ZX-calculus single-diagram formulation of gradient variance for barren plateau analysis in QuiZX

Andrias Cordeiro and Demerson N. Gonçalves

***Abstract*— We present a workflow that analyzes the optimization landscape of parameterized quantum models. The approach computes the gradient's expectation and variance entirely within the ZX calculus via diagrammatic differentiation and integration. In particular, the variance admits a single-diagram expression whose contraction yields exact values when the diagram is closed, avoiding large combinatorial sums and reducing reliance on sampling. When the resulting diagram has a dominant Clifford structure, stabilizer routines in QuiZX make contraction efficient in practice. We report preliminary simulations that reproduce known two-qubit variances and show scaling with the number of qubits at fixed depth using parameter-shift sampling. Figure 1 summarizes the pipeline from a circuit to the variance single diagram and its contraction. This is ongoing work: full single-diagram contractions and task-level studies in classification and regression are underway, with the goal of guiding model selection and reducing trial and error.**

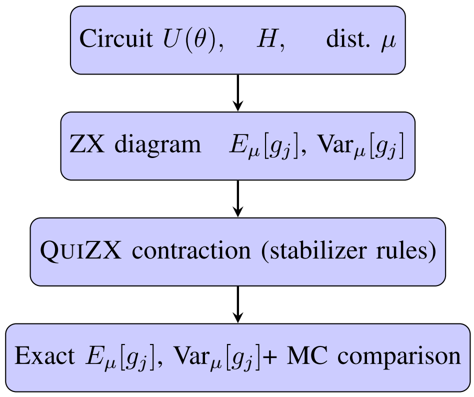

Fig. 1. Workflow of our approach. From a parametrized circuit $U(\theta)$ with observable H and distribution μ, we obtain the ZX diagram for $E_\mu[g_j]$ and, separately, the single-diagram expression for $\mathrm{Var}_\mu[g_j]$. The variance diagram is then contracted in QuiZX; since it is closed, the contraction returns a scalar that we compare against Monte Carlo baselines.

I. Introduction

Barren plateaus refer to the rapid decay of gradient variance as the number of qubits or the depth of a parameterized circuit increases, which makes gradient-based optimization nearly impossible [1]. The ZX-calculus provides a diagrammatic language to study this issue. Instead of writing gradients and variances as long algebraic sums, one can translate derivatives and integrals into graphical rules and represent the entire calculation as a diagram.

Recent work [2] shows how to compute the gradient's expectation and variance entirely within the ZX calculus through diagrammatic differentiation and integration. In particular, they demonstrate that the variance admits a single-diagram representation, whose contraction yields the desired value. This contrasts with the earlier approach of [3], where applying the standard ZX rules expanded into a sum that grows exponentially in the number of parameters, requiring general tensor-network methods. The formulation of [2] therefore avoids this exponential blow-up. By contraction we mean the systematic simplification of a ZX diagram or tensor network. If the diagram is closed, as in our expectation and variance cases, the contraction evaluates to a scalar. If the diagram is open, the contraction yields an equivalent but simpler diagram with the same inputs and outputs.

Andrias Cordeiro, Computation Engineering, CEFET-RJ, Petrópolis-RJ, e-mail: andrias.cordeiro@aluno.cefet-rj.br; Demerson N. Gonçalves, Department of Mathematics, CEFET-RJ, Petrópolis-RJ, e-mail: demerson.goncalves@cefet-rj.br.

II. Problem formulation

Let $U(\theta)$ be a parameterized ansatz, H an observable, and ρ_0 the input state. The cost is

$$C(\theta) = \mathrm{Tr}\big(H\, U(\theta)\rho_0 U^\dagger(\theta)\big). \tag{1}$$

For coordinate j, define the partial derivative $g_j(\theta) = \partial_{\theta_j} C(\theta)$. Given an initialization distribution μ over parameters, the expectation and the variance are

$$E_\mu[g_j] = \int g_j(\theta)\, \mathrm{d}\mu(\theta), \qquad \mathrm{Var}_\mu[g_j] = E_\mu\big[g_j^2\big] - \big(E_\mu[g_j]\big)^2. \tag{2}$$

When the generator of the parameterized gate has a binary spectrum, we validate gradients with the parameter-shift identity

$$g_j(\theta) = \kappa_j\Big(C\big(\theta + \tfrac{\pi}{2}e_j\big) - C\big(\theta - \tfrac{\pi}{2}e_j\big)\Big), \tag{3}$$

which we also use to build Monte Carlo baselines [1].

III. Method

The method has three conceptual steps that are carried out in one continuous workflow, as summarized in Figure 1. First, following the diagrammatic differentiation and integration rules of [2], we obtain the ZX diagram for $E_\mu[g_j]$ and, separately, the single-diagram expression for $\mathrm{Var}_\mu[g_j]$ for the chosen ansatz and observable. Second, we contract the variance diagram in QuiZX. In this step we incorporate the stabilizer-based decompositions of [4], which simplify structures such as triangles and star edges while keeping track of the non-Clifford content. Third, we study how the variance

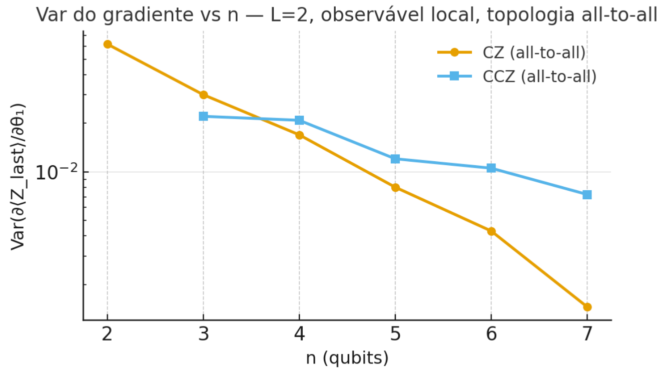

Fig. 2. Log-scale plot of $\mathrm{Var}_\mu[g_{\mathrm{last}}]$ versus the number of qubits n for depth L=2, local observable Z on the last qubit, and all-to-all connectivity. Each point is computed at a fixed sampling budget using the parameter-shift rule under uniform initialization. Both layouts show a near-exponential decay consistent with barren plateaus. The CCZ layout remains above the CZ baseline, indicating less collapsed gradients and a milder plateau.

regime scales with circuit size or depth, using it as a diagnostic of trainability before launching expensive optimization loops. The same pipeline is intended to support model selection for classification and regression, by ranking feature maps and ansatz layouts according to their variance regime as a surrogate of optimization difficulty, and then refining those choices with standard training.

IV. Preliminary progress

We report two complementary checks that anchor the approach. First, on a two-qubit test case we recovered the known gradient variances for depth one and two by Monte Carlo with parameter shift and by independent symbolic integration, which validates our conventions for angles, generators and normalization [3].

Second, we fixed the depth to L=2 and studied how the gradient variance scales with the number of qubits n for two entangling layouts under all-to-all connectivity: a CZ layout and a CCZ layout. Throughout this experiment the cost is $C(\theta) = \langle Z_{\mathrm{last}} \rangle$, the gradient component is $g_{\mathrm{last}} = \partial_{\theta_{\mathrm{last}}} C(\theta)$, and the parameters are initialized from μ as independent uniform angles in $[-\pi, \pi]$. Figure 2 plots $\mathrm{Var}_\mu[g_{\mathrm{last}}]$ on a logarithmic vertical scale as a function of n. Both layouts exhibit a nearly exponential decay with n, which is the hallmark of barren plateaus [1]. The CCZ curve stays consistently above the CZ curve, which means that the gradient signal is larger for the CCZ layout at the same depth. In practical terms, a larger variance indicates a milder plateau and suggests that gradient-based optimization should be easier for the CCZ layout at L=2.

At $n = 2$ our Monte Carlo estimates coincide with the exact values obtained from the single-diagram ZX construction, which serves as a sanity check. Exact single-diagram contractions are in progress; current points use fixed-budget sampling except where noted. Ongoing work replaces the sampled estimates with exact contractions in QuiZX for the full range of n, and reports wall-clock times side by side with sampling [2], [4].

V. Ongoing work and outlook

The next step is to turn the workflow in Fig. 1 into full single-diagram contractions for the ansatz families used throughout this paper. For each instance we will report the exact values of $E_\mu[g_j]$ and $\mathrm{Var}_\mu[g_j]$ obtained in QuiZX, together with wall-clock time and peak memory. These measurements will be shown side by side with Monte Carlo estimates at a fixed sampling budget, which provides a fair comparison between exact contraction and sampling-based evaluation [2], [4]. The analysis will follow the format of Fig. 2, so the reader can see how the variance behaves as the number of qubits grows at a fixed depth.

We will also study how non-Clifford content affects both the variance level and the cost of contraction. To do this we interpolate between layouts that use CZ blocks and layouts that use CCZ blocks, while keeping depth and observable fixed. For each configuration we record the fraction of non-Clifford vertices, the presence of triangles and star edges, and the number of stabilizer rewrites that QuiZX applies. This allows us to relate the structure of the diagram to the runtime and to the observed variance, and to explain why the CCZ curve in Fig. 2 stays above the CZ baseline.

The same routines are being brought to task-level objectives in quantum machine learning. For classification we pair the single-diagram analysis with costs and readouts that reflect accuracy on a labeled set. For regression we use mean-squared error style costs. In both cases the variance regime is used as a surrogate for trainability to rank feature maps and circuit templates before training, and then we verify the ranking by short training runs with the same initialization distribution [3]. The goal is to reduce trial and error during model selection and to give practitioners a compact procedure that links landscape diagnosis to concrete optimization choices.

All scripts and diagrams will be released with seeds, environment files and figure-generation notebooks so the results can be reproduced. This poster reports an undergraduate project that is still in progress. The simulations completed so far anchor the approach, and the planned contractions and task-level studies provide a clear path to broader experiments.

These results establish a reproducible pipeline that links ZX-calculus analysis to optimization choices in quantum machine learning.

Acknowledgments

Andrias Cordeiro gratefully acknowledges financial support from CNPq through the PIBIC program at CEFET/RJ.

Referências

[1] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, *Phys. Rev. Lett.* **121**, 210501 (2018).

[2] Y. Wang, R. W. Yeung, and B. Koch, Differentiating and integrating ZX diagrams with applications to quantum machine learning, *Quantum* (2024).

[3] Y. Zhao and X. Gao, Analyzing barren plateaus with ZX-calculus , *Quantum* **5**, 466 (2021).

[4] B. Koch, R. W. Yeung, and Y. Wang, Speedy contraction of ZX diagrams with triangles via stabiliser decompositions, arXiv:2307.01803 (2023).