

# Optimal partitioning of quantum Circuits using gate cuts and wire cuts.

Partitioning quantum circuits generally incudes an exponential increase in quantum computing runtimes, by repeatedly executing quantum subcircuits.

This is an optimal partitioning method based on quantum circuit knitting. By considering wire cuts and gate cuts in conjunction with ancilla qubit insertions and classical communication, this developed method can determine a minimal cost quantum circuit partitioning.

Minimizing the sampling overhead is a crucial objective for quantum circuit partitioning.

The more entangled a quantum gate is, as quantified by the robustness of entanglement, the higher the minimum incurred sampling overhead of quantum circuit knitting will be.

Dependency Resolution Techniques are able to reduce the sampling overhead of multiple, simultaneous cuts by using ancilla qubit and classical communication.

Minimal Sampling Overhead of cutting depending on cut type, ancilla qubits, and classical communication (CC)

	wire cuts	K-gate cuts			
Single	K-arbitrary	CNOT	SWAP	CR(0)	
CC and Ancilla	9	$(2^{k+1}-1)^2$	$(2^{k+1}-1)^2$	$16^k$	$(\leq 4)^k$

No CC and  
no ancilla.

16

$16^k$

$9^k$

$49^k$

$(1+2\sin\alpha)^{2k}$

quantum gates such as conditional-rotation gates may incur a sampling overhead smaller than  $O(4^k)$  depending on the rotation angles. A single individual wire cut incurs a multiplicative sampling overhead of 9 when CC and ancilla qubits are available and 16 otherwise.

In general, the number of applicable cuts is limited by the incurred exponential sampling overhead.

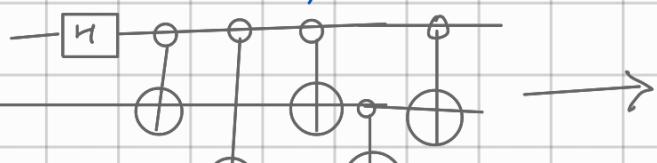
Sampling frequency	# cuts
1 kHz	5
1 MHz	10
1 GHz	15

will take runtimes of  
1 full day  
Current hardware range  
from 1 kHz to 20 GHz.

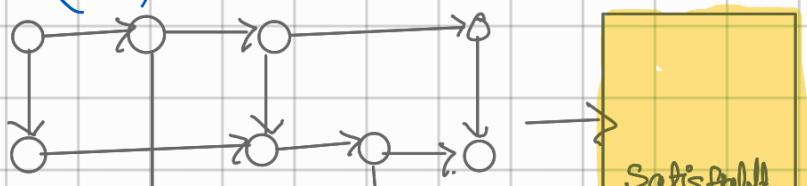
The limited number of cuts available at even large quantum computing budgets warrants a runtime-intensive and rigorous quantum circuit partitioning approach that can determine a suitable partitioning with provably minimal sampling overhead for a given quantum circuit.

Optimized Partitioning of Quantum Circuit  
Satisfiability modulo theories model.

(i)



(ii)



OPD

Satisfiable

Quantum circuit

cutting graph

(ii)

models  
Theories  
model

Quantum Computation  
Characteristics

(iv)

optimized  
cut  
partitions

(v)

SMT  
solver

objectives

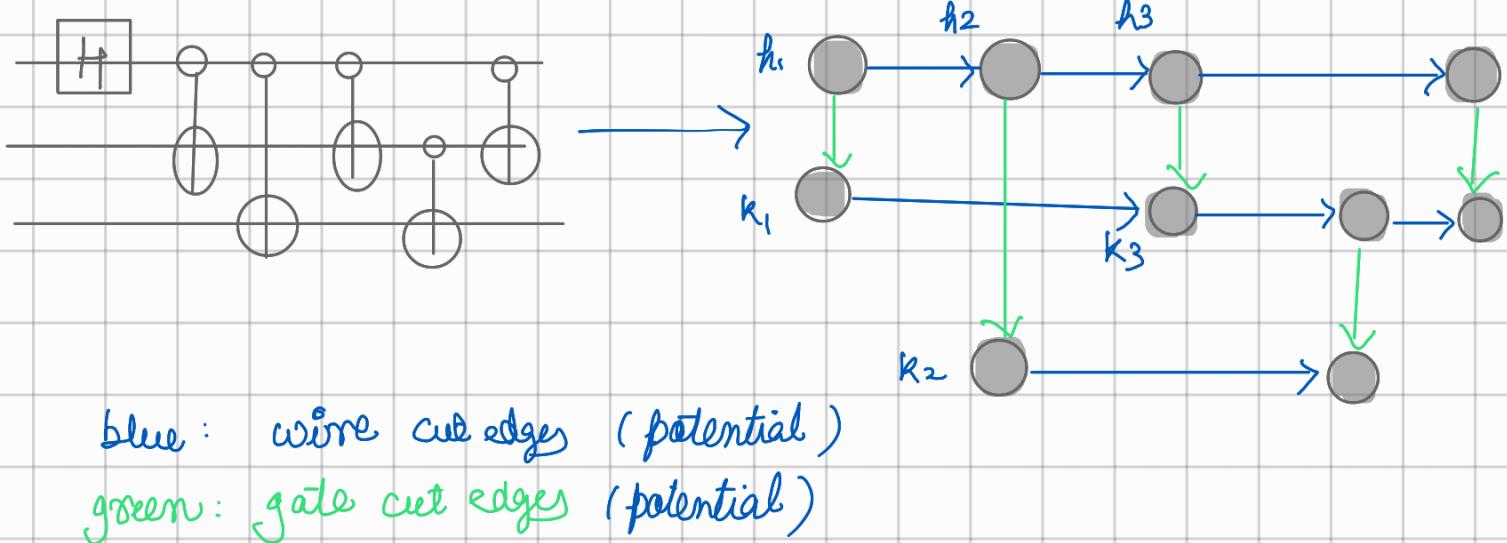
First, the target quantum circuit that should be partitioned is input to a pre-processing step that compiles the quantum circuit to two-qubit quantum gates and that generates the cutting graph used as a basis for the developed SMT model.

Then the generated graph is input to the SMT model together with a description of the QPDS including their T-factors as well as a specification of the available quantum computing resources in terms of qubits, computing time, and sampling frequency.

In the last step SMT solver (like Z3) is used to solve the input SMT model subject to defined objective functions

## Pre processing step

Target quantum circuit is compiled to two qubit gates and then to a graph  $G = (V, G \cup W)$  where  $W$  represents set of edges corresponding to wire cuts and  $G$  represents set of edges corresponding to gate cuts.



- i) single qubit gates are removed from the circuit and each two qubit gate  $g$  between qubits  $g_u$  and  $g_v$  is represented by two vertices.  $g_u$  and  $g_v$  in graph  $G$  and edge  $S_g = \{g_u, g_v\} \in G$

Then for each pair of two qubit gates  $(h, k)$ , where  $k$  is an immediate successor of  $h$ , an edge is added to edge set  $W$  for each overlapping qubit.

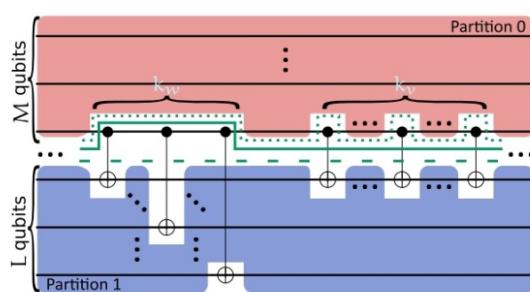


Figure 1: Example quantum circuit partitioning using gate cuts (dashed green line), wire cuts (dotted green line), and both (solid green line). The red (top) and blue (bottom) colored blocks indicate a high density of two-qubit gates where partitioning incurs a prohibitive cost in general.

