate of the primitive β-function from Hepp bounds in [36]; contribution of planar primitive decompletions, for $N = 1$; analytic beta function, including non-primitive graphs, to 3 digits, from [18].

above unity, see the hollow points in figure 18(a). Unfortunately, the large fluctuations of the non-complete samples severely distort the extrapolation and we can not determine with certainty how the extrapolation behaves for $14 \leq L \leq 18$. The fact that our numerical data coincides with the known coefficients of the beta function for $L \leq 7$ makes it improbable that the values of β_L^{prim} are severely wrong, at least for $L \leq 13$. Therefore, our findings suggest one of three different interpretations:

1. Either, the asymptotics eq. (5.6) is missing a factor ≈ 2, or
2. the conjecture is false that primitive graphs dominate the beta function in MS, or
3. the convergence to the asymptotic regime is very slow, such that eq. (5.7) reaches the value unity only for $L \gg 20$ loops.

JHEP11(2023)160

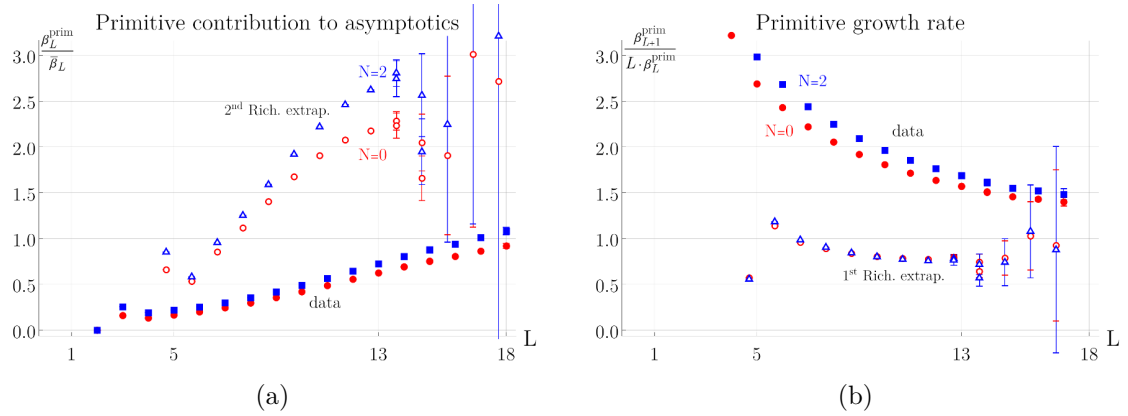


Figure 18. (a) Ratio eq. (5.7) between the observed primitive beta function and the supposed asymptotics, for $N = 0$ (red), and $N = 2$ (blue). Richardson extrapolation suggests a limit $\approx 2 \neq 1$ of this ratio, with significant uncertainty. (b) Ratio $\frac{\beta_{L+1}^{\text{prim}}}{L\cdot\beta_L^{\text{prim}}}$ and its Richardson extrapolation. Multiple data points at $L = 13$ and $L = 14$ loops represent the difference between samples "$\star$" and "s". The data is compatible with the expected limit unity (eq. (5.8)).

The first possibility can not be excluded, but we were unable to spot a concrete error, or un-adjusted difference in conventions, in neither of [36, 68, 69]. However, findings in section 5.4 support this hypothesis.

For the second possibility, note that our numerical values tend to be *larger* than the supposed asymptotics. Therefore, in this scenario, the non-primitive graphs would have to produce a relative contribution with opposite sign, which does not vanish at $L \to \infty$. However, the exact (non just primitive) beta function for $L \leq 7$ in table 14 suggests that non-primitive graphs contribute with the same sign, but smaller magnitude, compared to the primitive ones. From this data it is plausible that primitive graphs dominate in the limit, and the above second possibility seems unlikely.

To assess the third possibility, we consider the growth rate of coefficients. If eq. (5.6) describes the correct asymptotics of primitive graphs, the ratio of successive coefficients should be

$$\frac{\beta_{L+1}^{\text{prim}}}{L \cdot \beta_L^{\text{prim}}} \overset{?}{\sim} \frac{\bar{\beta}_{L+1}}{L \cdot \bar{\beta}_L} = 1 + \mathcal{O}\left(L^{-1}\right). \tag{5.8}$$

Figure 18(b) shows that for all our data points, the growth rate eq. (5.8) is well above unity, but it becomes smaller with growing loop order. We can use Richardson extrapolation [71] as explained in e.g. [72], to obtain a more accurate estimate of the asymptotic value of the ratio. The first order Richardson extrapolation (hollow points in figure 18(b)) lies below unity. This behavior is compatible with the third hypothesis above. Concretely, our data suggests that at first, β_L^{prim} grows too quickly, overshoots its supposed asymptotics, but then grows slower than the asymptotics. As we are merely considering rations, this observation does neither prove nor disprove the existence of a missing absolute factor of 2 in the asymptotics. Note that a slow convergence of β_L^{prim} to $\bar{\beta}_L$ has been observed already in [36]. In a slightly different analysis, the authors of [73] conclude that the asymptotic

JHEP11(2023)160

approximation at least is not useful below poles of order ϵ^{-10} in dimensional regularization, i.e. 10-loop graphs. It thus seems possible that our numerical data $L \leq 18$ is not yet representative for the truly asymptotic regime.

All in all, we do observe that primitive graphs contribute significantly to the conjectured asymptotic beta function in MS. But our data does not allow to clearly confirm, nor exclude, that the leading asymptotics of β_L^{prim} follows the concrete formula eq. (5.6). Note that the form eq. (5.6) is chosen to match [36, 69], but the same leading growth can be expressed by replacing the first two factors by, for example, $L^{4+\frac{N}{2}}\Gamma(L+1)$. The difference between the two is a subleading term $\propto L^{-1}$, but it amounts to a deviation of $\approx 40\%$ for $L = 13$. Consequently, even if eq. (5.6) is the correct leading growth, it is perfectly reasonable that subleading corrections dominate for our data. This situation is slightly tautological because for *any* finite sequence of coefficients, one can introduce subleading terms to make the coefficients match almost any supposed asymptotics. The question, therefore, should not be "Does our data follow the supposed leading asymptotics?", but rather "Given the data and the asymptotics, can we determine subleading coefficients and are they small?" We will see in section 5.4 that the answer to the second question is "yes" as soon as we introduce a factor of 2 into the asymptotics.

5.3 $O(N)$-dependence and planar graphs

In figure 18, the growth of coefficients differs visibly between the cases $N = 0$ and $N = 2$ of the $O(N)$-symmetry. To examine this effect, we normalize the coefficients with respect to the (supposed) leading growth in L eq. (5.6), according to

$$\frac{\bar{\beta}_L}{(L+1)^3\Gamma(L+2)} = \frac{36\sqrt{3}\cdot(L+1)^{\frac{N}{2}}3^{\frac{N}{2}}}{\pi\Gamma\left(2+\frac{N}{2}\right)A^{2N+4}}e^{-\frac{3}{2}-\frac{N+8}{3}\left(\frac{3}{4}+\gamma_{\text{E}}\right)}. \tag{5.9}$$

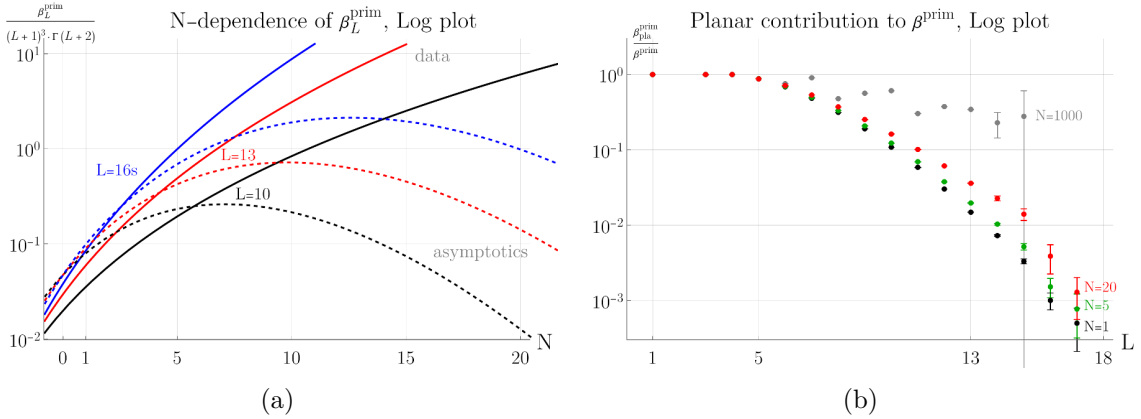
Firstly, we note that the right hand side is independent of L for $N = 0$, so all plots of this ratio for different L will intersect at $N = 0$. Secondly, for fixed L, eq. (5.9), has a finite maximum and approaches zero as $N \to \infty$. Conversely, for finite loop order L, the $O(N)$-factor $T(G, N)$ is a polynomial (eq. (5.5)) and hence unbounded for $N \to \infty$. Therefore, for large enough N, the finite-order coefficient $\beta_L(N)$ deviates from the asymptotics $\bar{\beta}_L$ arbitrarily much,

$$\lim_{N\to\infty}\frac{\beta_L(N)}{\bar{\beta}_L(N)} = \infty \quad \text{for fixed } L < \infty. \tag{5.10}$$

The asymptotic ratio eq. (5.9), together with our empirical data, is shown in figure 19(a). We see that the N-dependence of our data, even at $L = 13$, differs substantially from the predicted asymptotic behavior. The deviation grows with larger N according to eq. (5.10). Only at high loop order, $L = 16s$, our data begins to resemble the asymptotic shape at least for small N. We conclude that even knowledge of a sample of 18-loop graphs is insufficient to reach the large-L regime for values of N substantially different from unity.

In some theories which have an internal SU(N), $O(N)$, or similar gauge symmetry, the leading contribution to scattering amplitudes, in the limit $N \to \infty$, is given by planar (decompleted) graphs, see e.g. [74, 75]. It is interesting to examine whether we see such an effect for the $O(N)$-symmetry in our data, even if we have graphs at leading order in

JHEP11(2023)160



(a) (b)

Figure 19. (a) N-dependence of the ratio eq. (5.9) between $\beta_L^{\text{prim}}(N)$ and the conjectured asymptotics $\bar{\beta}_L$ for fixed loop order, log plot. Solid lines: data, dashed: asymptotic prediction. With growing loop order, the empirical data begins to resemble the asymptotic curve, but only near $N = 1$. (b) Relative contribution of planar decompletions to the primitive beta function. While this ratio is small for large L, the influence of planar graphs grows with N.

N which are not planar. Note that unlike eq. (5.10), this statement does not concern the $L \to \infty$ asymptotics, but the relation between graphs at fixed finite loop order as $N \to \infty$. Firstly, at $N = 1$, the contribution of planar decompletions is relatively small, and quickly decreases with growing loop order (table 14). This is not surprising since the number of planar graphs grows much slower than the total number of graphs (table 2).

Figure 19(b) shows the relative contribution of planar decompletions to the primitive beta function, as a function of the loop order for different values of N. Clearly, the relative importance of planar graphs increases with increasing N, but even as high as $N = 20$, planar graphs account for less than 1% of the total value at $L = 15$ loops. For illustration, figure 19(b) shows the data points for $N = 1000$, which contribute $\sim 10\%$ at 15 loops. Overall, it seems plausible that planar graphs will eventually dominate in the limit $N \to \infty$ for fixed L, but in the light of eq. (5.5) and figure 19(a), this regime is far outside the validity of our perturbative data, so we should not expect our concrete numerical values to be accurate in the limit.

The dominance of planar graphs in the limit $N \to \infty$ is caused entirely by their $O(N)$-factors $T(G, N)$. Regardless of these factors, we have seen in table 9 that the actual amplitude, i.e. the period, of planar graphs is systematically larger than the average by a factor $F^{(\text{Aut})} \sim 3$. This has the effect that even for $N = 1$, the contribution of planar graphs to the beta function, albeit small, is larger than what should be expected from merely the number of planar graphs. Conversely, if one were to extrapolate the amplitude of planar graphs to all graphs, the resulting beta function would come out significantly too large, even for $N = 1$.

5.4 Asymptotics of the period mean

The coefficients eq. (5.2) of the N-dependent primitive beta function is not directly related to the symmetry-factor weighted mean eq. (4.5) because the factor $T(G, N)$ is different for

JHEP11(2023)160

each graph,

$$\beta_L^{\text{prim}} = 2N_L^{(C)} \cdot 4!(L+2) \left\langle \frac{T(G,N) \cdot \mathcal{P}(G)}{|\text{Aut}(G)|} \right\rangle . \tag{5.11}$$

We have seen below eq. (5.10) that for $N > 1$, our results are far from their large-L-asymptotic values. We therefore restrict our attention to $N = 1$, where $T(G,1) = 1$. Then, combining the (conjectured) asymptotic expression for β_L^{prim} (eq. (5.6)) with the known asymptotic growth of $N_L^{(\text{Aut})}$ (theorem 2)) and the fact that the leading growth of $N_L^{(C)}$ coincides with that of $N_L^{(\text{Aut})}$ (eq. (2.6)), we obtain the following expression for the leading growth $\bar{\mathcal{P}}_L^{(\text{Aut})}$ of the mean period, up to terms $\mathcal{O}(L^{-1})$:

$$\left\langle \frac{\mathcal{P}}{|\text{Aut}|} \right\rangle \sim \bar{\mathcal{P}}_L^{(\text{Aut})} + \mathcal{O}\left(\frac{1}{L}\right), \qquad \bar{\mathcal{P}}_L^{(\text{Aut})} = \frac{\bar{\beta}_L}{2\bar{N}_L^{(\text{Aut})}} = \left(\frac{3}{2}\right)^{L+3} L^{\frac{5}{2}} \cdot \frac{2\sqrt{\frac{2}{\pi}} e^{-3\gamma_{\text{E}}}}{A^6}. \tag{5.12}$$

Owing to the vanishing influence of automorphisms as $L \to \infty$, the asymptotic growth $\bar{\mathcal{P}}_L$ of $\langle \mathcal{P} \rangle$ coincides with eq. (5.12) to leading order. In particular, the growth rate per loop order is

$$\frac{\bar{\mathcal{P}}_{L+1}}{\bar{\mathcal{P}}_L} = \frac{3}{2} + \delta \frac{1}{L}, \qquad \frac{\bar{\mathcal{P}}_{L+1}^{(\text{Aut})}}{\bar{\mathcal{P}}_L^{(\text{Aut})}} = \frac{3}{2} + \delta^{(\text{Aut})} \frac{1}{L}, \tag{5.13}$$

where we have introduced correction coefficients δ, $\delta^{(\text{Aut})}$.

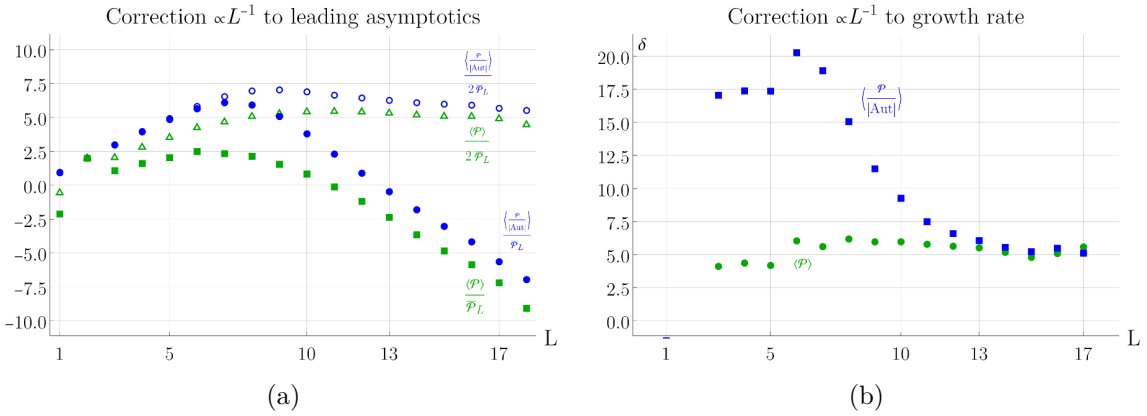
We can currently not determine subleading corrections to $\bar{\mathcal{P}}_L$ and $\bar{\mathcal{P}}_L^{(\text{Aut})}$, or the coefficients δ, $\delta^{(\text{Aut})}$, theoretically. Firstly, we are missing the corrections to various ratios, and secondly, β_L and β_L^{prim} are not expected to coincide beyond the leading term eq. (5.6). The naive $\frac{1}{L}$ correction to the right hand side of eq. (5.12), computed using $\bar{\beta}_L$ to order L^0 and $\bar{N}_L^{(\text{Aut})}$ to order L^{-1}, is $\frac{105}{16}\frac{1}{L}$, this corresponds to $\delta = \frac{15}{4}\frac{1}{L}$ in eq. (5.13). At $L = 18$, these terms still amount to a relative correction as large as 36%, or 21%, respectively. From this estimate, and our observations in section 5.2, we expect that our data considerably deviates from the leading asymptotic growth.

The period averages are reported in tables 6 and 7 and shown in the plot figure 2(a) in the introduction section. Much like the beta function (figure 18(a)), we see that the leading asymptotics eq. (5.12) matches the data only roughly, but the observed asymptotics appears to be larger than the expected one by a factor 2. In figure 20(a), we plot the subleading coefficient Δ, defined by

$$\frac{\langle \mathcal{P} \rangle}{\bar{\mathcal{P}}_L} = 1 + \Delta \frac{1}{L}. \tag{5.14}$$

The plot shows $L \cdot \left(\langle \mathcal{P} \rangle / \bar{\mathcal{P}}_L - 1 \right) = \Delta + \mathcal{O}(L^{-1})$. This quantity is expected to converge to a constant value Δ if eq. (5.12) is the correct leading term, but instead it seems to fall off linearly. However, if we normalize with respect to $2\bar{\mathcal{P}}_L$ (i.e. assume that eq. (5.12) is missing a factor of 2), the subleading correction comes out finite and we find $\Delta \approx 5$. Of course, it is perfectly possible that the true asymptotic regime is only reached at $L \gg 20$, and in

JHEP11(2023)160



(a) (b)

Figure 20. (a) Correction eq. (5.14) (filled symbols). The coefficient seems to fall off linearly, which suggests that the ratio $\langle \mathcal{P} \rangle / \bar{\mathcal{P}}_L$ is not unity. If we, however, introduce a factor of 2 into the asymptotics, the subleading correction converges to a value $\Delta \approx 5$ (hollow symbols). (b) Correction δ to the growth ratio of the period as defined in eq. (5.13). This ratio is not affected by overall factors in the asymptotics, the limiting value is $\delta \approx 5$.

that case there is no contradiction. But if we require that the asymptotics should be a good approximation for $L \approx 15$, then our finding suggests that a factor of two is missing in eq. (5.12). Note that, once we introduce that factor, the correction in figure 20(a) assumes its asymptotic value starting from $L_{\text{crit}} \approx 9$, which is approximately the same regime where the graph-theoretical quantities in section 2.3 start to follow their asymptotics.

The growth ratio eq. (5.13) is insensitive to overall factors, and it seems to be accurate, the leading correction is shown in figure 20(b). Empirically, we find that both coefficients approximately coincide, $\delta \approx \delta^{(\text{Aut})} \approx 5$.

6 Relations of the period to other graph properties

We have found above, in section 4.1, that a planar decompletion on average has larger period than a non-planar one. In the present section, we examine the correlation of the period with several other properties of the graph.

6.1 Symmetry factor

One the one hand, for $L \to \infty$, most graphs have trivial symmetry factor, $|\text{Aut}(G)| \to 1$ (section 2.2). On the other hand, in section 4.3, we found that the largest periods appear with graphs which "look symmetric". This motivates the study of the relation between the period and the symmetry factor of a graph.

In figure 21(a), we have plotted the period of completed graphs, normalized to its mean $\langle \mathcal{P} \rangle$ from table 6, as a function of the size of the automorphism group. There are some patterns, for example the graphs where $|\text{Aut}|$ is 2^n tend to have larger periods than others, but no significant overall trend. As a consequence, the contribution of a graph to physical observables, i.e. its period weighted by the symmetry factor, is small for graphs with large automorphism groups.

JHEP11(2023)160

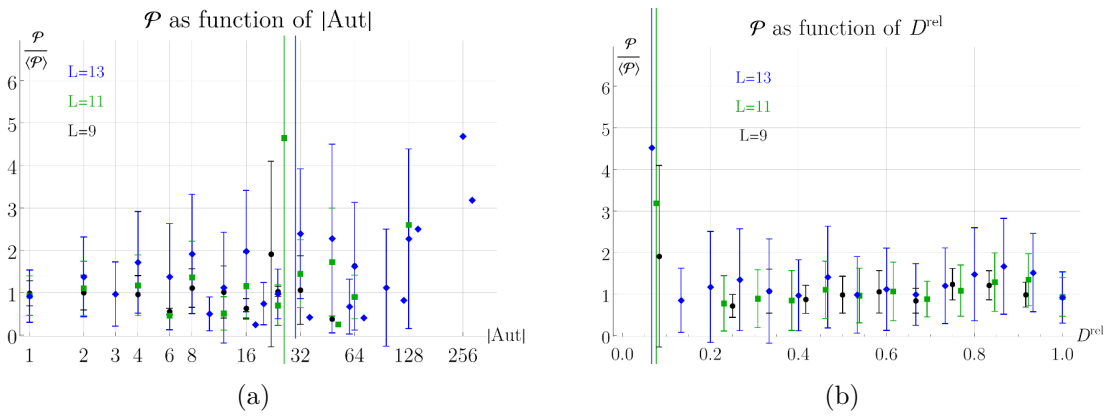


Figure 21. (a) Period, normalized to unit mean, as a function of the size of the automorphism group. Shown is the mean of the corresponding sets of graphs, error bars indicate the standard deviation within the set, which in almost all cases includes the value unity. (b) The same quantity, but as a function of non-isomorphic completions (eq. (2.7)). There is almost no correlation.

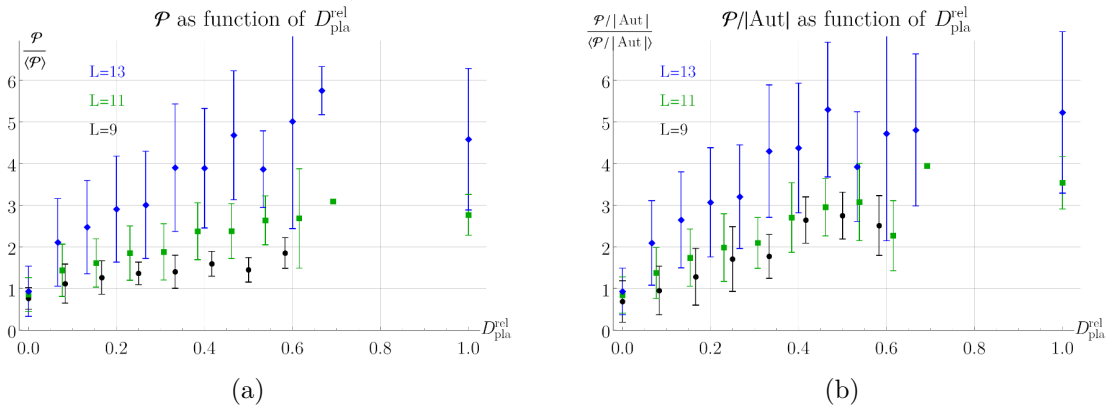


Figure 22. Period, rescaled to its average, as a function of the number of non-isomorphic planar decompletions (eq. (2.9)). There is an almost linear relation, which prevails even if the graphs are weighted by symmetry factors and gets stronger at higher loop orders.

In eq. (2.7), we introduced the number of non-isomorphic decompletions as an alternative measure for the automorphisms of a graph, it is inversely correlated with $\log|\text{Aut}|$ (figure 5(b)). Consequently, we find that the period is almost uncorrelated with D^{rel}, see figure 21(b). The large standard deviation at the smallest value of D^{rel} is caused by zigzag graphs. Note that the "bump" in figure 5(b) at $D^{\text{rel}} = \frac{1}{2}$ translates to the fact that graphs with $D^{\text{rel}} \approx \frac{1}{2}$ constitute a slightly larger contribution to the symmetry-factor weighted sum than what would be expected (i.e. those graphs have a smaller symmetry factor than expected from their value of D^{rel}, and hence contribute more).

The picture changes if we consider the ratio of *planar* decompletions $D^{\text{rel}}_{\text{pla}}$ as defined in eq. (2.9). We know from figure 6 that this quantity is essentially uncorrelated with the symmetry factor. Conversely, there is a significant, almost linear correlation of $D^{\text{rel}}_{\text{pla}}$ with the period, see figure 22. This correlation prevails even if the period is weighted with its

JHEP11(2023)160

symmetry factor. Those completed graphs which have a large number of non-isomorphic *planar* decompletions contribute significantly more to the perturbation series than the graphs with few planar decompletions.

6.2 Diameter and mean distance

The findings in section 4.3 can also be interpreted along the lines that, within the same loop number, the period of a graph tends to be larger if the graph is visually "larger". The *diameter* of a graph is defined as the maximum distance between any two vertices, where edges are assumed to have unit length:

$$\operatorname{diam}(G) := \max_{v_1,v_2 \in V_G} (\text{length of the shortest path from } v_1 \text{ to } v_2)\,.$$

Assuming, as in section 2.2, that the space of random 4-regular simple graph $G \in \mathcal{G}_{n,4}$ captures the leading asymptotic behavior of completed primitives, the diameter of an L-loop completion G is expected to grow like $\operatorname{diam}(G) \sim \frac{1}{\ln 3} \ln\left((L+2)\ln(L+2)\right)$ [76].

We have examined the relation between the diameter and the period, but it turns out to be of little use for our case. The reason is that the diameter of a graph is an integer and, for $L \to \infty$, it is known to lie almost surely within $[-2, +3]$ of its asymptotics, i.e. to only assume very few different values. For example, all 755643 completions with $L = 13$ have diameter 2, 3, or 4.

Instead of the maximum distance between two vertices, we rather consider the mean distance

$$d(G) := \operatorname*{mean}_{v_1,v_2 \in V_G} (\text{length of the shortest path from } v_1 \text{ to } v_2)\,.$$

This quantity is still granular, but much less so, compared to the diameter. For the 13-loop graphs, it takes 51 different values between 1.714 and 2.286. In particular, we find that the graph with the largest mean distance is the zigzag and the next smaller distances belong to the "almost-zigzag" shapes discussed in section 4.3.

Figure 23 shows the period as a function of the mean distance d of a graph. Clearly, the period is correlated with d, and both $\mathcal{P}$ and d grow with growing loop order L. What is more remarkable is that, for a fixed value of d, the graphs of different loop orders tend to have similar periods. This shows that the mean distance is not merely a proxy for the loop order, but it carries additional non-trivial information. To guide the eye, we have included in figure 23 the function $10^{-5}e^{12d}$. A few periods lie above this line, but it is remarkable that the overall distribution has a rather sharp upper boundary.

6.3 Girth and small circuits

Another straightforward property of a graph is the girth, that is, the length of the shortest circuit in the graph. Unfortunately, the girth is either 3 (the graph contains a triangle) or 4 (no triangles, but at least one square) for almost all graphs in our samples. This is expected from theorem 1 (3). Instead, it proves more interesting to consider n_3, the number of triangles in the graph.

JHEP11(2023)160

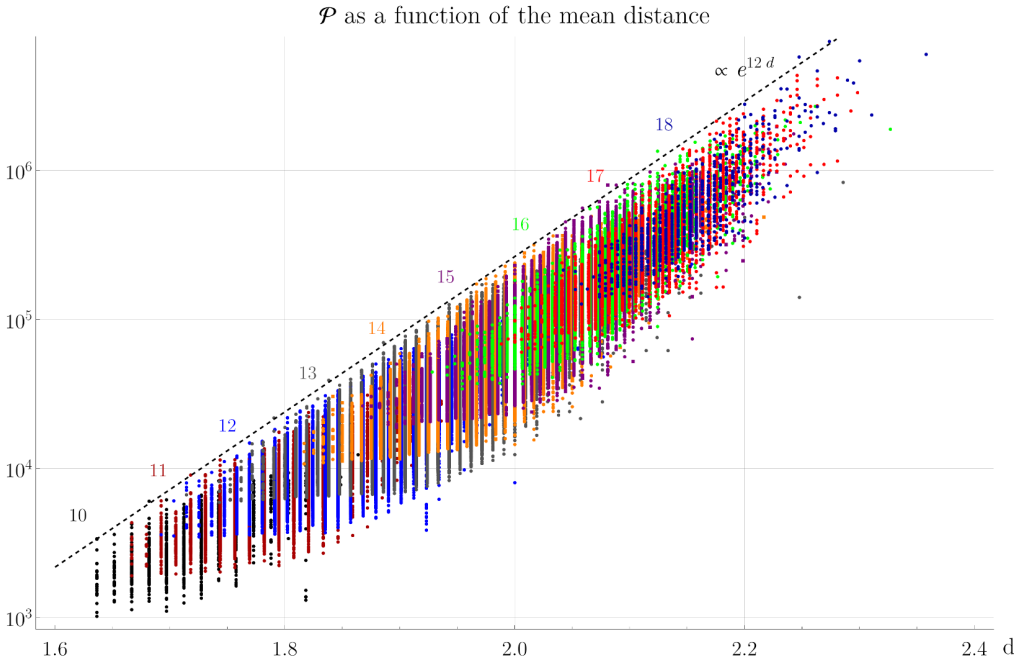


Figure 23. The period as a function of the mean distance d, log plot. Numbers indicate the loop order, graphs of the same order are plotted in the same color. Function $10^{-5}e^{12d}$ (black dashed) to guide the eye. Note that periods tend to be similar when their mean distance is similar, even if the loop order is different.

Unlike with the distance in figure 23, the plots depending on n_3 do not align for different loop orders. To eliminate the trivial dependence on loop order, we divide by $\left(\frac{3}{2}\right)^L L^{\frac{5}{2}}$, which is the leading growth rate eq. (5.12) (up to constants). The resulting distribution of periods is shown in figure 24. The plot suggests an exponential function as a lower bound, empirically we find that all our results for $L \geq 5$ satisfy

$$\mathcal{P} > 0.04 \cdot \left(\frac{3}{2}\right)^L L^{\frac{5}{2}} \exp\left(0.15 \cdot n_3\right). \tag{6.1}$$

One could suspect a similar correlation for the number of 4-circuits n_4 and the number of 5-circuits n_5. But this is not the case. The correlation of n_4 with the period is much weaker than that of n_3, and n_5 is even anticorrelated; graphs with many 5-circuits tend to have smaller periods.

6.4 6-edge cuts

Every 3-circuit passes through 3 vertices of the graph, each of which has 2 edges not in the circuit. Hence, every triangle is connected to the remainder of the graph by exactly 6 edges and it can be disconnected by a 6-edge cut. The converse is not true. In particular, removing from a completion any one of its edges, together with the two adjacent vertices, gives rise to a 6-edge cut. Denote the number of 6-edge cuts by c_6, then $c_6 \geq |E_G| = 2L_G+4$.

From a physics perspective, the number of 6-edge cuts is more interesting than the number of triangles. This is because the Feynman graphs in quantum field theory cor-

JHEP11(2023)160

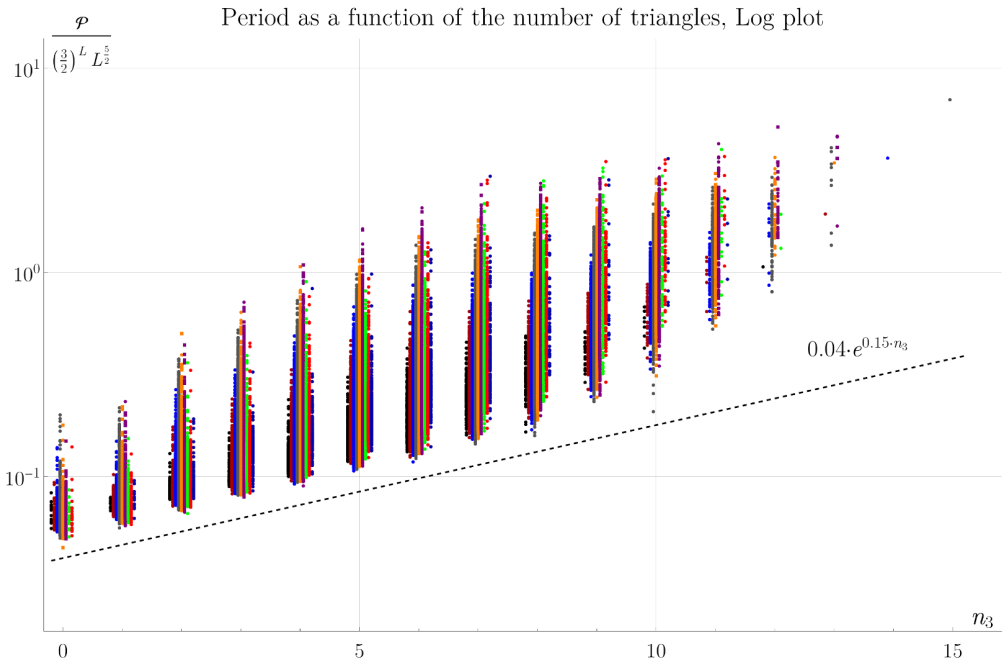


Figure 24. Period, scaled proportional to its leading growth, as a function of the number of triangles n_3. The plot contains all data points for $L \in \{10, \ldots, 18\}$, colors are the same as in figure 23. n_3 only takes integer values, the x-coordinate in the plot is slightly shifted to reduce overlapping.

respond to the same physical observable if they have the same number of external edges. Hence, c_6 can be understood as the number of subgraphs in G which contribute to a 6-point amplitude. In the spirit of Dyson-Schwinger equations, a primitive 4-point amplitude (i.e. a decompletion) can be written as an integral over the 6-point function, where 3 of the legs are joined by a new vertex. This consideration motivates us to examine the relation between the number of 6-edge cuts and the period.

As can be seen from figure 25, c_6 strongly correlates with the period. The precise relationship depends on the loop number, for a fixed loop number, it can be approximated by a function

$$\frac{\mathcal{P}}{\left(\frac{3}{2}\right)^L L^{\frac{5}{2}}} \approx f(c_6) := \left(\frac{c_6 - d}{s}\right)^p, \tag{6.2}$$

where the constants s, d, p are listed in table 15. With growing loop number, the relation between c_6 and $\mathcal{P}$ gets steeper. For our samples, the approximate dependence of the parameters of eq. (6.2) on the loop number is $s \approx 550/L$, $d \approx 30 \ln(L/8)$, and $p \approx 1.2e^{L/16}$, but no systematic approach has been made to justify these functional forms.

Even if the number of 6-edge cuts has a physical interpretation, its predictive power for the period is remarkable, keeping in mind that we are merely *counting* the number of 6-valent subgraphs, not actually modeling their contribution to recursive integrals. As indicated in table 15, simply counting n_6 allows to predict an individual period with an average uncertainty of less than 15%.

JHEP11(2023)160

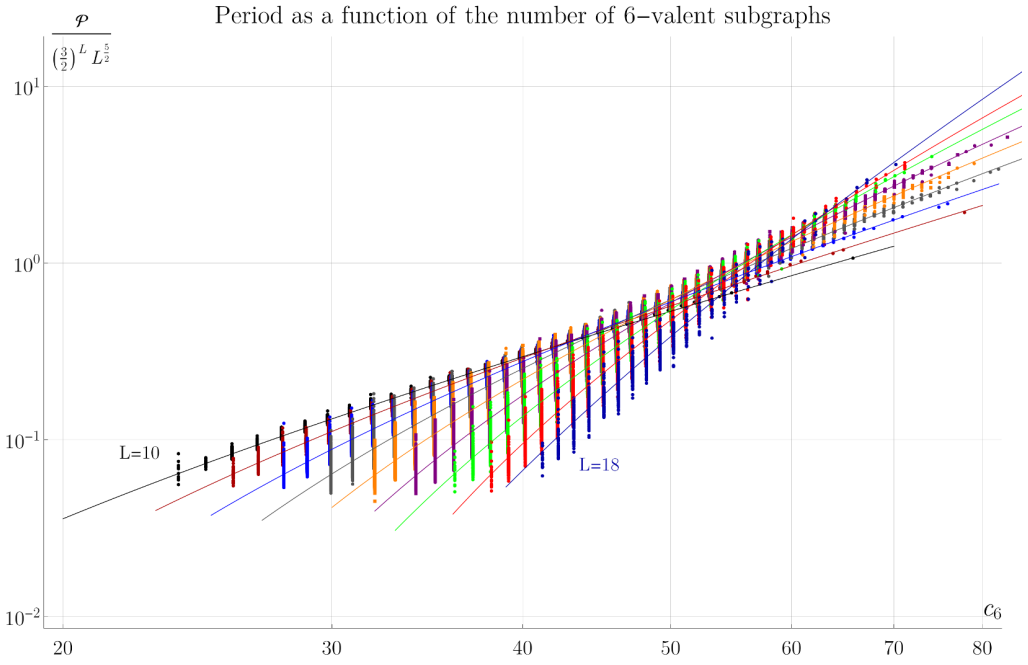


Figure 25. Period as a function of the number of 6-edge cuts c_6, double logarithmic plot. Colors and scaling as in figure 24. Solid lines show best fit functions eq. (6.2).

L	s	d	p	$\left\langle \left\lvert \frac{\mathcal{P}-f(c_6)}{\mathcal{P}} \right\rvert \right\rangle$	$\max \left\lvert \frac{\mathcal{P}-f(c_6)}{\mathcal{P}} \right\rvert$
8	71.810	1.2410	2.0045	0.007243	0.02946
9	63.566	4.6942	2.0804	0.013341	0.05797
10	56.368	7.6495	2.1934	0.020592	0.20688
11	50.268	10.475	2.3188	0.028544	0.25617
12	45.354	13.035	2.4653	0.039924	0.47640
13	41.446	15.410	2.6290	0.054229	0.50639
13s	41.570	15.244	2.6478	0.053728	0.29752
13⋆	41.700	15.156	2.6520	0.054431	0.31633
14s	38.356	17.613	2.8151	0.070503	0.41371
14⋆	38.662	17.342	2.8360	0.070642	0.43935
15s	35.923	19.770	2.9976	0.087821	0.48885
15⋆	36.047	19.671	3.0043	0.088337	0.45205
16s	35.412	20.452	3.3562	0.10536	0.49715
17s	34.041	22.312	3.5891	0.12145	0.55486
18s	39.530	17.736	4.7019	0.13513	0.55301

Table 15. The columns describe the following: best fit parameters for the function eq. (6.2) to approximate the period; average relative error of the approximation; maximum relative error.

JHEP11(2023)160

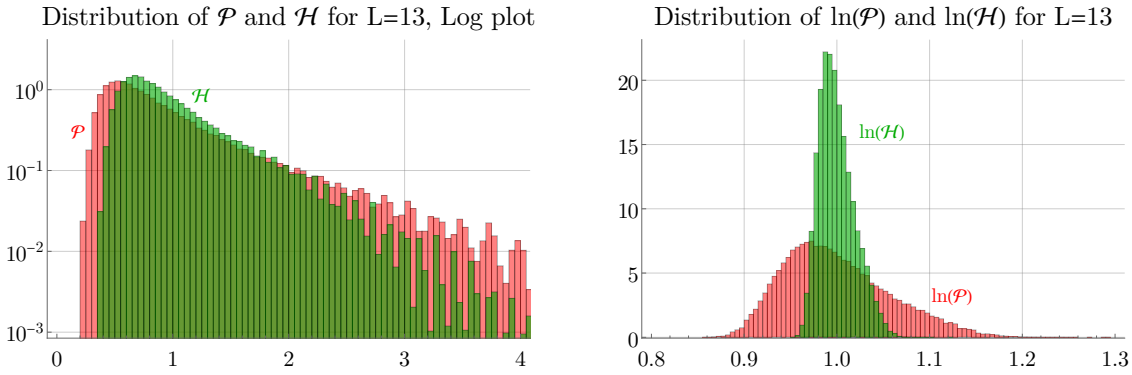


Figure 26. Histograms of the Hepp bound (green) and the period (red), both without symmetry factors.

6.5 Hepp bound

The Hepp bound $\mathcal{H}(G)$ enters the tropical integration algorithm [33] as a weighting factor for the contribution of sector integrals, consequently, we obtain this number automatically for every graph which is computed numerically. As explained in section 3.3, we also wrote our own program to compute Hepp bounds, based on the flag formula eq. (3.1).

Before we turn to the correlation with the period, it is interesting to inspect the distribution of the Hepp bound itself. In figure 26, we see that the Hepp bound, or its logarithm, are more concentrated around the mean, and have lower influence of tails, compared to the distribution of periods.

Much like the period in section 4, we can find mean and moments for the Hepp bound, they are reported in table 16, both for $\mathcal{H}$ and for $\ln \mathcal{H}$. The overall trend is similar to that of the period in tables 6 and 13, but the value of the Hepp bound is considerably larger, and grows faster, than that of the period. Conversely, the Hepp bound has smaller moments, indicating that its distribution is overall more narrow, less skew, and more concentrated around the mean. Nevertheless, moments grow monotonically with the loop order, indicating that the limiting distribution has infinite moments even for the Hepp bound. We note in passing that, similar to the period in section 4.5, the distribution of the transformed Hepp bound $\mathcal{H}^t$ comes close to a normal distribution, but for a power $t \approx 0.72$.

It is known that the Hepp bound constitutes a strong prediction for the period, and that their relationship takes almost the same form across different loop orders [54]. Concretely, the points

$$\left(\frac{\ln \mathcal{H} - \ln 2}{L-1}, \quad \frac{\ln(\mathcal{P})}{L-1} \right) \tag{6.3}$$

almost lie on the same curve for different loop orders L. In figure 27, we plot $\mathcal{P}$ vs. $\mathcal{H}$ in logarithmic coordinates, but without the rescaling eq. (6.3) in order to avoid overlap.

The almost-linear relation eq. (6.3) of the logarithms is equivalent to an almost-power-law relation $\mathcal{P} \sim a \cdot \mathcal{H}^b$. In [36], the function $\mathcal{P} \approx a\mathcal{H}^b(1-c\mathcal{H})$, with appropriate parameters a, b, c, was introduced to describe the relationship. This choice works well empirically, but

JHEP11(2023)160

L	$\langle\mathcal{H}\rangle$	C_2	c_3	c_4	$\langle\ln\mathcal{H}\rangle$	$C_2\cdot 10^6$	c_3	c_4
7	$1.29107\cdot 10^5$	0.0490	0.452	2.443	11.7442	349.6	0.06959	2.132
8	$7.42608\cdot 10^5$	0.0792	0.739	3.633	13.4796	421.4	0.05641	2.484
9	$4.24179\cdot 10^6$	0.0970	1.045	5.052	15.2152	384.8	0.1924	2.599
10	$2.33357\cdot 10^7$	0.1238	1.196	5.742	16.9091	384.2	0.2781	2.556
11	$1.25530\cdot 10^8$	0.1549	1.362	6.211	18.5800	378.4	0.3949	2.602
12	$6.58540\cdot 10^8$	0.1925	1.592	7.000	20.2248	371.5	0.5114	2.772
13	$3.38768\cdot 10^9$	0.2334	1.878	8.405	21.8507	358.7	0.6168	3.004
13s	$3.38236\cdot 10^9$	0.2332	1.812	7.411	21.8490	358.9	0.6284	2.986
13⋆	$3.39141\cdot 10^9$	0.2374	1.853	7.575	21.8505	362.4	0.6373	3.005
14s	$1.71560\cdot 10^{10}$	0.2762	2.170	9.944	23.4616	342.7	0.7031	3.237
14⋆	$1.71555\cdot 10^{10}$	0.2750	2.166	9.774	23.4621	340.8	0.7060	3.255
15s	$8.52439\cdot 10^{10}$	0.3139	2.594	13.86	25.0570	317.5	0.7676	3.508
15⋆	$8.54667\cdot 10^{10}$	0.3213	2.554	13.23	25.0574	322.7	0.7896	3.509
16s	$4.17260\cdot 10^{11}$	0.3345	2.661	13.94	26.6400	292.2	0.7910	3.564
17s	$2.03998\cdot 10^{12}$	0.3700	2.974	16.65	28.2210	269.3	0.8587	3.852
18s	$1.00353\cdot 10^{13}$	0.3984	3.084	17.05	29.8084	249.9	0.9120	3.934

Table 16. The columns describe the following: distribution parameters of the Hepp bound in our samples, without symmetry factors; parameters of the logarithm of the Hepp bound. In both cases, moments C_j, c_j (eq. (4.3)) refer to a normalized distribution with unit mean and the samples have been adjusted to be uniformly distributed (appendix A).

it has the drawback of parameters which are far from unity and strongly vary with loop order, making the fit procedure unstable and obscuring the regularity between different loop orders. Instead, we use the function

$$\mathcal{P}\approx g(\mathcal{H}) := \left(\mathcal{H}e^{-a(L+1)} - e^{c(L+1)}\right)^b . \tag{6.4}$$

This function is not equivalent to $a\mathcal{H}^b(1-c\mathcal{H})$, and we have found that our choice gives rise to lower average fit residues in all samples. As anticipated, the fit parameters of eq. (6.4) are of order unity, see table 17. We remark that although the relationship figure 27 of the logarithms is almost linear, we have determined the parameters of table 17 with respect to $\mathcal{H}$ and $\mathcal{P}$, not with respect to their logarithms. This is because fitting with respect to logarithms would put disproportional weight on the smaller periods, which have only small importance in the perturbation series eq. (1.1).

The residues of the approximation eq. (6.4) are shown in figure 28(a). The most remarkable feature of that plot is that the residues fall into discrete bands. As was pointed out by Erik Panzer,[4] all periods within one band share the same value of c_6, the number

[4]Talk at research seminar *structure of local quantum field theories*, Humboldt-Universität zu Berlin, December 20th, 2021.

JHEP11(2023)160

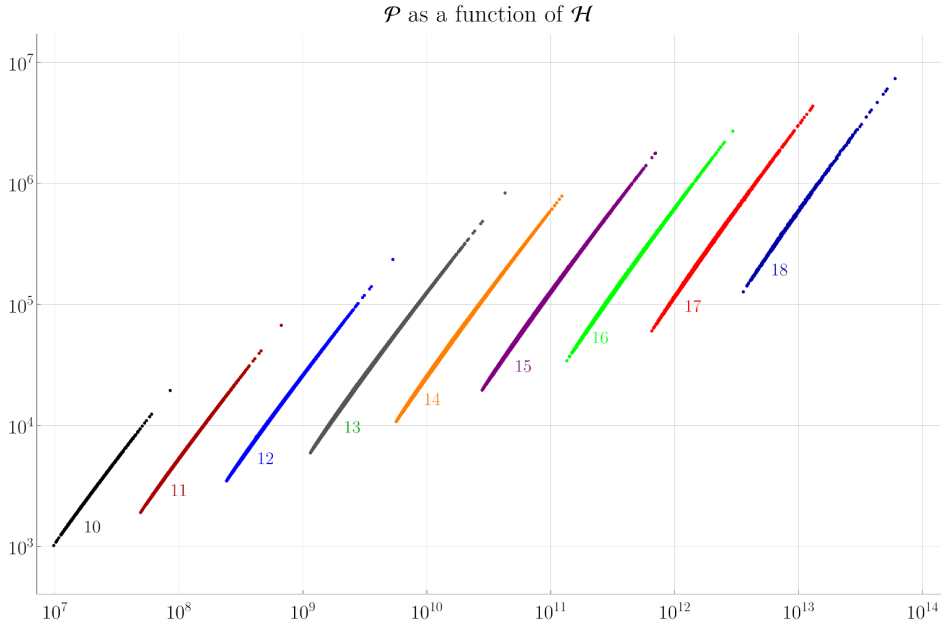
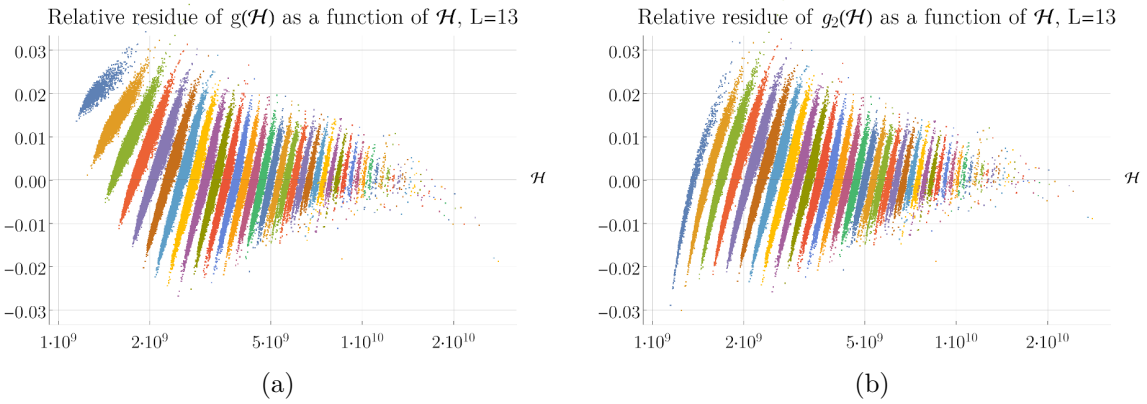


Figure 27. The period as a function of the Hepp bound, for various loop orders, Double-logarithmic plot. The single large outlier point for $10 \leq L \leq 13$ is the zigzag graph (eq. (4.11)).

L	a	b	c	$\left\langle \left\lvert \frac{\mathcal{P}-g(\mathcal{H})}{\mathcal{P}} \right\rvert \right\rangle$	$\max \left\lvert \frac{\mathcal{P}-g(\mathcal{H})}{\mathcal{P}} \right\rvert$
7	0.9098	1.2915	0.16039	0.0008	0.0031
8	0.9312	1.2909	0.22236	0.0012	0.0038
9	0.9480	1.2899	0.27444	0.0017	0.0071
10	0.9638	1.2930	0.30877	0.0022	0.0163
11	0.9786	1.2986	0.33197	0.0028	0.0197
12	0.9927	1.3063	0.34686	0.0035	0.0374
13	1.0056	1.3146	0.35726	0.0044	0.0405
13s	1.0070	1.3175	0.35294	0.0043	0.0236
13⋆	1.0067	1.3169	0.35348	0.0044	0.0272
14s	1.0183	1.3252	0.36197	0.0055	0.0350
14⋆	1.0179	1.3243	0.36276	0.0055	0.0336
15s	1.0284	1.3322	0.36929	0.0069	0.0385
15⋆	1.0278	1.3310	0.37085	0.0070	0.0390
16s	1.0399	1.3442	0.36836	0.0081	0.0405
17s	1.0489	1.3523	0.37138	0.0096	0.0477
18s	1.0611	1.3689	0.36250	0.0102	0.0398

Table 17. The columns describe the following: loop order; 3 columns best fit parameters for the approximation of the period by the Hepp bound according to eq. (6.4); average and maximum relative error.

JHEP11(2023)160



(a) (b)

Figure 28. (a) Residue of the best fit function g (eq. (6.4)) approximating the period in figure 27. The plots show the quantity $\frac{\mathcal{P}-g(\mathcal{H})}{\mathcal{P}}$. Colors indicate the number c_6 of 6-edge cuts, this number determines which "band" a residue lies in. Overall, the residues show an S-shaped pattern. (b) The S-shape is absent in the residues of g_2 (eq. (6.5)).

of 6-edge cuts. This is indicated by colors in figure 28(a). Besides the band structure, the residues in figure 28(a) display an overall S-shape which suggests to add a correction term cubic in $\ln(\mathcal{H})$ to the fit function eq. (6.4). We define

$$\mathcal{P} \approx g_2(\mathcal{H}) := \left(\mathcal{H}e^{-a(L+1)} - e^{c(L+1)}\right)^b - \left(d\ln(\mathcal{H}) - e(L+1)\right)^3 . \tag{6.5}$$

Best fit parameters for this function are shown in table 18. They refer to the samples "as is", without adjustment for non-uniform sampling.

The resulting residues are shown in figure 28(b). As expected, the S-shape has been removed, but in terms of mean error, the fit function eq. (6.5) is only marginally superior to eq. (6.4), whereas the fit is numerically much more unstable. The reason for this seeming contradiction is that, even if the extremely large or small values of $\mathcal{H}$ in the plots figure 28(b) have been moved towards zero, the large majority of data points lies in the central region and is unaffected by the introduction of a cubic term in eq. (6.5).

6.6 Martin invariant

As observed in [59], the Martin invariant (section 3.2) is correlated with the period.

For our data, the relation between the period and the Martin invariant is shown in figure 29. The correlation is close, but not quite as close as for the period and the Hepp bound (figure 27). Moreover, the slope is inverted, i.e. a large Martin invariant indicates a small period. Analogously to section 6.5, it is possible to define functions that depend on only few parameters and approximate the relationship $M \mapsto \mathcal{P}$ to good accuracy. Again, the residues fall into bands indexed by c_6, but the structure is not as pronounced as in figure 28. More details are subject of an upcoming article of the author.

This concludes our survey of statistical properties of Feynman periods in ϕ^4-theory. A summary as been given already in section 1.4 in the introduction, and a discussion in section 1.5.

JHEP11(2023)160

L	a	b	c	d	e	$\left\langle \left\lvert \frac{\mathcal{P}-g_2(\mathcal{H})}{\mathcal{P}} \right\rvert \right\rangle$	$\max \left\lvert \frac{\mathcal{P}-g_2(\mathcal{H})}{\mathcal{P}} \right\rvert$
7	0.6246	1.0019	0.62034	2.7385	3.0879	0.0007	0.0028
8	0.8052	1.1318	0.45202	2.5920	3.1158	0.0012	0.0042
9	0.9429	1.2813	0.33590	−0.0391	0.3138	0.0017	0.0070
10	0.9540	1.2765	0.39125	0.0670	0.6185	0.0022	0.0147
11	0.9669	1.2784	0.41731	0.1305	0.8447	0.0027	0.0193
12	0.9801	1.2841	0.43058	0.1644	1.0321	0.0034	0.0369
13	0.9931	1.2920	0.43684	0.1807	1.2030	0.0042	0.0417
13s	0.9962	1.2978	0.42483	0.1496	1.0959	0.0041	0.0233
13⋆	0.9963	1.2979	0.42323	0.1720	1.1180	0.0042	0.0256
14s	1.0070	1.3040	0.43235	0.2177	1.3740	0.0052	0.0356
14⋆	1.0070	1.3038	0.43251	0.2230	1.3837	0.0052	0.0389
15s	1.0168	1.3100	0.43816	0.3269	1.7507	0.0063	0.0388
15⋆	1.0161	1.3087	0.44043	0.3337	1.7798	0.0064	0.0384
16s	1.0291	1.3227	0.43285	0.3012	1.8896	0.0074	0.0386
17s	1.0374	1.3288	0.43523	0.4865	2.4404	0.0086	0.0375
18s	1.0510	1.3476	0.42446	−0.0372	1.8750	0.0098	0.0412

Table 18. The columns describe the following: loop order; 5 columns best fit parameters for the approximation of the period by the Hepp bound according to eq. (6.5); relative errors. Despite having two additional degrees of freedom, and looking visually more appropriate, the fit function g_2 produces only marginally smaller approximation errors than g in table 17.

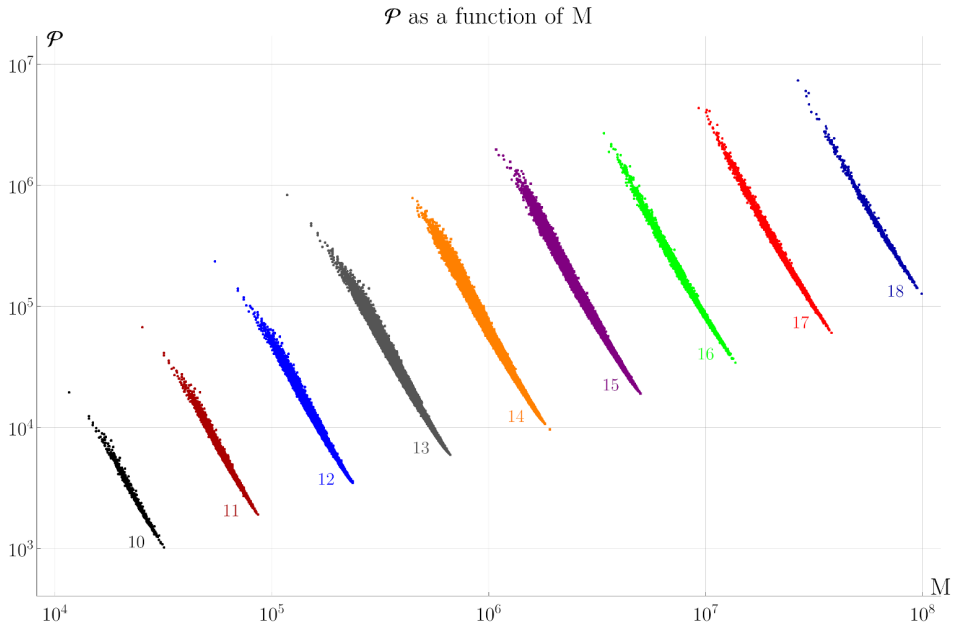


Figure 29. The period as a function of the (first) Martin invariant, for various loop orders, Double-logarithmic plot. Compared to the Hepp bound (figure 27), the relation is not as strict, and the slopes are reversed.

JHEP11(2023)160

Acknowledgments

Most of the numerical computations of this work were done while the author was affiliated with Humboldt-Universität zu Berlin.

The author thanks Karen Yeats, David Broadhurst, Gerald Dunne, Erik Panzer, Michael Borinsky, and Dirk Kreimer for discussion and comments.

A Non-uniform sampling

The samples generated by `genrang` are not distributed uniformly, but proportionally to their symmetry factor. The weight ρ of a graph G in the sample S is

$$\rho(G) = \frac{\frac{N}{|\mathrm{Aut}(G)|}}{\sum_{g\in S}\frac{1}{|\mathrm{Aut}(g)|}}.$$

Note that this weight is normalized to the total number of graphs N, and not to unity. Naively averaging over the sample S leads to the weighted mean

$$\langle \mathcal{P} \rangle_\rho := \frac{1}{N}\sum_{G\in S}\rho(G)\mathcal{P}(G).$$

To produce the un-weighted mean (eq. (4.1)), we can replace each graph G in the sample by $|\mathrm{Aut}(G)|$ copies of G. This gives rise to a "corrected" sample

$$S_{\mathrm{corr}} := \left\{ |\mathrm{Aut}(G)| \text{ copies of } G \,\middle|\, G \in S \right\}. \tag{A.1}$$

S_{corr} has uniform distribution and contains $N_{\mathrm{corr}} = \sum_{g\in S}|\mathrm{Aut}(g)|$ graphs, the mean is

$$\langle \mathcal{P} \rangle = \frac{1}{N_{\mathrm{corr}}}\sum_{G\in S_{\mathrm{corr}}}\mathcal{P}(G) = \frac{1}{N_{\mathrm{corr}}}\sum_{G\in S}\rho(G)\,|\mathrm{Aut}(G)|\,\mathcal{P}(G) = \left(\frac{N}{\sum_{g\in S}|\mathrm{Aut}(g)|}\right)\langle |\mathrm{Aut}|\cdot\mathcal{P}\rangle_\rho .$$

Using

$$\sum_{G\in S_{\mathrm{corr}}}\frac{1}{|\mathrm{Aut}(G)|} = N, \qquad \sum_{G\in S}|\mathrm{Aut}(G)| = N_{\mathrm{corr}} = |S_{\mathrm{corr}}|,$$

the correction factor can be identified as the average symmetry factor,

$$\frac{N}{\sum_{g\in S}|\mathrm{Aut}(g)|} = \frac{N}{N_{\mathrm{corr}}} = \frac{1}{|S_{\mathrm{corr}}|}\sum_{G\in S_{\mathrm{corr}}}\frac{1}{|\mathrm{Aut}(G)|} = \left\langle \frac{1}{|\mathrm{Aut}|}\right\rangle. \tag{A.2}$$

Adding copies of a graph to the sample as in eq. (A.1) has the disadvantage that we slightly underestimate the sampling uncertainty, because the copies are not independent additional samples. A different approach to obtain the correct mean is by scaling each individual period according to

$$\mathcal{P}(G) \mapsto \left\langle \frac{1}{|\mathrm{Aut}|}\right\rangle \cdot |\mathrm{Aut}(G)|\cdot\mathcal{P}(G). \tag{A.3}$$

JHEP11(2023)160

While this scaling does not influence the sample size, it significantly distorts the higher moments because it artificially introduces very large "effective periods". Hence, we use the first approach.

For the mean weighted by symmetry factors (eq. (4.5)), the graphs generated by `genrang` have the correct relative weighting, but still the normalization factor is wrong. To see this, note that if we were to weight the graphs in the set S_{corr} (eq. (A.1)) with respect to their symmetry factor, we reproduce the relative weight within S, but the number of elements is $N_{\text{corr}} = |S_{\text{corr}}| > |S| = N$. By eq. (A.2), the ratio of the two sizes is precisely the average symmetry factor, so eq. (4.5) can be computed as

$$\left\langle \frac{\mathcal{P}}{|\text{Aut}|} \right\rangle = \frac{1}{N_{\text{corr}}} \sum_{G \in S_{\text{corr}}} \frac{\mathcal{P}(G)}{|\text{Aut}(G)|} = \left\langle \frac{1}{|\text{Aut}|} \right\rangle \langle \mathcal{P} \rangle_\rho \,.$$

The beta function in section 5 requires not the average, but the sum of periods. Let U be the sought-after sum over *all* graphs G (not just the sample), a priori, it is

$$U := \sum_G \frac{\mathcal{P}(G)}{|\text{Aut}(G)|} = N_L^{(C)} \left\langle \frac{\mathcal{P}}{|\text{Aut}|} \right\rangle = N_L^{(C)} \left\langle \frac{1}{|\text{Aut}|} \right\rangle \langle \mathcal{P} \rangle_\rho \,.$$

However, the total number of completions $N_L^{(C)}$ is unknown for large L, therefore, it is useful to express U in terms of $N_L^{(\text{Aut})}$ (eq. (2.5)) according to

$$U = \frac{N_L^{(\text{Aut})}}{4!(L+2)} \cdot \langle \mathcal{P} \rangle_\rho \,. \tag{A.4}$$

Eq. (A.4) is exact, this means that it does not involve any additional uncertainty from e.g. approximating $N_L^{(C)}$ by $N_L^{(\text{Aut})}$ or $\langle |\text{Aut}| \rangle \approx 1$. These effects precisely cancel in the end result.

B Sampling uncertainties and higher moments

This section is a review of basic facts of sampling statistics, for more details see e.g. [77–80]. To simplify notation, we will call the random variable X.

It is important to distinguish between true properties of the underlying distribution, and the observed properties of finite samples. By $\mathbb{E}(X)$, we denote the expectation, by $\bar{M}_j(X)$ the j^{th} moment and by $\bar{C}_j(X)$ the j^{th} central moment,

$$\bar{M}_j = \bar{M}_j(X) = \mathbb{E}\left(X^j\right), \qquad \bar{C}_j = \bar{C}_j(X) = M_j\left(X - M_1(X)\right) = \mathbb{E}\left(X - \mathbb{E}(X)\right)^j .$$

These quantities are constant numbers (not random variables), and if the symbols appear without argument they refer to X. By construction $\bar{M}_0 = \bar{C}_0 = 1$, $\bar{C}_1 = 0$, and

$$\bar{C}_n = \sum_{j=0}^{n} \binom{n}{j} (-\bar{M}_1)^{n-j} \bar{M}_j, \qquad \bar{C}_2 = \bar{M}_2 - \bar{M}_1^2 \tag{B.1}$$

$$\bar{C}_3 = \bar{M}_3 - 3\bar{M}_1\bar{M}_2 + 2\bar{M}_1^3, \qquad \bar{C}_4 = \bar{M}_4 - 4\bar{M}_1\bar{M}_3 + 6\bar{M}_1^2\bar{M}_2 - 3\bar{M}_1^4.$$

JHEP11(2023)160

We further define the *skewness* and the *kurtosis* like in eq. (4.3)

$$\bar{c}_3 = \bar{C}_3 \bar{C}_2^{-\frac{3}{2}}, \qquad \bar{c}_4 = \bar{C}_4 \bar{C}_2^{-2}. \tag{B.2}$$

We refer to X as the *population*. Now consider a sample $X_n = \{x_1, \ldots, x_n\}$ of n elements drawn from that population *with replacement*. This means that each element follows the distribution of X and all elements are independent. In particular, for a finite population, an element is allowed to be present in the sample multiple times. By $\langle X_n \rangle$ we denote the sample average according to eq. (4.1), and by $M_j(X_n)$ and $C_j(X_n)$ the sample moments (note that in the main text eq. (4.3), we report sample moments scaled to the sample mean) by:

$$\langle X_n \rangle = \frac{1}{n} \sum_{k=1}^{n} x_k, \qquad M_j(X_n) = \left\langle X_n^j \right\rangle, \qquad C_j(X_n) = \left\langle (X_n - \langle X_n \rangle)^j \right\rangle. \tag{B.3}$$

We can probe the population only through samples, therefore, we need to know how the moments of the sample are related to the ones of the population. The *characteristic function* of a random variable is the Fourier transform of its density $\rho_Y(y)$, or the expectation of its complex exponential function,

$$\phi_X(t) = \int \mathrm{d}x \, \rho_X(x) e^{itx} = \mathbb{E} e^{itX} = \sum_{j=0}^{\infty} \mathbb{E}\left(\frac{(it)^j X^j}{j!} \right) = \sum_{j=0}^{\infty} \frac{(it)^j}{j!} \bar{M}_j. \tag{B.4}$$

We have used the linearity of the expectation to recognize the moments $\bar{M}_j = \mathbb{E}(X^j)$. For the sum of n copies of X, using the fact that these copies are independent and thus the expectation of the product is the product of expectations, one finds the characteristic function

$$\phi_{\langle X_n \rangle}(t) = \mathbb{E} e^{it\langle X_n \rangle} = \mathbb{E} e^{it\frac{1}{n}\sum_{j=1}^{n} x_j} = \mathbb{E} \prod_{j=1}^{n} e^{it\frac{1}{n}x_j} = \prod_{j=1}^{n} \mathbb{E} e^{i\frac{t}{n}X} = \left(\phi_X\left(\frac{t}{n}\right) \right)^n.$$

By eq. (B.4), the moments of a distribution are given by derivatives of the corresponding characteristic function. Evaluating the derivatives of $\phi_{\langle X_n \rangle}(t)$, we find the moments $\bar{M}_j\left(\langle X_n \rangle\right)$ of the n-element sample:

$$\bar{M}_j(\langle X_n \rangle) = (-i)^j \frac{\mathrm{d}^j \, \phi_{\langle X_n \rangle}(t)}{\mathrm{d}t^j}\bigg|_{t=0} = \frac{n!}{n^j} \sum_{k=0}^{j} \frac{1}{(n-k)!} B_{j,k}\left(\bar{M}_1, \bar{M}_2, \bar{M}_3, \ldots\right). \tag{B.5}$$

For the last equality, we have used Faà di Bruno's formula [81] and introduced the incomplete Bell polynomials $B_{j,k}$ [82]. The first moments are

$$\begin{aligned} \bar{M}_1(\langle X_n \rangle) &= \bar{M}_1, & \bar{M}_2(\langle X_n \rangle) &= \frac{n-1}{n}\bar{M}_1^2 + \frac{1}{n}\bar{M}_2 \\ \bar{M}_3(\langle X_n \rangle) &= \frac{(n-1)(n-2)}{n^2}\bar{M}_1^3 + \frac{3(n-1)}{n^2}\bar{M}_1\bar{M}_2 + \frac{1}{n^2}\bar{M}_3. \end{aligned} \tag{B.6}$$

We define the *sampling uncertainty* of the sample mean $\langle X_n \rangle$ as the square root of the variance of the sample mean, it is related to moments according to eq. (B.1):

$$(\Delta_{\text{samp}} \langle X_n \rangle)^2 = \bar{C}_2(\langle X_n \rangle) = \bar{M}_2(\langle X_n \rangle) - \left(\bar{M}_1(\langle X_n \rangle)\right)^2. \tag{B.7}$$

JHEP11(2023)160

By eq. (B.6), it is proportional to the population standard deviation $\sqrt{\bar{C}_2}$:

$$\Delta_{\text{samp}} \langle X_n \rangle = \sqrt{\frac{n-1}{n}\bar{M}_1^2 + \frac{1}{n}\bar{M}_2 - \bar{M}_1^2} = \frac{1}{\sqrt{n}}\sqrt{\bar{M}_2 - \bar{M}_1^2} = \frac{1}{\sqrt{n}}\sqrt{\bar{C}_2}. \tag{B.8}$$

We need to estimate $\sqrt{\bar{C}_2}$ from a finite sample. The naive sample variance according to eq. (B.3), $C_2(X_n) = \langle X_n^2 \rangle$, is systematically biased because it relies on the sample mean, which is not independent. The sample moments itself have the correct expectation,

$$\mathbb{E}\left(M_j(X_n)\right) = \mathbb{E}\left\langle X_n^j \right\rangle = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(x_k^j) = \frac{1}{n}\sum_{k=1}^{n}\bar{M}_j = \bar{M}_j,$$

but products of sample moments with individual sample elements are biased,

$$\mathbb{E}\left(x_i^l \left\langle X_n^j \right\rangle\right) = \mathbb{E}\left(x_i^l \frac{1}{n}\sum_{k=1}^{n} x_k^j\right) = \mathbb{E}\left(\frac{1}{n}\sum_{k\neq i} x_i^l x_k^j + \frac{1}{n}x_i^{j+l}\right) = \frac{n-1}{n}\bar{M}_l\bar{M}_j + \frac{1}{n}\bar{M}_{j+l}.$$

This implies that the expectation of the sample variance contains a non-trivial factor,

$$\mathbb{E}\left(C_2(X_n)\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{E}x_i^2 - 2\mathbb{E}x_i \langle X_n \rangle + \langle X_n \rangle^2\right) = \frac{n-1}{n}\bar{C}_2.$$

Including this correction factor, the unbiased estimator of the population variance is

$$\tilde{C}_2(X_n) = \frac{n}{n-1}C_2(X_n) = \frac{1}{n-1}\sum_{j=1}^{n}\left(x_j - \sum_{k=1}^{n} x_j\right)^2. \tag{B.9}$$

Similar corrections exist for higher moments. The sample moments reported in the paper contain these corrections, but the influence is minuscule since the sample sizes are $n \geq 900$.

The sample variance eq. (B.9) is itself a random variable, its value depends on the particular sample, while the population variance $\bar{C}_2$ is a constant. This implies that powers of the variance, such as the standard deviation or the denominators in eq. (B.2), are generally not equal to (the expectation of) the same powers of the sample variance. The necessary correction factors are known analytically if X is a normal distribution [83]. To find them for arbitrary distributions, rewrite the sample variance as

$$C_2(X_n)^r = \bar{C}_2^r\left(\frac{C_2(X_n)}{\bar{C}_2}\right)^r = \bar{C}_2^r\left(1 + \frac{C_2(X_n) - \bar{C}_2}{\bar{C}_2}\right) = \bar{C}_2^r\sum_{k=0}^{\infty}\binom{r}{k}\frac{1}{\bar{C}_2^k}\left(C_2(X_n) - \bar{C}_2\right)^k.$$

Consequently, the expectation of $C_2(X_n)^r$ is a power series of its k^{th} central moments,

$$\mathbb{E}\left(C_2(X_n)^r\right) = \bar{C}_2^r\sum_{k=0}^{\infty}\binom{r}{k}\frac{1}{\bar{C}_2^k}\mathbb{E}\left(C_2(X_n) - \bar{C}_2\right)^k = \bar{C}_2^r\sum_{k=0}^{\infty}\binom{r}{k}\frac{1}{\bar{C}_2^k}\bar{C}_k\Big(C_2(X_n)\Big).$$

Formulas for these moments can be found in [84]. The leading term $\bar{C}_2\left(C_2(X_n)\right)$ is determined by the population kurtosis $\bar{c}_4 = \frac{\bar{C}_4}{\bar{C}_2^2}$ (eq. (B.2)),

$$\mathbb{E}\left(C_2^r\right) = \bar{C}_2^r\left(1 + \frac{r(r-1)}{2}\frac{\bar{c}_4 - 1}{n} + \mathcal{O}\left(\frac{1}{n^2}\right)\right). \tag{B.10}$$

JHEP11(2023)160

Equation (B.10) is one example for how the knowledge of distribution parameters can be used to improve the error estimates of small samples. Corrections of this type are not used in the main text since they are vanishing small for our sample sizes, compare appendix C.

Everything above is valid for sampling *with replacement*. Conversely, sampling from a finite population of size N *without replacement* means that every element can be selected at most once. At $n = N$, the sample contains the entire population. In that setting, the individual sample elements are no longer independent because their possible value depends on which elements are contained in the sample already. The variance of the sample mean in this case requires a *finite population correction* factor [78] and eq. (B.8) becomes

$$\Delta_{\text{samp}} \langle X_n \rangle = \sqrt{\frac{N-n}{N-1}} \frac{1}{\sqrt{n}} \sqrt{\bar{C}_2}. \tag{B.11}$$

This factor is irrelevant for the main text since $n \ll N$ for all our samples, but we demonstrate its effect in appendix C.

C Practical example: sampling from $L = 13$ periods

In this section, we use the $L = 13$ data set, weighted with symmetry factors, to illustrate sampling accuracy. From tables 2 and 7 and eq. (4.4), the relevant numbers are

$$\bar{M}_1 = \left\langle \frac{\mathcal{P}}{|\text{Aut}|} \right\rangle \approx 26320, \qquad \sigma = \sqrt{\bar{C}_2} \approx 18079, \qquad N = N_L^{(C)} = 755643.$$

Concretely, we fix a sample size n and draw 1000 random samples of n periods without replacement. For each of the 1000 samples, we compute the sample mean $\langle X_n \rangle$ (eq. (B.3)) and the sample variance $\tilde{C}_2(X_n)$ (eq. (B.9)). We then compute the empirical standard deviation of the 1000 sample means, which is the sampling uncertainty $\Delta_{\text{samp}} \langle X_n \rangle$ (eq. (B.7)). This procedure is repeated for various sample sizes n.

The outcomes are shown in figure 30 as black data points, where the axes are scaled relative to the population size $N_L^{(C)}$ respective the population mean $\bar{M}_1$. For all $n \ll N_L^{(C)}$, these dots lie on the solid blue line, that is, the observed sampling uncertainty of the mean is described accurately by eq. (B.8). In particular, this formula is still true for very small samples as long as one uses the population standard deviation σ.

If the sample is very small, $n \approx 10$, then the naive sample standard deviation $\sqrt{\tilde{C}_2(X_n)}$ (eq. (B.9)) systematically underestimates the true standard deviation and consequently the sampling uncertainty is estimated too small. This is shown as a red line in figure 30. In this regime, we can use the kurtosis correction eq. (B.10) to obtain a much better estimate of the sampling uncertainty (green line in figure 30). Note that we used the true population kurtosis $\bar{c}_4 = 12.7$, not the kurtosis measured in the small sample itself. This highlights the importance of the population higher moments for estimating the reliability of small samples. For $n > 100$, this effect is barely visible.

The kurtosis correction might appear somewhat tautological because from a small sample it is very hard to obtain an accurate kurtosis. However, we do not necessarily need

JHEP11(2023)160

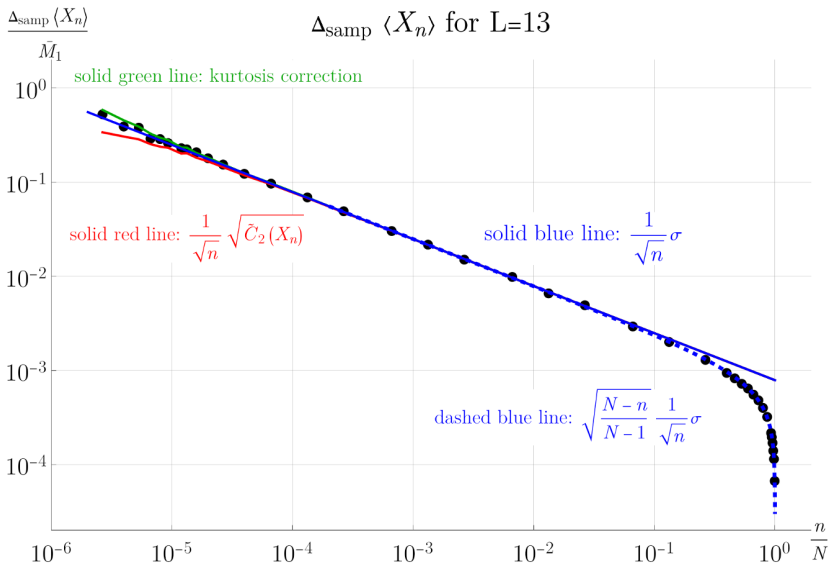


Figure 30. Black dots: empirical sampling uncertainty for different sample sizes of the data set $\frac{\mathcal{P}}{|\text{Aut}|}$ for $L = 13$ loops. As long as $n \ll N_L^{(C)}$, the observed accuracy is accurately described by eq. (B.8) (solid blue line). Only when n is similar to the population size $N = N_L^{(C)}$, the finite population correction eq. (B.11) (dashed blue line) is necessary. For very small samples $n \leq 10$, the sample standard deviation $\sqrt{\tilde{C}_2(X_n)}$ (eq. (B.9)) underestimates the true standard deviation and thus the sample uncertainty (red line). This can be corrected with the help of the population kurtosis according to eq. (B.10) (green line).

an accurate value. For example, suppose we want to sample periods at $L = 20$ loops. By eq. (4.6) and table 8, the kurtosis at higher loop order roughly grows like

$$c_4 = \frac{C_4}{C_2^2} \approx \frac{e^{a_4}\left(\frac{L}{8}\right)^{b_4}}{e^{2a_2}\left(\frac{L}{8}\right)^{2b_2}} \approx e^{1.5}\left(\frac{L}{8}\right)^{2.53}, \quad c_4(L=20) \approx 46.$$

This allows us to correct our sample standard deviation according to eq. (B.10).

We stress that the intuitive notion of standard deviation as "uncertainty" like in eq. (B.7) is often done with a normal distribution in mind. By the central limit theorem, the distribution of the mean of a large sample is almost normal, but the smaller the sample, the more the distribution of $\langle X_n\rangle$ resembles that of X itself, including outliers (indeed, we know this relation explicitly from eq. (B.5)). While our numerical data confirms that eq. (B.8) gives the correct expected standard deviation, one might not want to interpret this quantity as a reliable sampling accuracy for very small samples.

Conversely, for very large samples $n \approx N$, the observed uncertainty is increasingly smaller than estimated from eq. (B.8). In that case, the finite population correction eq. (B.11) gives an accurate description. We stress again that this effect only appears if one draws samples without replacement. As soon as replacement is allowed, eq. (B.8) stays correct even for $n > N_L^{(C)}$.

JHEP11(2023)160

We conclude that for all samples in the main part of this paper, eq. (B.8) is an accurate estimate of the sampling standard deviation. Note also the quantitative scale of the uncertainties: to reach an accuracy of just 1%, we need $n \geq \left(\frac{\sigma}{1\% \bar{M}_1}\right)^2 = 4120$ periods. In an upcoming paper, we examine how the empirical data from the present work allows to reduce the required sample sizes.

D Weighted averages and estimation of combined uncertainties

In this section, we discuss the influence of numerical uncertainty and the combination of uncertainties. The numerical integration procedure returns an estimate for the uncertainty $\Delta_{\text{num}}\mathcal{P}(G)$ of each period, measured as the standard deviation. These uncertainties are statistically independent since each period results from a different sum of Monte-Carlo summands. Comparing to the analytically known periods, we confirmed that the deviations are approximately normal distributed with standard deviation $\Delta_{\text{num}}\mathcal{P}(G)$ as expected. Let S_n be the sum of n periods, weighted by symmetry factor. The uncertainty $\Delta_{\text{num}}S_n$ is the square root of the variance, and the latter is the sum of the variances of the individual summands. Consequently, the uncertainty of the mean eq. (4.5) is given by

$$(\Delta_{\text{num}}S_n)^2 = \sum_{n \text{ completions } G} \left(\frac{\Delta\mathcal{P}(G)}{|\text{Aut}(G)|}\right)^2, \qquad \Delta_{\text{num}}\left\langle\frac{\mathcal{P}}{|\text{Aut}|}\right\rangle = \frac{\sqrt{(\Delta_{\text{num}}S_n)^2}}{n}. \tag{D.1}$$

The sum has n summands which all contribute roughly equally, hence the numerical uncertainty falls off like $n^{-\frac{1}{2}}$ with growing n. This explains why the means in tables 6 and 7 have much smaller uncertainty than the individual periods in table 1.

If we use a non-complete sample of n periods of a given loop order, the sample average is subject to a sampling uncertainty. In appendix B, we derived eq. (B.8),

$$\Delta_{\text{samp}}\langle X_n\rangle = \frac{1}{\sqrt{n}}\sigma, \tag{D.2}$$

where σ is the population standard deviation. In appendix C, we confirmed that this formula is accurate for all samples in the main text. The two uncertainties eqs. (D.1) and (D.2) are independent, therefore the total uncertainty is

$$\Delta\mathcal{P}_n = \sqrt{(\Delta_{\text{num}}\mathcal{P}_n)^2 + (\Delta_{\text{samp}}\mathcal{P}_n)^2}. \tag{D.3}$$

A third type of uncertainty can arise if the sampling itself is not uniform, for example due to a low-quality random number generator. We have included the $\star$-samples in order to verify the accuracy of `nauty`'s random graph generator, and found no significant systematic deviations as soon as the "s" samples are corrected with regard to their symmetry factor as explained in appendix A.

The procedure for combining multiple numerical values of *the same* period (that is, graphs which are either isomorphic or related by a known period symmetry) is similar to a sample mean, but the individual contributions are weighted with factors w_j:

$$\bar{\mathcal{P}} = \left(\sum_{j=1}^n w_j\right)^{-1} \cdot \sum_{j=1}^n w_j\mathcal{P}(G_j), \qquad \Delta\bar{\mathcal{P}} = \left(\sum_{j=1}^n w_j\right)^{-1} \cdot \sqrt{\sum_{j=1}^n (w_j\Delta\mathcal{P}(G_j))^2}.$$

JHEP11(2023)160

In principle, we can freely chose the w_j. The uncertainty of the average is minimized by choosing $w_i := (\Delta\mathcal{P}(G_j))^{-2}$. With this choice, the estimate and uncertainty are

$$\bar{\mathcal{P}} := \left(\sum_{i=1}^{n} w_i\right)^{-1} \cdot \sum_{i=1}^{n} w_i \mathcal{P}(G_i), \qquad \Delta\bar{\mathcal{P}} = \left(\sum_{i=1}^{n} w_i\right)^{-\frac{1}{2}}. \tag{D.4}$$

Note that this weighted mean is not equivalent to introducing non-trivial factors w_i for distinct periods, such as the $O(N)$ symmetry factors in section 5.3. In the latter case, different choices of weights w_i do change the value of the sum because we are summing *different* periods and the sum is normalized with respect to their number n,

$$\mathcal{P}_n = \frac{1}{n}\sum_{j=1}^{n} \frac{w_j \mathcal{P}(G_j)}{|\mathrm{Aut}(G_j)|} \neq \frac{1}{\sum_{j=1}^{n} w_j} \sum_{j=1}^{n} \frac{w_j \mathcal{P}(G_j)}{|\mathrm{Aut}(G_j)|}. \tag{D.5}$$

Conversely, in eq. (D.4), the sum is over periods which should be equal and the sum is normalized with respect to the sum of weights.

References

[1] G.V. Dunne and M. Unsal, *Resurgence and Trans-series in Quantum Field Theory: The $\mathbb{CP}^{N-1}$ Model*, *JHEP* **11** (2012) 170 [arXiv:1210.2423] [INSPIRE].

[2] M. Borinsky and G.V. Dunne, *Non-Perturbative Completion of Hopf-Algebraic Dyson-Schwinger Equations*, *Nucl. Phys. B* **957** (2020) 115096 [arXiv:2005.04265] [INSPIRE].

[3] M.P. Bellon and E.I. Russo, *Resurgent Analysis of Ward-Schwinger-Dyson Equations*, *SIGMA* **17** (2021) 075 [arXiv:2011.13822] [INSPIRE].

[4] M. Borinsky, G.V. Dunne and M. Meynig, *Semiclassical Trans-Series from the Perturbative Hopf-Algebraic Dyson-Schwinger Equations: ϕ^3 QFT in 6 Dimensions*, *SIGMA* **17** (2021) 087 [arXiv:2104.00593] [INSPIRE].

[5] P.J. Clavier, *Borel-Écalle Resummation of a Two-Point Function*, *Ann. Henri Poincaré* **22** (2021) 2103 [arXiv:1912.03237] [INSPIRE].

[6] M. Borinsky and D. Broadhurst, *Resonant resurgent asymptotics from quantum field theory*, *Nucl. Phys. B* **981** (2022) 115861 [arXiv:2202.01513] [INSPIRE].

[7] N. Nakanishi, *Graph Theory and Feynman Integrals*, in *Mathematics and Its Applications* **11**, Gordon and Breach, Science Publishers, Inc. (1971)

[8] E. Panzer, *Feynman integrals and hyperlogarithms*, Ph.D. Thesis, Humboldt-Universität zu Berlin, Berlin, Germany (2015) [arXiv:1506.07243] [INSPIRE].

[9] C. Bogner and S. Weinzierl, *Feynman graph polynomials*, *Int. J. Mod. Phys. A* **25** (2010) 2585 [arXiv:1002.3458] [INSPIRE].

[10] S. Bloch, H. Esnault and D. Kreimer, *On Motives associated to graph polynomials*, *Commun. Math. Phys.* **267** (2006) 181 [math/0510011] [INSPIRE].

JHEP11(2023)160

[11] O. Schnetz, *Quantum periods: A census of ϕ^4-transcendentals*, *Commun. Num. Theor. Phys.* **4** (2010) 1 [arXiv:0801.2856] [INSPIRE].

[12] F. Brown and D. Kreimer, *Angles, Scales and Parametric Renormalization*, *Lett. Math. Phys.* **103** (2013) 933 [arXiv:1112.1180] [INSPIRE].

[13] K.G. Wilson and J.B. Kogut, *The Renormalization group and the epsilon expansion*, *Phys. Rep.* **12** (1974) 75 [INSPIRE].

[14] D.J. Broadhurst and D. Kreimer, *Knots and Numbers in ϕ^4 Theory to 7 Loops and Beyond*, *Int. J. Mod. Phys. C* **06** (1995) 519 [hep-ph/9504352] [INSPIRE].

[15] F. Brown and O. Schnetz, *Single-valued multiple polylogarithms and a proof of the zig-zag conjecture*, *J. Number Theor.* **148** (2015) 478 [arXiv:1208.1890] [INSPIRE].

[16] E. Panzer, *On the analytic computation of massless propagators in dimensional regularization*, *Nucl. Phys. B* **874** (2013) 567 [arXiv:1305.2161] [INSPIRE].

[17] E. Panzer and O. Schnetz, *The Galois coaction on ϕ^4 periods*, *Commun. Num. Theor. Phys.* **11** (2017) 657 [arXiv:1603.04289] [INSPIRE].

[18] O. Schnetz, *Numbers and Functions in Quantum Field Theory*, *Phys. Rev. D* **97** (2018) 085018 [arXiv:1606.08598] [INSPIRE].

[19] O. Schnetz, *HyperlogProcedures*, (2023) https://www.math.fau.de/person/oliver-schnetz/.

[20] M. Kontsevich and D. Zagier, *Periods*, in *Mathematics Unlimited — 2001 and Beyond*, Springer (2001), pp. 771–808.

[21] P. Belkale and P. Brosnan, *Periods and Igusa Zeta functions*, math/0302090.

[22] F.C.S. Brown, *Multiple zeta values and periods of moduli spaces $\mathfrak{M}_{0,n}$*, *Annales Sci. Ecole Norm. Sup.* **42** (2009) 371 [math/0606419] [INSPIRE].

[23] F.C.S. Brown, *On the periods of some Feynman integrals*, arXiv:0910.0114 [INSPIRE].

[24] D. Kreimer, *Quantum fields, periods and algebraic geometry*, *Contemp. Math.* **648** (2015) 153 [arXiv:1405.4964] [INSPIRE].

[25] I. Todorov, *Polylogarithms and Multizeta Values in Massless Feynman Amplitudes*, in *Lie Theory and Its Applications in Physics*, *Springer Proceedings in Mathematics & Statistics* **111**, V. Dobrev ed., Springer, Tokyo, Japan (2014), pp. 155–176 [DOI:10.1007/978-4-431-55285-7_10] [INSPIRE].

[26] I. Crump, M. DeVos and K. Yeats, *Period preserving properties of an invariant from the permanent of signed incidence matrices*, *Ann. Inst. Henri Poincaré D* **3** (2016) 429.

[27] E. Nasrollahpoursamami, *Periods of Feynman Diagrams and GKZ D-Modules*, arXiv:1605.04970.

[28] I. Crump, *Properties of the Extended Graph Permanent*, arXiv:1608.01414.

[29] S. Hu, O. Schnetz, J. Shaw and K. Yeats, *Further investigations into the graph theory of ϕ^4-periods and the c_2 invariant*, *Ann. Inst. Henri Poincaré D* **9** (2022) 473.

[30] M. Laradji, M. Mishna and K. Yeats, *Some results on double triangle descendants of K_5*, *Ann. Inst. Henri Poincaré D* **8** (2021) 537 [arXiv:1904.06923] [INSPIRE].

[31] M. Borinsky and O. Schnetz, *Graphical functions in even dimensions*, *Commun. Num. Theor. Phys.* **16** (2022) 515 [arXiv:2105.05015] [INSPIRE].

JHEP11(2023)160

[32] M. Borinsky and O. Schnetz, *Recursive computation of Feynman periods*, *JHEP* **08** (2022) 291 [arXiv:2206.10460] [INSPIRE].

[33] M. Borinsky, *Tropical Monte Carlo quadrature for Feynman integrals*, *Ann. Inst. Henri Poincaré D* **10** (2023) 635 [arXiv:2008.12310] [INSPIRE].

[34] B.D. McKay and A. Piperno, *Practical graph isomorphism. Part II*, *J. Symb. Comput.* **60** (2014) 94.

[35] M. Borinsky, H.J. Munch and F. Tellander, *Tropical Feynman integration in the Minkowski regime*, *Comput. Phys. Commun.* **292** (2023) 108874 [arXiv:2302.08955] [INSPIRE].

[36] M.V. Kompaniets and E. Panzer, *Minimally subtracted six loop renormalization of $O(n)$-symmetric ϕ^4 theory and critical exponents*, *Phys. Rev. D* **96** (2017) 036016 [arXiv:1705.06483] [INSPIRE].

[37] B.G. Nickel, D.I. Meiron and G.A. Baker Jr., *Compilation of 2-pt and 4-pt Graphs for Continuous Spin Models*, University of Guelph, Guelph, ON, Canada (1977).

[38] N.J.A. Sloane, *The On-Line Encyclopedia of Integer Sequences*, (2023) https://oeis.org.

[39] M. Meringer, *Fast generation of regular graphs and construction of cages*, *J. Graph Theory* **30** (1999) 137.

[40] M. Meringer, *Regular Graphs Page*, (2009) https://www.mathe2.uni-bayreuth.de/markus/reggraphs.html.

[41] N.C. Wormald, *Models of Random Regular Graphs*, in *Surveys in Combinatorics, 1999*, *London Mathematical Society Lecture Note Series* **267**, Cambridge University Press, Cambridge, U.K. (1999), pp. 239–298 [DOI:10.1017/cbo9780511721335.010].

[42] B. Bollobás, *Random Graphs*, in *Cambridge Studies in Advanced Mathematics*, second edition, Cambridge University Press, Cambridge, U.K. (2001) [DOI:10.1017/cbo9780511814068].

[43] B. Bollobás, *Modern Graph Theory*, Springer, New York, NY, U.S.A. (1998).

[44] N.C. Wormald, *The asymptotic connectivity of labelled regular graphs*, *J. Combin. Theor. B* **31** (1981) 156.

[45] B. Bollobás, *A Probabilistic Proof of an Asymptotic Formula for the Number of Labelled Regular Graphs*, *Eur. J. Combinator.* **1** (1980) 311.

[46] N.C. Wormald, *The asymptotic distribution of short cycles in random regular graphs*, *J. Combin. Theor. B* **31** (1981) 168.

[47] E.A. Bender and E.R. Canfield, *The asymptotic number of labeled graphs with given degree sequences*, *J. Combin. Theor. A* **24** (1978) 296.

[48] B. Bollobás, *The Asymptotic Number of Unlabelled Regular Graphs*, *J. London Math. Soc.* **s2-26** (1982) 201.

[49] N. Bonichon, C. Gavoille, N. Hanusse, D. Poulalhon and G. Schaeffer, *Planar Graphs, via Well-Orderly Maps and Trees*, *Graphs Combin.* **22** (2006) 185.

[50] P. Cvitanović, B.E. Lautrup and R.B. Pearson, *The Number and Weights of Feynman Diagrams*, *Phys. Rev. D* **18** (1978) 1939 [INSPIRE].

[51] M. Borinsky, *Renormalized asymptotic enumeration of Feynman diagrams*, *Ann. Phys.* **385** (2017) 95 [arXiv:1703.00840] [INSPIRE].

JHEP11(2023)160

[52] B.D. McKay and N.C. Wormald, *Automorphisms of random graphs with specified vertices*, *Combinatorica* **4** (1984) 325.

[53] E.A. Bender, Z. Gao and N.C. Wormald, *The Number of Labeled* 2*-Connected Planar Graphs*, *Electron. J. Combinator.* **9** (2002) R43.

[54] E. Panzer, *Hepp's bound for Feynman graphs and matroids*, *Ann. Inst. Henri Poincaré D* **10** (2023) 31 [arXiv:1908.09820] [INSPIRE].

[55] D.J. Broadhurst, *Exploiting the* 1.440*-Fold Symmetry of the Master Two Loop Diagram*, *Z. Phys. C* **32** (1986) 249 [INSPIRE].

[56] M. Golz, E. Panzer and O. Schnetz, *Graphical functions in parametric space*, *Lett. Math. Phys.* **107** (2017) 1177 [arXiv:1509.07296] [INSPIRE].

[57] J. Siek, L.-Q. Lee and A. Lumsdaine, *The Boost Graph Library — 1.82.0*, (2001) https://www.boost.org/doc/libs/1_82_0/libs/graph/doc/index.html.

[58] O. Schnetz, *Quantum field theory over* F_q, *Electron. J. Combinator.* **18** (2011) P102 [arXiv:0909.0905] [INSPIRE].

[59] E. Panzer and K. Yeats, *Feynman symmetries of the Martin and* c_2 *invariants of regular graphs*, arXiv:2304.05299 [INSPIRE].

[60] A. Pelissetto and E. Vicari, *Critical phenomena and renormalization group theory*, *Phys. Rep.* **368** (2002) 549 [cond-mat/0012164] [INSPIRE].

[61] P.-H. Balduf, *Dyson-Schwinger equations in minimal subtraction*, *Ann. Inst. Henri Poincaré D 13 April 2023*.

[62] O. Schnetz, ϕ^4 *theory at seven loops*, *Phys. Rev. D* **107** (2023) 036002 [arXiv:2212.03663] [INSPIRE].

[63] P. Martin, *Enumérations eulériennes dans les multigraphes et invariants de Tutte-Grothendieck*, Ph.D. Thesis, Institut National Polytechnique de Grenoble, Université Joseph-Fourier, Grenoble I, France (1977).

[64] J.A. Ellis-Monaghan, *New Results for the Martin Polynomial*, *J. Combin. Theor. B* **74** (1998) 326.

[65] B. Bollobás, *Evaluations of the Circuit Partition Polynomial*, *J. Combin. Theor. B* **85** (2002) 261.

[66] A. Bouchet and L. Ghier, *Connectivity and* β*-invariants of isotropic systems and* 4*-regular graphs*, *Discrete Math.* **161** (1996) 25.

[67] H. Kleinert and V. Schulte-Frohlinde, *Critical Properties of* ϕ^4*-Theories*, World Scientific (2001).

[68] A.J. McKane, D.J. Wallace and O.F. de Alcantara Bonfim, *Non-perturbative renormalisation using dimensional regularisation: Applications to the* ϵ *expansion*, *J. Phys. A* **17** (1984) 1861 [INSPIRE].

[69] A.J. McKane, *Perturbation expansions at large order: Results for scalar field theories revisited*, *J. Phys. A* **52** (2019) 055401 [arXiv:1807.00656] [INSPIRE].

[70] H. Kinkelin, *Ueber eine mit der Gammafunction verwandte Transcendente und deren Anwendung auf die Integralrechnung*, *J. Reine Angew. Math.* **1860** (1860) 122.

JHEP11(2023)160

[71] L.F. Richardson and R.T. Glazebrook, *The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam*, *Philos. Trans. Roy. Soc. Lond. Ser. A* **210** (1911) 307.

[72] I. Aniceto, G. Başar and R. Schiappa, *A Primer on Resurgent Transseries and Their Asymptotics*, *Phys. Rep.* **809** (2019) 1 [arXiv:1802.10441] [INSPIRE].

[73] M.V. Komarova and M.Y. Nalimov, *Asymptotic Behavior of Renormalization Constants in Higher Orders of the Perturbation Expansion for the* $(4-\epsilon)$*-Dimensionally Regularized* $O(n)$ *symmetric* ϕ^4 *theory*, *Theor. Math. Phys.* **126** (2001) 339 [INSPIRE].

[74] G. 't Hooft, *A Planar Diagram Theory for Strong Interactions*, *Nucl. Phys. B* **72** (1974) 461 [INSPIRE].

[75] R. Gurau and J.P. Ryan, *Colored Tensor Models — a review*, *SIGMA* **8** (2012) 020 [arXiv:1109.4812] [INSPIRE].

[76] B. Bollobás and W.F. de la Vega, *The diameter of random regular graphs*, *Combinatorica* **2** (1982) 125.

[77] L.R. Shenton and K.O. Bowman, *The Development of Techniques for the Evaluation of Sampling Moments*, *Int. Stat. Rev.* **43** (1975) 317.

[78] A.C. Davison and D.V. Hinkley, *Bootstrap Methods and their Application*, in *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, Cambridge, U.K. (1997) [DOI:10.1017/cbo9780511802843].

[79] Y. Tillé, *Variance Estimation by Linearization*, John Wiley & Sons, Ltd. (2020) [DOI:10.1002/9781119071259.ch15].

[80] M.J. Evans and J.S. Rosenthal, *Probability and Statistics: The Science of Uncertainty*, (2023) https://www.utstat.toronto.edu/mikevans/jeffrosenthal/.

[81] W.P. Johnson, *The Curious History of Faa Du Bruno's Formula*, *Am. Math. Mon.* **109** (2002) 217.

[82] E.T. Bell, *Exponential Polynomials*, *Ann. Math.* **35** (1934) 258.

[83] W.H. Holtzman, *The Unbiased Estimate of the Population Variance and Standard Deviation*, *Am. J. Psychol.* **63** (1950) 615.

[84] J.A. Angelova, *On Moments of Sample Mean and Variance*, *Int. J. Pure Appl. Math.* **79** (2012) 67.

JHEP11(2023)160