Algoritmos Quânticos para a Geração de Unitários Pseudo-Aleatórios

por

Igor Tuche de Almeida Diniz

Orientador: Professor Dr. Daniel Jonathan

Dissertação
apresentada na
Universidade Federal Fluminense
como requisito na
obtenção do título de
Mestre em Física

Niterói, Rio de Janeiro, Brasil , 2009

INSTITUTO DE FÍSICA
Universidade Federal Fluminense
CURSO DE PÓS-GRADUAÇÃO EM FÍSICA
RUA GAL MILTON TAVARES DE SOUZA, SN
24210-346 – NITERÓI - RIO DE JANEIRO
TEL: (21)2629-5878 - FAX: 2629-5887
E-MAIL: cpg@ if.uff.br

Ata dos trabalhos finais da Comissão Examinadora da dissertação de mestrado apresentada por ***Igor Tuche de Almeida Diniz.*** *Aos nove dias do mês de outubro de dois mil e nove, às 13:30h, reuniram-se no Instituto de Física da Universidade Federal Fluminense, os membros da Comissão Examinadora constituída pelos professores doutores: Daniel Jonathan – IF/UFF, Roberto Imbuzeiro Moraes Felinto de Oliveira – IMPA, Marcelo de Oliveira Terra Cunha – DM/UFMG e Marcelo Silva Sarandy – IF/UFF, sob a presidência do primeiro, para prova pública de apresentação de dissertação de mestrado intitulada* ***"Algoritmos Quânticos para a Geração de Unitários Pseudo-Aleatórios"****, tendo em vista as exigências contidas no Regulamento Específico do curso de Física, relacionadas com a conclusão do Mestrado em Física pela Universidade Federal Fluminense. A dissertação foi elaborada sob a orientação do professor Daniel Jonathan. O aluno fez a exposição do seu trabalho durante 57 minutos. A seguir, respondeu às questões formuladas pelos integrantes da Comissão Examinadora, que apresentou parecer no sentido de aprová-lo. Para constar, foi lavrada a presente ata, que vai assinada por mim, secretária da Pós-graduação em Física em exercício, pelos membros da Comissão Examinadora e pelo mestrando.*

Niterói, nove de outubro de dois mil e nove.

Valéria Vanda Azevedo de Lima
Dr. Daniel Jonathan
Dr. Roberto I. Moraes Felinto de Oliveira
Dr. Marcelo de Oliveira Terra Cunha
Dr. Marcelo Silva Sarandy
Igor Tuche de Almeida Diniz

Autorizo a Universidade Federal Fluminense a ceder essa dissertação para outras instituições ou indivíduos para o uso na pesquisa acadêmica.

Igor Tuche de Almeida Diniz

Resumo

Investigamos a geração de operações unitárias quânticas aleatórias sobre certos aspectos. Mostramos que um determinado algoritmo de interação de pares gera unitários pseudo-aleatórios e mostramos como esses unitários podem substituir unitários completamente aleatórios na implementação de outros algoritmos quânticos, com economia exponencial de recursos. Investigamos em detalhe como a interação entre pares afeta o tempo de convergência do algoritmo.

Palavras-Chave: Computação Quântica, Unitários Aleatórios, K-desenhos, Pseudo-Aletoriedade.

Agradecimentos

Eu gostaria de agradecer ao meu orientador Daniel Jonathan pelos inúmeros momentos de discussão, pelos horas dedicadas a me ensinar tanto quanto eu poderia aprender e por me ajudar tantas vezes a tornar minhas idéias mais claras. A sua exigência incansável em cada detalhe dessa dissertação é que é responsável pela qualidade que ela possui. Entretanto o resultado dessa orientação não termina nessa tese pois o que pude aprender, dentro e fora da Física, levarei para minha vida. Essa dissertação não poderia existir sem o seu esforço embora ele seja totalmente isento de qualquer imperfeição que ela possua. Espero que possamos trabalhar juntos novamente.

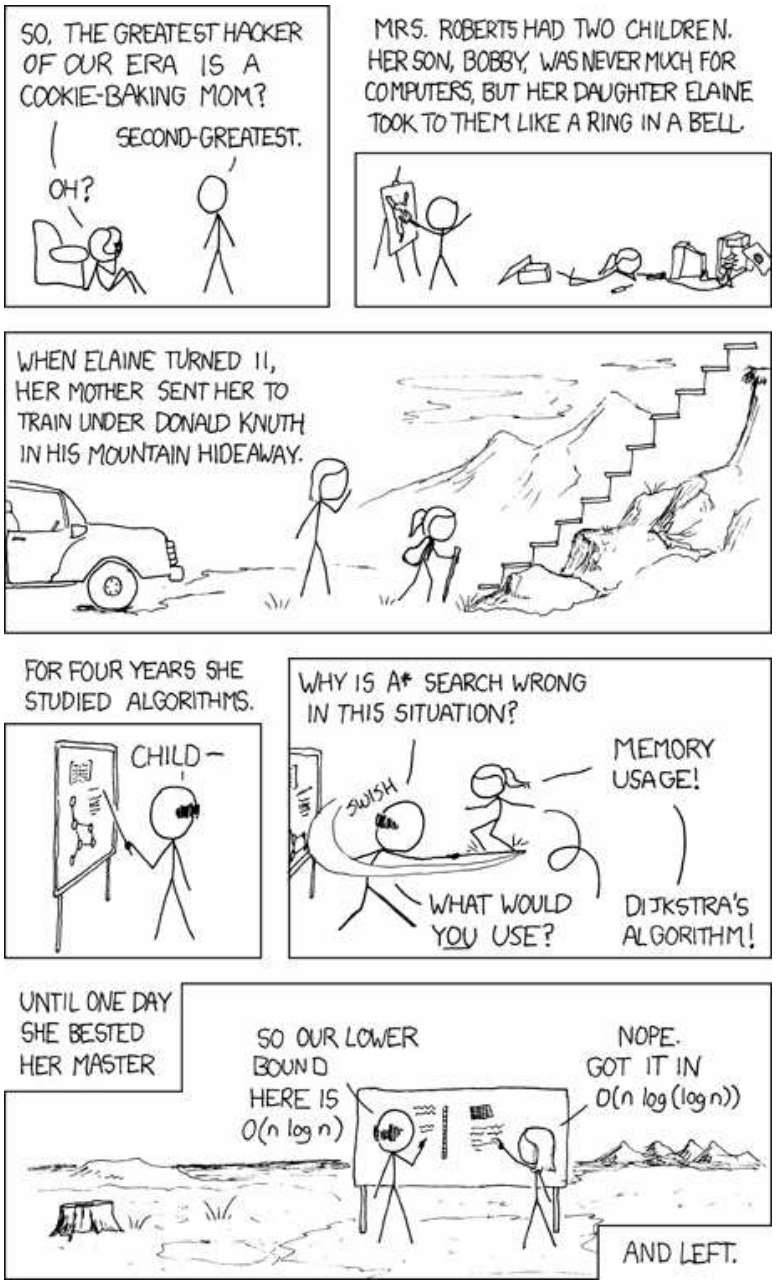
Também gostaria de agradecer ao professor Márcio Argollo que participou com uma visão diferentes em discussões que levaram a essa dissertação e a outros trabalhos embrionários que ficaram pelo caminho.

Colaboraram indiretamente para a minha formação e para o meu trabalho todos os membros do grupo de Ótica e Informação Quântica da Universidade Federal Fluminense. Todos eles foram importantes em algum momento do meu trabalho, em especial agradeço aos professores Ernesto, Kaled, Zelaquett e aos colegas Cadu e Raphael. Sem eles Paraty e Florianópolis não teriam a mesma graça.

Agradeço aos professores e aos demais membros da pós-graduação do IF-UFF pela qualidade e estrutura do curso que me foi oferecido. Em especial agradeço aos professores Thadeu Penna e Antônio Costa, sem eles não teríamos a infraestrutura de rede e a sala-de-micros que servem a todos nós membros da pós-graduação.

Agradeço a toda minha família por seu suporte, pela confiança nas minhas escolhas e o amor incondicional. É impossível retribuir tudo o que eles me deram.

Por fim agradeço àquela que esteve sempre ao meu lado, como se não bastasse todo o apoio que ela me dá em todos os outros aspectos da minha vida, foi ela quem leu tantas versões preliminares dessa dissertação corrigindo meus erros e me ajudando a melhorar. Caca's, que boa sorte é ter você comigo.



xkcd.com tirinha 432, Randall Munroe.

Sumário

Capítulo 1

Introdução

1.1 Motivação

Os números aleatórios têm um papel fundamental na teoria da informação clássica, onde são usados em inúmeros protocolos e algoritmos. Entre os seus variados usos podemos destacar os algoritmos que dependem de uma amostragem em um espaço amostral muito grande, algoritmos estocásticos e protocolos de criptografia e segurança [1, 2, 3, 4]. Felizmente existem maneiras eficientes de obter números aleatórios com boa qualidade, ou seja, existem geradores de números pseudo-aleatórios que criam sequências quase indistinguíveis de uma sequência realmente aleatória [5]. Além disso, recentemente, algumas empresas passaram a comercializar geradores de números aleatórios produzidos diretamente por medidas quânticas (por exemplo, pela passagem ou não de um fóton através de um divisor de feixe balanceado). Pelo menos em teoria, estes números são de fato perfeitamente aleatórios, algo que parece ser confirmado por diversos testes de aleatoriedade [6].

De maneira análoga, na teoria da Informação Quântica também é muito útil dispor de estados e operações aleatórias. Entretanto, nesse caso não temos a mesma facilidade para gerar sequências de estados e operações pseudo-aleatórias, veremos adiante por que isso acontece.

Essa dificuldade nos leva a avaliar mais cuidadosamente quais os recursos realmente necessários para cada algoritmo e protocolo que dependem de objetos quânticos aleatórios. Como veremos adiante, muitas vezes é possível realizar estas tarefas usando estados ou operações que, sem ser perfeitamente aleatórios e independentes, sejam ainda *suficientemente* aleatórios para a tarefa em questão. Antes de mostrar como isso pode ser feito no caso quântico, vamos ilustrar a idéia com um exemplo clássico relacionado, onde observaremos que nem sempre números totalmente aleatórios e independentes são necessários.

Exemplo: independência aos pares

Seja a sequência de bits aleatórios $Y_1, Y_2, ..., Y_k$. Dizemos que ela é independente aos pares se:

$$\forall i, j \;\; \mathbb{P}(Y_i = y_i, Y_j = y_j) = \mathbb{P}(Y_i = y_i)\mathbb{P}(Y_j = y_j) \; ,$$

onde $\mathbb{P}$ é a probabilidade de ocorrência dos eventos. Note que esta propriedade não exclui a possibilidade de haver correlações entre conjuntos contendo mais do que dois bits da sequência.

Há um método bem-conhecido que permite gerar deterministicamente, a partir de um conjunto de n bits aleatórios $a_1, \ldots, a_n$ completamente independentes, $2^n - 1$ bits Y_i que são independentes aos pares [7]. Cada bit Y_i é gerado da seguinte forma: escreve-se o índice i em forma binária ($i \in \{0,1\}^n / \{0\}^n$), e define-se

$$Y_i = \sum_{m=1}^{n} a_m \, i_m \pmod 2 \; .$$

Para ver que Y_i é independente de Y_j basta considerar uma coordenada m em que $i_m \neq j_m$. Como um desses dois números vale 0, se a *string* a mudar nessa coordenada o valor de apenas Y_i ou Y_j mudará.

Assim, se a exigência de um algoritmo que utilize números aleatórios for relaxada da independência total para independência aos pares, há uma economia exponencial do recurso de aleatoriedade.

Um exemplo de um algoritmo deste tipo se aplica na seguinte situação: considere um grafo[1] $G = (V, E)$, suponha que cada vértice x possa ser colorido com uma cor $D(x)$ escolhida entre duas cores possíveis. Um problema interessante[2], conhecido como o problema de 'máximo corte', é: como colorir o grafo de modo a maximizar o número c de arestas ligando vértices de cores distintas?

Uma estratégia 'aleatória', na qual simplesmente se escolhe uma cor ao acaso para cada vértice, produz em média

$$\langle c \rangle = \sum_{(x,y) \in E} \mathbb{P}(D(x) \neq D(y)). \tag{1.1}$$

pares adjacentes com cores distintas, ao 'custo' de $|V|$ bits aleatórios. Por construção, porém, o mesmo valor de $\langle c \rangle$ é obtido se estes $|V|$ bits são apenas independentes aos pares.

[1] Grafo é uma estrutura G(V,E) onde V é um conjunto não vazio de objetos denominados vértices e E é um conjunto de pares não ordenados de V, chamado arestas. Um exemplo de grafo é uma rede de computadores onde cada computador é um vértice e cada cabo de rede ligando dois computadores é uma aresta. Para uma introdução à física-estatística dos grafos veja [8].

[2] Esse é um problema NP-completo muito comum na literatura de algoritmos clássicos [9].

Em outras palavras, o mesmo resultado pode ser alcançado com um custo exponencialmente menor, de apenas $\lceil \log_2(|V|) \rceil$ bits independentes. Neste caso, a busca exaustiva por um corte "bom"pode ser feita testando-se todas as $2^{\lceil \log_2 |V| \rceil} \leq 2|V|$ possibilidades, o que mostra que uma algoritmo determinístico linear pode achar um corte passando pelo menos $|V|/2$ arestas.

Estados quânticos aleatórios

Suponha que um determinado protocolo de informação quântica necessite em certo ponto de um estado quântico puro aleatório de n q-bits (veremos a seguir exemplos de tais protocolos). Por 'aleatório', queremos dizer aqui um estado escolhido de forma uniforme do conjunto de todos os estados possíveis. A pergunta é: será possível, na prática, realizar este protocolo?

Se for realmente necessário um estado completamente aleatório com n arbitrariamente grande, a resposta é *não*. O motivo é que o espaço de Hilbert é simplesmente grande demais. Para alcançar com probabilidade uniforme todos os estados possíveis de n q-bits a partir de um estado inicial dado, é necessário poder aplicar uma rotação unitária também escolhida uniformemente (ou seja, a partir da medida invariante ou de Haar). Sabe-se, porém, que para gerar (com precisão ε) um unitário genérico de n qubits é necessário um circuito contendo [3]

$$\Omega\left(\frac{2^n \log(1/\varepsilon)}{\log\log 2^n}\right)$$

portas de 1- e 2- q-bits [11]. Em outras palavras, gerar um unitário uniformemente distribuído num espaço de dimensão 2^n requer um número também exponencial de portas de 1- e 2-q-bits, o que é na prática inviável.

Veremos abaixo uma maneira de relaxar esse requisito em determinadas circunstâncias, e que nos permitirá obter uma diminuição drástica do número de portas de 1 e 2 q-bits usadas.

1.1.1 Uso de 1-desenho em protocolos Quânticos

Um conceito central nesta dissertação é o de um *k-desenho quântico* [12]. Antes de definirmos formalmente o seu significado, é útil introduzi-lo em sua forma mais simples com um típico problema de Alice e Bob dado em [13]. Suponha que Alice queira enviar um q-bit

[3] Nessa dissertação usaremos a notação de ordem usada tradicionalmente na literatura da ciência da computação [10]. Nessa notação $f = \Omega(g)$ indica que, assintoticamente, f é inferiormente limitada por g. Também usaremos $f = O(g)$ (f é superiormente limitada por g) e $f = \Theta(g)$ (f é superior e inferiormente limitada por g) .

para Bob de maneira sigilosa, encriptado de maneira que apenas ele possa decifrá-lo. Em outras palavras, ela quer que qualquer espião que porventura intercepte a comunicação não possa obter qualquer informação (clássica ou quântica) útil. Para não destruir a informação contida no estado ρ, o protocolo de encriptação de Alice deve estar restrito a operações unitárias. Ou seja, antes de enviar seu estado pelo canal, Alice faz uma rotação no estado ρ e assim

$$\rho \mapsto \rho' = U\rho U^\dagger \ .$$

Repare, entretanto, que se Alice usa sempre o mesmo unitário U, toda a informação ainda está contida em ρ', só que em outra base. Para ver isso repare que ρ e ρ' têm os mesmos coeficientes em bases distintas. A situação muda se Alice escolher um unitário aleatoriamente a cada q-bit que mandar. Vamos ver o que acontece se ela escolhe um unitário com probabilidade dada pelo ensemble($\mu(U)$). Considere, por exemplo, que o ensemble $\mu(U)$ seja o de Haar($\mathcal{H}$), ou seja, distribuído sobre todos os unitários de 1 q-bit de maneira unitariamente invariante ($\mathcal{H}(U) = \mathcal{H}(VU) = \mathcal{H}(UV)$), para qualquer unitário V. Nesse caso

$$\rho \to \rho' = \int d\mathcal{H}(U)\ U\rho U^\dagger \ , \tag{1.2}$$

mas repare que $V\rho' V^\dagger = \rho'$ para qualquer V e portanto $\rho' = \mathbb{I}/2$. Ou seja, Alice realiza um unitário desconhecido para o espião, e portanto para ele o estado parece completamente misto. Se Bob não souber qual unitário Alice aplicou, Bob também não obterá informação nenhuma. Uma maneira de contornar esse problema é Bob e Alice possuírem bits clássicos aleatórios correlacionados, previamente compartilhados, de maneira que o unitário escolhido seja parametrizado pelo bit clássico.

Entretanto, esse protocolo com o ensemble $\mathcal{H}$ tem alguns problemas: para encriptar um só q-bit Alice deverá ser capaz de realizar qualquer porta de 1 q-bit, e além disso ela deverá possuir muitos bits clássicos aleatórios compartilhados com Bob para parametrizar, até uma certa precisão, o unitário escolhido. Felizmente existe uma escolha diferente de μ que pode solucionar esses problemas. Considere que agora Alice usa 2 bits aleatórios (b_1 e b_2), sem viés, e realiza o unitário $\sigma_x^{b_1}\sigma_z^{b_2}$. Chame essa distribuição de μ_{b_1,b_2}. As vantagens desse protocolo são claras: só são necessários 2 bits compartilhados e a gama de portas realizadas é bem mais simples, entretanto:

$$\rho' = \frac{1}{4}\sum_{i=1}^{2}\sum_{j=1}^{2} \sigma_x^j\sigma_z^i\rho\sigma_z^i\sigma_x^j = \mathbb{I}/2 \ .$$

Para ver isso basta calcular termos como

$$\frac{1}{4}\sum_{i=1}^{2}\sum_{j=1}^{2} \sigma_x^j\sigma_z^i \left|0\right\rangle\left\langle 0\right| \sigma_z^i\sigma_x^j = \frac{1}{2}\sum_{j=1}^{2} \sigma_x^j \left|0\right\rangle\left\langle 0\right| \sigma_x^j = \mathbb{I}/2 \ .$$

Essa distribuição é o que chamamos de um 1-desenho. Repare que ela não é uniforme, mas reproduz o mesmo valor para a eq. (1.2). Essa estratégia contém a essência dos k-desenhos: podemos poupar recursos usando distribuições diferentes do ensemble uniforme mas que reproduzem alguma propriedade útil desta. Naturalmente, nem todas as propriedades podem ser reproduzidas com um 1-desenho como o apresentado nessa Seção. Em particular um 1-desenho reproduz a média ou o o momento de ordem 1 do estado ρ enquanto um k-desenho reproduz um momento estatístico de ordem k de ρ. Isso ficará claro na próxima Seção onde veremos exemplos onde surge a necessidade do uso de 2-desenhos, que serão o alvo central dessa dissertação.

Em outra situações veremos que pode ser útil obter desenhos de estados ao invés de desenhos de unitários. Repare ainda que o ensemble de unitários μ aplicado em um estado fixo gera um estado que também é bem distribuído sobre o espaço de estados. Em particular se $\mu = \mathcal{H}$ o estado vai para uma distribuição que também é invariante por rotações, e se μ é um k-desenho de unitários os estados gerados serão k-desenhos de estados.

1.1.2 Uso de 2-desenhos em Algoritmos Quânticos

Nessa Seção apresentaremos tarefas onde será suficiente obter 2-desenhos para reproduzir propriedades da medida uniforme. O primeiro protocolo visa estimar a fidelidade média da implementação física de uma porta lógica. Abordaremos em seguida outros algoritmos que exibem potencialidades do uso de 2-desenhos.

Estimativa da Fidelidade média com estados aleatórios

Um algoritmo quântico no modelo de circuitos corresponde à aplicação de uma operação unitária U em um estado inicial $|\psi\rangle$. Em experiências reais, sempre existem imperfeições que se traduzem na realização de um circuito diferente do almejado. Chamaremos a operação quântica efetivamente realizada de $\mathcal{E}$, a qual pode ser em principio um mapa completamente positivo qualquer. Para o desenvolvimento de componentes que realizem computação quântica, é imperativo que possamos medir o quão bom é esse componente. Uma medida que é frequentemente usada é a fidelidade [11]. Dado o estado $|\psi\rangle$, a fidelidade entre o estado desejado e o efetivamente gerado é

$$F_{|\psi\rangle}(U, \mathcal{E}) = F(U|\psi\rangle\langle\psi|U^\dagger, \mathcal{E}(|\psi\rangle\langle\psi|)) = \langle\psi|U^\dagger\mathcal{E}(|\psi\rangle\langle\psi|)U|\psi\rangle .$$

Para medir o desempenho do dispositivo como um todo, precisamos tornar esta medida independente do estado de entrada. Uma maneira conveniente é calcular a sua média sobre

os possíveis estados de entrada:

$$F(U,\mathcal{E}) = \int \langle\psi| U^\dagger \mathcal{E}(|\psi\rangle\langle\psi|) U |\psi\rangle \, d\psi.$$

Nesta equação, a integral é tomada numa medida no espaço de Hilbert que é unitariamente invariante (medida Fubini-Study). Assim, como na Seção anterior, poderíamos resolver o problema se fosse possível obter o ensemble uniforme sobre unitários $\mathcal{H}$. Esta distribuição seria aplicada em um estado fixo($|0\rangle$) e o estado gerado seria dado pela medida unitariamente invariante de estados. Podemos então reescrever

$$F(U,\mathcal{E}) = \int \langle 0| V^\dagger U^\dagger \mathcal{E}(V|0\rangle\langle 0|V^\dagger) UV|0\rangle \, d\mathcal{H}(V). \tag{1.3}$$

Queremos estimar F experimentalmente, o que à primeira vista pode parecer requerer que sejam realizadas tomografias nos estados $\mathcal{E}(V|0\rangle\langle 0|V^\dagger)$. Fazendo isso, porém, o número de experimentos e passos computacionais clássicos para o pós-processamento é exponencialmente crescente com o número de q-bits [11]. Podemos eliminar a necessidade da tomografia de estados de duas maneiras. Vamos começar com a maneira que é mais simples mas que exige um outro dispositivo capaz de realizar $U^\dagger$ com grande fidelidade. Nesse caso podemos realizar o circuito descrito pela figura 1.1 e medir a projeção do estado final no estado $|0\rangle$.

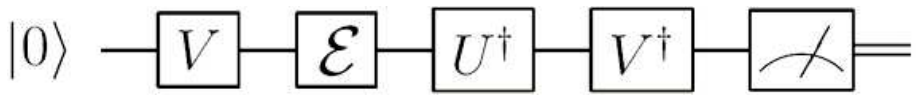


Figura 1.1: Circuito para estimativa da fidelidade de processo.

Dessa maneira, probabilidade do resultado de uma medida projetiva no final ser $|0\rangle$ é dado por

$$p = \langle 0| V^\dagger U^\dagger \mathcal{E}(V|0\rangle\langle 0|V^\dagger) UV|0\rangle \ . \tag{1.4}$$

Comparando com a eq. (1.3) vemos que essa probabilidade é justamente a fidelidade média da porta que queremos estudar. A média pode ser calculada por uma média simples com 'r' realizações e o erro na estimativa é $O(\frac{1}{\sqrt{r}})$, ou seja, não tem nenhuma dependência com o número de q-bits em que $\mathcal{E}$ age.

Como mencionamos anteriormente, porém, a geração de unitários aleatórios é em geral ineficiente, e portanto novamente o algoritmo pode parecer inviável de se realizar à medida que o número de q-bits aumenta. Entretanto nota-se que algo similar ao problema da eq. (1.1) acontece, ou seja, a probabilidade p depende de um polinômio de grau $(2,2)$

em $(V, V^\dagger)$. Assim, não é necessário na verdade selecionar V a partir de uma distribuição completamente uniforme, mas apenas de uma que seja indistinguível daquela no cálculo de médias de funções que sejam polinômios do grau desejado. Um ensemble com essa propriedade é chamada de 2-desenho de unitários (*unitary 2-design*). Formalmente, podemos definir:

Definição 1.1.1. *Um 2-desenho de unitários é um ensemble de operadores unitários μ sobre o conjunto desses unitários tal que*

$$\int U^\dagger MUNU^\dagger OU d\mu(U) = \int U^\dagger MUNU^\dagger OU \, d\mathcal{H}(U) \, , \tag{1.5}$$

para quaisquer operadores lineares M, N e O e onde $\mathcal{H}$ é o ensemble uniforme(Haar).

É fácil ver que qualquer distribuição que satisfaz essa definição será suficiente para calcular a média da eq. (1.4) sobre escolhas de V. Também é evidente que um 1-desenho como o da Seção anterior não é em geral suficiente para reproduzir o efeito da medida uniforme na eq. (1.5), e a razão é que agora é necessário reproduzir os momentos de segunda ordem da distribuição uniforme. Em geral definimos um k-desenho (de unitários) como uma distribuição sobre o conjunto de unitários que tem os momentos de k-ésima ordem iguais aos da distribuição uniforme. Na Seção 2.3 introduziremos uma versão mais formal desta definição que nos será mais útil.

O protocolo que acabamos de apresentar parece conter um argumento circular: para estimarmos a fidelidade $F(U, \mathcal{E})$ do equipamento, parece ser necessário já possuir uma outra máquina capaz de realizar $U^\dagger$ com alta fidelidade. Em realidade isto não verdade: mesmo uma máquina que realiza $U^\dagger$ de forma imperfeita já basta. O ponto é que o erro introduzido pela realização de $U^\dagger$ imperfeita será descorrelacionado do erro introduzido por U, e por isso podemos ter uma cota inferior da fidelidade média. Esse argumento intuitivo é justificado em [14].

Tomografia seletiva de processos (SEQPT) [15]

No algoritmo acima vimos como é possível calcular a média da fidelidade de um processo. Entretanto ela não dá nenhuma informação sobre a operação $\mathcal{E}$. Esse problema pode ser atacado eficientemente com o protocolo proposto em [15]. A estratégia do protocolo SEQPT é mapear a tomografia num problema de estimativa de fidelidade média. E como vimos esse problema pode ser solucionado eficientemente com um 2-desenho.

Seja a decomposição de $\mathcal{E}$ em uma base de operadores unitários $\{E_m\}$ tais que $\mathrm{Tr}(E_m E_{m'}^\dagger) = 2^n \delta_{mm'}$ e $\mathrm{Tr}(E_m) = 2^n \delta_{m0}$(essa base pode ser formada pelos produtos de operadores de

Pauli, por exemplo):

$$\mathcal{E}(\rho) = \sum_{mm'} \chi_{mm'} E_m \rho E_{m'}^\dagger \ .$$

Fazer uma tomografia deste processo equivale a estimar os coeficientes $\chi_{mm'}$. E a tomografia será dita eficiente se puder ser realizada com recursos polinomiais em n. No protocolo SEQPT cada $\chi_{mm'}$ pode ser obtido de forma seletiva e independente. Vamos nos restringir aqui aos termos diagonais já que o procedimento é mais simples nesse caso, entretanto, note o protocolo SEQPT não está limitado a esse elementos. Vamos supor que queremos ober o elemento χ_{mm}; então considere o canal quântico($\mathbb{E}_m$) gerado pela aplicação consecutiva do canal que queremos estudar($\mathbb{E}$) e do conjugado do m-ésimo elemento da base($E_m^\dagger$) . A fidelidade média desse canal, em relação à identidade, é dada simplesmente por:

$$F_m(\mathcal{E}_m, \mathbb{I}) = \int \langle\psi| E_m^\dagger \mathcal{E}(|\psi\rangle \langle\psi|) \ E_m |\psi\rangle \, d\psi = \frac{2^n \chi_{mm} + 1}{2^n + 1} \ .$$

A prova desse resultado é simples e decorre diretamente das relações de ortogonalidade da base e da identidade [16]:

$$\int \langle\psi| O_1 |\psi\rangle \langle\psi| O_2 |\psi\rangle \, d\psi = \frac{Tr[O_1]Tr[O_2] + Tr[O_1 O_2]}{2^n(2^n + 1)} \ .$$

Ou seja, com esse circuito conseguimos mapear os termos diagonais da decomposição de $\mathcal{E}$ na fidelidade média de um processo simples. Como temos um jeito eficiente de calcular a fidelidade média, podemos então fazer a tomografia do processo de maneira eficiente. Um circuito similar é capaz de mapear os termos não-diagonais de $\mathcal{E}$ em fidelidades médias e assim podemos completar toda a tomografia. É claro que como o número de termos de $\mathcal{E}$ cresce exponencialmente com n, o processo de tomografia completo continua ineficiente. A grande diferença desse método é a possibilidade de calcular seletivamente os termos desejados de maneira eficiente.

Emaranhamento "típico" e codificação super-densa

Mesmo para funções que não são (2,2) polinômios, o uso de 2-desenhos pode ser útil. Por exemplo: para estados puros de um sistema quântico bipartite, uma das mais úteis medidas de emaranhamento entre as duas partes é a entropia de von Neumann

$$S(\rho_A) = -Tr(\rho_A \ln \rho_A) \ ,$$

onde ρ_A é a matriz densidade reduzida de qualquer uma das partes [11]. Sabe-se há tempos que, quando a dimensão n do sistema como um todo é grande, o valor médio de S sobre

todos os estados puros fica próximo do seu valor máximo possível (ou seja, de $\ln n_A$, onde n_A é a dimensão da menor parte) [17]. Em outras palavras, um estado 'típico' deste sistema deve ter um emaranhamento quase máximo.

Este resultado supões que a média é tomada usando-se a medida unitariamente invariante no conjunto de estados. Na prática, mais uma vez temos de levar em conta que gerar um estado distribuído de fato desta maneira requer uma quantidade de recursos que cresce exponencialmente com n. Uma pergunta relevante então é: será possível, usando apenas recursos polinomiais em n, gerar estados que sejam ainda "tipicamente" emaranhados, mesmo que não sejam de fato completamente aleatórios?

Uma resposta positiva foi dada em [18], onde se demonstrou que estados distribuídos de acordo com um 2-desenho também possuem emaranhamento 'típico'. Como a entropia de von Neumann não é um polinômio, esta conexão não é óbvia. Estes autores usaram porém o fato de que

$$S(\rho_A) \geq \log_2 Tr(\rho_A^2).$$

Como $tr(\rho_A^2)$ é uma função polinomial de grau 2 em ρ, seu valor médio tomado usando um 2-desenho é o mesmo obtido com o ensemble uniforme - ou seja, quase máximo. A desigualdade acima garante então que o mesmo é verdade para S. Para mostrar explicitamente que o valor médio de $Tr(\rho_A^2)$ pode ser calculado com um 2-desenho vamos expandir o estado global ($|\psi\rangle\langle\psi|$) na base de produtos de Pauli, veja Seção 2.2 para definições e propriedades dessa base, obtendo $|\psi\rangle\langle\psi| = \sum_{\vec{P}\in\{0,X,Y,Z\}^n} \sigma_{\vec{p}}\, \xi(\vec{p})$. Assim:

$$\begin{aligned} \left\langle Tr(\rho_A^2)\right\rangle =& \int \mathrm{Tr}(Tr_B(|\psi\rangle\langle\psi|))d\mathcal{H} \\ =& 2^{n-n_A} \sum_{\vec{p}\,:\,\forall j\notin A, p_j=0} \left\langle \xi^2(\vec{p})\right\rangle \ , \end{aligned} \tag{1.6}$$

onde aparecem os momentos de segunda ordem de em $\xi(\vec{p})$. Para os detalhes veja [18].

Por sua vez, a geração eficiente de estados quase maximamente emaranhados entre bipartições é útil para diversas outras tarefas. Um exemplo é o algoritmo de codificação superdensa alheia de q-bits (*oblivious superdense coding of qubits*), descrito em [19]. Recorde que no protocolo usual de codificação super-densa[11] Alice utiliza um estado de Bell preexistente mais o envio de um q-bit para transmitir 2 bits de informação clássica para Bob. Da mesma forma, um total de $2n$ bits clássicos podem ser enviados se Alice compartilha com Bob n pares de Bell, e envia n qubits. Um aspecto importante deste protocolo é que as ações de Alice podem ser realizadas de forma 'alheia ao seu conhecimento', isto é, sem que ela precise conhecer os bits exatos a serem transmitidos.

No protocolo de [19], o objetivo é enviar, com grande probabilidade, um estado $|\psi\rangle$ de $2n$ bits *quânticos* para Bob utilizando os mesmos n pares de Bell previamente compartilhados, juntamente com o envio de n q-bits. Note que, ao contrário do protocolo de teletransporte, aqui não se permite o envio de informação clássica de forma 'gratuita', ou seja, qualquer transmissão de informação clássica só pode ser feita por meio do envio de mais q-bits, os quais também devem ser contabilizados. Neste caso, realizar esta tarefa de forma 'alheia' revela-se impossível, mas os autores de [19] mostram que ela pode ser realizada, probabilisticamente, quando o estado $|\psi\rangle$ é conhecido por Alice. Esta probabilidade tende a 1 para estados maximamente emaranhados, mas é baixa para estados não emaranhados. Sabemos porém que, aplicando-se uma rotação aleatória U sobre $|\psi\rangle$, obtém-se com grande probabilidade um estado quase maximamente emaranhado. Se Alice e Bob compartilham também previamente números aleatórios, podem usá-los para gerar o mesmo U em ambos os lados, sem comunicação. Assim, Alice pode transmitir $U|\psi\rangle$, com grande probabilidade, e Bob pode então aplicar $U^\dagger$ para obter $|\psi\rangle$. O problema com isso é que, mais uma vez, para gerar um unitário U aleatório seriam necessários recursos exponenciais em n (tanto o tempo para gerar este operador, como o número de bits aleatórios que Alice e Bob precisariam compartilhar para poderem escolher o mesmo U). No entanto, basta na verdade que eles possam gerar estados quase maximamente emaranhados com grande probabilidade. E para isso, como vimos acima, basta que U seja escolhido a partir de um 2-desenho, o que pode ser feito apenas com recursos polinomiais em n.

Além das aplicações acima de 2-desenhos, existem ainda diversas outras que não descreveremos. Podemos citar entre elas a destilação de emaranhamento[20] , a criação de canais privados(*private channels*)[13], etc. Mais aplicações continuam sendo descobertas.

1.2 Circuitos Quânticos Aleatórios

Apesar das dificuldades de se gerar objetos quânticos perfeitamente aleatórios, uma amostragem eficiente a partir de um k-desenho é algo possível. No caso de k-desenhos para estados, muitos autores deram construções, aproximadas ou exatas, que são eficientes para um k arbitrário. Em outras palavras, existem conjuntos de estados de cardinalidade polinomial, tanto em n como em k, que são k-desenhos (de estados) $\mu(\psi)$ [13].

Por outro lado, para construir um k-desenho de unitários μ precisamos que o próprio circuito resulte em um operador unitário pseudo-aleatório U, distribuído de acordo com $\mu(U)$. Até agora, não foi encontrada nenhuma construção desse tipo que seja eficiente para todo k. O mais próximo que já se conseguiu foi uma construção recente baseada em *tensor product expanders* [21], a qual, para um dado n, gerava eficientemente um k-desenho aproximado de unitários para valores de k até $O(n/\log n)$. Outros autores também

criaram construções eficientes de 1- e 2-desenhos aproximados, usando técnicas que são específicas para esses valores de k. Vale lembrar que qualquer k-desenho de unitários μ pode ser usado para gerar o k-desenho correspondente de estados, aplicando-se um operador pseudo-aleatório tirado de μ a um estado inicial fixo. Entretanto, nem todos os k-desenhos de estados precisam ser gerados dessa forma.

Nessa dissertação nós analisamos um esquema particular para gerar operadores quânticos pseudo-aleatórios através de *circuitos quânticos aleatórios* (CQAs) [22]. De maneira geral, essa é a classe de circuitos de n q-bits que resulta da aplicação repetida de portas de 2 q-bits escolhidas aleatoriamente segundo um determinado ensemble. Mais especificamente, definimos aqui um CQA como sendo um circuito obtido do seguinte algoritmo simples:

1. Escolha de forma aleatória um par de q-bits (i, j) a partir de um dado conjunto Γ de pares possíveis. Escolha ainda de forma aleatória uma ordem para este par.

2. Aplique sobre esse par uma porta de 2 q-bits W_{ij}, selecionada aleatoriamente de um ensemble μ sobre o conjunto de todas as portas desse tipo.

3. Repita os passos anteriores.

No primeiro passo acima, a escolha de Γ deve refletir as condições geométricas ou topológicas que porventura restrinjam a aplicação de portas entre q-bits, como por exemplo o caso de q-bits dispostos sobre uma rede regular de dimensão d e interagindo somente com primeiros vizinhos. Tecnicamente, podemos enxergar Γ como um grafo cujos vértices representam os n q-bits e cujas arestas ligam os pares que podem ser conectados por uma porta de 2 q-bits. Neste caso, em cada passo do CQA selecionamos de forma uniforme e aleatória uma aresta de Γ, e ainda uma de duas ordenações possíveis.

Grande parte das técnicas usadas nesta dissertação são independentes da forma de Γ, desde que este seja um grafo conexo. Neste caso, pode-se mostrar por argumentos bastante gerais de teoria de grupos que, se μ tiver suporte não-nulo num conjunto de portas universais para computação quântica, então a distribuição sobre os operadores unitários de n q-bits resultante de circuitos desse tipo converge uniformemente para a distribuição de Haar [22, 23].

Por simplicidade, porém, a análise detalhada de convergência que faremos no Capítulo 5 será restrita ao caso em que Γ é o grafo completo, ou seja, quando a escolha é uniforme sobre todos os $n(n-1)$ pares ordenados possíveis. Esta é uma suposição razoável se o CQA for fisicamente implementado em uma arquitetura que permite a aplicação direta de portas de 2 q-bits entre qualquer par de q-bits (tal como armadilhas de íons, por exemplo [24, 25]). Existem até modelos físicos realistas onde tais circuitos surgiriam naturalmente

(por exemplo, um gás de q-bits a alta temperatura interagindo através de colisões de 2 corpos [26]).

Como já foi dito acima, em geral a convergência à distribuição uniforme levará um número de passos exponencial em n para ser atingida com determinada precisão. Entretanto, começando com o trabalho seminal de Emerson e colaboradores [22], diversos autores usaram ferramentas numéricas e analíticas para desvendar quais propriedades da distribuição de Haar podem ser reproduzidas usando um CQA com profundidade apenas polinomial [27, 21, 18, 28, 29, 30, 31].

Um importante avanço nesse estudo foi obtido por Oliveira, Dahlsten e Plenio [18], que mostraram que pelo menos alguns aspectos do problema podem ser abordados analiticamente usando-se ferramentas da teoria de cadeias de Markov clássicas [32]. Especificamente, esses autores estudaram a geração, através de CQAs, de emaranhamento entre subdivisões bipartites dos n q-bits. Eles provaram que, após no máximo $O(n^3)$ passos, esse emaranhamento se torna indistinguível daquele de um estado "típico" de n q-bits tirado da medida de Haar. A abordagem baseada em cadeias de Markov foi estendida por Harrow e Low [27], que usaram-na para argumentar que um CQA gera um 2-desenho de unitários aproximado. Nesse trabalho eles analisam duas noções de 2-desenho de unitários ε-aproximados: a primeira, baseada na norma diamante, que chamaremos simplesmente de 2-desenho aproximado e a segunda, baseada no *twirling* de canais quânticos, que chamaremos de 2-desenho **de canal**[4]. Eles obtêm que depois de rodar o CQA por $O(n(n+\log \varepsilon^{-1}))$ passos é gerado um 2-desenhos de unitários para ambas as definições. Eles ainda obtém que o CQA com portas de 2 q-bits uniformes(U(4))[5] o tempo de convergência para 2-desenho **de canal** pode ser melhorado para $O(n \log(n/\varepsilon))$. Esses autores também sugeriram, sem demonstrar, que CQAs poderiam gerar eficientemente k-desenhos aproximados para k's mais altos.

A análise feita por Harrow e Low é intrincada, usando de forma sofisticada várias técnicas da teoria de cadeias de Markov. Entretanto, como mostramos na Seção 5.1 dessa dissertação, o principal argumento deles (para o caso $U(4)$) está na verdade incompleto, faltando um último passo não-trivial. Um dos principais objetivos dessa dissertação é apresentar um argumento alternativo (e mais simples) que confirma o resultado e as escalas de tempo de convergência para um 2-desenho **de canal** aproximado reivindicado na ref. [27].

Também provamos diversas outras propriedades de CQAs, muitas das quais já haviam sido conjecturadas por outros autores, baseados em evidências numéricas. Primeiramente, nós mostramos que essas escalas de tempo são robustas a mudanças na escolha precisa do

[4] Essas definições serão abordadas em detalhe na Seção 2.4.

[5] O ensemble U(4) é formada com uma medida uniforme(Haar) por todas as portas de 2 q-bits.

ensemble de unitários de 2 q-bits μ utilizada para gerar a porta W_{ij} no passo 2 acima. Escolhas diferentes de μ foram utilizadas na literatura. Por exemplo: em [18], W_{ij} foi gerada aplicando-se primeiro portas aleatórias de 1 q-bit U_i e V_j em cada q-bit, e a seguir uma porta CNOT. Em [27], os resultados para a escala de tempo de convergência de 2-desenhos citadas acima foram obtidos no caso onde W_{ij} é um operador unitário escolhido uniformemente e aleatoriamente a paritr de $U(4)$. Outras definições ligeiramente diferentes também foram utilizadas [33]. Foi conjecturado [27] que, para essencialmente qualquer ensemble μ onde as portas emaranhantes tenham probabilidade não-zero, deve-se observar tempos de convergência de mesma ordem.

Nessa dissertação nós tratamos tanto do caso de portas de 2 q-bits escolhidas uniformemente como em [27], quanto do caso onde só está disponível uma única porta de 2 q-bits fixa C (não necessariamente a CNOT), e portas de 1 q-bit escolhidas uniformemente. Vale notar que, enquanto o primeiro caso é matematicamente mais simples, o último é uma suposição mais razoável na maioria das situações experimentais, onde em geral é difícil implementar uma variedade de portas de 2 q-bits diferentes. Mostraremos que, sob determinadas condições, os dois ensembles de fato resultam na mesma escala de tempo de convergência, como conjecturado em [27]. Além disso, no caso das distribuições com "porta-fixa', nós conseguimos investigar a dependência detalhada da escala de tempo de convergência com a escolha de C. Nós também mostramos que, com a escolha correta de C, um CQA pode de fato convergir ligeiramente mais rápido do que um baseado nas portas arbitrárias de 2 q-bits. Isso confirma resultados numéricos obtidos em [28]. Mais ainda, nós podemos mostrar que a convergência do ensemble gerado pelo CQA para um 2-desenho **de canal** ocorre numa janela de tempo bem estreita, exibindo um efeito de *cutoff*. Essa propriedade é uma característica bem conhecida de muitas cadeias de Markov [34, 35, **?**], e foi observada numericamente no estudo da convergência para o emaranhamento "típico"em [18].

Todas as escalas de tempo de convergência citadas acima assumem que cada passo do circuito quântico aleatório é aplicado em sequência. Entretanto, como cada passo afeta apenas um único par de q-bits, é natural questionar se algum ganho de escala pode ser obtido por paralelização do circuito. Nós apresentamos fortes evidências analíticas e numéricas de que isso de fato ocorre: se as portas forem aplicadas em paralelo, a profundidade do circuito pode ser reduzida de um fator $O(n)$. Então, o tempo necessário para a convergência para um 2-desenho pode ser tão rápido quanto $O(\log n)$ o tempo de aplicação de uma porta. Essa é, de fato, a convergência mais rápida existente na literatura. Ela é compartilhada com o algoritmos proposto por [30] de mesmo tempo total e que necessita de $O(n \log(1/\varepsilon))$ portas de 1 e 2 q-bits mas que é, entretanto, específica para 2-desenhos. Vale notar que uma conjectura similar quanto à paralelização de CQAs foi também feita em [33], baseada em outros argumentos heurísticos e também numéricos.

1.3 Organização da dissertação

Essa dissertação está organizada da seguinte forma:

No Capítulo 2 introduzimos diversas ferramentas matemáticas e conceituais que serão utilizadas ao longo da dissertação. Embora a maior parte deste material não seja original, utilizaremos uma notação diferente da de outros autores, a qual nos parece simplificar consideravelmente a descrição e o entendimento dos cálculos de diversos resultados.

No Capítulo 3 mostramos como a evolução dos segundos momentos do CQA pode ser mapeada num problema de cadeias de Markov, e derivamos a dependência dos coeficientes desta cadeia em função do ensemble μ utilizado para gerar as portas de 2 q-bits. O material deste capítulo segue em linhas gerais o da referência [27], mas generalizando os resultados para uma classe de ensembles que chamamos de '*localmente randomizados*'. A dedução será simplificada por meio da notação introduzida no Capítulo 2.

No Capítulo 4 analisamos esta cadeia em detalhe e de vários pontos de vista. Em particular, mostramos como ela pode ser reinterpretada como um passeio aleatório cujas características gerais são independentes do ensemble utilizado. Ainda, mostramos como decompor a cadeia de uma forma que evidencia sua relação com problemas do tipo Ising clássico.

No Capítulo 5 estudamos a convergência desta cadeia de Markov, buscando uma cota para o seu tempo de convergência ao estado estacionário (ou 'tempo de mistura'). Discutiremos várias normas que têm sido utilizadas para medir a distância da estacionaridade e consequentemente o tempo de convergência. Explicamos onde o argumento da referência [27] falha, e apresentamos a nossa abordagem alternativa, baseada na decomposição encontrada no capítulo 4. Com isso obtemos o tempo de convergência da cadeia de Markov e obtemos as cotas de convergência a um 2-desenho aproximado segundo as duas diferentes noções discutidas na Seção 5.7. Procuraremos ainda obter o resultado final usando uma variedade de técnicas diferentes, entre elas a teoria de passeios aleatórios em grupos [36], com a qual investigaremos a existência de 'corte abrupto' em parte da cadeia.

No Capítulo 6 consideraremos a possibilidade de paralelização do algoritmo com a realização de mais de uma porta de 2 q-bits no mesmo intervalo de tempo, obtendo fortes evidências de que é possível ganhar um fator de ordem n no tempo de convergência.

No Capítulo 7 concluímos discutindo alguns problemas em aberto, entre eles a extensão da análise para os momentos superiores, e também a sua relação com descobertas recentes na comunidade de cadeias de Markov acerca das propriedades de cutoff da chamada dinâmica de Glauber de cadeias tipo Ising. [37, 38]

Capítulo 2

Ferramentas matemáticas

Neste capítulo introduzimos várias ferramentas que serão úteis para a análise dos CQAs. Iniciamos introduzindo um notação análoga à de Dirac, mas aplicada a espaços vetoriais formados por operadores. Prosseguimos relembrando algumas propriedades essenciais dos operadores de Pauli e seus produtos tensoriais. Finalmente, discutimos a noção de super-operadores de *twirl* (giro), e sua relação com a definição dos k-desenhos.

2.1 Notação de Dirac para Operadores e Super-operadores

Nosso CQA atua em um sistema quântico de n q-bits, com espaço de Hilbert $\mathcal{H}_n$. Chamemos de $\mathcal{O}_n$ o conjunto de operadores lineares em $\mathcal{H}_n$. Como se sabe, $\mathcal{O}_n$ pode ser visto como um espaço vetorial complexo, de dimensão 2^{2n}. Nós representamos *super*-operadores lineares atuando em $\mathcal{O}_n$ com letras maiúsculas com um "chapéu" ($\hat{}$) sobre-escrito. Por exemplo:

$$\hat{C}_X(A) \equiv XAX^\dagger \tag{2.1}$$

é o super-operador linear que mapeia, por conjugação pelo operador X, cada operador A.

Quando trabalharmos com super-operadores, será útil introduzir a seguinte notação análoga à de Dirac:

- Um "ket duplo" $|A\rangle\rangle$ representará um operador comum A. Em particular, o operador densidade $|\psi\rangle\langle\psi|$ associado com o estado quântico puro $|\psi\rangle$ será comumente denotado por $|\psi\rangle\rangle$.

- U "bra duplo" $\langle\langle A|$ vai representar o funcional linear $\mathrm{Tr}(A^\dagger \cdot\ )$, que mapeia operadores em escalares.

- Um "bracket duplo" $\langle\langle A|B\rangle\rangle$ vai representar o produto escalar de Hilbert-Schmidt: $A \cdot B = \mathrm{Tr}(A^\dagger B)$. Assim como com os brackets comuns de Dirac, é verdade que

$$\langle\langle A|B\rangle\rangle = \langle\langle B|A\rangle\rangle^* . \tag{2.2}$$

  Quando tanto A quanto B são Hermiteanos, o produto é real. Em particular, estados puros $|\psi\rangle$ normalizados no sentido usual, também o são com respeito a esse produto

$$\langle\langle \psi|\psi\rangle\rangle = \mathrm{Tr}\Big[\ |\psi\rangle\langle\psi|\ \ |\psi\rangle\langle\psi|\ \Big] = 1 \tag{2.3}$$

- Um "produto externo duplo" $|A\rangle\rangle\langle\langle B|$ vai representar o super-operador linear $\big[\mathrm{Tr}(B^\dagger \cdot\ )\big]\, A$. Assim, podemos identificar, como de praxe, $|A\rangle\rangle\langle\langle B|\ |X\rangle\rangle = |A\rangle\rangle\ \langle\langle B|X\rangle\rangle$ sem a necessidade de parênteses. Pode se checar diretamente que todas as outras propriedades familiares da notação usual de Dirac também se mantêm.

Apesar dessa notação ser particularmente útil quando trabalhamos com a natureza vetorial dos operadores, ela pode se tornar inconveniente em expressões que envolvem produtos usuais de operadores (produto de matrizes). Nesses casos nós retornaremos à notação usual (sem kets) para operadores, isto é, escrevendo A em vez de $|A\rangle\rangle$.

2.2 Base de Pauli

Uma base conveniente para $\mathcal{O}_n$ é o conjunto de produtos tensoriais de operadores de Pauli de 1 q-bit (a "base de Pauli"): $|\sigma_{\vec{p}}\rangle\rangle \equiv |\sigma_{p_1}\rangle\rangle \otimes \ldots \otimes |\sigma_{p_n}\rangle\rangle$. Aqui, $\vec{p} \equiv (p_1 \ldots p_n)$, $p_j \in \{0, X, Y, Z\}$ e $\big|\sigma_{p_j}\big\rangle\big\rangle$ são os operadores de Pauli usuais de 1 q-bit. Essa base é ortogonal, mas não normalizada, sob a norma de Hilbert-Schmidt. As relações de ortonormalidade e completude são

$$\langle\langle \sigma_{\vec{p}}\ |\sigma_{\vec{q}}\rangle\rangle = 2^n \delta_{\vec{p},\vec{q}}\ , \tag{2.4}$$

$$\sum_{\vec{p}} |\sigma_{\vec{p}}\rangle\rangle\langle\langle \sigma_{\vec{p}}\ | = 2^n \hat{I}\ . \tag{2.5}$$

Expandindo um operador densidade $|\rho\rangle\rangle$ na base de Pauli, podemos escrever

$$|\rho\rangle\rangle = 2^{-n/2} \sum_{\vec{p}} \xi(\vec{p})\ |\sigma_{\vec{p}}\rangle\rangle\ , \tag{2.6}$$

onde os coeficientes numéricos $\xi(\vec{p})$ podem ser extraídos usando a eq. (2.4):

$$\xi(\vec{p}) = 2^{-n/2} \langle\langle \sigma_{\vec{p}} \, | \rho \rangle\rangle \, . \tag{2.7}$$

Em particular, o coeficiente do operador identidade $|\sigma_{\vec{0}}\rangle\rangle$ é sempre

$$\xi\left(\vec{0}\right) = 2^{-n/2} \langle\langle \sigma_{\vec{0}} \, | \rho \rangle\rangle = 2^{-n/2} \mathrm{Tr}(\rho) = 2^{-n/2} \, . \tag{2.8}$$

Como produtos de operadores de Pauli e operadores densidade são ambos Hermiteanos, os demais coeficientes são sempre números reais. Além disso, a soma dos seus quadrados é

$$\sum_{\vec{p}} \xi^2(\vec{p}) = 2^{-n} \sum_{\vec{p}} \langle\langle \rho | \sigma_{\vec{p}} \rangle\rangle \, \langle\langle \sigma_{\vec{p}} \, | \rho \rangle\rangle = \langle\langle \rho | \rho \rangle\rangle = \mathrm{Tr}\rho^2 \tag{2.9}$$

onde usamos eq. (2.2) e a relação de completude (2.5).

Em particular, quando $|\rho\rangle\rangle$ é um estado puro $|\psi\rangle\rangle$, os coeficientes quadrados $\xi^2(\vec{p})$ formam uma distribuição de probabilidade sobre o conjunto $\{0, X, Y, Z\}^n$ [18]. Nas seções seguintes nós usaremos **vetores de probabilidade** formados pela coleção desses 4^n números num vetor linha.

Para evitar confusão vale notar que, embora nesses casos o estado seja puro, as distribuições $\xi^2(\vec{p})$ são sempre 'mistas', ou seja, têm coordenadas não-nulas sobre um conjunto de vários valores de $\vec{p}$. Por exemplo, o estado $|0 \ldots 0\rangle$ tem operador densidade $\left[\frac{1}{2}\left(I + Z\right)\right]^{\otimes n}$, e portanto seus coeficientes de Pauli ao quadrado são iguais a 2^{-n} para todos os *strings* $\vec{p}$ com $p_i \in \{0, Z\}, \forall i$.

2.3 Super-operadores de twirl

Em protocolos de Informação Quântica, frequentemente é útil apagar informação de forma controlada, aplicando-se uma rotação randomizada a um ou mais q-bits. Essa técnica, conhecida como *twirling*, é conhecida há bastante tempo, por exemplo foi já utilizada nos protocolos de destilação de emaranhamento bipartite de Bennett e colaboradores [20].

Formalmente, a forma mais simples de twirling é realizada aplicando-se (por exemplo, a um operador densidade num conjunto de m q-bits) um super-operador da forma

$$\hat{G}_{\mu}^{(1)}(\cdot) \equiv \int W \cdot W^{\dagger} d\mu(W). \tag{2.10}$$

Nessa expressão, W é uma rotação unitária dos m q-bits, e $\mu(W)$ é uma medida de probabilidade sobre o conjunto desses operadores. Nos referiremos a $\hat{G}_{\mu}^{(1)}$ como um "1-twirl de

m q-bits" (um ligeiro abuso de linguagem, já que esse super-operador não atua sobre o espaço de Hilbert $\mathcal{H}_m$, mas sobre $\mathcal{O}_m$). Em muitas aplicações de twirling, onde se deseja um apagamento completo de informação, μ é tomada como a medida uniforme (Haar) $\mathcal{H}$. Nessa dissertação nós também faremos uso de twirlings sobre medidas de probabilidades mais gerais.

Uma generalização útil do conceito de 1-twirl é a aplicação de rotações aleatórias *correlacionadas* a k conjuntos idênticos de m q-bits. Formalmente, podemos definir o seguinte super-operador:

$$\hat{G}_{\mu}^{(k)}(\cdot) \equiv \int W^{\otimes k} \cdot W^{\dagger \otimes k} d\mu(W) \; . \tag{2.11}$$

Nós iremos nos referir a isso, novamente com algum abuso de linguagem, como um "k-twirl de m q-bits" (deve ser entendido que isso se refere a um super-operador atuando em $\mathcal{O}_m^{\otimes k}$, ou k cópias do espaço de operadores em m q-bits).

Podemos usar este conceito para dar uma nova definição de k-desenho, um pouco diferente mas equivalente à que vimos na Seção 1.5.

Definição 2.3.1. *O ensemble μ é um k-desenho unitário exato se o seu k-twirl $\hat{G}_{\mu}^{(k)}$ reproduz o k-twirl $\hat{G}_{\mathcal{H}}^{(k)}$ gerado pela distribuição uniforme $\mathcal{H}$, ou seja, se*

$$\int W^{\otimes k} \cdot W^{\dagger \otimes k} d\mu(W) = \int W^{\otimes k} \cdot W^{\dagger \otimes k} d\mathcal{H}(W) \; . \tag{2.12}$$

O ensemble μ que satisfaz a eq. (2.3.1) pode ser usado em todos os algoritmos apresentados na Seção 1.1. Essa definição equivale a da eq. (1.5), ver Corolário 5.2.2 de [30].

Em seções futuras nós iremos precisar de algumas propriedades de super-operadores de k-twirl, a maioria das quais foram obtidas na ref. [27], Seção 3. Abaixo apresentamos as demonstrações numa forma mais simples, explorando as propriedades convenientes da notação introduzida na Seção 2.1.

Lema 2.3.1. *Se $|A\rangle\rangle \in \mathcal{O}_n^{\otimes k}$ é um operador Hermiteano, então $\hat{G}_{\mu}^{(k)} |A\rangle\rangle$ é também um operador Hermiteano.*

Prova: Imediata pela definição (2.11).

Lema 2.3.2. *Se $\mu(W) = \mu(W^{\dagger})$ para todos os operadores $W \in \mathcal{O}_n$, então $\hat{G}_{\mu}^{(k)}$ é um super-operador Hermiteano, e sua matriz na base de Pauli é real e simétrica.*

Prova: Considere o super-operador de "k-conjugação" $\hat{C}_W^{(k)} \equiv W^{\otimes k} \cdot W^{\dagger \otimes k}$. Usando a propriedade cíclica do traço, assim como eq. (2.2), é fácil ver que $\hat{C}_W^{(k)\,\dagger} = \hat{C}_{W^\dagger}^{(k)}$. Portanto

$$\hat{G}_\mu^{(k)\,\dagger} = \int \hat{C}_W^{(k)\,\dagger} d\mu(W) = \int \hat{C}_{W^\dagger}^{(k)} d\mu(W) = \int \hat{C}_W^{(k)} d\mu(W^\dagger).$$

Quando $\mu(W) = \mu(W^\dagger)$, isso é então igual a $\hat{G}_\mu^{(k)}$. Dado isso, para a segunda afirmação nós apenas precisamos mostrar que a matriz de $\hat{G}_\mu^{(k)}$ na base de Pauli é real. Como os operadores de Pauli são Hermiteanos, isso ocorre diretamente usando a eq. (2.2) e o Lema 2.3.1.

Lema 2.3.3. *Seja $|\Pi\rangle\!\rangle \in \mathcal{O}_n^{\otimes k}$ qualquer operador de permutação que permute k cópias de $\mathcal{H}_n$ entre elas próprias. Para qualquer escolha de medida μ, $|\Pi\rangle\!\rangle$ e o seu bra associado $\langle\!\langle\Pi|$ são respectivamente autovetores à direita e à esquerda do super-operador de k-twirl $\hat{G}_\mu^{(k)}$, ambos com autovalor $\lambda = 1$, i.e.:*

$$\hat{G}_\mu^{(k)} |\Pi\rangle\!\rangle = |\Pi\rangle\!\rangle\,; \tag{2.13}$$

$$\langle\!\langle\Pi| \hat{G}_\mu^{(k)} = \langle\!\langle\Pi| \tag{2.14}$$

Prova: Pela definição de operadores de permutação, $\left[\Pi, W^{\otimes k}\right] = 0$. O primeiro resultado segue substituindo-se isto diretamente em eq. (2.11). O segundo resultado segue imediatamente no caso em que $\hat{G}_\mu^{(k)}$ é Hermiteano, mas é de fato verdade para qualquer $\hat{G}_\mu^{(k)}$. Para ver isso, note que, para qualquer operador $|A\rangle\!\rangle \in \mathcal{O}_n^{\otimes k}$,

$$\begin{aligned}\langle\!\langle\Pi| \hat{G}_\mu^{(k)} |A\rangle\!\rangle &= \int \mathrm{Tr}\left[\Pi W^{\otimes k} A W^{\dagger \otimes k}\right] d\mu(W) = \int \mathrm{Tr}\left[W^{\otimes k} \Pi A W^{\dagger \otimes k}\right] d\mu(W) \\ &= \int \mathrm{Tr}\left[\Pi A\right] d\mu(W) = \mathrm{Tr}\left[\Pi A\right] = \langle\!\langle\Pi|A\rangle\!\rangle\end{aligned}$$

onde, na primeira linha nós comutamos novamente Π e $W^{\otimes k}$, e na segunda nós usamos a propriedade cíclica do traço.

Twirling Uniforme

No caso especial onde μ é a medida uniforme (Haar) $\mathcal{H}$, a simetria rotacional leva a propriedades adicionais para o super-operador de k-twirl $\hat{G}_{\mathcal{H}}^{(k)}$ [27]. Essas propriedades estão ligadas à chamada dualidade Schur-Weyl, uma construção para as representações de produtos tensoriais do grupo linear que são simultaneamente representações do grupo de permutações [39].

Considere novamente os operadores de permutação $|\Pi\rangle\!\rangle$ definidos acima. Estes $k!$ operadores formam uma representação do grupo de permutações S_k. Considerados como vetores, eles geram também um sub-espaço vetorial $\mathcal{S}_{k,n} = \text{span}\{|\Pi\rangle\!\rangle\} \subset \mathcal{O}_n^{\otimes k}$. O lema 2.3.3 significa que $\hat{G}_\mu^{(k)}\mathcal{S}_{k,n} = \mathcal{S}_{k,n}$ para qualquer ensemble μ. No caso do ensemble uniforme, porém, pode-se dizer mais (vide [27] para demonstrações):

Teorema 2.3.1. *[27] Os super-operadores Haar-twirl $\hat{G}_{\mathcal{H}}^{(k)}$ são projetores em $\mathcal{S}_{k,n}$.*

Em outras palavras, $\hat{G}_{\mathcal{H}}^{(k)}|V\rangle\!\rangle = 0$ para todo operador $|V\rangle\!\rangle$ ortogonal a $\mathcal{S}_{k,n}$.

Na maior parte desta dissertação, estaremos interessados particularmente no caso $k = 2$. Vamos então escrever uma expressão explícita para a matriz de $\hat{G}_{\mathcal{H}}^{(2)}$ na base de Pauli, usando o teorema 2.3.1. Para isto precisamos escrever os operadores $|\Pi\rangle\!\rangle$ nesta base. Como $k = 2$, só há dois operadores: a identidade $|I\rangle\!\rangle = |\sigma_{\vec{0}}\rangle\!\rangle$ e o operador de transposição (SWAP) $|S\rangle\!\rangle$. Podemos então usar

Lema 2.3.4. *[27] Para qualquer dimensão d, o operador SWAP S entre dois sistemas d-dimensionais pode ser escrito como*

$$S = \frac{1}{d}\sum_{p=0}^{d^2-1} \alpha_p \otimes \alpha_p \tag{2.15}$$

onde $\{\alpha_p\}$ é qualquer base de operadores Hermiteanos e ortogonais com $\langle\!\langle \alpha_p|\alpha_p\rangle\!\rangle = d$.

Em particular, para dois sistemas de n q-bits, $d = 2^n$ e podemos tomar $\alpha_p = \sigma_{\vec{p}}$.

É importante notar que os operadores de permutação $|I\rangle\!\rangle$ e $|S\rangle\!\rangle$ não são ortonogonais: $\langle\!\langle I|S\rangle\!\rangle = \text{Tr}(S)$. Uma base ortonormal $\{|\gamma_i\rangle\!\rangle\}$ para o subespaço $\mathcal{S}_{2,n}$ gerado por esses operadores pode ser construída usando o procedimento de Gram-Schmidt. Defina

$$\bar{S} \equiv 2^n S - I = \sum_{\vec{p}\neq\vec{0}} \sigma_{\vec{p}} \otimes \sigma_{\vec{p}} \equiv \sum_{\vec{p}\neq\vec{0}} \sigma_{\vec{p},\vec{p}} \tag{2.16}$$

onde, na última equação, nós apenas simplificamos a notação para maior clareza. Então $\langle\!\langle I|\bar{S}\rangle\!\rangle = 0$ e

$$\{|\gamma_i\rangle\!\rangle\} \equiv \left\{\frac{1}{2^n}I, \frac{1}{2^n\sqrt{2^{2n}-1}}\bar{S}\right\} \tag{2.17}$$

é uma base ortonormal, e pelo teorema 2.3.1

$$\hat{G}_{\mathcal{H}}^{(2)} = \sum_i |\gamma_i\rangle\!\rangle\langle\!\langle\gamma_i| \ . \tag{2.18}$$

Em outras palavras, a matriz de $\hat{G}_{\mathcal{H}}^{(k)}$ na base de Pauli é formada por dois blocos diagonais: um 1×1 correspondendo ao projetor sobre $|\sigma_{\vec{0}}\rangle\rangle$, e outro $(4^n - 1) \times (4^n - 1)$, sendo este último a matriz com todas as entradas uniformes e iguais a $\frac{1}{4^n - 1}$.

2.4 k-Desenhos de unitários ε-aproximados

O nosso objetivo nesta dissertação é usar um CQA para gerar um k-desenho de n qubits. Entretanto, a convergência do circuito para um k-desenho exato se dará apenas no limite em que o número de passos do circuito vai a infinito. Precisamos assim de critérios para decidir quando é que já temos um 'k-desenho aproximado' , ou seja, uma aproximação suficientemente boa de um k-desenho exato. Não há uma única forma de fazer isso, a escolha de qual é mais apropriada depende do fim prático que se quer dar ao k-desenho. Uma maneira é através de uma norma apropriada no espaço de super-operadores, por exemplo a norma diamante [40], dada por:

Definição 2.4.1. *A norma diamante do super-operador $\hat{G}$ é*

$$\left|\left|\hat{G}\right|\right|_{\diamond} = \sup_{d_{aux}} \sup_{\rho} \; Tr\left(\hat{G} \otimes \mathbb{I}_{d_{aux}}\left(\rho\right)\right) \; ,$$

onde d_{aux} é a dimensão de um espaço auxiliar e o segundo supremo é tomado sobre todos os operadores positivos semi-definidos de traço unitário (matrizes densidade) no espaço conjunto.

Esta norma tem uma interpretação operacional direta em termos da probabilidade de distinguir fisicamente dois super-operadores. Mais especificamente: dados dois super-operadores $\hat{G}_1$ e $\hat{G}_2$ suponha que aplicamos um ou outro sobre parte de um sistema (possivelmente de alta dimensão) preparado em um estado inicial ρ. Se fazemos então uma medição generalizada no estado de saída, então a probabilidade de obter um dado resultado final difere por no máximo $\left|\left|\hat{G}_1 - \hat{G}_2\right|\right|_{\diamond}$. Para o uso em algoritmos quânticos essa norma garante, por exemplo, que um k-twirl ε-aproximado pode substituir um k-twirl exato quando este for usado como uma sub-rotina em certo conjunto de q-bits (os quais estão em geral emaranhado com outros), com probabilidade no máximo ε de alterar o resultado da computação.

Podemos então estender a Definição 2.3.1 acima para:

Definição 2.4.2. *O ensemble μ é um k-desenho unitário ε-aproximado se o seu k-twirl $\hat{G}_{\mu}^{(k)}$ satisfaz*

$$\left|\left|\hat{G}_{\mu}^{(k)} - \hat{G}_{\mathcal{H}}^{(k)}\right|\right|_{\diamond} \leq \varepsilon \; . \tag{2.19}$$

Outra definição, específica para 2-desenhos aproximados, foi proposta por Dankert et al em [30]:

Definição 2.4.3. *Seja $\mu(U)$ um ensemble de operadores unitários. Então esse ensemble é um 2-desenho* ***de canal*** *ε-aproximado se*

$$\max_{\Lambda} \left|\left| \int U(\Lambda(U^\dagger \rho U))U^\dagger d\mu(U) - \int V(\Lambda(V^\dagger \rho V))V^\dagger d\mathcal{H}(V) \right|\right|_\diamond \leq \frac{\varepsilon}{d^2} \ , \tag{2.20}$$

onde a maximização é feito sobre todos os canais quânticos Λ e d é a dimensão do espaço de Hilbert (2^n para n q-bits).

Em outras palavras, na definição destes autores um ensemble μ é um 2-desenho ε-aproximado se ele tem o mesmo efeito que o ensemble uniforme (Haar) quando usado para realizar *twirling* em um canal qualquer. Essa definição é menos rígida do que a primeira, mas é suficiente por exemplo para garantir que o ensemble em questão possa ser usado nos algoritmos discutidos na Seção 1.1.

Veremos no Capítulo 5 que um CQA necessita de menos passos para atingir um 2-design ε-aproximado de canal do que no sentido mais rígido da Definição 2.4.2.

2.5 Noções de Cadeias de Markov

Nesta dissertação usaremos técnicas da literatura de cadeias de Markov. Boas referências incluem [41] e [32]; usaremos principalmente a nomenclatura desta última. Nessa Seção resumimos para fins de referência alguns resultados e definições que serão úteis, especialmente no capítulo 5.

2.5.1 Definição

Intuitivamente, uma cadeia de Markov finita é uma evolução probabilística em um dado conjunto finito de estados Ω, na qual cada passo da evolução não depende da história precedente. Assim, se o sistema está no elemento $x \in \Omega$, ele evolui em um passo para o elemento $y \in \Omega$ com probabilidade fixa $P(x, y)$. Chamaremos a matriz P, cuja dimensão é $|\Omega| \times |\Omega|$, de matriz de transição. Note que, para preservar a soma das probabilidades, esta deve ser uma matriz estocástica ou seja, a soma de cada linha vale 1.

Alternativamente, pode ser conveniente pensar nessa dinâmica como uma sequência de distribuições de probabilidades tomando valores em Ω. Assim temos:

$$\nu^{(t+1)}(y) = \sum_{x \in \Omega} \nu^{(t)}(x) \ P(x, y) \ ,$$

ou na notação mais frequente na literatura de cadeias de Markov:

$$\nu^{(t+1)} = \nu^{(t)} P \ .$$

(Observe que, nessa convenção, $\nu^{(t)}(x)$ é considerado um vetor linha, que multiplica a matriz de Markov pela esquerda).

A diferença da segunda para a primeira interpretação é a seguinte: se o sistema evolui a partir da condição inicial $X^{(0)} = x$, na primeira interpretação, após t passos vamos encontrar o sistema num elemento $z \in \Omega$. Se repetirmos esse procedimento várias vezes, o sistema pode evoluir para elementos diferentes, e poderemos calcular a frequência com que cada elemento é atingido. Essa frequência é justamente a distribuição de probabilidade da segunda interpretação $\nu_t = \nu_0 P^t$.

2.5.2 Convergência à estacionariedade

Uma distribuição π é dita estacionária para P quando satisfaz a equação $\pi = \pi P$. Estamos interessados em matrizes de transição que possuem distribuições estacionárias únicas, que são atingidos independentemente da distribuição inicial. Abaixo definiremos certas propriedades que são importantes para que possamos garantir que a matriz de transição é desse tipo.

Uma propriedade básica de muitas cadeias de Markov é a *ergodicidade*. Se uma cadeia P é *ergódica* ela possui uma única distribuição de probabilidade estacionária π, para a qual qualquer condição inicial ν converge (ver Capítulo 4 de [32]), ou seja: $\nu P^t \to \pi$. Algebricamente, π é o único autovetor (à esquerda) de P com autovalor de módulo 1, sendo todos os demais autovalores com módulo < 1.

Uma condição suficiente para uma cadeia ser ergódica (em espaços finitos) é quando ela é *irredutível* (não pode ser escrita como a soma direta de sub-cadeias independentes) e *aperiódica* (nenhuma condição inicial leva a evoluções repetitivas). Para provar irredutibilidade de P, basta mostrar que, para quaisquer dois elementos $p, q \in \Omega$, haverá uma probabilidade não nula de transição de p para q após um número suficiente de passos, ou seja: $\exists t$ tal que $P^t(p, q) \neq 0$. Para provar aperiodicidade, é suficiente mostrar que, para cada elemento p, há uma probabilidade não-nula da cadeia ficar parada em cada passo, ou seja: $P(p, p) \neq 0$. Se uma cadeia é aperiódica mas não irredutível, então pode-se dividir o espaço de estados Ω em diversos sub-espaços, em cada um dos quais a cadeia é irredutível.

2.5.3 Noções de tempo de convergência

Tradicionalmente, o estudo de cadeias de Markov se concentrou em suas propriedades assintóticas, ou seja, no comportamento da cadeia para tempos grandes. Nos últimos

30 anos, porém, desenvolveu-se uma teoria capaz de responder também a questões mais precisas (e mais úteis na prática) sobre o comportamento convergente das cadeias ainda em tempos finitos. Mencionamos a seguir algumas das ferramentas principais usadas nestes dois pontos de vista.

Tempo de mistura

Para estudos de convergência em tempo finito, é útil introduzir a seguinte noção de distância entre distribuições de probabilidade.

Definição 2.5.1 (**Distância de Variação Total**). *Sejam ν, π duas distribuições de probabilidade sobre um espaço finito Ω. Definimos a sua distância de variação total (TV) de duas formas equivalentes:*

$$\|\nu - \pi\|_{TV} \equiv \max_{A \subset \Omega} |\nu(A) - \pi(A)| = \frac{1}{2} \sum_{x \in \Omega} |\nu(x) - \pi(x)| \ . \tag{2.21}$$

Esta distância tem uma interpretação operacional no cenário Bayesiano de *teste de hipóteses*. Suponha que uma fonte (Alice) prepara um elemento $a \in \Omega$ de acordo com uma das duas distribuições (ou hipóteses) ν ou π, com probabilidade *a priori* $1/2$ para cada. Bob precisa distinguir qual das duas distribuições Alice usou. Uma estratégia possível é da forma: para algum subconjunto A, se $a \in A$ escolha ν; se não escolha π. Então, com a melhor escolha de A, a probabilidade de Bob acertar é $(1 + \|\nu - \pi\|_{TV})/2$. Não é difícil mostrar que não há outra estratégia melhor.

Podemos usar esta definição para obter uma noção precisa de convergência de cadeias de Markov em tempo finito. Se a distância TV entre a distribuição gerada após t passos e a distribuição estacionária for muito pequena, então o estado do seu sistema é essencialmente indistinguível do estacionário.

Definição 2.5.2 (**Distância à estacionaridade**). *Seja $\nu_t = \nu_0 P^t$ a distribuição de probabilidade obtida após o sistema evoluir segundo a cadeia P a partir da condição inicial ν_0. Seja π a distribuição estacionária de P. A distância à estacionariedade é*

$$d(t) \equiv \sup_{\nu_0} \|\nu_t - \pi\|_{TV} \ , \tag{2.22}$$

onde o supremo é sobre todas as distribuições iniciais possíveis.

Observe que, com essa definição, garantimos a convergência mesmo no pior caso, e não apenas para 'a maioria' das condições iniciais.

A quantidade chave para entender a convergência de uma cadeia é o tempo necessário para que $d(t)$ fique pequeno. Esse tempo é chamado de tempo de mistura (*mixing time*).

Definição 2.5.3 (**Tempo de Mistura**). *O tempo de mistura da cadeia P é*

$$t_{mixP} \equiv \min\{t : d(t) \leq 1/4\} \tag{2.23}$$

O valor $d = 1/4$ usado nesta definição é arbitrário, na verdade poderia ser qualquer constante $< 1/2$. Isso se justifica porque pode se mostrar que

$$t_{mixP}(\varepsilon) \equiv \min\{t : d(t) \leq \varepsilon\} \leq \lceil \log_2(\varepsilon^{-1}) \rceil \cdot t_{mixP} \ . \tag{2.24}$$

Gap espectral

Outra quantidade frequentemente estudada em cadeias de Markov é:

Definição 2.5.4 (**Gap espectral**). *Seja P uma cadeia de Markov ergódica. Nesse caso, como vimos, só há um autovalor igual a 1, e pode-se mostrar também que todos os demais autovalores de P satisfazem $1 > \lambda > -1$. O* gap *espectral Δ é então a menor distância desses autovalores a 1 (ou -1):*

$$\Delta = \min_{\lambda_i < 1}\{1 - |\lambda_i|\} \ . \tag{2.25}$$

A interpretação do *gap* está ligada ao comportamento assintótico da cadeia. Essencialmente, para tempos suficientemente longos, todos as condições iniciais exceto a ligada ao autovalor que minimiza o *gap* já convergiram. Neste regime a cadeia converge a uma taxa exponencial, ou seja a distância para a distribuição estacionária decresce de forma aproximadamente proporcional a $e^{-t\,\Delta}$. Essa heurística indica que o inverso do gap espectral é um tempo característico da dinâmica das cadeias de Markov. A quantidade $t_{rel} = 1/\Delta$ também é chamada de *tempo de relaxação*. Uma noção precisa que liga a convergência ao gap é dada pela seguinte desigualdade[1] (Capítulo 1 de [41]):

$$t_{2-mix}(\varepsilon) \leq \frac{1}{\Delta} ln\left(\frac{1}{\varepsilon}\right) \ , \tag{2.26}$$

onde $t_{2-mix}(\varepsilon) = \min\{t : \max_\mu \|\mu P^t - \pi\|_2 \leq \varepsilon\}$. Note que distância que aparece aqui não é a TV, mas aquela proveniente da 2-norma:

$$\|\nu - \pi\|_2 \equiv \sum_x (\nu(x) - \pi(x))^2 \ .$$

[1] Válida para cadeias reversíveis.

De modo geral, é mais fácil calcular o *gap* (ou obter cotas) do que o tempo de mistura. Porém, como vimos acima, é este último que tem o sentido operacional que muitas vezes se deseja quando se quer garantir convergência. Pode-se também usar cada uma dessas quantidades para obter cotas para a outra. Por exemplo (Corolário 1.15 de [41]):

$$t_{mix}(\varepsilon) \leq \frac{1}{\Delta} \ln\left(\frac{1}{\varepsilon\ \pi_{min}}\right) \ , \tag{2.27}$$

onde π_{min} é menor valor que a distribuição estacionária assume em todos os elementos do espaço Ω. Ainda (eq. (4.13) de [41]):

$$\frac{\Delta}{1-\Delta} \geq \frac{\log(1/2\varepsilon)}{t_{mix}(\varepsilon)} \quad . \tag{2.28}$$

Ambas essas desigualdades não são úteis para a obtenção de cotas justas para o tempo de mistura pois quando o tamanho do espaço de estados cresce exponencialmente com o tamanho da entrada (n) ($|\Omega| \sim (\text{constante})^n$) , pois nesse caso $\pi_{min} \leq (\text{constante})^{-n}$. Com isso, há um fator n adicional no tempo de mistura.

Capítulo 3

Mapeamento da Dinâmica do CQA em uma cadeia de Markov

Neste capítulo analisamos a evolução gerada, em média, por um circuito quântico aleatório do tipo descrito na Seção 1.2. Um fato crucial, ressaltado pela primeira vez em [18], é que certos aspectos dessa evolução podem ser mapeados em uma *cadeia de Markov* [32]. Como mostramos em detalhes no Teorema 3.3.1 abaixo, para que isso aconteça basta que a distribuição μ da qual as portas de 2 q-bits W_{ij} são sorteadas seja invariante por rotações locais. Nossa discussão segue em boa parte a abordagem da ref. [27], onde estudou-se em detalhe o caso em que $\mu = \mathcal{H}$ é a distribuição uniforme (Haar) sobre o grupo $U(4)$ formado por todas as portas de 2 q-bits (que pode ser substituído por um 2-desenho de 2 q-bits, ver Seção 3.4).

No que se segue, estendemos esses resultados para uma classe mais ampla de distribuições, definidas na próxima Seção, as quais chamamos de distribuições "localmente invariantes".

3.1 Ensembles Localmente Invariantes

Se μ é um ensemble qualquer sobre as portas de 2 q-bits, definimos sua versão *localmente randomizada* μ_L como sendo a aplicação de uma porta extraída de μ precedida e seguida de operações locais aleatórias em cada q-bit. Em outras palavras, cada porta W_{ij} de um CQA baseada no ensemble μ_L tem a forma

$$W_{ij} = (U_i' \otimes V_j')C_{ij}(U_i \otimes V_j) \otimes I_{k\neq i,j}. \tag{3.1}$$

onde C_{ij} é escolhida de acordo com o ensemble μ, respeitando a ordenação dos q-bits, e onde U_i, V_j, U_i' e V_j' são portas aleatórias de 1 q-bit. Aqui "aleatórias" significa portas

sorteadas do ensemble de Haar uniforme de 1 q-bit. Na prática, pode-se substituir esse ensemble por um k-desenho de 1 q-bit (veja Corolário 3.4.1 abaixo para detalhes).

Um conceito a princípio ligeiramente diferente é de um ensemble *localmente invariante*, definido da seguinte forma

Definição 3.1.1. *Um ensemble μ é* localmente invariante *se $(U'_i \otimes V'_j)\mu(U_i \otimes V_j) = \mu$ para quaisquer portas de 1 q-bit U_i, V_j, U'_i e V'_j.*

É evidente que todo ensemble localmente randomizado é também localmente invariante. Também é imediato ver que todo ensemble localmente invariante é a sua própria versão localmente randomizada, de modo que os dois conceitos coincidem.

Um caso particular importante, que chamaremos de um ensemble *porta-fixa*, é quando a porta C é sempre a mesma, ou seja, quando $\mu(C) = \delta(C - C_0)$. Nesse caso escreveremos μ_C no lugar de μ_L. Outro caso particular é o ensemble uniforme $\mu_{\mathcal{H}}$ que, por ser invariante por qualquer rotação, é também localmente invariante.

Pela eq. (2.11), o k-twirl resultante de um ensemble localmente invariante é

$$\hat{G}^{(k)}_{\mu^L_{ij}}(\cdot) = \int W_{ij}^{\otimes k} \cdot W_{ij}^{\dagger \otimes k} dU_i dU'_i dV_j dV'_j d\mu(C) \; . \tag{3.2}$$

Este é um super-operador atuando em $\mathcal{O}_2^{(i_1 j_1)} \otimes \mathcal{O}_2^{(i_2 j_2)} \otimes \ldots \otimes \mathcal{O}_2^{(i_k j_k)}$. Substituindo W_{ij} da eq. (3.1) e reunindo os termos contendo U_i, V_j, U'_i, V'_j, podemos reescrevê-lo como uma composição de três super-operadores

$$\begin{aligned} &\hat{G}^{(k)}_{\mu^L_{ij}} = \\ &= \left(\hat{G}^{(k)}_{\mathcal{H}_{i_1 i_2 \ldots i_k}} \otimes \hat{G}^{(k)}_{\mathcal{H}_{j_1 j_2 \ldots j_k}}\right) \left(\int \hat{C}_{i_1 j_1} \otimes \hat{C}_{i_2 j_2} \otimes \ldots \otimes \hat{C}_{i_k j_k} d\mu(C)\right) \left(\hat{G}^{(k)}_{\mathcal{H}_{i_1 i_2 \ldots i_k}} \otimes \hat{G}^{(k)}_{\mathcal{H}_{j_1 j_2 \ldots j_k}}\right). \end{aligned} \tag{3.3}$$

Aqui, cada $\hat{G}^{(k)}_{\mathcal{H}_{a_1 a_2 \ldots a_k}}$ é um super-operador k-twirl de Haar atuando em (k cópias de) um único q-bit, e $\hat{C}$ é a abreviação para o super-operador $\hat{C}_C$ que realiza a conjugação por C, definido na eq. (2.1). Os índices indicam em quais q-bits cada super-operador atua.

Note que, como os super-operadores uniformes $\hat{G}^{(k)}_{\mathcal{H}}$ são Hermiteanos, e como $\hat{C}^\dagger_C = \hat{C}_{C^\dagger}$, então $\hat{G}^{(k)}_{\mu^L_{ij}}$ é Hermiteano sempre que o ensemble original μ satisfaça $\mu(C) = \mu(C^\dagger)$ (vide Lema 2.3.2). Em particular, para ensembles "porta-fixa", $\hat{G}^{(k)}_{C_{ij}}$ é Hermiteano se C também for. De modo geral, entretanto, $\hat{G}^{(k)}_{\mu^L_{ij}}$ *não* é necessariamente Hermiteano.

A simetria dos ensembles localmente invariantes se manifesta de duas formas complementares. Se $A = A_i \otimes A'_j$ e $B = B_i \otimes B'_j$ são operadores unitários fixos formados por produtos de portas de 1 q-bit então

i. Qualquer ensemble localmente invariante μ_L tem o mesmo valor para portas que são equivalentes a menos de rotações locais:

$$\mu_L(BW_{ij}A) = \mu_L(W_{ij}) \; ; \tag{3.4}$$

ii. De forma semelhante, se dois ensembles μ, ν são iguais a menos de rotações locais ($\mu(C) = B\nu(C)A$), então suas versões localmente invariantes serão idênticas:

$$\mu_L(W_{ij}) = \nu_L(W_{ij}) \; . \tag{3.5}$$

Em particular, duas portas fixas $C, C' = BCA$ que sejam iguais a menos de rotações locais resultam no mesmo ensemble "porta-fixa":

$$\mu_{BCA}(W_{ij}) = \mu_C(W_{ij}) \; . \tag{3.6}$$

Observação: Na definição de W_{ij} dada acima existem duas etapas separadas onde portas randomizantes de 1 q-bit são aplicadas (antes e depois de C_{ij}). Ambas são necessárias para que as simetrias enunciadas acima sejam válidas. Essas por sua vez são necessárias em algumas das provas dadas em seções futuras (veja o Comentário 2 da Seção 3.4 para detalhes).

Entretanto, quando realmente rodamos um CQA, só precisamos na verdade aplicar uma das etapas randomizantes, desde que correções adequadas sejam feitas no início ou no fim do circuito como um todo. Veja fig. 3.2.

Suponha que removamos (por exemplo) as portas U'_i e V'_j da definição de W_{ij}, e que no passo t do circuito aplicamos essa porta aos q-bits i, j. Sempre que, em tempos posteriores (possivelmente diferentes) $t + k$, $t + l$ cada um desses q-bits for novamente selecionado, a primeira coisa que fazemos com eles é aplicar neles uma porta aleatória U_i ou V_j. Como aplicar dois unitários aleatórios independentes em sequência é equivalente a aplicar apenas um, o resultado é o mesmo do que se tivéssemos de fato aplicado U'_i e V'_j ao final do passo t. Há um pequeno detalhe: para gerar exatamente a mesma evolução geral, precisamos aplicar rotações aleatórias independentes em todos os q-bits, seja no início (se removemos as portas U_i, V_j) ou no fim (se removemos as portas U'_i, V'_j) do CQA como um todo. Para certas escolhas de C a necessidade do passo do circuito onde são aplicadas portas de 1 q-bit em todos os sítios é eliminada, ver 3.4.

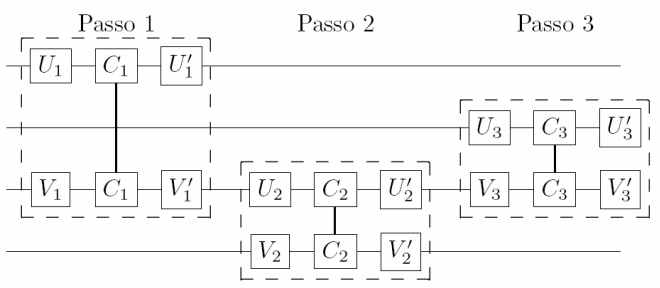


Figura 3.1: Circuito referente a uma possível implementação do CQA, com a definição original, onde as rotações locais são aplicadas antes e depois da porta de 2 q-bits.

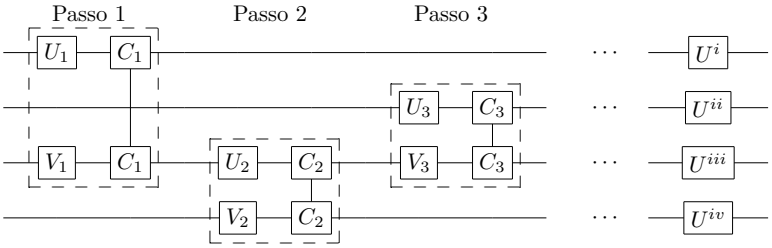


Figura 3.2: Modificação do CQA original que só realiza uma rodada de rotações locais por passo. Note entretanto a necessidade da aplicação de rotações locais em todos os q-bits no fim do circuito.

3.2 Evolução dos momentos da distribuição de Pauli

A evolução de um sistema de n q-bits sobre o qual se realiza o CQA pode ser pensada como um passeio aleatório do espaço de estados. Assuma, por simplicidade, um estado inicial puro $|\psi_0\rangle$. Após cada passo do circuito, podemos considerar que o sistema está num outro estado puro determinado pelo estado anterior e a operação W_{ij} escolhida naquele passo:

$$|\psi_{t+1}\rangle = W_{ij}\,|\psi_t\rangle \ . \tag{3.7}$$

Os coeficientes de Pauli destes estados (mais precisamente, do seu operador densidade $|\psi_t\rangle\rangle$) podem ser vistos como variáveis aleatórias. Usando eqs. (2.1), (2.6), (2.7), sua evolução

pode ser descrita como:

$$\xi_{t+1}(\vec{q}) = 2^{-n} \sum_{\vec{p}} \langle\langle \sigma_{\vec{q}} | \hat{C}_{W_{ij}} | \sigma_{\vec{p}} \rangle\rangle \, \xi_t(\vec{p}) \; . \tag{3.8}$$

Estaremos interessados nas propriedades estatísticas dessas variáveis depois de realizar a média, em cada passo, sobre os possíveis W_{ij} e todos os pares ordenados (i, j). Essas médias, que dependem da escolha da distribuição μ, serão denotadas por $\langle \, \rangle_\mu$.

Por exemplo, o valor médio de $\xi_{t+1}(\vec{q})$ é

$$\begin{aligned} \langle \xi_{t+1}(\vec{q}) \rangle_\mu &\equiv \frac{1}{n(n-1)} \sum_{i \neq j} \int \xi_{t+1}(\vec{q}) d\mu(W_{ij}) \\ &= \frac{2^{-n}}{n(n-1)} \sum_{i \neq j} \sum_{\vec{p}} \langle\langle \sigma_{\vec{q}} | \hat{G}^{(1)}_{\mu_{ij}} | \sigma_{\vec{p}} \rangle\rangle \, \langle \xi_t(\vec{p}) \rangle_\mu \, , \end{aligned} \tag{3.9}$$

onde $\hat{G}^{(1)}_{\mu_{ij}}$ é um super-operador de 1-twirl nos q-bits i, j (ver eq. (2.10)). Note que para cada t, $\langle \xi_t(\vec{q}) \rangle_\mu$ pode ser vista como os coeficientes de Pauli do operador densidade ρ_t, obtido pela média entre todos os caminhos possíveis do passeio aleatório até o tempo t.

Já que $\hat{G}^{(1)}_{\mu_{ij}}$ não afeta outros q-bits que não o par i, j, podemos escrever a equação acima como

$$\langle \xi_{t+1}(\vec{q}) \rangle_\mu = \frac{2^{-2}}{n(n-1)} \sum_{i \neq j} \sum_{\vec{p}} \delta_{\vec{p}_{/ij}, \vec{q}_{/ij}} \left\langle\left\langle \sigma_{q_i q_j} \right| \hat{G}^{(1)}_{\mu_{ij}} \left| \sigma_{p_i p_j} \right\rangle\right\rangle \langle \xi_t(\vec{p}) \rangle_\mu \, . \tag{3.10}$$

onde $\vec{p}_{/ij}$, $\vec{q}_{/ij}$ denotam os vetores obtidos pela remoção das entradas i, j de $\vec{p}$ e $\vec{q}$. Vê-se então que, independentemente do número de q-bits no circuito, a evolução da média dos coeficientes em cada passo do CQA depende apenas no elementos de matriz do super-operator de 1-twirl em 2 q-bits ($\hat{G}^{(1)}_{\mu}$) na base de Pauli.

Analogamente, os momentos de segunda ordem da distribuição dos coeficiente de Pauli depois de $t+1$ passos do CQA são

$$\begin{aligned} \langle \xi_{t+1}(\vec{q}) \xi_{t+1}(\vec{q}\,') \rangle_\mu &= \frac{2^{-2n}}{n(n-1)} \sum_{i \neq j} \sum_{\vec{p}, \vec{p}\,'} \langle\langle \sigma_{\vec{q}, \vec{q}\,'} | \hat{G}^{(2)}_{\mu_{ij}} | \sigma_{\vec{p}, \vec{p}\,'} \rangle\rangle \, \langle \xi_t(\vec{p}) \xi_t(\vec{p}\,') \rangle_\mu \\ &= \frac{2^{-4}}{n(n-1)} \sum_{i \neq j} \sum_{\vec{p}, \vec{p}\,'} \delta_{\vec{p}_{/ij} \vec{p}\,'_{/ij}, \vec{q}_{/ij}, \vec{q}\,'_{/ij}} \left\langle\left\langle \sigma_{q_i q_j q'_i q'_j} \right| \hat{G}^{(2)}_{\mu_{ij}} \left| \sigma_{p_i p_j p'_i p'_j} \right\rangle\right\rangle \langle \xi_t(\vec{p}) \xi_t(\vec{p}\,') \rangle_\mu \end{aligned} \tag{3.11}$$

Aqui usamos abreviações do tipo $|\sigma_{\vec{p},\vec{q}}\rangle\!\rangle \equiv |\sigma_{\vec{p}}\rangle\!\rangle \otimes |\sigma_{\vec{q}}\rangle\!\rangle$, e

$$\hat{G}^{(2)}_{\mu_{ij}}(\cdot) \equiv \int W_{ij}^{\otimes 2} \cdot W_{ij}^{\dagger \otimes 2} d\mu(W_{ij}) \tag{3.12}$$

é um super-operador de 2-twirl operando em (duas cópias de) um par de q-bits i, j. Assim, cada passo da evolução depende dos $4^4 \times 4^4$ elementos de matriz de $\hat{G}^{(2)}_{\mu}$. Novamente, entretanto, no caso de ensembles localmente invariantes, quase todos esses elementos são nulos. De fato, mostraremos no Teorema 3.3.1 abaixo que mesmo os elementos não-nulos restantes dependem de apenas dois parâmetros independentes.

Essas relações podem ser generalizadas para valores arbitrários de k, embora seja cada vez mais inconveniente escreve-las explicitamente. Colocando em palavras, o fato essencial é que a evolução dos momentos de ordem k dos coeficientes de Pauli são funções lineares dos elementos de matriz do super-operador de k-twirl $\hat{G}^{(k)}_{\mu}$ na base de Pauli.

3.3 Evolução Markoviana dos momentos de segunda ordem

Mostraremos agora como a evolução média gerada por um circuito quântico aleatório fica muito simplificada quando μ é uma distribuição localmente invariante (de agora em diante, deixamos de usar o índice L. Todos os ensembles μ que encontraremos serão localmente invariantes exceto quando explicitamente assinalado).

Em particular, nos concentraremos na evolução dos segundos momentos definida na eq. (3.11). Como já foi dito, em muitos casos essa evolução pode ser mapeada em uma *cadeia de Markov* clássica. Nossa prova generaliza a que foi dada em [27], onde mostrou-se que esse resultado vale para o caso em que μ é a distribuição de Haar $\mathcal{H}$ sobre $U(4)$. Aqui estendemos esse resultado para todas as distribuições localmente invariantes (isso também generaliza o caso especial, $C = CNOT$, estudado em [18]). Para melhor ressaltar como a nossa demonstração generaliza a de [27], seguimos em paralelo tanto o caso geral quanto o caso particular da distribuição uniforme, o qual é mais simples.

Teorema 3.3.1. *Se um CQA usa portas sorteadas de um ensemble localmente invariante μ, então os momentos de segunda ordem dos coeficientes de Pauli do sistema evoluem segundo*

$$\langle \xi_{t+1}(\vec{q}) \xi_{t+1}(\vec{q}\,') \rangle_{\mu} = \delta_{\vec{q},\vec{q}\,'} \sum_{\vec{p}} \langle \xi_t^2(\vec{p}) \rangle_{\mu} P_{\mu}(\vec{p},\vec{q}) \ , \tag{3.13}$$

onde

$$P_\mu(\vec{p},\vec{q}) \equiv \frac{2^{-2n}}{n(n-1)} \sum_{i\neq j} \langle\langle \sigma_{\vec{q},\vec{q}} | \hat{G}^{(2)}_{\mu_{ij}} | \sigma_{\vec{p},\vec{p}} \rangle\rangle \tag{3.14}$$

é uma ***matriz de Markov bi-estocástica****, i.e. uma matriz de números reais e não negativos, que satisfazem*

$$\sum_{\vec{q}} P_\mu(\vec{p},\vec{q}) = \sum_{\vec{p}} P_\mu(\vec{p},\vec{q}) = 1 \; . \tag{3.15}$$

Em particular, observando que momentos com $\vec{q} \neq \vec{q}\,'$ se cancelam:

> *Os vetores de probabilidade médios $\langle \xi^2(\vec{p}) \rangle$ evoluem segundo uma cadeia de Markov com matriz P_μ.*

Observação sobre convenções: a eq. (3.13) está escrita de acordo com as convenções comumente usadas na literatura de cadeias de Markov (veja por exemplo [32]), onde vetores de probabilidade são considerados vetores linha, que multiplicam a matriz de Markov pela esquerda. Portanto, a definição da matriz P_μ em eq. (3.14) é a transposta daquela que um físico usaria normalmente.

Prova:

A prova está organizada em duas partes: inicialmente mostramos que, para **qualquer** ensemble μ (mesmo que não seja localmente invariante), os coeficientes $P_\mu(\vec{p},\vec{q})$ definidos na eq. (3.14) formam uma matriz de Markov bi-estocástica. Demonstramos em seguida que, nos casos específicos dos ensembles localmente invariantes, os momentos com $\vec{p} \neq \vec{p}\,'$ na eq. (3.13) se anulam quando tomamos a média.

Para mostrar que $P_\mu(\vec{p},\vec{q})$ é bi-estocástica, é suficiente verificar que isto é verdade para cada uma das $n(n-1)$ matrizes

$$P_{\mu_{ij}}(\vec{p},\vec{q}) \equiv 2^{-2n} \langle\langle \sigma_{\vec{q},\vec{q}} | \hat{G}^{(2)}_{\mu_{ij}} | \sigma_{\vec{p},\vec{p}} \rangle\rangle = 2^{-4} \delta_{\vec{p}_{/ij}\vec{q}_{/ij}} \left\langle\left\langle \sigma_{q_iq_j,q_iq_j} \right| \hat{G}^{(2)}_{\mu_{ij}} \left| \sigma_{p_ip_j,p_ip_j} \right\rangle\right\rangle \tag{3.16}$$

cuja média forma $P_\mu(\vec{p},\vec{q})$ (vide eq. (3.14)).

Primeiro somamos os elementos em cada linha:

$$\sum_{\vec{q}} P_{\mu_{ij}}(\vec{p},\vec{q}) = 2^{-2n} \sum_{\vec{q}} \langle\langle \sigma_{\vec{q},\vec{q}} | \hat{G}^{(2)}_{\mu_{ij}} | \sigma_{\vec{p},\vec{p}} \rangle\rangle = 2^{-2n} \langle\langle S | \hat{G}^{(2)}_{\mu_{ij}} | \sigma_{\vec{p},\vec{p}} \rangle\rangle = 2^{-2n} \langle\langle S | \sigma_{\vec{p},\vec{p}} \rangle\rangle = 1. \tag{3.17}$$

onde $|S\rangle\!\rangle = \sum_{\vec{q}} |\sigma_{\vec{q},\vec{q}}\rangle\!\rangle$ é o operador de troca (SWAP) que permuta as duas cópias do sistema de n q-bits, e onde usamos os Lemas 2.3.4 e 2.3.3. As somas em cada coluna seguem de forma semelhante:

$$\sum_{\vec{p}} P_{\mu_{ij}}(\vec{p},\vec{q}) = 2^{-2n} \langle\!\langle \sigma_{\vec{q},\vec{q}}| \hat{G}^{(2)}_{\mu_{ij}} |S\rangle\!\rangle = 2^{-2n} \langle\!\langle \sigma_{\vec{q},\vec{q}}|S\rangle\!\rangle = 1 \tag{3.18}$$

Agora checamos que os elementos de matriz $P_{\mu_{ij}}(\vec{p},\vec{q})$ precisam ser não-negativos. Usando a eq. (3.12):

$$\langle\!\langle \sigma_{\vec{q},\vec{q}}| \hat{G}^{(2)}_{\mu_{ij}} |\sigma_{\vec{p},\vec{p}}\rangle\!\rangle = \int \mathrm{Tr}^2 \left(\sigma_{\vec{q}} W_{ij} \sigma_{\vec{p}} W_{ij}^\dagger\right) d\mu(W_{ij});. \tag{3.19}$$

Precisamos mostrar que o traço do operador entre parênteses é um número real. De fato, usando a identidade $\mathrm{Tr}^* A = \mathrm{Tr} A^\dagger$ e também a propriedade cíclica do traço:

$$\mathrm{Tr}^* \left[\sigma_{\vec{q}} W_{ij} \sigma_{\vec{p}} W_{ij}^\dagger\right] = \mathrm{Tr}\left[\sigma_{\vec{q}} W_{ij} \sigma_{\vec{p}} W_{ij}^\dagger\right]^\dagger = \mathrm{Tr}\left[W_{ij} \sigma_{\vec{p}} W_{ij}^\dagger \sigma_{\vec{q}}\right] = \mathrm{Tr}\left[\sigma_{\vec{q}} W_{ij} \sigma_{\vec{p}} W_{ij}^\dagger\right] . \tag{3.20}$$

Concluimos assim que $P_{\mu_{ij}}$ (e P_μ) são de fato bi-estocásticas para qualquer ensemble μ.

Retornemos agora à demonstração da eq. (3.13). Comparando esta equação com a eq. (3.11), fica claro que precisamos apenas mostrar que os momentos com $\vec{q} \neq \vec{q}\,'$ se cancelam nos casos em que μ é um ensemble localmente invariante. No que se segue, denotaremos o super-operador de 2-twirl $\hat{G}^{(2)}_{\mu_{ij}}$ e a matriz correspondente P_μ alternativamente como $\hat{G}^{(2)}_{\mathcal{H}_{ij}}$ ($P_\mathcal{H}$) ou $\hat{G}^{(2)}_{C_{ij}}$ (P_C) nos casos particulares do ensemble uniforme e de ensembles "porta-fixa".

Começamos escrevendo os super-operadores de 2-twirl. No caso $\mu = \mathcal{H}$, usando o Teorema 2.3.1 podemos escrevê-lo explicitamente como o projetor

$$\hat{G}^{(2)}_{\mathcal{H}_{ij}} = \left(\frac{1}{16}|\sigma_{00,00}\rangle\!\rangle\langle\!\langle\sigma_{00,00}| + \frac{1}{240}\sum_{(kl),(k'l')\neq 00} |\sigma_{kl,kl}\rangle\!\rangle\langle\!\langle\sigma_{k'l',k'l'}|\right) \otimes \hat{I}_{\text{qubits} \neq i,j} \tag{3.21}$$

onde os operadores de Pauli atuam nos q-bits $(i_1 j_1, i_2 j_2)$.

De modo geral, para um ensemble localmente invariante μ a eq. (3.3) fornece, no caso $k = 2$

$$\hat{G}^{(2)}_{\mu_{ij}} = \left(\hat{G}^{(2)}_{\mathcal{H}_{i_1 i_2}} \otimes \hat{G}^{(2)}_{\mathcal{H}_{j_1 j_2}}\right) \left(\int \hat{C}_{i_1 j_1} \otimes \hat{C}_{i_2 j_2} d\mu(C)\right) \left(\hat{G}^{(2)}_{\mathcal{H}_{i_1 i_2}} \otimes \hat{G}^{(2)}_{\mathcal{H}_{j_1 j_2}}\right). \tag{3.22}$$

O significado dessa expressão pode ser entendido usando mais uma vez o Teorema 2.3.1: nesse caso, ele implica que o produto $\hat{G}^{(2)}_{\mathcal{H}_{i_1 i_2}} \otimes \hat{G}^{(2)}_{\mathcal{H}_{j_1 j_2}}$ é o projetor no subespaço

4-dimensional de $\mathcal{O}_2^{(i_1j_1)} \otimes \mathcal{O}_2^{(i_2j_2)}$ gerado pelo conjunto de operadores ortogonais:

$$\left\{ |I_{i_1i_2}\rangle\rangle\, |I_{j_1j_2}\rangle\rangle \,;\; |I_{i_1i_2}\rangle\rangle\, |\bar{S}_{j_1j_2}\rangle\rangle \,;\; |\bar{S}_{i_1i_2}\rangle\rangle\, |I_{j_1j_2}\rangle\rangle \,;\; |\bar{S}_{i_1i_2}\rangle\rangle\, |\bar{S}_{j_1j_2}\rangle\rangle \right\}. \tag{3.23}$$

Assim, $\hat{G}^{(2)}_{\mu_{ij}}$ representa a restrição de $\int \hat{C}_{i_1j_1} \otimes \hat{C}_{i_2j_2} d\mu(C)$ a esse subespaço 4-dimensional.

Mostramos agora que tomar a média sobre a distribuição μ anula todos elementos da matriz $\hat{G}^{(2)}_{\mu_{ij}}$ exceto quando $\vec{q} = \vec{q}\,'$ e também $\vec{p} = \vec{p}\,'$, ou seja:

$$\langle\langle \sigma_{\vec{q},\vec{q}\,'} |\, \hat{G}^{(2)}_{\mu_{ij}}\, |\sigma_{\vec{p},\vec{p}\,'}\rangle\rangle = \delta_{\vec{q},\vec{q}\,'}\delta_{\vec{p},\vec{p}\,'}\, \langle\langle \sigma_{\vec{q},\vec{q}} |\, \hat{G}^{(2)}_{\mu_{ij}}\, |\sigma_{\vec{p},\vec{p}}\rangle\rangle \ . \tag{3.24}$$

Para $\mu = \mathcal{H}$ o resultado é imediato: as eqs. (2.4) e (3.21) implicam que $\hat{G}^{(2)}_{\mathcal{H}_{ij}} |\sigma_{\vec{p},\vec{p}\,'}\rangle\rangle = 0$ se $\vec{p} \neq \vec{p}\,'$ e analogamente $\langle\langle \sigma_{\vec{q},\vec{q}\,'} | \hat{G}^{(2)}_{\mathcal{H}_{ij}} = 0$ se $\vec{q} \neq \vec{q}\,'$. O caso geral segue de forma semelhante: substituindo a eq. (2.16) na eq. (3.23) e rearranjando a ordem dos q-bits, podemos reescrever esta base não-normalizada na forma

$$\left\{ \left|\sigma_{00}^{i_1j_1}\right\rangle\right\rangle \left|\sigma_{00}^{i_2j_2}\right\rangle\right\rangle \,;\; \sum_{k\neq 0} \left|\sigma_{0k}^{i_1j_1}\right\rangle\right\rangle \left|\sigma_{0k}^{i_2j_2}\right\rangle\right\rangle \,;\; \sum_{k\neq 0} \left|\sigma_{k0}^{i_1j_1}\right\rangle\right\rangle \left|\sigma_{k0}^{i_2j_2}\right\rangle\right\rangle \,;\; \sum_{k,l\neq 0} \left|\sigma_{kl}^{i_1j_1}\right\rangle\right\rangle \left|\sigma_{kl}^{i_2j_2}\right\rangle\right\rangle \right\}. \tag{3.25}$$

Claramente, em cada um desses quatro elementos de base, um dado operador de Pauli σ_{ab} que atua no par de q-bits i_1j_1 está sempre emparelhado com o mesmo operador de Pauli atuando nos q-bits i_2j_2. Segue-se então como acima que $\hat{G}^{(2)}_{\mathcal{H}_{i_1i_2}} \otimes \hat{G}^{(2)}_{\mathcal{H}_{j_1j_2}} |\sigma_{\vec{p},\vec{p}\,'}\rangle\rangle = 0$ se $\vec{p} \neq \vec{p}\,'$, e portanto $\hat{G}^{(2)}_{\mu_{ij}} |\sigma_{\vec{p},\vec{p}\,'}\rangle\rangle = 0$ se $\vec{p} \neq \vec{p}\,'$. Novamente, por um argumento análogo, $\langle\langle \sigma_{\vec{q},\vec{q}\,'} | \hat{G}^{(2)}_{\mu_{ij}} = 0$ se $\vec{q} \neq \vec{q}\,'$. □

3.4 Comentários sobre o Teorema 3.3.1

Nesta Seção apresentamos diversos corolários e extensões simples do Teorema 3.3.1. Ela não é essencial para o restante da dissertação, e pode ser deixada de lado em uma primeira leitura.

1. **Usando 2-desenhos de 1 e 2 q-bits ao invés de unitários distribuídos segundo a medida de Haar**.

   Gerar unitários distribuídos de acordo com os ensembles $\mu = \mathcal{H}$ ou qualquer outro ensemble localmente invariante μ_L pode ser inconveniente na prática. Como ambos estes ensembles supõem unitários uniformemente distribuídos (de 1 ou 2 q-bits, conforme o caso), fazer isto requer um aparato experimental com parâmetros continuamente ajustáveis. Felizmente, esta dificuldade pode ser eliminada [18, 27]:

Corolário 3.4.1. *O Teorema 3.3.1 permanece válido se substituirmos os unitários distribuídos segundo a medida de Haar, necessários nas definições de $\mathcal{H}$ ou μ_L, por 2-desenhos de $n = 2$ ou $n = 1$ q-bits, respectivamente.*

Em particular, podemos usar 2-desenhos discretos, como os descritos em [42].

Prova: No caso em que $\mu = \mathcal{H}$, não há quase nada a provar: o resultado segue de forma direta da eq. (2.3.1), já que os elementos de matriz do 2-twirl $\left\langle\left\langle \sigma_{\vec{q}_{ij},\vec{q}\,'_{ij}} \middle| \hat{G}^{(2)}_{\mathcal{H}_{ij}} \middle| \sigma_{\vec{p}_{ij},\vec{p}\,'_{ij},} \right\rangle\right\rangle$ permanecem iguais quando trocamos $\mathcal{H}$ por qualquer 2-desenho de 2 q-bits. Para ensembles localmente invariantes μ, a eq. (3.22) implica em

$$\begin{aligned}
&\left\langle\left\langle \sigma_{\vec{q}_{ij},\vec{q}\,'_{ij}} \middle| \hat{G}^{(2)}_{\mu_{ij}} \middle| \sigma_{\vec{p}_{ij},\vec{p}\,'_{ij}} \right\rangle\right\rangle = \\
&= \sum_{\substack{\vec{r}_{ij},\vec{r}\,'_{ij} \\ \vec{s}_{ij},\vec{s}\,'_{ij}}} \alpha_{\vec{q}_{ij},\vec{q}\,'_{ij},\vec{s}_{ij}\vec{s}\,'_{ij}} \left\langle\left\langle \sigma_{\vec{s}_{ij},\vec{s}\,'_{ij}} \middle| \int \hat{C}_{i_1 j_1} \otimes \hat{C}_{i_2 j_2} d\mu(C) \middle| \sigma_{\vec{r}_{ij},\vec{r}\,'_{ij}} \right\rangle\right\rangle \alpha_{\vec{r}_{ij}\vec{r}\,'_{ij},\vec{p}_{ij},\vec{p}\,'_{ij}} \ , \qquad (3.26)
\end{aligned}$$

onde inserimos super-operadores identidade $\hat{I} = N \sum_{\vec{r}_{ij}\vec{r}\,'_{ij}} \left| \sigma_{\vec{r}_{ij},\vec{r}\,'_{ij}} \right\rangle\!\!\right\rangle \left\langle\!\!\left\langle \sigma_{\vec{r}_{ij},\vec{r}\,'_{ij}} \right|$ e definimos

$$\begin{aligned}
\alpha_{\vec{r}_{ij}\vec{r}\,'_{ij},\vec{p}_{ij},\vec{p}\,'_{ij}} &\equiv N \left\langle\left\langle \sigma_{\vec{r}_{ij},\vec{r}\,'_{ij}} \middle| \hat{G}^{(2)}_{\mathcal{H}_{i_1 i_2}} \otimes \hat{G}^{(2)}_{\mathcal{H}_{j_1 j_2}} \middle| \sigma_{\vec{p}_{ij},\vec{p}\,'_{ij}} \right\rangle\right\rangle \\
&= N \left\langle\left\langle \sigma_{r_i,r'_i} \middle| \hat{G}^{(2)}_{\mathcal{H}_{i_1 i_2}} \middle| \sigma_{p_i,p'_i} \right\rangle\right\rangle \left\langle\left\langle \sigma_{r_j,r'_j} \middle| \hat{G}^{(2)}_{\mathcal{H}_{j_1 j_2}} \middle| \sigma_{p_j,p'_j} \right\rangle\right\rangle \qquad (3.27)
\end{aligned}$$

Novamente usando a eq. (2.3.1), esses elementos da matriz de 2-twirl permanecem inalterados se cada média local de Haar em μ for substituída por um 2-desenho de 1 q-bit.

2. **Abandonando um dos passos randomizantes em μ_L**

Como foi dito no fim da Seção 3.1, apesar de a definição de μ_L envolver dois passos separados onde portas de 1 q-bit randomizantes são aplicadas, um deles pode ser abandonado desde que seja aplicada uma correção apropriada. Em outras palavras:

Corolário 3.4.2. *O Teorema 3.3.1 continua válido para ensembles localmente invariantes μ_L mesmo se removemos um dos pares de portas (U_i, V_j) ou (U'_i, V'_j) da definição de W_{ij} na eq. (3.1). Isto pode ser feito desde que rotações aleatórias independentes sejam aplicadas a cada q-bit no início (ou no fim, respectivamente) do CQA inteiro.*

É interessante notar que ambos os passos randomizantes são necessários para a prova do Teorema 3.3.1 funcionar: cada um leva a um dos super-operadores $\hat{G}^{(2)}_{\mathcal{H}_{i_1 i_2}} \otimes \hat{G}^{(2)}_{\mathcal{H}_{j_1 j_2}}$ na eq. (3.22). Seguindo o argumento após eq. (3.23) fica claro que, em geral, ambos os passos precisam existir para que os cancelamentos entre valores diferentes de $\vec{p}, \vec{q}$ e $\vec{p}\,', \vec{q}\,'$ ocorram na eq. (3.24). Por sua vez, isso é necessário para a evolução dos segundos momentos ser descritível como uma cadeia de Markov.

Uma exceção ocorre para escolhas específicas de C, pertencentes ao chamado *grupo de Clifford*, isto é, o conjunto de portas para as quais o super-operador de conjugação correspondente $\hat{C}$ mapeia o conjunto de operadores de Pauli nele mesmo, a menos de um sinal:

$$\hat{C}_{Clifford}|\sigma_{ab}\rangle\rangle = \pm|\sigma_{cd}\rangle\rangle \ , \ \forall \sigma_{ab}. \tag{3.28}$$

Exemplos bem conhecidos de portas desse tipo incluem CNOT, CZ e XY. Discutiremos este tipo de porta em mais detalhe nos exemplos da Seção 4.2.

Nesse caso, podemos remover um dos passos randomizantes do ensemble porta-fixa μ_C sem a necessidade de correções. Para ver isso note que, para tais portas de Clifford, aplicar $\hat{C}_{i_1 j_1} \otimes \hat{C}_{i_2 j_2}$ a cada um dos elementos da base na eq. (3.25) produz novos operadores na forma $\sum |\sigma_{ab}^{i_1 j_1}\rangle\rangle |\sigma_{ab}^{i_2 j_2}\rangle\rangle$. Esses operadores guardam a propriedade de que um dado operador de Pauli σ_{ab}, que age sobre os q-bits $i_1 i_2$, sempre acompanha um mesmo operador que age nos q-bits $j_1 j_2$. Consequentemente, se por exemplo removemos as rotações (U'_i, V'_j) da definição de W_{ij} na eq. (3.1), então eq. (3.24) continua se cancelando para $\vec{p}\,' \neq \vec{q}\,'$, mesmo que $\langle\langle \sigma_{\vec{p}\,',\vec{q}\,'}|\hat{G}^{(2)}_C \neq 0$. Podemos concluir então que

Corolário 3.4.3. *Para ensembles porta-fixa μ_C com portas de Clifford C, o Teorema 3.3.1 continua válido, sem correções, se removemos um dos pares de portas (U_i, V_j) ou (U'_i, V'_j) da definição de W_{ij}.*

Vale notar que os primeiros autores a reduzirem a evolução do segundo momento de um CQA para uma cadeia de Markov [18] tomaram uma distribuição μ_C precisamente desse tipo, com $C = CNOT$ e sem as portas (U'_i, V'_j).

3. **Robustez do resultado e implicações práticas**

Para aplicar na prática um CQA baseado em um dado ensemble μ, é necessário poder realizar experimentalmente todas as portas no suporte de μ. A dificuldade ou não dessa realização em geral dependerá da escolha de μ. Por exemplo, para realizar o ensemble uniforme $\mu_{\mathcal{H}}$ é necessário ter a capacidade *experimental* de realizar qualquer porta de 2 q-bits, o que é algo muito difícil. Mesmo um 2-desenho de 2 q-bits,

que como vimos pode ser usado no lugar do $\mu_{\mathcal{H}}$, não é tão simples, pois ainda requer a capacidade de realizar diversas portas emaranhantes bastante distintas uma da outra. Na maioria dos casos, implementações experimentais de processadores quânticos procuram ser capazes de implementar diretamente um número muito reduzido de portas de 2-qbits específicas (por exemplo, apenas $C = \text{CNOT}$ ou Control-Z). Para estes casos, ensembles tipo 'porta-fixa' seriam mais apropriados.

Nestas situações, é interessante o fato de que o Teorema 3.3.1 continua válido mesmo se, ao invés de uma porta C perfeita, gera-se um ensemble ν com um pico ao redor de C. Isto é justamente o que pode esperar se levarmos em conta erros experimentais como por exemplo pulsos de radiação cuja duração ou intensidade podem variar com alguma probabilidade. Em outras palavras, a obtenção de uma cadeia de Markov (e as consequências que derivaremos disso nos próximos capítulos) será robusta a pelo menos alguns tipos de erros experimentais.

Capítulo 4

Análise da cadeia de Markov P_μ

Neste capítulo analisamos em detalhe as propriedades da matriz de Markov P_μ obtida no Teorema 3.3.1. Começamos, na Seção 4.1, apontando várias simetrias e características gerais da matriz. Na Seção 4.2, utilizamos essas simetrias para obter uma descrição simples do passeio aleatório correspondente a P_μ. Mostramos que este passeio tem sempre a mesma forma geral, qualquer que seja o ensemble localmente invariante utilizado para gerar as portas do CQA, havendo apenas dois parâmetros independentes que variam de ensemble para ensemble. Na Seção 4.5 simplificamos ('reduzimos') a cadeia, e obtemos uma decomposição da cadeia reduzida que mostra que ela é, essencialmente, equivalente a um modelo tipo 'campo médio' combinado com transposições aleatórias. Esta decomposição será a chave usada, no próximo capítulo, para obter o tempo de convergência da cadeia.

4.1 Propriedades gerais da matriz de Markov

A eq. (3.14), juntamente com as eqs. (3.12) e (3.1), fornece expressões explícitas, mas aparentemente bastante complicadas, para os elementos da matriz de Markov $P_\mu(\vec{p},\vec{p}\,')$. Na verdade, entretanto, grande parte da estrutura dessa matriz decorre de simetrias do CQA e/ou dos ensembles porta-fixa e uniforme. Por exemplo, as cinco propriedades simples de $P_\mu(\vec{p},\vec{p}\,')$ que se seguem são consequências diretas da estrutura do circuito quântico aleatório, e são independentes da escolha de μ:

1. **Matriz esparsa:** como cada porta W_{ij} só conecta 2 q-bits, $P_\mu(\vec{p},\vec{p}\,') = 0$ para os vetores $\vec{p},\vec{p}\,'$ que diferem em mais de duas coordenadas.

2. **Invariância da identidade:** O operador densidade completamente misto do sistema de n q-bits , $I = 2^{-n}\sigma_{\vec{0}}$, é invariante sob qualquer rotação unitária, incluindo, é

claro aquelas geradas pelo CQA. No nível de segundos momentos, isso implica que o vetor formado pelos coeficientes quadrados $\xi^2(\vec{p}) = 2^{-2n}\delta_{\vec{p}\vec{0}}$ também é invariante sob cada passo do circuito. É portanto um autovetor (à esquerda) da matriz de Markov $P_\mu(\vec{p}, \vec{p}\,')$ com autovalor 1. Normalisando esse vetor (ele não é normalizado, já que I é um estado misto), concluímos que o vetor de probabilidade $\xi^2(\vec{p}) = \delta_{\vec{p}\vec{0}}$ é sempre uma distribuição estacionária de P_μ.

3. **Bi-estocasticidade**: A matriz P_μ faz parte de uma categoria particular de matrizes de Markov, a saber aquelas que são bi-estocásticas (cujas colunas somam 1). Uma consequência imediata disto é que o vetor uniforme $\pi = (1 \dots 1)$ satisfaz $\pi P = \pi$, ou seja é também uma distribuição estacionária. Veremos no próximo item que este e $(\delta_{\vec{p}\vec{0}})$ são os únicos estados estacionários da cadeia.

   Pode-se dizer bem mais ainda: por exemplo, o chamado Teorema de Birkhoff [43] afirma que qualquer matriz bi-estocática pode ser escrita como uma combinação convexa de permutações. Heuristicamente, isto quer dizer que, em cada passo da cadeia P_μ, qualquer vetor de probabilidade $\langle \xi_t^2(\vec{p}) \rangle$ é mapeado em um novo vetor $\left\langle \xi_{t+1}^2(\vec{p}) \right\rangle = \langle \xi_t^2(\vec{p}) \rangle P_\mu$ que é 'mais misturado' que $\langle \xi_t^2(\vec{p}) \rangle$. Esta idéia é capturada de forma precisa pela relação de majoração $\left\langle \xi_{t+1}^2(\vec{p}) \right\rangle \prec \langle \xi_t^2(\vec{p}) \rangle$ [43], a qual implica entre outras coisas que a entropia de Shannon $S = -\sum_i x_i \ln x_i$ é não-decrescente em cada passo da cadeia: $S(\langle \xi_{t+1}^2(\vec{p}) \rangle) \geq S(\langle \xi_t^2(\vec{p}) \rangle)$. Pode-se esperar então que a tendência da cadeia é de aumentar a entropia até seu valor máximo possível, o qual é atingido para a distribuição uniforme de probabilidade. Isto é de fato o que ocorre, excetuando-se a condição inicial $\delta_{\vec{p}\vec{0}}$.

4. **Invariância sob permutações de q-bits:** Pela definição do CQA dada na Seção 1.2, cada par de q-bits (i, j) que corresponde a uma aresta do grafo Γ tem igual probabilidade de ser escolhido, em qualquer ordem, como aquele em que a porta W_{ij} atua. Desta forma, as probabilidades $P(\vec{p}, \vec{q})$ devem ser invariantes sob qualquer permutação Λ dos índices de $\vec{p}, \vec{q}$ que corresponda uma simetria do grafo:

$$P(q_1, \dots, q_n; p_1, \dots, p_n) = P(q_{\Lambda(1)}, \dots, q_{\Lambda(n)}; p_{\Lambda(1)}, \dots, p_{\Lambda(n)})$$

   Por exemplo, no caso em que qualquer par pode ser escolhido, correspondendo ao grafo cheio, então Λ pode ser qualquer permutação de $(1 \dots n)$.

Nos casos específicos dos ensembles localmente invariantes, existe ainda uma simetria adicional que simplifica consideravelmente a análise:

5. **Invariância sob permutações de eixo de cada q-bit:** Ambas as distribuições de probabilidade $\mu = \mathcal{H}$ e $\mu = \mu_C$ são invariantes sob rotações locais independentes

de qualquer um dos q-bits (veja eq. (3.4), que se aplica aos dois ensembles). Em particular, elas são invariantes sob permutações independentes dos eixos x, y, z de cada q-bit. Formalmente, isso pode ser visto escolhendo A_i, A'_j, B_i, B'_j na eq. (3.4) como sendo operadores que mapeiam o conjunto de operadores de Pauli X, Y, Z nele mesmo sob conjugação.

Conclui-se que cada $P_{\mu_{ij}}(\vec{q}, \vec{p})$, e portanto também $P_\mu(\vec{q}, \vec{p})$, devam permanecer invariantes se qualquer um dos elementos p_i, q_j que sejam iguais a X, Y ou Z for trocado por outro desses valores. Em outras palavras, os elementos de matriz $P_\mu(\vec{q}, \vec{p})$ só são sensíveis ao fato de p_i, q_j serem iguais a 0 ou não.

4.2 Passeio aleatório

Podemos entender melhor a cadeia de Markov P_μ mudando nosso ponto de vista para o passeio aleatório correspondente. Nesse caso, ao invés de seguirmos a evolução de uma distribuição de probabilidade sobre $\{0, X, Y, Z\}^n$, fazemos em cada passo de tempo uma transição aleatória de um ponto a outro deste espaço, sendo $P_\mu(\vec{p}, \vec{q})$ a probabilidade de transição do ponto $\vec{p}$ para o $\vec{q}$.

Para ilustrar a idéia, examinemos primeiro o caso do ensemble uniforme $\mu = \mathcal{H}$, e por simplicidade assumindo ainda que todos os pares de q-bits podem ser escolhidos em cada passo do CQA. Pela eq. (3.21), a matriz de $\hat{G}^{(2)}_{\mathcal{H}_{ij}}$ na base de Pauli tem uma forma simples bloco-diagonal, sendo um bloco 1×1 (correspondendo ao elemento $00, 00$) e o outro 15×15. Ambos os blocos podem ser escritos na forma $\frac{1}{m} F_m$ onde F_m é a matriz $m \times m$ contendo todas as entradas iguais a 1. Podemos interpretar então a matriz P_μ como representando o passeio aleatório com as seguintes regras [27]: dada a posição $\vec{p}$,

- escolha um par de coordenadas p_i, p_j de $\vec{p}$ de forma aleatória e uniforme
- se $p_i = p_j = 0$, não faça nada
- se $(p_i, p_j) \neq (0, 0)$, substitua o valor desta dupla por qualquer elemento de $\{0, X, Y, Z\}^2/(0, 0)$, com probabilidade uniforme para as 15 possibilidades.

Obteremos a seguir um algoritmo simples que gera o passeio aleatório correspondente para todo ensemble localmente invariante. Ele contém como casos especiais o procedimento acima para o ensemble uniforme, e também a regra semelhante encontrada em [18] para o ensemble porta-fixa com $C = CNOT$.

Uma característica importante do algoritmo geral é que, em decorrência das várias simetrias discutidas na Seção anterior, ele só tem dois parâmetros independentes a e b, cujos valores dependem dos detalhes de cada ensemble. Assim, esses parâmetros contêm toda a informação necessária no cálculo dos momentos de Pauli de segunda ordem.

Teorema 4.2.1. *Para qualquer ensemble localmente invariante μ, a cadeia de Markov com matriz P_μ é equivalente a um passeio aleatório gerado pelo seguinte algoritmo: começando pelo ponto $\vec{p}$, escolha o próximo ponto $\vec{q}$ como se segue*

1. *Primeiro, escolha um par de coordenadas p_i, p_j de $\vec{p}$ de forma aleatória e uniforme dentre todos os pares presentes no grafo Γ. Para todas as outras coordenadas $k \neq i, j$, mantenha $q_k = p_k$.*

2. *Em seguida, escolha os valores para q_i, q_j da seguinte forma:*

   a) Se $p_i = p_j = 0$ faça $q_i = q_j = 0$

   b) Se $p_i = 0, p_j \neq 0$ então $\begin{cases} \text{com probabilidade } a, & \text{faça } q_i = r, q_j = 0; \\ \text{com probabilidade } b, & \text{faça } q_i = r, q_j = s; \\ \text{com probabilidade } 1-a-b, & \text{faça } q_i = 0, q_j = r; \end{cases}$

   onde r, s são escolhidos uniformemente e independentemente de $\{X, Y, Z\}$, e

$$a = \frac{3}{32}\left[\langle\langle \sigma_{X0,X0}|\hat{G}^{(2)}_\mu|\sigma_{0X,0X}\rangle\rangle + \langle\langle \sigma_{0X,0X}|\hat{G}^{(2)}_\mu|\sigma_{X0,X0}\rangle\rangle\right] \tag{4.1}$$

$$b = \frac{9}{32}\left[\langle\langle \sigma_{XX,XX}|\hat{G}^{(2)}_\mu|\sigma_{0X,0X}\rangle\rangle + \langle\langle \sigma_{XX,XX}|\hat{G}^{(2)}_\mu|\sigma_{X0,X0}\rangle\rangle\right] \tag{4.2}$$

   c) Se $p_i \neq 0, p_j = 0$ então faça como acima, trocando os índices $i \leftrightarrow j$

   d) Se $p_i, p_j \neq 0$ então $\begin{cases} \text{com probabilidade } b/3, & \text{faça } q_i = r, q_j = 0; \\ \text{com probabilidade } b/3, & \text{faça } q_i = 0, q_j = r; \\ \text{com probabilidade } 1-2b/3, & \text{faça } q_i = r, q_j = s; \end{cases}$

   onde r, s são escolhidos uniformemente e independentemente de $\{X, Y, Z\}$.

Prova: Começamos reescrevendo eq. (3.14) de forma que fique explícita sua simetria sob permutações dos índices de $\hat{G}^{(2)}_{\mu_{ij}}$:

$$P_\mu(\vec{p}, \vec{q}) = \frac{2^{-2n}}{n(n-1)} \sum_{(i,j)\in\Gamma} \frac{1}{2}\langle\langle \sigma_{\vec{q},\vec{q}}|\hat{G}^{(2)}_{\mu_{ij}} + \hat{G}^{(2)}_{\mu_{ji}}|\sigma_{\vec{p},\vec{p}}\rangle\rangle \tag{4.3}$$

P_μ é uma média uniforme das matrizes de Markov $\bar{P}_{\mu_{ij}} = \frac{1}{2}\left(P_{\mu_{ij}} + P_{\mu_{ji}}\right)$, onde $P_{\mu_{ij}}$ é dado na eq. (3.16). Para cada valor dos índices (i, j), os elementos de $\bar{P}_{\mu_{ij}}$ definem probabilidades de transição para um passeio aleatório que só afeta as coordenadas i, j de $\vec{p}$. Portanto, o passo (1) do algoritmo está justificado.

Para verificar o passo (2), precisamos checar que a probabilidade de cada transição $p_ip_j \to q_iq_j$ possível, de acordo com as regras acima, é a mesma que aquela na cadeia de Markov com matriz $\bar{P}_{\mu_{ij}}$. Em outras palavras, precisamos checar se

$$\begin{aligned}\mathbb{P}(p_ip_j \to q_iq_j) &= 2^{-4} \left\langle\!\left\langle \sigma_{q_iq_j,q_iq_j} \right| \frac{1}{2}\left(\hat{G}^{(2)}_{\mu_{ij}} + \hat{G}^{(2)}_{\mu_{ji}}\right) \left| \sigma_{p_ip_j,p_ip_j} \right\rangle\!\right\rangle \\ &= \frac{1}{32}\left(\left\langle\!\left\langle \sigma_{q_iq_j,q_iq_j} \right| \hat{G}^{(2)}_{\mu_{ij}} \left| \sigma_{p_ip_j,p_ip_j} \right\rangle\!\right\rangle + \left\langle\!\left\langle \sigma_{q_jq_i,q_jq_i} \right| \hat{G}^{(2)}_{\mu_{ij}} \left| \sigma_{p_jp_i,p_jp_i} \right\rangle\!\right\rangle\right) \end{aligned} \tag{4.4}$$

(na última linha reordenamos os índices da matriz e dos estados de base). Comparando essa equação com as eqs. (4.1) e (4.2), e levando em conta o número de possíveis valores de r e s, vemos que as transições $0X \to X0$ e $0X \to XX$ no passo (2b) têm de fato as probabilidades corretas. Todas as transições restantes podem ser levadas em conta de forma semelhante, apelando-se para simetrias gerais da matriz de Markov $\bar{P}_{\mu_{ij}}$. Para começar, o passo (2a) segue do fato que, para qualquer ensemble μ, $|\sigma_{00,00}\rangle\!\rangle$ é um autovetor de $\hat{G}^{(2)}_{\mu_{ij}}$ com autovalor 1. Isso é consequência do Lema 2.3.3, e também pode ser visto das eqs. (3.21) e (3.25). Portanto, $\left\langle\!\left\langle \sigma_{q_iq_j,q_iq_j} \right| \left(\hat{G}^{(2)}_{\mu_{ij}} + \hat{G}^{(2)}_{\mu_{ji}}\right) |\sigma_{00,00}\rangle\!\right\rangle$ se anula exceto quando $q_i = q_j = 0$.

Indo agora para os passos (2b)-(2d), primeiro notamos que, novamente pelo Lema 2.3.3, $\langle\!\langle \sigma_{00,00}| \hat{G}^{(2)}_{\mu_{ij}} \left|\sigma_{p_ip_j,p_ip_j}\right\rangle\!\rangle = 0$ exceto se $p_i = p_j = 0$. Isso explica porque nesses casos não pode haver transição para $q_i = q_j = 0$. Além disso, a simetria de permutação axial discutida na Seção 4.1.4 implica que todas as probabilidades de transição onde p_i, p_j, q_i ou q_j são $\neq 0$ devem ser insensíveis ao fato de essas coordenadas serem iguais a X, Y ou Z. Portanto, no passo (2b) por exemplo, a probabilidade de $p_i = 0, p_j = X$ trocar para $q_i = X, q_j = 0$ deve ser a mesma que (digamos) a de $p_i = 0, p_j = Y$ trocar para $q_i = Z, q_j = 0$. Segue que todas as outras transições com probabilidades a e b no passo (2b) também estão corretas. As transições restantes no passo (2b) requerem

$$1 - a - b = 3 \cdot \mathbb{P}(0X \to 0X) = \frac{3}{32}\left[\langle\!\langle \sigma_{0X,0X}| \hat{G}^{(2)}_{\mu} |\sigma_{0X,0X}\rangle\!\rangle + \langle\!\langle \sigma_{X0,X0}| \hat{G}^{(2)}_{\mu} |\sigma_{X0,X0}\rangle\!\rangle\right] \tag{4.5}$$

mas isso é automaticamente verdadeiro já que $\bar{P}_{\mu_{ij}}$ é uma matriz estocástica com linhas que somam 1.

O passo (2c) segue simplesmente da invariância de $\bar{P}_{\mu_{ij}}$ sob trocas de i, j.

O passo (2d) é menos trivial. Segundo eq. (4.4), a probabilidade de $XX \to X0$ deve ser a mesma que $XX \to 0X$, e igual a

$$c \equiv \frac{1}{32}\left[\langle\!\langle \sigma_{X0,X0}| \hat{G}^{(2)}_{\mu} |\sigma_{XX,XX}\rangle\!\rangle + \langle\!\langle \sigma_{0X,0X}| \hat{G}^{(2)}_{\mu} |\sigma_{XX,XX}\rangle\!\rangle\right] \tag{4.6}$$

Entretanto, as regras para o passo (2d) dão

$$\mathbb{P}(XX \to X0) = \frac{b}{9} = \frac{1}{32}\left[\langle\langle \sigma_{XX,XX}|\hat{G}^{(2)}_{\mu}|\sigma_{0X,0X}\rangle\rangle + \langle\langle \sigma_{XX,XX}|\hat{G}^{(2)}_{\mu}|\sigma_{X0,X0}\rangle\rangle\right] \tag{4.7}$$

Se $\hat{G}^{(2)}_{\mu}$ é simétrico (e é o caso, por exemplo, para o ensemble uniforme e também para ensembles porta-fixa com C Hermiteano), as duas expressões claramente concordam. Para ensembles localmente invariantes em geral, entretanto, não podemos garantir que $\hat{G}^{(2)}_{\mu}$ será simétrico (veja Seção 3.1). Apesar disso, as duas expressões acima são ainda iguais, pela seguinte razão: do Teorema 3.3.1, sabemos que $P_{\mu_{ij}}$ é bi-estocástica - logo o mesmo é verdade para $\bar{P}_{\mu_{ij}}$. Em particular, somando os elementos de matriz na coluna $0X$:

$$\begin{aligned}
1 =& \sum_{p_ip_j} \bar{P}_{\mu_{ij}}(0X, p_ip_j) = \frac{1}{32}\sum_{p_ip_j} \langle\langle \sigma_{0X,0X}|\left(\hat{G}^{(2)}_{\mu_{ij}} + \hat{G}^{(2)}_{C_{ji}}\right)|\sigma_{p_ip_j,p_ip_j}\rangle\rangle \\
=& \frac{1}{32}\sum_{p_ip_j} \langle\langle \sigma_{0X,0X}|\hat{G}^{(2)}_{\mu_{ij}}|\sigma_{p_ip_j,p_ip_j}\rangle\rangle + \langle\langle \sigma_{X0,X0}|\hat{G}^{(2)}_{\mu_{ij}}|\sigma_{p_jp_i,p_jp_i}\rangle\rangle \\
=& \frac{3}{32}\langle\langle \sigma_{0X,0X}|\hat{G}^{(2)}_{\mu_{ij}}|\sigma_{X0,X0}\rangle\rangle + \langle\langle \sigma_{X0,X0}|\hat{G}^{(2)}_{\mu_{ij}}|\sigma_{0X,0X}\rangle\rangle \\
&+ \frac{3}{32}\langle\langle \sigma_{0X,0X}|\hat{G}^{(2)}_{\mu_{ij}}|\sigma_{0X,0X}\rangle\rangle + \langle\langle \sigma_{X0,X0}|\hat{G}^{(2)}_{\mu_{ij}}|\sigma_{X0,X0}\rangle\rangle \\
&+ \frac{9}{32}\langle\langle \sigma_{0X,0X}|\hat{G}^{(2)}_{\mu_{ij}}|\sigma_{XX,XX}\rangle\rangle + \langle\langle \sigma_{X0,X0}|\hat{G}^{(2)}_{\mu_{ij}}|\sigma_{XX,XX}\rangle\rangle \\
=& \ a + (1 - a - b) + 9c
\end{aligned}$$

onde na última linha foram usadas as eqs. (4.1), (4.5) e (4.6). Logo, $c = b/9$ para todos os ensembles localmente invariantes. □

4.2.1 Exemplos

Vimos na Seção 2.3 que $\hat{G}^{(2)}_{\mathcal{H}_{i_1i_2}} \otimes \hat{G}^{(2)}_{\mathcal{H}_{j_1j_2}}$ é o projetor no subespaço gerado pela base eq. (3.25). Substituindo esse projetor na eq. (3.22) (depois de normalizar a base), e então nas eqs. (4.1) e (4.2), obtemos depois de um pouco de álgebra

$$a = \frac{1}{96}\int \sum_{j,k\in\{X,Y,Z\}} \mathrm{Tr}^2\left[\sigma_{0j}C\sigma_{k0}C^\dagger\right] + \mathrm{Tr}^2\left[\sigma_{j0}C\sigma_{0k}C^\dagger\right] d\mu(C) \tag{4.8}$$

$$b = \frac{1}{96}\int \sum_{i,j,k\in\{X,Y,Z\}} \mathrm{Tr}^2\left[\sigma_{ij}C\sigma_{k0}C^\dagger\right] + \mathrm{Tr}^2\left[\sigma_{ij}C\sigma_{0k}C^\dagger\right] d\mu(C) \tag{4.9}$$

Vamos agora calcular essas expressões explicitamente em vários casos de interesse.

Ensemble Uniforme

Nesse caso, ao invés de calcular a e b usando as eqs. acima, é mais rápido substituir diretamente a eq. (3.21) nas eqs. (4.1) e (4.2). Usando também a eq. (2.4), obtemos $a = 1/5$, $b = 3/5$ [27]. É fácil conferir que, com estes valores, o algoritmo geral do Teorema 4.2.1 se reduz ao caso particular visto antes.

Ensemble porta-fixa

No caso de ensembles porta-fixa μ_C, simplesmente ignoramos as integrais nas eqs. (4.8) e (4.9). Podemos então encontrar expressões fechadas para a e b usando o fato de que qualquer porta de 2 q-bits C é localmente equivalente a uma porta na "forma canônica' [44, 45]

$$C(\alpha_X, \alpha_Y, \alpha_Z) = \exp\left(i\left[\alpha_X \sigma_{XX} + \alpha_Y \sigma_{YY} + \alpha_Z \sigma_{ZZ}\right]\right) \ , \tag{4.10}$$

onde os parâmetros α_X,α_Y e α_Z são números reais.

Já que mostramos na eq. (3.6) que portas localmente equivalentes geram o mesmo ensemble porta-fixa, segue-se que os coeficientes a, b são eles próprios invariantes por rotações locais, e podem portanto ser escritos em função dos parâmetros $\alpha_X, \alpha_Y, \alpha_Z$. De fato, calculando os comutadores

$$[C(\alpha_X, \alpha_Y, \alpha_Z), \sigma_{0X} \pm \sigma_{X0}] = 2(\alpha_Y \mp \alpha_Z)(\sigma_{ZY} \pm \sigma_{YZ})$$
$$[C(\alpha_X, \alpha_Y, \alpha_Z), \sigma_{ZY} \pm \sigma_{YZ}] = -2(\alpha_Y \mp \alpha_Z)(\sigma_{X0} \pm \sigma_{0X})$$

podemos usar a identidade de Baker-Campbell-Hausdorff e a simetria de eq. (4.10) sob permutação de X, Y, Z para reescrever as eqs. (4.8) e (4.9) como

$$a = \frac{1}{6} \sum_{i \neq j} \operatorname{sen}^2(2\alpha_i) \operatorname{sen}^2(2\alpha_j) \tag{4.11}$$

$$b = \frac{1}{3} \sum_{i \neq j} \operatorname{sen}^2(2\alpha_i) \cos^2(2\alpha_j) \tag{4.12}$$

Vale observar ainda que a "forma canônica" dada pela eq. (4.10), não é única, ou seja, o mesmo operador pode em geral corresponder a conjuntos diferentes de coeficientes $\alpha_{X,Y,Z}$. Este problema é reduzido em parte notando que esses 3 parâmetros podem ser restritos à região $\pi/4 \geq \alpha_X \geq \alpha_Y \geq |\alpha_Z| \geq 0$ sem perda de generalidade [44, 46]. Entretanto mesmo essa restrição não é suficiente, de fato sabe-se que as propriedades não-locais de uma porta de 2 q-bits podem ser caracterizadas por apenas 2 parâmetros independentes [46, 47], sendo um real e um complexo. É possível, portanto, escrever a e b como função de apenas 2 parâmetros, embora as expressões sejam complicadas.

Os valores de a e b têm que estar entre 0 e 1, de maneira a respeitar a desigualdade $a+b \leq 1$. A estrutura do grupo de Pauli impõe restrições adicionais, a mais importante é:

Lema 4.2.1. *Para qualquer ensemble de porta-fixa μ_C, $b \leq 2/3$.*

Prova: A ação de C por conjugação sobre um operador de Pauli é mapeá-lo numa nova combinação de Paulis. Seja $C\sigma_{A0}C^\dagger = \sum_{ij} x^A_{ij}\sigma_{ij}; A = X, Y, Z$. Defina ainda $\gamma_A = \sum_{i,j\neq 0}(x^A_{ij})^2$. Como o super-operador associado a conjugação por C é unitário, $\gamma_A \leq 1$. Comparando com eq. (4.9) e usando eq. (2.4), vemos que o primeiro termo de b é igual a $\frac{1}{6}\sum_A \gamma_A$.

Agora, devido à estrutura do grupo de Pauli, as expansões acima não são independentes:

$$C\sigma_{Y0}C^\dagger = i\sum_{ij} x^X_{ij}\sigma_{ij}\sum_{kl} x^Z_{kl}\sigma_{kl} = i\sum_{ij,kl} x^X_{ij}x^Z_{kl}\sigma_i\sigma_k \otimes \sigma_j.\sigma_l \ .$$

Considerando ainda que a conjugação preserva a hermiticidade, alguns termos da soma acima terão de ser nulos. Em particular, se $\sigma_i, \sigma_j, \sigma_k, \sigma_l$ são todos $\neq \sigma_0$, os termos resultantes só podem existir se exatamente 1 dos $\sigma_i\sigma_k$ e 1 dos $\sigma_j\sigma_l$ é igual a σ_0. Em outras palavras, o produto dos termos que contribuem para γ_X com os que contribuem para γ_Z não contribuirão para γ_Y:

$$\gamma_Y \leq 1 - \sum_{\substack{i\neq j\\ k\neq l}} \left(x^X_{ij}x^Z_{kl}\right)^2 = 1 - \gamma_X\gamma_Z \ ,$$

Consequentemente

$$\gamma_X + \gamma_Y + \gamma_Z \leq 1 + \gamma_X + \gamma_Z - \gamma_X\gamma_Z \leq 2 \ ,$$

e o primeiro termo de b é menor ou igual a $1/3$. O mesmo argumento pode ser usado no segundo termo de b concluindo a prova. □

Vamos agora investigar vários casos especiais da porta C que são relevantes. Muitos podem ser obtidos diretamente da ref. [48], que lista valores de $\alpha_{X,Y,Z}$ para várias portas interessantes:

1. **Portas não-emaranhantes**

   Lema 4.2.2. *$b = 0$ se e só se C é não-emaranhante.*

   Prova: Da eq. (9) na ref. [48], um porta é não-emaranhante (i.e., mapeia estados produto puros em estados produto puros) se e somente se $\alpha_i \equiv 0$ ou $\alpha_i \equiv \pi/4$ para todo $i = X, Y, Z$. Esses casos correspondem respectivamente a portas localmente

Tabela 4.1: Permutações induzidas por algumas portas Clifford sobre Paulis

CNOT		
$00 \leftrightarrow 00$	$0X \leftrightarrow 0X$	$YZ \leftrightarrow XY$
$X0 \leftrightarrow XX$	$0Y \leftrightarrow ZY$	$XZ \leftrightarrow YY$
$Y0 \leftrightarrow YX$	$0Z \leftrightarrow ZZ$	$ZX \leftrightarrow ZX$
$Z0 \leftrightarrow Z0$		

XY		
$00 \leftrightarrow 00$	$0X \leftrightarrow YZ$	$XX \leftrightarrow XX$
$X0 \leftrightarrow ZY$	$0Y \leftrightarrow XZ$	$YY \leftrightarrow YY$
$Y0 \leftrightarrow ZX$		$ZZ \leftrightarrow ZZ$
$Z0 \leftrightarrow 0Z$		$XY \leftrightarrow YX$

equivalentes à identidade e à porta SWAP. Substituindo em eq. (4.12) conclui-se a prova.

No contexto dos circuitos quânticos aleatórios esse resultado pode ser entendido pelo fato já comentado acima que, para qualquer porta emaranhante C, o circuito converge para uma distribuição uniforme sobre os unitários de n q-bits [23]. Nesse caso a cadeia descrita pela matriz P_μ tem que convergir para a distribuição uniforme sobre todos $\vec{p} \neq 0$, independentemente da posição inicial. Entretanto, se $b = 0$, então em cada passo o passeio aleatório leva $\vec{p}$ em $\vec{q}$ sem alterar o número de componentes iguais a zero, o que significa que a distribuição uniforme nunca é atingida. Por outro lado, se $b \neq 0$, veremos (Seção 5.3) que o circuito converge para um 2-desenho em tempo polinomial. Em particular, um 2-desenho é capaz de gerar estados com muito emaranhamento [18]. E portanto, nesse caso, C tem que ser emaranhante.

2. **Portas Clifford**

Muitas das mais conhecidas portas de 2 q-bits são elementos do grupo de Clifford [49], formado por operadores que mapeiam (sob conjugação) os Paulis neles mesmos:

$$C\sigma_{ij}C^\dagger = \pm\sigma_{kl} \text{ para algum } k, l. \tag{4.13}$$

A maioria das portas usadas em estudos anteriores de CQAs [28, 18] são desse tipo, incluindo as portas CNOT , CZ (Z-controlada) e XY. Calculando-se os coeficientes α_i pode-se ver que as primeiras duas são na verdade localmente equivalentes. Assim, pela eq. (3.6) CQAs construídos com uma ou outra serão completamente equivalentes. Já a XY (que não é o produto σ_{XY}, mas a porta com a matriz

$$XY = \begin{pmatrix} 1 & & & \\ & & -i & \\ & -i & & \\ & & & 1 \end{pmatrix}$$

na base computacional) é localmente equivalente à porta DCNOT (uma sequência de duas portas CNOT onde se inverte o papel do controle e do alvo) [48]

Embora seja possível, naturalmente, calcular a, b para as portas Clifford fazendo uso de eq. (4.11) e eq. (4.12), também pode-se usar o seguinte método mais intuitivo: observe que, a menos do sinal ($\pm$) na eq. (4.13), qualquer porta Clifford pode ser considerada como uma permutação no conjunto dos Paulis. Na tabela 4.1 listamos essas permutações para os dois casos acima. Note que as portas mostradas acima são Hermiteana (CNOT, CZ) ou anti-Hermiteana (XY), de maneira que $C^2 = \pm I$. Assim as permutações são reflexivas i.e. contêm apenas ciclos de comprimento 1 e 2. Para outras portas de Clifford isso pode não acontecer.

Os valores de a e b podem ser "lidos" da estrutura da permutação induzida nos Paulis da seguinte maneira: veja quantos operadores da forma σ_{0k} são levados para a outra forma σ_{j0} (onde $j, k \neq 0$). Faça o mesmo para a direção oposta (esses 2 números, que chamaremos de n_1 e n_2, são iguais quando C é Hermiteana ou anti-Hermiteana). Agora veja quantos operadores da forma σ_{0k} ou σ_{k0} são permutados para outro da forma σ_{ij} com $i, j \neq 0$ (chame esses números de n_3 e n_4). Os valores de a e b são então

$$a = \frac{1}{6}(n_1 + n_2); \quad b = \frac{1}{6}(n_3 + n_4)$$

Para os exemplos acima obtemos

$$\textbf{CNOT (ou CZ)}: \quad a = 0, b = \frac{2}{3}; \quad \textbf{XY (ou DCNOT)}: \quad a = \frac{1}{3}, b = \frac{2}{3} \tag{4.14}$$

(Os valores de CNOT foram obtidos previamente em [18]). Nota-se que todas essas portas saturam a desigualdade (4.2.1). Um exemplo não-trivial com um valor diferente de b é a porta SWAP, para a qual $a = 1, b = 0$. De fato, podemos provar que

Lema 4.2.3. *Para qualquer porta Clifford, b só pode valer 0 ou 2/3. Em particular, dado o Lema 4.2.2, todas as portas Clifford que não são equivalentes à identidade ou à SWAP devem ter $b = 2/3$.*

Prova: Começamos mostrando que n_3 e n_4 são números pares. Da prova do Lema 4.2.1, sabemos que $n_3, n_4 \leq 2$, o que elimina o valor 3 .

Suponha $n_3 \geq 1$. Como $n_3 \neq 3$, podemos assumir sem perda de generalidade[1] que

$$C\sigma_{X0}C^\dagger = \pm\sigma_{kl};\ k, l \neq 0 \quad \text{e} \quad C\sigma_{Y0}C^\dagger = \pm\sigma_{mn};\ \text{onde } m = 0 \text{ ou } n = 0.$$

[1] Qualquer outra permutação pode ser obtida por um porta C' que é equivalente a C por rotações Clifford locais [49]

Consequentemente

$$C\sigma_{Z0}C^\dagger = \pm i\sigma_k \otimes \sigma_l\sigma_n \text{ ou } \pm i\sigma_k\sigma_m \otimes \sigma_l$$

No primeiro caso ou $l = n$, e o operador resultante não é Hermiteano, ou $l \neq n$, o que implica em $n_3 \neq 1$. O outro caso é similar. Finalmente, argumentos análogos funcionam para n_4, que é também $\neq 1$ ou 3. Com isso, b só pode valer $0, 1/3$ ou $2/3$. Resta então mostrarmos que é necessário que $n_3 = n_4$, eliminando o caso intermediário.

Para isso, note que $n_1 \neq 2$ já que, se $C\sigma_{X0}C^\dagger = \pm\sigma_{0k}$ e $C\sigma_{Y0}C^\dagger = \pm\sigma_{0l}$, então necessariamente $C\sigma_{Z0}C^\dagger = \pm\sigma_{0m}$. Defina então um número n_5, que conta quantos Paulis $\sigma_{0k}, k \neq 0$, são mapeados a outros Paulis da mesma forma. Por uma argumento análogo ao que acabamos de dar, podemos concluir que $n_5 \neq 2$.

De volta ao argumento principal, mostraremos que se $n_3 = 0$, então $n_4 \neq 2$. Suponha por absurdo que $n_3 = 0$ e $n_4 = 2$. Como $n_1 + n_3 + n_5 = 3$, mas n_1 e $n_5 \neq 2$, segue que ou $n_1 = 3, n_5 = 0$ ou vice-versa. Vamos supor $n_1 = 3$, o outro caso é similar. Sem perda de generalidade, a permutação induzida por C pode ser tomada da forma $\sigma_{A0} \to \pm\sigma_{0A}, \forall A \in X, Y, Z$; $\sigma_{0X} \to \pm\sigma_{j0}; \sigma_{0Y} \to \pm\sigma_{kl}$, onde $j, k, l \neq 0$. Mas então

$$C\sigma_{YY}C^\dagger = \pm\sigma_k \otimes \sigma_Y\sigma_l \text{ e também } \quad C\sigma_{ZY}C^\dagger = \pm\sigma_k \otimes \sigma_Z\sigma_l$$

A primeira expressão é Hermiteana somente se $l = Y$, e a segunda somente se $l = Z$. Essa contradição leva a $n_3 = 0 \Rightarrow n_4 = 0$. Com o mesmo argumento na direção oposta obtemos $n_4 = 0 \Rightarrow n_3 = 0$ e finalmente obtemos $b \neq 1/3$.

3. **Portas tipo Ising**

   Outra subclasse importante de unitários são as geradas pela interação tipo Ising: $C = \exp(i\alpha_Z\sigma_{ZZ})$. Da eq. (4.11) e eq. (4.12) obtemos

$$\textbf{Ising}: \quad a = 0; \quad b = \frac{2}{3}\,\text{sen}^2(2\alpha_Z)$$

   Nesse caso b vai continuamente de 0 (quando $\alpha = 0$, C é a identidade) a $2/3$ (quando $\alpha = \pi/4$, C é localmente equivalente a CZ e $CNOT$).

4. **Porta B**

   A porta B [47] é a porta na forma 'canônica' da eq. (4.10) com $\alpha_X = \pi/4, \alpha_Y = \pi/8, \alpha_Z = 0$. Ela possui a propriedade especial [47] de gerar qualquer porta de 2 q-bits com apenas 2 usos (enquanto, por exemplo, 3 usos de $CNOT$ são necessários

[50, 51]). Isto poderia sugerir que o uso da porta B como a porta fixa levaria ao circuito de convergência mais rápida. Entretanto, veremos na Seção 5.4 que, em geral, essa porta tem desempenho semelhante às portas Clifford. Para a porta B temos os parâmetros

$$\mathbf{B}: \quad a = \frac{1}{6}; \quad b = \frac{2}{3}$$

Caso geral

As eqs. (4.8) e (4.9) significam que, para qualquer ensemble localmente invariante μ, os coeficientes a e b são simplesmente médias sobre os valores de $a(C)$ e $b(C)$ de cada ensemble porta-fixa μ_C.

$$a = \int a(C) d\mu(C) \quad b = \int b(C) d\mu(C)$$

Podemos então imediatamente generalizar os Lemas 4.2.1 e 4.2.2 acima.

Lema 4.2.4. *Para qualquer ensemble localmente invariante, $b \leq 2/3$*

Lema 4.2.5. *Um ensemble localmente invariante μ tem $b = 0$ se e só se $\int_{em} d\mu(C) = 0$, onde a integral é sobre o conjunto de portas emaranhantes.*

Em resumo, qualquer ensemble localmente invariante com probabilidade não-nula de gerar uma porta emaranhante produz uma cadeia de Markov com $0 < b \leq 2/3$. Veremos no próximo capítulo que toda cadeia desta forma converge para o mesmo estado estacionário, em um número de passos que escalona inversamente com b. Desta forma, as cadeias baseadas em ensembles porta-fixa com $b = 2/3$ são as mais rápidas possíveis entre todos os ensembles. Em particular, são mais rápidas do que as baseadas no ensemble uniforme. Esse resultado confirma dados numéricos obtidos em [28], conforme figura 5.1.

4.3 Reversibilidade

Uma matriz de transição é dita reversível quando

$$\pi(x)P(x,y) = \pi(y)P(y,x), \quad \text{onde } \pi P = \pi \ .$$

Pela definição da matriz de transição P_μ dada pelo passeio aleatório do Teorema 4.2.1 é fácil ver que ela será sempre reversível para qualquer ensemble μ. Repare que esse resultado é válido mesmo que o operador $\hat{G}_\mu^{(2)}$ seja não-Hermiteano. De acordo com o Teorema Perron-Frobenius[41] uma matriz de transição reversível é diagonalizável e tem autovetores e autovalores reais. Essa propriedade será particularmente útil em 5.2.

4.4 Ergodicidade

Nessa Seção mostraremos que, sob certas condições, uma distribuição de probabilidade(ν) nas *strings* $\{0,X,Y,Z\}^{(n)}$ que evolui sob a ação da cadeia P_μ converge sempre para uma mesma distribuição de probabilidade(π).

A condição que deve ser observada para que isso ocorra é que a distribuição inicial deve ter seu suporte em $\{0,X,Y,Z\}^{(n)}/\ \{0\}^{(n)}$, pois a *string* $\{0\}^{(n)}$, associada à matriz densidade completamente mista, é sempre levada a si mesma por P_μ. Mostraremos então que a cadeia P_μ com espaço de estados formado pelas *strings* $\{0,X,Y,Z\}^{(n)}/\{0\}^{(n)}$ é ergódica. Um resultado básico da literatura de cadeias de Markov, veja a Seção 2.5.2, é que cadeias ergódicas possuem somente um estado estacionário para o qual convergem partindo de qualquer condição inicial. Daqui por diante a cadeia P_μ será sempre restrita a $\{0,X,Y,Z\}^{(n)}/\ \{0\}^{(n)}$. Conforme comentários da Seção 4.1, a cadeia P_μ é bi-estocática e portanto a distribuição uniforme é estacionária sobre P_μ. E portanto:

Lema 4.4.1. *A distribuição $\pi(\vec{p}) = 1/(4^n-1)$ é a única distribuição estacionária de P_μ. Qualquer distribuição converge no limite de muitos passos para π.*

Prova : Basta provar que a cadeia P_μ é ergódica. Uma condição suficiente para uma cadeia ser ergódica é ser *irredutível* (não pode ser escrita como a soma direta de sub-cadeias independentes) e *aperiódica* (nenhuma condição inicial leva a evoluções repetitivas) (veja Capítulo 1 de [32] para definições mais precisas). A aperiodicidade é trivialmente satisfeita pela cadeia P_μ já que $P(\vec{q},\vec{q}) > 0, \forall \vec{q} \in \Omega_Q$. Para garantir a irredutibilidade precisamos mostrar que existe um t (que pode depender de $\vec{q}$ e $\vec{p}$) tal que $P^t(\vec{q},\vec{p}) > 0$ para qualquer par $(\vec{q},\vec{p})$. Para ver que isso sempre acontece vamos definir o conjunto $B_0 = \{i \in [1,n] | q_i \neq p_i\}$. Para construir um caminho de $\vec{p}$ para $\vec{q}$ com probabilidade positiva basta ver que sempre existe um $j \in B_0$ para o qual a transição de $\vec{p}$ para um estado intermediário $\vec{q^{(1)}}$, que é diferente de $\vec{p}$ apenas na componente j, tem sempre probabilidade positiva. A iteração desse processo leva de $\vec{p}$ para $\vec{q}$ com probabilidade positiva. □

Se houver restrição aos pares de q-bits escolhidos para interagir formando um grafo Γ, como definido na introdução do CQA (1.2), a irredutibilidade se verifica se e só se Γ for um grafo conexo. Se Γ for desconexo, o argumento acima só continua válido dentro de cada cluster conexo de q-bits, os quais convergem separadamente para estados estacionários uniformes. Fisicamente, após a convergência obtemos um 2-desenho para unitários restritos a cada cluster.

4.5 Redução e decomposição da cadeia

Nesta Seção vemos como a matriz de Markov P_μ pode ser mapeada em cadeias 'reduzidas', mais simples, e ainda como ela pode ser escrita como uma combinação de cadeias que comutam, e cujas propriedades são também relativamente simples. De agora em diante, por conveniência omitimos o índice μ, escrevendo apenas P ao invés de P_μ exceto quando necessário.

4.5.1 Redução da Cadeia P

A primeira técnica que usaremos para simplificar a análise é a redução da cadeia. Diz-se que uma cadeia de Markov pode ser *reduzida* quando podemos subdividir o espaço de *strings* Ω em subconjuntos $\Omega_a, \Omega_b, \ldots \subset \Omega$ tais que as probabilidades de transições entre subconjuntos

$$\mathbb{P}(\Omega_a, \Omega_b) = \sum_{\substack{x \in \Omega_a \\ y \in \Omega_b}} P(x, y)$$

formem elas próprias uma cadeia de Markov. Uma condição necessária e suficiente para isto ser possível é que a soma das probabilidades de transição $p(x, y)$ entre um elemento de Ω_μ e todos os elementos de Ω_ν é a mesma para qualquer elemento de Ω_μ [33].

Por exemplo: vimos na Seção 4.1 que a probabilidade de transição $P(\vec{p_1}, \vec{p_2})$ é invariante por permutações axiais dos componentes de $\vec{p_1}$ e $\vec{p_2}$. Assim, como a matriz de transição P só distingue se a i-ésima componente do elemento $\vec{p}$ é 0 ou X, Y, Z, podemos reduzi-la. Cada subconjunto de Ω será indexado por uma *string* $\vec{q} \in \{0, 1\}^n = \Omega_Q$. Cada subconjunto $\Omega_{\vec{q}}$ conterá todas as *strings* de Ω que possuírem zeros nas coordenadas nulas de $\vec{q}$ e não-zeros nas coordenas de $\vec{q}$ que valham 1. Essa redução também é uma matriz de Markov, que chamaremos de Q, porque a probabilidade de transição entre elementos de conjuntos distintos só depende dos índices dos conjuntos.

Em geral, a cadeia reduzida não contém todos os autovalores da cadeia original, e pode apresentar convergência mais rápida. Isso não acontece aqui pois a dinâmica dentro dos subconjuntos $\Omega_{\vec{q}}$ não é relevante. Isso pode ser visto como uma consequência direta da aplicação do unitário local antes da interação entre pares, pois após a rotação local não há diferença se a i-ésima (ou a j-ésima) componente do estado era X, Y ou Z. Ou seja, depois de que todos os q-bits foram escolhidos todos os vetores dentro de um mesmo subconjunto $\Omega_{\vec{q}}$ tem a mesma probabilidade.

A cadeia Q, reduzida de P, no espaço das *strings* $\Omega_Q = \{0, 1\}^n / \{0\}^n$, tem a matriz de

$Q^{(i,j)}$	00	01	10	11
00	1	0	0	0
01	0	1-a-b	a	b
10	0	a	1-a-b	b
11	0	b/3	b/3	1-2b/3

Tabela 4.2: Tabela de elementos de $Q^{(i,j)}$. Na coluna mais à esquerda temos os valores iniciais $q_i q_j$ e na linha superior os valores finais $q'_i q'_j$.

transição

$$Q(\vec{q},\vec{q}') = \sum_{\substack{\vec{p}\in\Omega\vec{q} \\ \vec{p}'\in\Omega\vec{q}'}} P(\vec{p},\vec{p}') \ .$$

Assim como P, Q pode ser escrita como uma soma convexa sobre todos os pares de coordenadas de uma mesma matriz que só altera os valores de um par específico, ou seja, $Q = \frac{1}{n(n-1)} \sum_{i\neq j} Q^{(i,j)}$. Para $\vec{q},\vec{q'} \in \Omega_Q$ que só diferem nas coordenadas i e j a matriz $Q^{(i,j)}$ assume a forma dada pela tabela 4.2.

4.5.2 Decomposição da cadeia reduzida Q

Vamos continuar nossa análise da cadeia de Markov associada ao CQA introduzindo uma decomposição da matriz Q que simplificará consideravelmente a análise dos seus efeitos. Para cada par de q-bits (i,j), escrevemos

$$Q^{(i,j)} = aT^{(i,j)} + (1-a)M^{(i,j)}. \tag{4.15}$$

onde $T^{(i,j)}$ é a matriz de transposição (SWAP) das entradas i e j, e $M^{(i,j)}$ a matriz dada na tabela 4.3.

Vê-se então da eq. (4.15) que a cadeia $Q = \sum_{i,j} Q^{(i,j)}$ pode ser escrita como uma média, com pesos definidos pela constante a, de duas outras cadeias:

$$Q = aT + (1-a)M \tag{4.16}$$

onde $T \equiv \frac{1}{n(n-1)} \sum_{(i,j)\in\Gamma} T^{(i,j)}$ e $M \equiv \frac{1}{n(n-1)} \sum_{(i,j)\in\Gamma} M^{(i,j)}$, respectivamente. A primeira representa uma transposição aleatória [52] das componentes de $\vec{q}$. Já a segunda é mais fácil de compreender através da sua versão 'preguiçosa', ou seja, a sua média com a operação

$M^{(i,j)}$	00	01	10	11
00	1	0	0	0
01	0	$\left(\frac{1-a-b}{1-a}\right)$	0	$\left(\frac{b}{1-a}\right)$
10	0	0	$\left(\frac{1-a-b}{1-a}\right)$	$\left(\frac{b}{1-a}\right)$
11	0	$\left(\frac{b/3}{1-a}\right)$	$\left(\frac{b/3}{1-a}\right)$	$\left(\frac{1-a-2b/3}{1-a}\right)$

Tabela 4.3: Tabela de elementos de $M^{(i,j)}$. Na coluna mais à esquerda temos os valores iniciais $q_i q_j$, e na linha superior os valores finais $q'_i q'_j$.

identidade:

$$L = \frac{M + \mathbb{I}}{2} \, . \tag{4.17}$$

Os passeios aleatórios gerados por L e M são essencialmente os mesmos, exceto pelo fato de que aquele baseado em L tem 50% de chance de não se mover em cada passo. Assim, todas as escalas temporais relevantes do passeio L (em particular, seu tempo de mistura) ficam dobradas em relação às de M. Vamos tornar essa noção mais precisa na Seção 5.2, onde mostraremos que podemos obter o tempo de mistura de Q a partir do tempo de mistura de L.

AVISO: até agora, todos os resultados que derivamos são válidos qualquer que seja o grafo Γ. A partir de agora, exceto onde explicitamente assinalado, assumiremos por simplicidade que Γ é o grafo cheio, ou seja, que todos os pares de q-bits podem interagir em cada passo do CQA. Com esta suposição, mostramos a seguir que a cadeia L pode ser escrita como uma média de cadeias idênticas $L^{(i)}$, cada uma atuando sobre uma única componente de $\vec{p}$. No caso de outros grafos Γ, L continua sendo uma média de cadeias individuais, mas já não necessariamente idênticas, e os detalhes destas cadeias também serão diferentes do que obteremos abaixo. Embora existam diversos casos interessantes de Γ que podem também ser analisados (por exemplo, interações de primeiros vizinhos sobre uma rede regular de dimensão d), não faremos isto nesta dissertação.

Lema 4.5.1. *A cadeia L pode ser escrita na forma*

$$L = \frac{1}{n} \sum_i L^{(i)} \, , \tag{4.18}$$

onde $L^{(i)}$ é uma matriz estocástica 2×2 que só afeta a componente p_i, com

$$L^{(i)}_{0\to 1} = \frac{b}{1-a}\ \frac{H}{n-1},\ \ L^{(i)}_{1\to 0} = \frac{b}{3(1-a)}\ \frac{H-1}{n-1}, \quad e \quad L^{(i)}_{0\to 0} = 1 - L^{(i)}_{0\to 1};\ L^{(i)}_{1\to 1} = 1 - L^{(i)}_{1\to 0}, \tag{4.19}$$

e H é o peso de Hamming de $\vec{p}$ (ou seja, o número de coordenadas p_i que são diferentes de zero).

Prova: Observe inicialmente na tabela 4.3 que a matriz $M^{(i,j)}$ não permite transições que alterem ambos os bits i e j. Assim, M, e portanto L, só têm probabilidade de transição não-nula entre vetores $\vec{p}$ e $\vec{q}$ que diferem em no máximo uma única coordenada. Vamos verificar que a probabilidade $\mathbb{P}$ desta transição é idêntica quando usamos a eq. (4.17) ou a eq. (4.18).

Considere por exemplo dois vetores diferindo somente por $p_k = 0$ e $q_k = 1$. Para calcular a probabilidade desta transição no ponto de vista da eq. (4.17) podemos considerar que primeiro escolhemos um par (i, j) e em seguida usamos a versão preguiçosa de $M^{(i,j)}$. Da tabela 4.3 vê-se que, para a transição ocorrer, precisamos que o sítio k pertença ao par escolhido, e ainda que o outro sítio do par esteja com valor 1. A escolha de um par (i, j) com essas propriedades acontece com probabilidade $\frac{2H}{n(n-1)}$. Dado este caso, a probabilidade de ocorrer a transição é $b/2(1-a)$ (onde o fator meio aparece devido à versão preguiçosa). Assim probabilidade total da transição é $\mathbb{P} = \frac{1}{2}\left(\frac{2H}{n(n-1)}\frac{b}{1-a}\right)$. Por outro lado, para calcular $\mathbb{P}$ usando a eq. (4.18), temos simplesmente que multiplicar $L^{(i)}_{0\to 1}$ na primeira equação de (4.19), por um fator $\frac{1}{n}$ correspondendo à probabilidade da escolha da componente i. O mesmo resultado é obtido. Um raciocínio análogo vale também no caso $p_k = 1$ e $q_k = 0$. □

É importante notar que, apesar das probabilidades na eq. (4.19) serem iguais para todos os sítios, L não é uma soma de cadeias independentes, pois H é uma função global do vetor $\vec{p}$. Pode-se fazer uma analogia direta com modelos de física estatística de 'campo médio' para spins clássicos interagentes, no qual a probabilidade de um spin trocar de sinal é sensível à magnetização média do conjunto de todos os spins. Estudaremos mais explicitamente esta conexão no Capítulo 7.

Ainda assim, a simetria de eq. (4.18) com respeito a qualquer permutação de sítios, e em particular qualquer transposição, implica que $[L, T] = 0$. Usando ainda a eq. (4.17) podemos concluir também que:

$$[M, T] = 0 \tag{4.20}$$

Por fim, notamos que o estado estacionário π de L também é um estado estacionário de T, o que implica que as duas cadeias Q e L convergem para π:

Lema 4.5.2. *Partindo de qualquer elemento inicial $\vec{q}$ diferente de $\vec{0}$, as cadeias de Markov Q e L convergem para o mesmo estado estacionário*

$$\pi(\vec{q}) = \frac{1}{4^n - 1} 3^{H(\vec{q})};. \tag{4.21}$$

Prova O fato de π ser o único estado estacionário $\neq \vec{0}$ de Q pode ser obtido aplicando-se a redução definida no início desta Seção, e considerando que o estado estacionário de P é a distribuição uniforme. Como a distribuição uniforme é invariante por permutações, π também é um estado estacionário de T, e segue-se então imediatamente das eqs. (4.16) e (4.17) que o mesmo é verdade para M e L. Para mostrar que esta última cadeia não possui outros estado estacionários exceto π e $\vec{0}$, precisamos apenas verificar que a sua restrição a $\{0,1\}^n/\vec{0}$ é aperiódica e irredutível (veja Seção 2.5.2). O primeiro fato segue de L ser uma cadeia "preguiçosa", e o segundo de existir probabilidade não-nula de conversão entre qualquer par de vetores diferindo em um único sítio. □

Retornemos agora à eq. (4.16). Os resultados acima nos permitem uma visão intuitiva de como a cadeia Q se comportará. Em cada passo de Q, escolhe-se aplicar (com probabilidade a) uma transposição aleatória ou (com probabilidade $1-a$) a evolução 'campo médio' M. Como, porém, estas duas cadeias comutam e compartilham um mesmo estado estacionário, intuitivamente podemos esperar que Q se comporte de forma semelhante a uma versão 'preguiçosa' de M, com pesos a para a identidade e $(1-a)$ para M. Como, ao contrário da identidade, a cadeia T em geral reduz a distância com respeito ao estado estacionário, pode-se esperar ainda que o tempo de convergência de Q deve ser no máximo igual ao dessa cadeia preguiçosa. Veremos que isso realmente ocorre na Seção 5.2, onde mostraremos que podemos obter o tempo de mistura de Q a partir do de M (e este a partir do de L).

4.6 Cadeia do peso de Hamming

Na eq. (4.19) acima vemos que a probabilidade da cadeia L não ficar parada no *string* $\vec{q}$ é essencialmente proporcional ao peso de Hamming de $\vec{q}$. Uma estratégia interessante para analisar a evolução da cadeia Q, introduzida em [27], consiste em primeiro ver como se comporta a evolução destes pesos sob a ação do CQA.

É fácil ver que esta evolução pode ser descrita por uma nova cadeia de Markov Z, com espaço de estados $\Omega_\zeta = \{1, 2, \ldots, n\}$. Os elementos $Z(n,m)$ correspondem à probabilidade do número de Hamming mudar de n para m em cada passo do CQA. Pela tabela 4.3, vemos que um par de coordenadas $(q_i, q_j) = (0,0)$ nunca passa a $(1,1)$ e vice-versa. Por isso, em

cada passo o valor de H pode variar apenas de ± 1, ou então se manter constante. Este tipo de cadeia de Markov é conhecida na literatura como uma cadeia "nascimento-e-morte" (*birth and death chain*), pois pode ser usada para representar uma população na qual, em cada intervalo de tempo, uma pessoa pode nascer ou morrer. Também é conhecida às vezes como um 'passeio de bêbado' (*drunk man's walk*), pois representaria um bêbado que, ao caminhar, pode dar um passo para frente ou para trás, ou ficar parado.

No nosso caso, os únicos elementos não-nulos de Z são:

$$\begin{aligned} Z(H, H+1) &= b\, H\, (n-H) \binom{n}{2}^{-1}, \\ Z(H, H-1) &= \frac{b}{3}\, H\, (H-1) \binom{n}{2}^{-1}, \\ Z(H, H) &= 1 - Z(H, H-1) - Z(H, H+1) = 1 - \frac{2bH(3n-2H-1)}{3n(n-1)}, \end{aligned} \tag{4.22}$$

onde $H \in \Omega_\zeta$, e b é a constante introduzida na Seção 4.2, e é dependente da porta de dois q-bits escolhida.

Para obter estas expressões, basta reunir os elementos de transição entre conjuntos de *strings* com mesmo peso de Hamming. Da tabela 4.2 vemos que se uma *string* ($\vec{p}$) possui H não-zeros, a única transição para uma *string* com H maior é aquela na qual o par (i,j) é tal que $(p_i, p_j) = (0, 1)$ ou $(p_i, p_j) = (1, 0)$. A fração de pares dessa forma para o número total de pares é $H(n-H)\binom{n}{2}^{-1}$ e o elemento de matriz de Q para esse caso é b. Podemos obter o elemento de matriz de Z para a diminuição de H de maneira análoga. Vale notar que, como esta probabilidade depende apenas do peso de Hamming de $\vec{p}$, a cadeia Z é também uma redução da cadeia Q (esse tipo de redução é recorrente na literatura de cadeias de Markov [32] onde diz-se que Z é uma projeção de Q).

O estado estacionário de Z é trivialmente obtido a partir do estado estacionário da cadeia Q da qual é projetada. Basta somar o número de estados com o mesmo peso de Hamming H, a saber $\binom{n}{H}$, e considerar o estado estacionário de Q dado pela eq. (4.21),

$$\pi_\zeta(H) = \frac{3^H \binom{n}{H}}{4^n - 1}. \tag{4.23}$$

Agora que temos uma compreensão melhor da cadeia de segundos momentos, passaremos, a discutir como o seu tempo de convergência escalona com o número de q-bits. Vê-se neste sentido a importância da cadeia Z: a sua convergência é uma condição necessária (mas que pode não ser suficiente) para a convergência da cadeia Q como um todo. Em outras palavras, o tempo de mistura de Z fornece uma cota inferior para o de Q. Em [27] Harrow e Low obtiveram $t_{mixZ} = \Theta(n \log n)$ e $\Delta_Z = \Omega(1/n)$ para o caso $b = 3/5$, esses resultados se estendem trivialmente para $b \neq 3/5$. Isso acontece pois todas as probabilidades

de transição contém um fator constante($b \leq 2/3$) podemos interpretar um passeio aleatório nessa cadeia da seguinte maneira: o caminhante fica parado com probabilidade $1-b$ e anda com probabilidade b, quando anda o caminhante transita com as probabilidades da matriz original com $b = 2/3$. Ou seja, esperamos que os passeios com valores de b diferentes tenham tempo de mistura relacionado por uma constante, e teremos $t_{mix} = \Theta(n \log n)$ para qualquer $b \in (0, 2/3]$. Ver Apêndice A ou Seção 5.3 para uma discussão mais detalhada.

É interessante observar neste ponto que as probabilidades de transição Z dependem apenas do parâmetro b (mas não de a). Assim, as cadeias Z de CQAs baseados em ensembles com o mesmo valor de b terão rigorosamente o mesmo tempo de mistura, independente de seus coeficientes a. Por exemplo, a eq. (4.14) implica então que as cadeias Z provenientes de ensembles porta-fixa com portas $CNOT$ e XY convergirão no mesmo tempo.

Ainda, como as probabilidades de transição entre valores diferentes de H são proporcionais a b, pode-se suspeitar intuitivamente (e, como veremos, é verdade), que o tempo de mistura será inversamente proporcional a b. Assim, as cadeias Z de CQAs baseados no ensemble uniforme (com $b = 3/5$) convergirão ligeiramente mais lentamente que as de CQAs baseados em ensembles 'porta-fixa' com portas de Clifford ($b = 2/3$).

Capítulo 5

Convergência

Este capítulo contém os resultados de convergência do CQA a um 2-desenho aproximado. Mostraremos que o tempo necessário para a convergência escalona de forma polinomial no número n de q-bits. A potência exata deste polinômio dependerá da noção exata de 2-desenho aproximado que se usa (vide Seção 2.4).

Ao longo desse capítulo analisaremos a cadeia de Markov P obtida na Seção 3.3,[1] onde vimos que a evolução dos momentos de segunda ordem dos coeficientes de Pauli pode ser mapeada por essa cadeia. Na última Seção desse capítulo mostraremos como a convergência desta cadeia implica na convergência do CQA como um todo a um 2-desenho.

Antes de prosseguir com a nossa análise de convergência vamos expor os trabalhos anteriores de outros autores na mesma área traduzindo para definições comuns os diferentes resultados obtidos. Nesse ponto mostraremos que um resultado importante obtido em [27] se baseia em argumentos incorretos, entretanto poderemos recuperar o mesmo resultado com argumentos distintos. Esse será o objetivo das duas seções consecutivas. Na primeira delas mostraremos que podemos garantir a convergência da cadeia Q analisando a cadeia L. Em seguida, na Seção 5.3, analisaremos a cadeia L usando a técnica de acoplamento de cadeias de Markov. Assim recuperaremos o resultado questionado de [27] e ainda o estenderemos a qualquer ensemble localmente invariante.

Em seguida, na Seção 5.4, vamos iniciar uma análise mais detalhada do tempo de convergência em função do ensemble utilizado. Mesmo que não possamos obter o tempo exato de convergência de cada um poderemos ordenar parcialmente os diferentes ensembles em função do tempo de mistura da cadeia de Markov associada ao CQA.

Apresentaremos ainda a análise de uma cadeia semelhante à cadeia associada ao CQA que utilizará um método muito elegante de obtenção de tempo de mistura. Essa cadeia

[1] Frequentemente analisaremos diretamente a cadeia Q, que como discutido na Seção 4.5, apresenta exatamente o mesmo tempo de convergência.

de Markov será um passeio em grupo e suas propriedades especiais permitirão também que mostremos a existência de corte abrupto nesse caso. Especulamos que esses resultados possam ser utilizados para fornecer uma demonstração alternativa à demonstração de [27] do tempo de mistura da cadeia dos pesos de Hamming.

Finalmente responderemos à pergunta: Podemos garantir que CQA converge a um 2-desenho através da análise de P? Veremos que a resposta é positiva para ambas as definições de 2-desenho aproximado mas que o polinômio com o qual o tempo de convergência escalona é diferente.

5.1 Resumo e crítica de análises anteriores do problema

O problema da convergência de um CQA para um 2-desenho foi proposto (de forma não inteiramente explícita) por Oliveira et al em [18] , e depois analisado em detalhe por Harrow e Low em [27]. Nesta Seção resumimos rapidamente as estratégias utilizadas por estes autores, e em particular identificamos uma falha importante na demonstração apresentada pelos últimos. Outra contribuição importante para a análise dos CQAs foi dada por Znidaric que usa um método alternativo para obtenção do *gap* espectral da matriz de transição associada ao CQA. Por fim apresentamos o trabalho de Dankert et al, que propõe um algoritmo que não é um CQA para geração de um 2-desenho.

5.1.1 Oliveira et al.

O uso de ferramentas de cadeias de Markov para analisar modelos de CQAs foi proposto por Oliveira, Dahlsten e Plenio em [18]. Como já mencionamos na Seção 1.2, o objetivo principal destes autores era investigar o uso de CQAs para a geração de estados com emaranhamento bipartido 'típico', ou seja, próximo do máximo possível sob qualquer bipartição do sistema em dois sub-sistemas. A questão principal de interesse era se para chegar neste ponto seria necessário um CQA com profundidade exponencial em n (ou seja, um circuito 'não-operacional'), ou apenas polinomial em n.

Assumindo um ensemble tipo 'porta-fixa' com $C = CNOT$ [2], Oliveira et al reduziram o problema à determinação do tempo de mistura da cadeia de Markov reduzida que chamamos acima de Q. Utilizando a técnica de acoplamento de Bubley e Dyer[53], diferente da técnica de acoplamento que usaremos, e o argumento de comparação de Diaconis e Saloff Coste [54] foram capazes de encontrar uma cota superior de ordem $O(n^3)$ para

[2] com as ligeiras diferenças já comentadas na Seção 3.4

o tempo de mistura, provando assim que o emaranhamento bipartite 'típico' é de fato fisicamente atingível.

Como já descrevemos na Seção 1.1, o emaranhamento bipartite 'típico' pode ser visto na verdade como uma propriedade de um 2-desenho de estados. Em outras palavras, o que Oliveira et al fizeram essencialmente foi demonstrar que o seu CQA gera um 2-desenho de estados em tempo polinomial quando aplicado em um estado inicial fixo. Pelo menos aparentemente, porém, eles não chegaram a notar que de fato a mesma demonstração também prova a convergência do circuito como um todo a um 2-desenho de unitários aproximado, em tempo polinomial. Vale notar que a própria linguagem de 'k-desenhos' não chegou a ser utilizada explicitamente por estes autores, exceto em uma referência de passagem em uma nota de rodapé (p. 16 do preprint).

5.1.2 Harrow e Low

Em um longo artigo [27], Harrow e Low (HL) definiram explicitamente o objetivo de demonstrar a convergência de um CQA a um 2-desenho de unitários aproximado (e também, conjecturaram, a desenhos de ordem superior). Eles desenvolveram um formalismo sistemático para tanto, focando no caso onde as portas de 2 q-bits são extraídas do ensemble uniforme sobre $U(4)$. Nesta dissertação estivemos essencialmente seguindo a abordagem destes autores, embora generalizando-a para tratar de qualquer ensemble localmente invariante.

Para o caso geral, isto é, ensembles possivelmente diferentes do uniforme, Harrow e Low apresentaram uma demonstração de que um 2-desenho ε-aproximado é gerado depois de $O(n(n + \log \varepsilon^{-1}))$ passos do CQA para ambas as definições da Seção 2.4. No caso específico do ensemble uniforme, eles argumentaram ainda que o tempo de convergência para um 2-desenho **de canal** poderia ser melhorado para $O(n \log(n/\varepsilon))$.

Veremos adiante, porém, que os argumentos apresentados contêm um erro que os invalida. Mesmo assim, conseguiremos mostrar, por argumentos alternativos, que os resultados estão essencialmente corretos. Conseguiremos inclusive estender a validade da cota $O(n \log(n/\varepsilon))$ dada acima para todo ensemble localmente invariante.

Para entender a natureza do erro e motivar nossas estratégias para contorná-lo, precisamos explicar em maior detalhe a estratégia da prova proposta por HL. A demonstração se inicia com a dedução da cadeia de Markov para segundos momentos (a que chamamos de P acima) no caso do ensemble uniforme. HL procuram então obter uma cota para o tempo de mistura dessa cadeia, usando os seguintes passos:

Passo 1: convergência da cadeia Z

Primeiro, eles estudam a convergência da cadeia de peso de Hamming (cadeia Z) até o estado estacionário da eq. (4.23). A estratégia escolhida para esta análise é bastante elaborada, tomando boa parte do volume do artigo. A intuição básica por trás da prova é de que condições iniciais com valores baixos de H serão as que demorarão mais para evoluir (e atingir a convergência), pois para estes casos a eq. (4.19) indica que há uma probabilidade muito pequena de H aumentar. Como a definição de tempo de mistura (eq. (2.21)) envolve uma maximização sobre todas as condições iniciais, esses 'piores casos' acabam dominando o tempo de convergência da cadeia. Para lidar com isto, HL dividem a evolução da cadeia Z em três fases, de acordo com a progressão de H: uma primeira, em que, para as condições iniciais mais desfavoráveis, H ainda está em média bem abaixo do valor estacionário ($H \in [1, n^\delta]$, onde $0 < \delta < 1/2$); uma segunda, em que $H \in [n^\delta/2, \theta n]$; e uma terceira, em que $H \in [\theta n/2, n]$. Em cada fase, diferentes e variadas técnicas de análise de cadeias de Markov são utilizadas para encontrar cotas para o tempo máximo necessário para completá-las. O resultado final, após estes cálculos laboriosos, é que a cadeia Z converge após $O(n \log n)$ passos.

Passo 2: convergência da cadeia P. Erro no argumento de Harrow e Low

Assumindo que a cadeia Z já tenha convergido, HL tentam demonstrar que a cadeia P como um todo também convergirá (até uma distância ε) em no máximo mais $O(n \ln n/\varepsilon)$ passos. É esta parte do argumento, descrita no Corolário 5.1[3] de [27], que **está incorreta**. Como este é um ponto importante para justificar o trabalho desenvolvido nesta dissertação, convém explicá-lo em algum detalhe.

O argumento proposto por HL utiliza um conceito conhecido na literatura de cadeias de Markov como um 'tempo estacionário forte' (*strong stationary time*, ou SST). Grosso modo, um SST existe para um passeio aleatório quando após cada passo é possível decidir, dado o histórico do passeio até aquele momento, se a cadeia de Markov correspondente já se encontra em seu estado estacionário. Uma definição mais precisa pode ser encontrada em [55, 32], mas o conceito pode ser ilustrado intuitivamente por um exemplo:

Considere um passeio aleatório no conjunto $\mathbb{Z}_2^n$ dos vetores binários de n coordenadas, o qual pode ser identificado com os vértices de um hipercubo n-dimensional. A regra de movimento do passeio é simples: em cada passo, escolhe-se uma das n coordenadas, e troca-se o seu valor para 0 ou 1 com 50% de probabilidade cada. Geometricamente, isto equivale a escolher uma das n arestas ligadas ao vértice em que se está, e mover-se para o vértice adjacente com 50% de chance. É fácil ver que o estado estacionário desta

[3] Ou Corolário 5.6 na versão Arxiv.

cadeia de Markov é a distribuição uniforme, em que todos os 2^n vértices do hipercubo são equiprováveis.

Um SST para este passeio ocorre assim que cada coordenada tiver sido selecionada pelo menos uma vez. Para ver por que, é só considerar que, a cada vez que selecionamos uma coordenada, ela é randomizada, independente do ponto exato do cubo em que se esteja. Assim, após todas as coordenadas terem sido selecionadas, todos as 2^n posições possíveis no hipercubo são igualmente prováveis, o que é justamente o estado estacionário.

A questão então é: quantos passos é preciso realizar para que a probabilidade de este evento ter ocorrido ficar maior do que, digamos, $1-\varepsilon$? Este é, na verdade, um problema clássico em cadeias de Markov, conhecido como *coupon collecting*, em referência ao fato de que é o mesmo problema enfrentado por alguém que tenta completar um album de n figurinhas, comprando figurinhas aleatórias. Pode-se mostrar que a probabilidade do álbum não ter sido completado ainda após a compra de $n(\ln n + c)$ figurinhas é $< e^{-c}$ [32].

Finalmente, cotas para a probabilidade de ocorrência de um SST podem ser convertidas em cotas para o tempo de mistura da cadeia, veja a Proposição 6.10 de [32]. Por exemplo, para o passeio no hipercubo / *coupon collecting*, a probabilidade acima implica que $t_{mix}(\varepsilon) < n\ln(n/\varepsilon)$.

Após essa breve introdução a SSTs, retornemos agora ao argumento de Harrow e Low. Basicamente, eles procuram adaptar o argumento de *coupon collecting* para o caso do passeio com matriz P. Um fato que complica essa adaptação é o seguinte: ao contrário do passeio no hipercubo, onde em cada passo no máximo uma coordenada p_i de $\vec{p}$ é modificada, na cadeia P em cada passo selecionamos duas coordenadas p_i, p_j, ambas as quais podem vir a mudar. Porém, para esta mudança ter uma chance de ocorrer, é necessário que pelo menos uma dessas coordenadas seja diferente de zero (vide Teorema 4.2.1). HL definem então que uma coordenada p_i conta como 'coletada' para efeito de *coupon collecting* apenas quando ela faz parte de um par selecionado com $(p_i, p_j) \neq (0,0)$. Eles ainda mostram cuidadosamente que a probabilidade de que tais pares realmente ocorram é alta quando a cadeia Z já tiver convergido[4]. Com isto, são capazes de mostrar que, assim como no passeio do hipercubo, $O(n\ln n)$ passos são suficientes para que a probabilidade de que todas as n coordenadas tenham sido 'coletadas' no sentido acima fique próxima de 1.

Até este ponto o argumento está perfeito. O problema aparece, no entanto, quando HL tentam replicar a propriedade essencial do passeio do hipercubo, que é o fato de que o tempo de *coupon collecting* é um SST. Citando diretamente:

> Once each site of the full chain has been hit, meaning it is chosen and paired

[4]Isto ocorre pois a distribuição estacionária da eq. (4.23) para o peso de Hamming é fortemente concentrada ao redor do valor $H = 3n/4$, o que implica que podemos garantir que haverá muitos sítios diferentes de zero na cadeia.

> with another site so not both equal zero, the chain has mixed. This is because, after each site has been hit, the probability distribution over the states is uniform.

Esta frase está incorreta. Recorde da Seção 4.2 que, após um par de coordenadas (p_i, p_j) ser 'coletado', ele tem probabilidade uniforme de assumir qualquer valor $\in \{0, X, Y, Z\}^2$ que seja *diferente de (0,0)*. A ausência desta última possibilidade faz com que o estado geral da cadeia *não* fique uniformemente distribuído após todos os sítios terem sido 'coletados'. Em outras palavras, para a cadeia P o tempo de 'coupon collecting' *não* é um SST, e o resto do argumento não se aplica.

Podemos ilustrar estes fatos com um exemplo, usando uma cadeia de 3 sítios apenas. Para simplificar a notação, vamos usar '1' para significar "X,Y ou Z, com iguais probabilidades para cada um".

Começamos o passeio no estado (0 X X). Suponha que, no primeiro passo, escolhemos os dois primeiros sítios. Como estes não são ambos iguais a zero, ambos contam como 'coletados'. Pelo Teorema 4.2.1, o novo estado da cadeia será uma das possibilidades de forma (0 1 X), (1 0 X) ou (1 1 X). Suponha ainda que, no passo seguinte, escolhemos o 2o e 3o sítios. Em todos os três casos, essas coordenadas não são ambas iguais a zero, de modo que o 3o sítio conta como 'coletado', completando assim o 'álbum'. O novo estado será um dos seguintes: (0 1 1), (0 1 0), (0 0 1), (1 0 1), (1 1 0) ou (1 1 1), sendo que os três últimos podem surgir tanto do segundo quanto do terceiro caso anterior. Vê-se assim que neste instante não temos uma distribuição uniforme sobre todas as sequências $\neq (000)$: por um lado, a opção (1 0 0) não aparece; por outro, mesmo entre as opções possíveis a probabilidade não é uniforme. Por exemplo, sequências da forma (0 1 1) não tem a mesma probabilidade que as de forma (1 1 0). Este instante não é, portanto, um SST.

Passo 3: extensão a outros ensembles via *gap* espectral

Após concluirem sua demonstração para o caso do ensemble uniforme, HL estendem sua validade para outros ensembles. Para fazer isso, primeiro usam relações gerais entre tempo de mistura e *gap* espectral (eq. (2.28)) para obter uma cota para esta última quantidade, a saber $\Delta = O(1/n)$. Note que esta cota depende do resultado anterior para tempos de mistura, e portanto também não se sustenta dado o erro apontado. Dada porém uma cota para o *gap* do ensemble uniforme, HL usam a chamada técnica de comparação de cadeias de Markov [54], a qual permite obter cotas para o *gap* de uma cadeia a partir do *gap* de uma outra cadeia semelhante, obtendo finalmente que a mesma ordem $\Delta = O(1/n)$ também vale para o *gap* de outros ensembles. Vale notar que, para valer, este resultado presume a *existência* de uma cadeia de Markov para esses outros ensembles, algo que HL

não parecem demonstrar explicitamente (ao contrário do que fizemos aqui no Teorema 3.3.1). Retornando então para tempos de mistura usando a eq. (2.27), obtêm finalmente uma cota $t_{mix} = O(n(n + \log_2(1/\varepsilon)))$ para estes outros ensembles.

Passo 4: passando de cadeias de Markov para 2-desenhos

Por último, HL discutem como se pode usar uma cota para a convergência da cadeia P (seja em termos de tempos de mistura, seja em termos de *gap*), para obter cotas para a convergência a um 2-desenho aproximado segundo os dois sentidos apresentados na Seção 2.4. Discutiremos em mais detalhe como funciona esta passagem final na Seção 5.7 adiante.

5.1.3 Znidaric

Em [29], Znidaric mostra de forma interessante como mapear a evolução Markoviana do CQA em um Hamiltoniano de cadeia de spin. Nesse mapeamento, o *gap* da cadeia P se transforma na diferença entre a maior e segunda maior autoenergias da cadeia de spin. Usando então métodos conhecidos de física de muitos corpos, Znidaric obtém exatamente o gap espectral para os casos do ensemble uniforme $U(4)$ e para os ensembles "portas-fixa" com portas XY e $CNOT$, sempre obtendo $\Delta = \Theta(1/n)$. Embora Znidaric não mencione este fato, o *gap* obtido demonstra através da eq. (2.26) e do Teorema 5.7.2 que nesses casos o CQA de fato converge para um 2-desenho ε-aproximado em $O(n(n + \log \varepsilon^{-1}))$ passos. Porém, como o método não fornece informação direta sobre o tempo de mistura, então não é suficiente para comprovar o tempo de convergência a um 2-desenho de canal em tempo $O(n \log(n/\varepsilon))$ alegado por HL.

5.1.4 Dankert et al

Em [30, 56] esses autores propuseram um algoritmo que não é um CQA como definido na Seção 1.2, mas que também gera um 2-desenho aproximado.

As técnicas de análise usadas são bastantes distintas das que usaremos, entretanto podemos comparar a profundidade do circuito exigido. Em [30] demonstra-se que o algoritmo citado necessita de $O(n \log \varepsilon^{-1})$ portas de 2 q-bits para gerar um 2-desenho **de canal** ε-aproximado, o que se compara favoravelmente com a cota de $O(n \log(n\varepsilon^{-1}))$ que obtemos para um CQA. Para a outra definição de convergência, Harrow e Low afirmam que são necessários $O(n(n + \log \varepsilon^{-1}))$ portas de 2 q-bits, ou seja a mesma cota do CQA que analisamos.

Vale notar que a construção destes autores é específica para desenhos de ordem 2.

5.2 Preparando argumentos alternativos

Apesar do argumento de Harrow e Low estar incorreto, mostramos nas seções que se seguem que as suas conclusões são de fato válidas. Apresentamos um argumento distinto na Seção 5.3, onde analisamos a cadeia P usando uma técnica de acoplamento. Obtemos um tempo de mistura de ordem $O(n\log(n/\varepsilon))$, confirmando o resultado alegado por HL. Nossa análise se aplica apenas na situação do 'Passo 2' acima, ou seja, apenas quando a cadeia Z já estiver suficientemente próxima de convergir. Por este motivo, depende de uma prova independente da convergência da cadeia Z no 'Passo 1'. Por outro lado, ela é válida para todo ensemble localmente invariante.

Já na Seção 5.5 utilizamos uma outra técnica, baseada na teoria de passeios aleatórios em grupos [36], para estudar uma cadeia ligeiramente diferente da cadeia Q. Obtemos novamente um tempo de mistura de ordem $O(n\log n)$. Apesar de não conseguirmos estender este método para dar informações diretas sobre a cadeia Q, especulamos que talvez seja possível obter uma prova parcialmente independente do tempo de mistura da cadeia Z.

Para podermos realizar estas análises, precisamos primeiro simplificar o problema. Mostramos agora que, para entender a convergência das cadeias P (ou Q), basta estudar a da cadeia 'tipo Campo Médio' L.

5.2.1 A análise da cadeia L garante a convergência do CQA

Vamos mostrar que o tempo de mistura de L fornece uma cota superior para o tempo de mistura de Q.

A nossa estratégia será dividida em dois passos: primeiro vamos mostrar que a convergência de L implica a de M e depois que a de M, por sua vez, implica a de Q. Como L é uma versão "preguiçosa" de M, para o primeiro passo precisamos entender como o tempo de mistura de cadeias "preguiçosas" se comporta em relação ao tempo de mistura da cadeia original.

Intuivamente, os passeios aleatórios gerados por L e M são essencialmente os mesmos, exceto pelo fato de que aquele baseado em L tem metade da probabilidade de se mover em cada passo. Assim, esperamos que as escalas temporais relevantes do passeio L aumentem por um fator 2. Estranhamente, não encontramos uma demonstração rigorosa deste fato em livros-texto de cadeias de Markov.

No Apêndice A estabelecemos condições para que o tempo de mistura de uma cadeia "preguiçosa" seja, o dobro da cadeia original. Uma das condições é que o *gap* espectral venha de um autovalor positivo da matriz de transição. Os lemas a seguir estabelecem que a cadeia M tem essa propriedade e por isso o resultado apresentado no apêndice é válido.

Lema 5.2.1. *A matriz de transição M não possui nenhum autovalor menor do que* $-2/3 - 1/3(n-1)$.

Prova: Primeiro vamos demonstrar que os elementos da diagonal de M são da forma $\frac{1}{6}(1-\frac{1}{n-1}) + \delta_i$ com $\delta_i \geq 0 \; \forall \, i$.

$$M = \begin{pmatrix} \frac{1}{6}(1-\frac{1}{n-1}) + \delta_1 & & \cdots \\ & \frac{1}{6}(1-\frac{1}{n-1}) + \delta_2 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

Esses elementos correspondem às probabilidades do passeio aleatório ficar parado na *string* $\vec{p}$ em cada passo de M. Vamos agora calcular uma cota inferior para essa probabilidade lembrando que a evolução de M pode ser vista como a escolha de um par que evoluirá com $M^{(i,j)}$ dada pela tabela 4.3. Repare que a probabilidade de escolher um par da forma $(0,0)$ ou $(1,1)$ é no mínimo[5] $(1/2)(1-1/(n-1))$ para qualquer *string*. Nesse caso, dada pela tabela 4.3, a probabilidade de não haver mudança maior ou igual a $1 - 2b/3(1-a)$. Mas, como mostramos na Seção 4.2, $a+b \leq 1$ e portanto $1-2b/3(1-a) \geq 1/3$. Com isso, a probabilidade de não haver mudança numa *string* $\vec{p}$ sob ação de uma passo de M, que é igual ao elemento diagonal $M(\vec{p},\vec{p})$, é maior do que $\alpha \equiv (1/6)(1-1/(n-1))$.

Para ver porque isso implica no enunciado do Lema note que podemos pensar em M como a combinação convexa, com coeficientes $1-\alpha$ e α, de uma certa cadeia de Markov, e da identidade. Ou seja,

$$M = \alpha \mathbb{I} + (1-\alpha)B, \quad \text{onde } B \equiv (M - \alpha\mathbb{I})/(1-\alpha) \; .$$

Repare que B é uma matriz estocástica, que terá autovalores entre -1 e 1, pois tem todos elementos não-negativos e suas linhas somam 1. Com isso o menor autovalor de M será $\geq \alpha + (1-\alpha)(-1) = -1 + (2/6)(1-1/(n-1))$. □

Lema 5.2.2. *Para n suficientemente grande, o* gap *de M vem de um autovalor positivo.*

Prova: Como, em cada passo da cadeia M, o número de Hamming só aumenta de no máximo 1, e seu estado estacionário é de ordem n (vide eq. (4.23)), então no pior caso precisamos esperar $\Omega(n)$ passos até a convergência. Em particular, isso implica que o *gap* espectral de M satisfaz $\Delta = O(1/n)$. Por outro lado acabamos de mostrar que a região $[-1, -2/3 - 1/3(n-1)]$ nunca contém um autovalor de M. Com isso fica claro que *gap* só pode vir de autovalores positivos. □

Podemos então demonstrar o resultado que nos interessa

[5] A probabilidade de escolher um par da forma $(0,0)$ ou $(1,1)$ numa *string* de peso de Hamming H é $[H(H-1) + (n-H)(n-H-1)]/[n(n-1)]$ que tem o valor mínimo $(1/2)(1-1/(n-1))$.

Teorema 5.2.1. *Podemos obter uma cota superior para o tempo de mistura de Q a partir do tempo de mistura de L : $t_{mixQ} \leq \frac{1}{2(1-a)} t_{mixL}$.*

Prova: Definimos na eq. (4.17) que $L = \frac{1}{2}\mathbb{I} + \frac{1}{2}M$, e pelo lema 5.2.2 temos que o *gap* de M vem de um autovalor positivo. Podemos então usar o resultado do Apêndice A e obter $t_{mixL} = 2t_{mixM}$. Ainda, como da eq. (4.16) $Q = aT + (1-a)M$ e T, M comutam, Q tem justamente a forma discutida no fim do Apêndice A e, portanto, $t_{mixQ} \leq \frac{1}{(1-a)} t_{mixM}$ o que completa a demonstração. □

5.3 Análise por acoplamento

Nessa Seção vamos usar a técnica de acoplamento (*coupling*), e obteremos o tempo de mistura $O(n \log n)$.

A técnica de acoplamento é extremamente útil na teoria de cadeias de Markov, tanto como uma ferramenta teórica, como uma ferramenta prática na prova de tempo de mistura. Indicamos o Capítulo 5 de [32] para uma introdução no assunto, o qual contém todos os resultados usados nesta Seção.

Definição e propriedades do Acoplamento

Intuitivamente, a idéia de um acoplamento de cadeias de Markov é a seguinte: dada a cadeia L no espaço de estados Ω, define-se uma nova cadeia A cujo espaço de estados é o produto $\Omega \times \Omega$, de tal forma que as probabilidades marginais de A em cada cópia de Ω evoluem de acordo com a cadeia L. Em outras palavras, a cadeia A representa duas cópias de L, as quais podem porém em geral evoluir de forma correlacionada. Formalmente, em cada passo t temos uma distribuição de probabilidade $\nu^{(t)}(x, y)$ em $\Omega \times \Omega$, com marginais $\nu_1^{(t)}(x) = \sum_{y \in \Omega} \nu^{(t)}(x, y)$ e $\nu_2^{(t)}(y) = \sum_{x \in \Omega} \nu^{(t)}(x, y)$, tais que

$$\nu^{(t+1)} = \nu^{(t)} A; \quad \nu_1^{(t+1)} = \nu_1^{(t)} L; \quad \nu_2^{(t+1)} = \nu_2^{(t)} L \tag{5.1}$$

É possível em geral escolher uma cadeia acoplada $A \neq L \otimes L$ mas ainda obedecendo à eq. (5.1). Ou seja, em geral teremos $\nu(x, y) \neq \nu_1(x)\nu_2(y)$. Nesse caso há correlações entre as duas cadeias marginais. Pode-se ainda escolher acoplamentos com a seguinte propriedade: se X_t, Y_t são os passeios aleatórios percorridos em cada espaço local, então X_t, Y_t permanecem juntos em todos os tempos posteriores à sua primeira visita simultânea a um mesmo sítio. Em outras palavras, $X_s = Y_s \Rightarrow X_t = Y_t, t \geq s$. O primeiro instante em que este encontro acontece é uma variável aleatória, conhecida como *tempo de acoplamento*.

O seu valor em geral depende das condições iniciais: escreveremos $\tau_{\text{acopla}}(x, y)$ para denotar o tempo de acoplamento partindo da condição inicial $X_0 = x$ e $Y_0 = y$.

O nosso interesse em estudar um acoplamento de L com essas características é que essa é uma maneira de obter cotas para o tempo de mistura. Pode-se mostrar (vide por exemplo a Seção (5.1) de [32]) que a distância à estacionaridade $d(t)$ definida na eq. (2.22) satisfaz:

$$d(t) \leq \max_{x,y \in \Omega} \mathbb{P}(\tau_{\text{acopla}}(x, y) \geq t) \, . \tag{5.2}$$

Escolhendo de forma conveniente a cadeia A pode-se obter cotas para a probabilidade de acoplamento da eq. 5.2. Se, por exemplo, esta cota for inferior a $1/4$ para um certo tempo t, então pela Definição 2.5.3 este instante será uma cota superior para o tempo de mistura de L. Em particular, é suficiente encontrar cotas para o valor médio $\langle \tau_{\text{acopla}} \rangle$ e em seguida usar a chamada desigualdade de Markov:

$$\mathbb{P}(\tau_{\text{acopla}} > a) \leq \frac{\langle \tau_{\text{acopla}} \rangle}{a} \, . \tag{5.3}$$

Acoplamento de L

Como vimos na Seção 4.5, a ação da cadeia L pode ser interpretada como: escolha uma coordenada e atualize-a segundo as probabilidades da eq. (4.19). O acoplamento que usaremos é tal que ambas as cadeias locais escolhem a mesma coordenada i. Se o valor dessa coordenada é diferente entre as duas variáveis aleatórias, então evoluimos cada cadeia de forma independente. Isto faz com que, com boa chance, essa coordenada passe a ter o mesmo valor nos dois estados depois da iteração. Se as coordenadas já são iguais usamos a liberdade de escolha da cadeia A para maximizar a chance de ambas permanecerem na mesma *string* ou que ambas mudem, minimizando a probabilidade de que elas se tornem diferentes.

Sejam então os estados iniciais $\vec{p}^{(1)}$ e $\vec{p}^{(2)}$ com peso de Hamming H_1 e H_2 respectivamente. Seja D o número de componentes diferentes. Definimos então a evolução acoplada da coordenada i escolhida aleatoriamente de acordo com a regra: se $p_i^{(1)} \neq p_i^{(2)}$, cada coordenada evolui independentemente de acordo com as probabilidades da eq. (4.19). Caso $p_i^{(1)} = p_i^{(2)}$ a evolução é correlacionada, com as probabilidades dadas na tabela 5.1.

É fácil checar que as probabilidades marginais de cada coordenada evoluem de acordo com a cadeia L. Por exemplo, dado que $p_i^{(1)} = p_i^{(2)}$, então $\mathbb{P}_1(0 \to 1) = \mathbb{P}_1(0 \to 1|00) = \mathbb{P}(00 \to 11) + \mathbb{P}(00 \to 10) = \min\{L_{01}^{(1)}, L_{01}^{(2)}\} + \max\{0, L_{01}^{(1)} - L_{01}^{(2)}\} = L_{01}^{(1)}$.

Buscaremos uma cota para o tempo médio que as duas *strings* levam para se acoplar. Como vimos, isso será suficiente para obter uma cota para o tempo de mistura. Mostraremos

$\mathbb{P}(00 \to 00) \equiv 1 - \max\{L_{01}^{(1)}, L_{01}^{(2)}\}$	$\mathbb{P}(11 \to 00) \equiv \min\{L_{10}^{(1)}, L_{10}^{(2)}\}$
$\mathbb{P}(00 \to 11) \equiv \min\{L_{01}^{(1)}, L_{01}^{(2)}\}$	$\mathbb{P}(11 \to 11) \equiv 1 - \max\{L_{10}^{(1)}, L_{10}^{(2)}\}$
$\mathbb{P}(00 \to 10) \equiv \max\{0, L_{01}^{(1)} - L_{01}^{(2)}\}$	$\mathbb{P}(11 \to 10) \equiv \max\{0, L_{10}^{(1)} - L_{10}^{(2)}\}$
$\mathbb{P}(00 \to 01) \equiv \max\{0, L_{01}^{(2)} - L_{01}^{(1)}\}$	$\mathbb{P}(11 \to 01) \equiv \max\{0, L_{10}^{(2)} - L_{10}^{(1)}\}$

Tabela 5.1: Definição do acoplamento.

a seguir que, no caso de duas *strings* que já tiverem o peso de Hamming estacionário, este acoplamento ocorrerá após $O(n \log n)$ passos.

Primeiro calculamos a probabilidade $\mathbb{P}(D \to D+1) \equiv \mathbb{P}_+$ de que num passo da dinâmica acoplada haja aumento da distância (D) entre as *strings* acopladas. Obviamente isso só pode ocorrer no caso em que escolhemos um par no qual inicialmente $p_i^{(1)} = p_i^{(2)}$. Já que a escolha de coordenada é equiprovável a chance de isso ocorrer é $(n-D)/n$. Definamos γ como a probabilidade de que ambas as entradas sejam iguais a 1, condicionada ao fato de serem iguais. Chamemos ainda $b' \equiv b/(1-a)$. Agora com a definição de acoplamento de tabela 5.1 e a eq. (4.19), temos para $\mathbb{P}_+$:

$$\begin{aligned} \mathbb{P}_+ &= \frac{n-D}{n}\left\{\gamma\frac{b'}{3}\frac{|H_1-H_2|}{n-1} + (1-\gamma)b'\frac{|H_1-H_2|}{n-1}\right\} \\ &= \frac{n-D}{n}\,\frac{b'|H_1-H_2|}{n-1}\left(1-\frac{2\gamma}{3}\right) \ . \end{aligned} \tag{5.4}$$

Por outro lado, D só pode diminuir quando inicialmente $(p_i^{(1)}, p_i^{(2)}) = (0,1)$ ou $(1,0)$, em cujo caso a evolução é independente e dada por L_i. A chance de escolher um par dessa forma é D/n. Para o caso $(0,1)$ temos que a distância D diminui com probabilidade

$$\begin{aligned} \mathbb{P}_- &= \frac{D}{n}\left\{\frac{b'H_2}{n-1}\left(1-b'\frac{H_1-1}{3(n-1)}\right) + \left(1-\frac{b'H_2}{n-1}\right)b'\frac{H_1-1}{3(n-1)}\right\} \\ &= \frac{D}{n}\frac{b'}{3(n-1)}\left\{3H_2 + H_1 - 1 - \frac{2b'}{(n-1)}\left(H_2(H_1-1)\right)\right\} \\ &= \frac{D}{n}\frac{b'}{3(n-1)}\left\{3H_2 + (H_1-1)\left(1-2b'\frac{H_2}{n-1}\right)\right\} \ . \end{aligned} \tag{5.5}$$

Para o caso $(1,0)$ basta fazer $H_1 \to H_2$ e $H_2 \to H_1$, por isso vamos usar a notação H_{12} para representar H_1 ou H_2 e englobar os dois casos em uma equação.

$$\mathbb{P}_- = \frac{D}{n}\frac{b'}{3(n-1)}\left\{3H_{12} + (H_{21}-1)\left(1 - 2b'\frac{H_{12}}{n-1}\right)\right\} \ . \tag{5.6}$$

Considere agora o tempo esperado, entre caminhos do acoplamento, necessário para que D diminua por uma unidade, que denotamos por $E(D \to D-1) \equiv E_-$. Como em cada passo a distância só pode mudar por ± 1 ou 0 (permanecer fixa), podemos escrever uma expressão recursiva para esta quantidade:

$$\begin{aligned} E_- &= \mathbb{P}_-(1) + \mathbb{P}_+(1 + E(D+1 \to D) + E_-) + (1 - \mathbb{P}_- - \mathbb{P}_+)(1 + E_-) \\ &\leq \mathbb{P}_-(1) + \mathbb{P}_+(1 + 2E_-) + (1 - \mathbb{P}_- - \mathbb{P}_+)(1 + E_-) \qquad (5.7) \\ &= E_- - E_-(\mathbb{P}_- - \mathbb{P}_+) + 1 \ . \qquad (5.8) \end{aligned}$$

Na passagem (5.7), usamos que $E(D+1 \to D) \leq E(D \to D-1)$, que é consequência do fato de que quanto menor a diferença D mais lento é o acoplamento. Ou seja, se D diminuir, em cada passo da dinâmica acoplada a chance de escolher uma coordenada com valores diferentes e eles se tornarem iguais é menor, por outro lado, a chance de escolher coordenadas iguais e estas se tornarem diferentes cresce. Agora manipulando ambos os lados da eq. (5.8), temos

$$E_- \leq 1/(\mathbb{P}_- - \mathbb{P}_+) \ . \tag{5.9}$$

Mas da eq. (5.6) e da eq. (5.4) temos,

$$\begin{aligned} &\mathbb{P}_- - \mathbb{P}_+ \\ &= \frac{D}{n}\frac{b'}{3(n-1)}\left\{3H_{12} + (H_{21}-1)\left(1 - 2b'\frac{H_{12}}{n-1}\right)\right\} - \frac{n-D}{n}\,\frac{b'|H_1 - H_2|}{n-1}\left(1 - \frac{2\gamma}{3}\right) \ . \end{aligned} \tag{5.10}$$

Para prosseguir na nossa análise vamos agora considerar que os *strings* iniciais já eram ambos tais que sua cadeia Z já havia convergido. Como discutido no passo 1 de HL da Seção 5.1 isso pode levar um tempo $O(n \log n)$, mas como veremos adiante, o tempo de acoplamento também é $O(n \log n)$ e portanto essa suposição não piora a ordem da convergência. Assim podemos usar que com alta probabilidade, $\gamma \geq 3/4 - \alpha$, $H_{12}/n \geq 3/4 - \alpha$ [6]. Isso acontece pois a distribuição estacionária do peso de Hamming é concentrada

[6] α será uma constante fixa para todo n, por exemplo $\alpha = 1/10$.

em $3n/4$ com dispersão $\Theta(\sqrt{n})$. Assim a probabilidade de que em todos os passos t do acoplamento a desigualdade se verifique é maior do que $1 - te^{-O(\alpha^2 n^2)} - \varepsilon'$, onde ε' é a distância da cadeia do peso de Hamming à estacionariedade e onde usamos a desigualdade de Hoeffding [57]. Além disso $b' \leq 1$ e $|H_1 - H_2| \leq D$ e temos da equação acima,

$$\begin{aligned} \mathbb{P}_- - \mathbb{P}_+ &\leq \frac{b'D}{3n}\left(3(3/4-\alpha) + (3/4-\alpha)(1-2(3/4-\alpha)) - 3 + 2(3/4-\alpha)\right) \\ &= b'D\,\Omega(1/n)\ , \end{aligned} \tag{5.11}$$

e da eq. (5.9),

$$E_- = \frac{1}{b'D}O(n). \tag{5.12}$$

Para obter o tempo esperado total até o acoplamento vamos somar os tempos médios para cada queda na distância D, desde seu valor inicial até zero. Note que a maximização entre as condições iniciais equivale aqui à maximização da distância inicial. E portanto:

$$\max_{x,y\in\Omega} \langle \tau_{\text{acopla}} \rangle \leq \sum_{D=n}^{0} E_-(D),$$

agora usando o fato de que $|\sum_{D=0}^{n} 1/D - \log n| \leq 1$, temos:

$$E(\tau_{\text{acopla}}) = \frac{1}{b'}O(n\log n)\ . \tag{5.13}$$

Lema 5.3.1. *O tempo médio de acoplamento, maximizado entre todas as condições iniciais em que a cadeia Z já convergiu é $E(\tau_{acopla}) = \frac{1}{b'}O(n\log n)$.*

Aqui termina nossa análise por acoplamento. Agora que obtivemos o tempo esperado para o acoplamento vamos aplicar a desigualdade de Markov (eq. (5.3)) e ainda a eq. (5.2) para obter o tempo de mistura de L.

Lema 5.3.2. *O tempo de mistura da cadeia L é $O(n\log n)$.*

Prova: Com a desigualdade de Markov temos para a eq. (5.13)

$$\mathbb{P}\left(t_{\text{acopla}} > 4E(\tau_{\text{acopla}})\right) \leq \frac{1}{4}\ ,$$

e com a eq. (5.2) conectamos esse resultado ao tempo de mistura, fornecendo:

$$t_{mixL} \leq \frac{1}{b'}O(n\log n) + t_{mixZ}\ . \tag{5.14}$$

Onde incluímos o tempo de mistura da cadeia Z, já que usamos essa condição para a obtenção do lema 5.3.1. Como já mencionamos acima o tempo de mistura de Z é $O(n \log n)$ e portanto o tempo de mistura de L será da mesma ordem. □

Finalmente, juntando o Lema 5.3.2 com o Teorema 5.2.1 podemos concluir que

Teorema 5.3.1. *O tempo de convergência da cadeia Q é $\Theta(n \log n)$.*

Prova : Já provamos acima que $t_{mixQ} = O(n \log n)$, resta mostrar que ele é também $t_{mixQ} = \Omega(n \log n)$, ou seja que o tempo de mistura de Q também é inferiormente limitado por $n \log n$. Vamos recorrer novamente a um argumento de *coupon collecting* como o visto Seção 5.1.2. Pode-se mostrar que a probabilidade do álbum já ter sido completado após a compra de $\frac{n}{2}(\ln n - c)$ figurinhas vai a zero [32] à medida em que c cresce. Ou seja com pouco menos do que $n \log n$ de passos da cadeia Q alguma coordena não terá sido escolhida para interagir, se essa for a única coordenada igual a 1 então a cadeia não terá dado nenhum passo. Mas o tempo de mistura é justamente maximizado entre todas as condições iniciais possíveis, portanto o tempo de mistura de Q não pode ser menor do que $n \log n$. □

5.4 Como os parâmetros a e b afetam a convergência?

Até agora estudamos a convergência do CQA em termos da ordem de crescimento do número de passos necessários em função do tamanho da entrada (número de q-bits). Como vimos na Seção 5.3, qualquer ensemble localmente invariante (μ) tem o mesmo desempenho em termos da ordem de crescimento. Entretanto, é de se esperar que o número exato de passos necessários para a convergência deve depender do ensemble usado. Em particular um ensemble com portas pouco emaranhantes ($b \simeq 0$) deve ser mais lento do que um ensemble altamente emaranhante.

Nessa Seção vamos comparar o desempenho de diferentes ensembles e tentar introduzir um "hierarquia" de desempenho. Vimos na Seção 4.1 que a cadeia de Markov P associada a um ensemble qualquer μ tem sempre a mesma forma e depende dos parâmetros a e b dados pelas eqs. (4.8) e (4.9). Ou seja, se pudermos mostrar como o tempo de convergência de P depende de a e b estaremos comparando o desempenho de diferentes ensembles no que diz respeito à geração de um 2-desenho aproximado.

Nessa comparação, mais uma vez, a decomposição introduzida na Seção 4.5 será útil. Nas seções anteriores, em especial no Teorema 5.2.1, vimos que é possível obter cotas para o tempo de mistura de Q só analisando L. Por isso vamos começar analisando o efeito da mudança dos parâmetros em L e depois veremos como essa análise se estende para Q.

Na cadeia L, cada probabilidade de transição não-trivial contém um fator b. Com isso é fácil ver que qualquer valor de b menor do que o máximo ($b = 2/3$, veja Lema 4.2.1) simplesmente diminui a chance das transições e torna a convergência mais lenta. Assim o tempo de mistura é afetado por b simplesmente como

$$t_{mix(L|b\leq 2/3,a)} = \frac{2}{3b} t_{mix(L|b=2/3,a)} \ . \tag{5.15}$$

Para justificar esse resultado note que para uma cadeia em que todas as probabilidades de transição contém um fator constante($b \leq 2/3$) podemos interpretar um passeio aleatório nessa cadeia da seguinte maneira: o caminhante fica parado com probabilidade $1-b$ e anda com probabilidade b, quando anda o caminhante transita com as probabilidades da matriz original com $b = 2/3$. Ou seja, é como se o passeio fosse uma versão preguiçosa de uma outra matriz que não depende de b. Assim podemos usar o Apêndice A para justificar a eq. (5.15).

Mais uma vez podemos ver nesse resultado que mesmo que o ensemble contenha portas pouco emaranhantes($b \simeq 0$) a ordem de crescimento do tempo de convergência não será afetada, a razão entre o número de passos necessários para a convergência de duas portas diferentes será sempre uma constante. Note que o caso de portas não-emaranhantes($b = 0$), que leva a divergência do tempo convergência, está incluído em eq. (5.15).

Ainda na cadeia L, o efeito do parâmetro a é similar. Isto acontece pois cada probabilidade de transição não-trivial contém uma fator $1/(1-a)$. O mesmo argumento acima se aplica e podemos dizer

$$t_{mix(L|b\leq 2/3,a\leq 1-b)} = \frac{2(1-a)}{3b} t_{mix(L|b=2/3,a=0)} \ , \tag{5.16}$$

ou seja, a convergência de L é mais rápida quando a aumenta. Qual será a implicação da eq. (5.16) para a convergência de Q? Essa pergunta pode ser respondida com auxílio do Teorema 5.2.1, de onde obtemos que a cota

$$t_{mix(Q|b\leq 2/3,a\leq 1-b)} = \frac{1}{3b} t_{mix(L|b=2/3,a=0)} \ , \tag{5.17}$$

que não depende de a. Uma maneira de entender porque isso ocorre é notando que, se $a \neq 0$ o número de passos do CQA necessário para que tenhamos t passos de L é $\frac{t}{1-a}$. Mas a cadeia L é acelerada pelo fator $(1-a)$ que cancela o outro fator.

A eq. (5.17) se baseia numa cota onde estamos efetivamente ignorando a possibilidade de que a matriz T possa também ajudar na convergência. Isso quer dizer que em geral

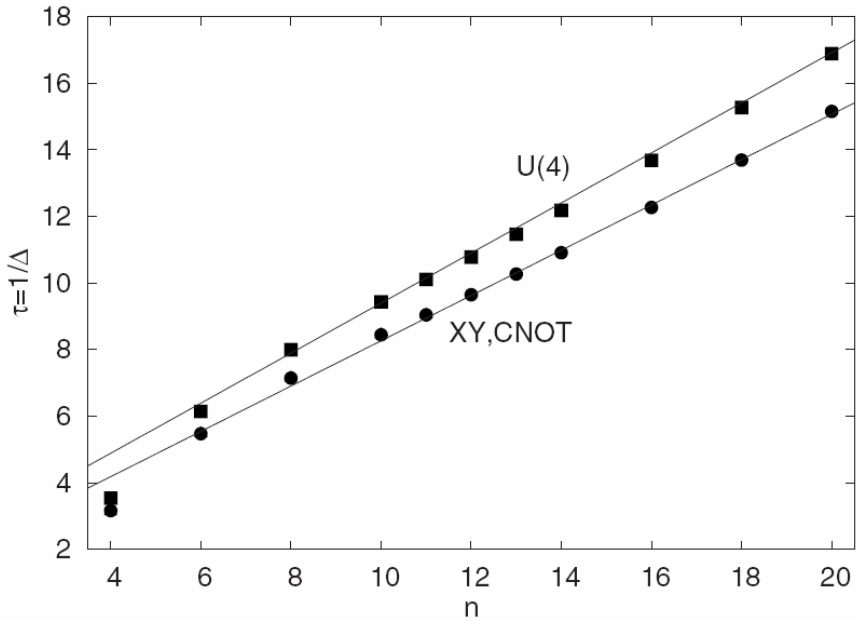


Figura 5.1: Reproduzido de [28]. Cálculo numérico do inverso do gap espectral usando XY, CNOT ou U(4). XY e CNOT, que tem o mesmo valor de $b = 2/3$, tem desempenhos equivalentes. O valor de b maior faz com que ambas sejam mais rápidas do que U(4) ($b = 3/5$).

t_{mixQ} pode em princípio depender de a, pois a é a chance da matriz T ser escolhida em cada passo de Q. Entretanto note que nem T nem a chance de escolha de T (num passo de Q) dependem de b e portanto o comportamento com b na eq. (5.17) é o correto.

Intuitivamente esperamos que a matriz T, que só realiza transposições, seja em geral pouco relevante para diminuir a distância $d(t)$, note que em particular ela não pode ajudar na convergência do peso de Hamming. Ainda assim sabemos que um valor maior de a, sem alterar b, nunca torna a convergência mais lenta, pois seu único efeito é realizar mais transposições. E portanto podemos fornecer a seguinte ordem parcial em relação ao desempenho de diferentes ensembles: dados dois ensembles com mesmo valor de a, será mais rápido o que tiver maior valor de b. Por outro lado se dois ensembles têm mesmo valor de b e valores diferentes de a, aquele com maior valor de a poderia em princípio vir a ser mais rápido. Em particular, por exemplo, o ensemble uniforme (que, como vimos nos exemplos do Teorema 4.2.1, tem $b = 3/5$ e $a = 1/5$) convergirá mais lentamente do que um ensemble como o 'porta-fixa' com porta XY (ou $DCNOT$), o qual tem $b = 2/3 \simeq 0.67$ e $a = 1/3$, um resultado que confirma uma análise numérica feita em [28], ver figura 5.1.

Estes mesmos resultados numéricos também apóiam o argumento heurístico dado acima de que a seja na verdade praticamente irrelevante para o tempo de mistura de Q. Isto pode ser mostrado estritamente para certas condições iniciais específicas: repare que se a distribuição inicial sobre as *strings* for invariante por permutações das coordenadas, o que acontece por exemplo com o estado $|\psi\rangle = |0\ldots0\rangle$ (veja Seção 2.2) , as transposições são desnecessárias e a convergência é com certeza totalmente independente do parâmetro a.

5.5 Tempo de mistura via grupos

Nessa Seção vamos estudar um outro problema, semelhante ao que estivemos estudando até o momento, no qual podemos utilizar uma técnica elegante de análise de convergência baseada na teoria de passeios aleatórios em grupos. Essencialmente, estudaremos uma cadeia semelhante à L, para a qual conseguiremos obter uma cota para o tempo de convergência. Discutiremos a semelhança dos dois problemas e a possibilidade de usar o resultado desta Seção para fornecer informação sobre o problema anterior.

5.5.1 Passeios aleatórios em Grupos

Dado um grupo finito $\mathbb{G}$ composto pelos elementos $\{g_i\}$ e pela operação de grupo $(\cdot)$, e ainda uma distribuição de probabilidade $\gamma(g)$ com $g \in \mathbb{G}$, podemos definir um passeio aleatório sobre $\mathbb{G}$ da seguinte forma: em cada passo do passeio, o elemento g vai para o elemento $h \cdot g$ com probabilidade $\gamma(h)$.

Esse passeio nada mais é do que uma cadeia de Markov no espaço de estados $\mathbb{G}$, com a matriz de transição $G(g,h) = \gamma(h \cdot g^{-1})$. [7] Observe que a aplicação seguida dessa matriz pode ser escrita em termos da convolução de γ com ela própria, e.g.: $G^2(g,s) = \sum_h \gamma(s \cdot h^{-1})\gamma(h \cdot g^{-1}) \equiv \gamma^{*2}(g,s)$.

Toda cadeia de Markov que se encaixa nessa definição tem as seguintes propriedades, de fácil demonstração (Seção 2.6 de [32])

- A distribuição uniforme $\pi(g) \equiv |\mathbb{G}|^{-1}$ é uma distribuição estacionária.
- A cadeia G é irredutível se, e somente se, o conjunto $\{g|\gamma(g) > 0\}$ gera $\mathbb{G}$. E nesse caso a distribuição uniforme é a única distribuição estacionária.
- A cadeia G é transitiva [8].

[7] γ é uma distribuição *fixa*, e não deve ser confundida com uma distribuição de probabilidade $\nu^t = \nu^0 G^t$ evoluindo sob a ação da cadeia de Markov G.

[8] Essa propriedade será definida e utilizada na Seção 5.6.3 .

O mapeamento em uma passeio de grupo é vantajoso pois além das propriedades acima existem técnicas poderosas para a análise de tempo de mistura. Essas técnicas, introduzidas por Persi Diaconis e David Aldous [35], se baseiam na teoria de representações irredutíveis de um grupo e na generalização da Transformada de Fourier. Não temos espaço nesta dissertação para uma explicação destes métodos, que são introduzidos de forma admiravelmente clara em [36], por isso daremos apenas alguns conceitos e resultados-chave para sua utilização.

Começamos com a idéia de Transformada de Fourier em grupos: dada uma função $f(g)$ definida em $\mathbb{G}$, a sua transformada é a função

$$\hat{f}(\rho) \equiv \sum_{g \in \mathbb{G}} f(g)\rho(g). \tag{5.18}$$

Aqui, ρ é uma representação irredutível (irrep) de $\mathbb{G}$. Assim, f tem como domínio o *conjunto de irreps* de $\mathbb{G}$, e sua imagem, para cada ρ, está no espaço de matrizes que realizam essa irrep. Repare que a eq. (5.18) é a Transformada de Fourier discreta usual se $\mathbb{G}$ é o grupo $\mathbb{Z}_n$ formado pelos inteiros (mod n) sob soma (mod n).

Para o estudo de passeios em grupos, o resultado que mais nos será útil é o que dá uma cota superior para a distância TV, após t passos a partir de qualquer distribuição inicial:

$$\left\|\gamma^{*t} - \pi\right\|_{TV}^2 \leq \frac{1}{4} \sum_{irreps \neq \mathbb{I}} d_\rho \, Tr\left(\hat{\gamma}^t(\rho)\hat{\gamma}^t(\rho)^\dagger\right) \ . \tag{5.19}$$

Nesta equação, a soma é sobre todas as irreps ρ não triviais de $\mathbb{G}$, e d_ρ é a dimensão de ρ.

Para grupos abelianos, como é o grupo que analisaremos, todas as representações são unidimensionais e as eqs (5.18) e (5.19) podem ser agrupadas na forma

$$\left\|\gamma^{*t} - \pi\right\|_{TV}^2 \leq \frac{1}{4} \sum_{irreps \neq \mathbb{I}} \left(\sum_{g \in \mathbb{Z}} \gamma(g)\rho(g)\right)^{2t} \ . \tag{5.20}$$

5.5.2 Mapeamento num passeio de grupo

Retornemos agora ao nosso problema de interesse. Nessa Seção vamos mostrar como uma cadeia $\tilde{L}$ semelhante a L pode ser mapeada num passeio de grupo.

Para motivar a definição de $\tilde{L}$, vamos olhar novamente para a dinâmica da cadeia L. Vimos na Seção 4.5 que esta dinâmica pode ser interpretada da seguinte maneira: escolha uma coordenada e mude-a com as probabilidades dadas pela eq. (4.19). Repare que $L^{(i)}_{0\to1}/L^{(i)}_{1\to0} = 3H/(H-1)$. Em outras palavras, para estados com pesos de Hamming

$H \gg 1$, a razão entre essas duas probabilidades é praticamente constante e igual a 3. Como, para sistemas de tamanho $n \gg 1$, a quase totalidade dos estados $\vec{p}$ satisfaz esta propriedade, podemos esperar que a cadeia L se comporte de forma semelhante a uma outra cadeia em que fixamos o valor dessa razão em 3 exatamente. Temos de ter algum cuidado, porém, já que os estados que não satisfazem à propriedade (ou seja, têm $H \simeq 1$) são justamente os que demoram mais para convergir até o estado estacionário de L.

Para simplificar nossa análise, fazemos ainda mais uma modificação: tomamos $\tilde{L}$ como uma versão mais rápida de L, ou seja, diminuindo a probabilidade da cadeia não avançar em cada passo. Dadas essas considerações, definimos então a cadeia modificada $\tilde{L}$ da seguinte forma: em cada passo, escolhe-se uma das coordenadas i, e a mudamos com as seguintes probabilidades (compare a eq. (4.19)):

$$\tilde{L}^{(i)}_{0\to 1} = 1; \quad \tilde{L}^{(i)}_{1\to 0} = 1/3; \quad \tilde{L}^{(i)}_{0\to 0} = 0; \quad \tilde{L}^{(i)}_{1\to 1} = 2/3 \tag{5.21}$$

Note que, ao contrário do que ocorre com L, as probabilidades de transição para $\tilde{L}$ são independentes do estado.

A cadeia $\tilde{L}$ foi estudada por Oliveira et al. em [18], os quais usaram um argumento de acoplamento para mostrar que ela tem *gap* $\Delta = \Omega(1/3n)$. A partir daí, estes autores usaram então um argumento de *comparação de cadeias* [54] para obter um limite de ordem $\Delta = \Omega(1/3n(n-1))$ para o *gap* da cadeia Q [9]. A eq. (2.27) garante então uma cota de $O(n^3)$ para t_{mixQ} neste caso. O Teorema 5.2.1 implica que este argumento, bem como a cota obtida, pode ser estendido para todos os ensembles localmente invariantes, bastando comparar-se a cadeia $\tilde{L}$ com a L.

Pela eq. (2.27), a cota obtida em [18] para o *gap* de L implica que esta cadeia tem um tempo de mistura de ordem $O(n^2)$. Mostramos agora um outro argumento, baseado num mapeamento em um passeio de grupo, que melhora esta cota para $O(n \log n)$.

Note primeiro que a cadeia $\tilde{L}$ em si não é um passeio de grupo, entre outras coisas porque não tem estado estacionário uniforme. Podemos porém mapeá-la num passeio de grupo expandindo a cadeia de volta para um espaço de estados maior. Especificamente, voltamos a associar o valor 1 em cada coordenada a três valores possíveis, que chamaremos agora de 1,2 ou 3, e definimos a cadeia expandida G sobre o espaço $\mathbb{Z}_4^n = \{0,1,2,3\}^n$ da seguinte maneira: escolha uma coordenada i. Se ela for 0, mude-a para $1, 2$ ou 3 com probabilidades iguais (1/3 para cada). Se for igual a $1, 2$,ou 3, mude-a para 0 com probabilidade 1/3, e com probabilidade 2/3 mude para 1 dos valores diferentes de 0, que não seja o mesmo do valor anterior, uniformemente. Vale notar que, embora estejamos voltando a um espaço de estados isomorfo ao original, a cadeia G não será igual, nem próxima, à cadeia original P.

[9] Estes autores estudaram caso do ensemble 'porta-fixa' com $C = CNOT$, para o qual vimos na eq. (4.14) que $a = 0$. Neste caso $Q = M \sim L$.

Para ver que essa nova cadeia G em $\mathbb{Z}_4^n$ é uma passeio aleatório de grupo precisamos exibir a operação de grupo e definir uma distribuição fixa $\gamma(\vec{g})$) que reproduza o passeio descrito acima. A operação de grupo será a soma (módulo 4) coordenada a coordenada. A distribuição γ será uniforme sobre o subconjunto de *strings* $\vec{g}$ que têm exatamente uma coordenada não-nula, e zero para todas as outros strings. Existem $3n$ *strings* dessa forma (cada coordenada podendo valer 1,2 ou 3), e portanto a probabilidade de escolha de cada um desses é $1/3n$. Temos assim:

$$\gamma(\vec{g}) = \begin{cases} 1/3n & \text{se existe } j \text{ tal que } g_i = 0 \forall i \neq j \text{ e } g_j \neq 0 \ , \\ 0 & \text{outros casos .} \end{cases} \tag{5.22}$$

É fácil ver que essa cadeia reproduz a dinâmica desejada. Note que o conjunto $\{\vec{g}|\gamma(\vec{g}) > 0\}$ gera o grupo $\mathbb{Z}_4^n$. Portanto, a cadeia é irredutível, e seu único estado estacionário é a distribuição uniforme.

Transformada de Fourier de γ

Para usar as técnicas de Transformada de grupo que introduzimos acima o primeiro passo é a obtenção das representações irredutíveis do grupo $\mathbb{Z}_4^n$, esse é o objetivo dessa Seção. Usaremos a teoria elementar de representações em grupos, uma boa referência para o tema é [58].

Defina os "vetores de base cartesiana" (elementos com exatamente um 1 que está na i-ésima coordenada) como $\vec{e}_i$, essa "base" gera todos os elementos $\vec{g}$ do grupo pois

$$\vec{g} = \prod_{i=1}^{n} (\vec{e}_i)^{g_i} \ .$$

O grupo $\mathbb{Z}_4^n$ é abeliano, e por isso tem tantas representações irredutíveis ρ quanto elementos, e todas são unidimensionais. Isso torna simples a obtenção das irreps.

Como cada representação deve manter a estrutura do grupo temos que

$$\rho(\vec{g}) = \prod_{i=1}^{n} (\rho(\vec{e}_i))^{g_i}$$

e portanto uma representação fica definida para todo o grupo se soubermos $\rho(\vec{e}_i) \, \forall \, i$. Além disso, como $(\vec{e}_i)^4 = id$ então $\rho^4(\vec{e}_i) = \rho(\mathbb{I}) = 1$. Portanto $\rho(\vec{e}_i)$ é uma das raízes de quarta ordem da unidade. Temos uma representação diferente para cada uma das 4 escolhas para $\rho(\vec{e}_i)$ (1,-1, i ou $-i$) para cada uma das n componentes, e assim temos todas as 4^n representações irredutíveis de $\mathbb{Z}_4^n$.

Agora podemos calcular a Transformada de Fourier de grupo da eq. (5.18),

$$\begin{aligned}\hat{\gamma}(\rho) &= \sum_{g\in\mathbb{Z}_4^n} \gamma(g)\rho(g) \\ &= \frac{1}{3n}\sum_{i=1}^{n}(\rho(\vec{e}_i)+\rho^2(\vec{e}_i)+\rho^3(\vec{e}_i)) \ ,\end{aligned} \tag{5.23}$$

os termos da soma na eq. (5.23) valem 3 se $\rho(\vec{e}_i)=1$ e valem -1 nos demais casos. Se denotarmos por $|\rho|$ o número de componentes em que $\rho(\vec{e}_i)\neq 1$ na representação ρ poderemos escrever (5.23) como

$$\hat{\gamma}(\rho) = 1-\frac{4|\rho|}{3n} \ . \tag{5.24}$$

Com a eq. (5.24) poderemos obter o principal resultado dessa Seção:

Teorema 5.5.1. *Seja $t=\frac{3n}{8}(\ln(3n)+2\theta)$, para $\theta>0$. A distância TV à estacionariedade após t passos de G é*

$$d(t) = \left\|\gamma^{*t}-\pi\right\|_{TV} \leq \frac{e^{-\theta}}{\sqrt{2}}. \tag{5.25}$$

Em consequência, o tempo de mistura de G a uma distância ε satisfaz

$$t_{mix_G}(\varepsilon) = O(n\ln(n/\varepsilon)) \tag{5.26}$$

Prova: Basta introduzir a eq. (5.24) na eq. (5.19) para obter:

$$\begin{aligned}\left\|\gamma^{*t}-\pi\right\|_{TV}^2 &\leq \frac{1}{4}\sum_{irreps\neq\mathbb{I}}\left(1-\frac{4|\rho|}{3n}\right)^{2t} \\ &= \frac{1}{4}\sum_{|\rho|=1}^{n}\binom{n}{|\rho|}3^{|\rho|}\left(1-\frac{4|\rho|}{3n}\right)^{2t} \ ,\end{aligned}$$

agora usando $\binom{n}{|\rho|}\leq\frac{n^{|\rho|}}{|\rho|!}$ e $1-x\leq e^{-x}$ temos:

$$\begin{aligned}\left\|\gamma^{*t}-\pi\right\|_{TV}^2 &\leq \frac{1}{4}\sum_{|\rho|=1}^{n}\frac{n^{|\rho|}}{|\rho|!}3^{|\rho|}e^{-8t|\rho|/3n}\\ &= \frac{1}{4}\sum_{|\rho|=1}^{n}\frac{[\exp(\ln(3n)-8t/3n)]^{|\rho|}}{|\rho|!}\\ &\leq \frac{1}{4}\left\{\exp[\exp(\ln(3n)-8t/3n)]-1\right\}.\end{aligned}$$

Agora se $t=\frac{3n}{8}(\ln(3n)+2\theta)$, o lado direito desta equação fica igual a $\exp(e^{-2\theta})-1$. Podemos então usar o fato que $e^x-1\leq 2x$ para $0\leq x\leq 1$ e obtemos a cota da eq. (5.25). Finalmente, escolhendo $\varepsilon=e^{-\theta}/\sqrt{2}$, de modo que $d(t)\leq\varepsilon$, então o tempo acima é uma cota superior para $t_{mixG}(\varepsilon)$ pela eq. (2.24). Substituindo o valor de θ pode se ver que este t satisfaz

$$t=O\left(n\ln\left(n/\varepsilon^2\right)\right)=O(n\ln\left(n/\varepsilon\right)).\ \square \tag{5.27}$$

Projeção em uma cadeia de peso de Hamming

Da mesma forma como, na Seção 4.6, projetamos a cadeia Q em uma cadeia de peso de Hamming Z, podemos fazer o mesmo para as cadeias $\tilde{L}$ ou G. É fácil ver que essa projeção, que chamaremos de Z_G, também é uma cadeia do tipo *Birth-and-Death*, com os seguintes elementos não-nulos:

$$\begin{aligned}Z_G(H,H+1)&=\frac{n-H}{n},\\ Z_G(H,H-1)&=\frac{1}{3}\frac{H}{n},\\ Z_G(H,H)&=\frac{2}{3}\frac{H}{n}.\end{aligned} \tag{5.28}$$

Evidentemente, Z_G também converge com tempo de mistura $t_{mix}=O(n\log n)$. Na Seção 5.6 veremos que na verdade G e Z_G têm exatamente o mesmo tempo de mistura.

5.5.3 Discussão comparativa

Nessa Seção esboçamos algumas idéias sobre a possível relevância dos resultados acima para a análise da cadeia de Markov L que governa o problema de nosso real interesse. Não

conseguimos obter resultados conclusivos a esse respeito, portanto elas devem ser tomadas apenas como especulações.

Recorde que, quando definimos $\tilde{L}$, propositalmente fizemos com que esta cadeia tivesse uma probabilidade menor de ficar parada em cada passo, comparada com L. Heuristicamente, pode-se esperar então a princípio que $\tilde{L}$ convirja mais rapidamente que L. Neste caso, porém, a cota derivada na eq. (5.27) não nos diria nada sobre esta última cadeia.

Podemos, entretanto, fornecer um outro argumento heurístico no sentido oposto. Para isto, precisamos de um conceito adicional, a saber o de uma 'versão acelerada' de uma cadeia. Dada uma cadeia N qualquer definimos sua versão acelerada N^A como a cadeia condicionada ao fato de ter andado em cada passo. Em outras palavras, a cadeia acelerada é obtida eliminando-se os elementos diagonais da matriz de transição (ou seja, eliminando-se a possibilidade do passeio não andar), mas mantendo as razões entre as demais probabilidades de transição. Formalmente, se N^A é a versão acelerada de N então $N^A(x,y) = N(x,y)/(1-N(x,x))$ para $x \neq y$ e $N^A(x,x) = 0$.[10].

Com este procedimento, podemos verificar que as cadeias Z e Z_G possuem versões aceleradas semelhantes:

$$\begin{aligned} Z^A(H, H+1) &= \frac{3(n-H)}{3n-2H-1} \\ Z^A(H, H-1) &= \frac{H-1}{3n-2H-1} \\ Z^A(H, H) &= 0 \ , \end{aligned} \tag{5.29}$$

$$\begin{aligned} Z_G^A(H, H+1) &= \frac{3(n-H)}{3n-2H} \\ Z_G^A(H, H-1) &= \frac{H}{3n-2H} \\ Z_G^A(H, H) &= 0 \ . \end{aligned} \tag{5.30}$$

Pode se observar porém que a cadeia Z_G^A tem um viés na direção positiva que é estritamente menor que a de Z^A. Assim, se ela partir de uma condição inicial com $H = O(1)$, esperamos então que ela leve um tempo $\left\langle \tau_{Z_G^A} \right\rangle$ *maior*, em média, que o de Z^A para atingir um valor de H próximo do estacionário, $3n/4$. Vale notar que são essas as condições iniciais que demoram mais tempo, em média para atingir este valor. Esperamos então que, no pior caso, esse tempo médio de chegada (ou *hitting time*) para as cadeias aceleradas seja no

[10]Esta cadeia em geral não será aperiódica. Quando isto for um problema, pode-se sempre definir $N^A(x,x) = \alpha$ para α suficientemente pequeno

máximo igual ao tempo de mistura da respectiva cadeia não-acelerada Z ou Z_G. Neste caso teríamos, por exemplo $\max \langle \tau_Z \rangle \leq \max \leq \langle \tau_{Z_G} \rangle \leq t_{mixZ_G} = O(n \log n)$

Por outro lado, se somarmos o número médio de passos que a cadeia não-acelerada fica parada com o tempo médio de chegada da cadeia acelerada, temos o tempo médio de chegada para a cadeia não-acelerada. E é possível em muitos casos, vide por exemplo Teorema 10.15 de [32], limitar o tempo de mistura de uma cadeia pelos máximos tempos médios de chegada de um ponto a outro do espaço de estados. Se um resultado deste tipo valer aqui, teríamos então $t_{mixZ} \leq \max \langle \tau_Z \rangle + \max \langle \tau_{esperaZ} \rangle$, onde $\langle \tau_{esperaZ} \rangle$ é o número de passos que a cadeia fica parada, em média, até ocorrer a chegada ao ponto-alvo. Pelo argumento anterior, o primeiro termo desta soma é $O(n \log n)$. E, segundo parte da demonstração de convergência de Z feita em [27], o segundo termo também é $O(n \log n)$.

Em conclusão: se este argumento heurístico puder ser formalizado, forneceria uma prova alternativa do fato de que a convergência de Z como um todo leva $O(n \log n)$ passos. Como já mencionado anteriormente, a prova deste fato apresentada em [27] é muito mais longa. (Entretanto, como dito acima, também necessitaríamos de parte dela).

5.6 Cutoff

5.6.1 Definição e propriedades gerais

O algoritmo que estudamos leva à geração de um unitário aleatório e nesses processos de randomização ocorre frequentemente o fenômeno conhecido como *corte abrupto* [52]. Intuitivamente, um corte abrupto ocorre quando o processo converge abruptamente de um estado longe do randomizado para um randomizado em uma estreita janela de tempo. Um exemplo prático bem conhecido se dá no embaralhar de um baralho de cartas inicialmente ordenado. Se realizamos apenas uma ou duas embaralhadas, ainda podemos observar bastante ordem remanescente nas cartas; assim, o baralho ainda não está bem randomizado. Porém, basta um número finito de embaralhadas para que as cartas fiquem completamente aleatorizadas. Este fenômeno foi estudado matematicamente pela primeira vez por Aldous e Diaconis no artigo [35], onde demonstraram que sete embaralhadas (de um determinado tipo) são suficientes.

No contexto de cadeias de Markov, um corte abrupto ocorre quando a convergência da cadeia se dá com um número de passos muito bem delimitado. Ou seja, a distância TV passa de valores $\simeq 1$ para valores $\simeq 0$ em um número de passos pequeno se comparado ao número total de passos necessários. Daremos a seguir algumas noções e resultados precisos já conhecidos, para uma introdução a esse fenômeno veja o Capítulo 18 de [32] ou o Capítulo 3 de [36] para uma introdução dada por um dos responsáveis pelo desenvolvimento da área.

Mais precisamente, podemos definir um conceito de *corte abrupto* quando temos uma família de cadeias semelhantes definidas em espaços cujo tamanho depende de um parâmetro n. Nesse caso definimos que ocorre um *corte abrupto* quando os tempos de mistura dessas cadeias satisfazem

$$\lim_{n\to\infty} \frac{t_{mix}^{(n)}(\varepsilon)}{t_{mix}^{(n)}(1-\varepsilon)} = 1. \tag{5.31}$$

Ou seja, há *corte abrupto* quando, para n suficientemente grande, o tempo de mistura é essencialmente o mesmo para qualquer valor de $\varepsilon \in (0,1)$. Do ponto de vista da distância ao estado estacionário, há um comportamento abrupto na passagem do seu valor máximo para seu valor mínimo num tempo desprezível frente a $t_{mix}^{(n)}$. Esse comportamento está ilustrado qualitativamente na figura 5.2.

Essa figura motiva uma definição equivalente (demonstrada no Capítulo 18 de [32]). Há um *corte abrupto* se e só se a distância à estacionariedade da cadeia n satisfaz

$$\lim_{n\to\infty} d_n(c\, t_{mix}^{(n)}) = \begin{cases} 0 \text{ se } c < 1 \\ 1 \text{ se } c > 1 \end{cases} . \tag{5.32}$$

Diz-se ainda que a sequência de cadeias de Markov tem *janela* $\{\omega_n\} = o(t_{mix}^{(n)})$ se existe uma constante c_ε,tal que, para n qualquer

$$t_{mix}^{(n)}(\varepsilon) - t_{mix}^{(n)}(1-\varepsilon) \leq c_\varepsilon \omega_n.$$

Claramente a existência de uma *janela* com essas propriedades implica que a cadeia tem *corte abrupto*.

Note que demonstrar um *corte abrupto* significa dar uma noção bem mais precisa do tempo de mistura de uma cadeia do que simplesmente dizer que $t_{mix}^{(n)}$ é de alguma ordem $O(f(n))$. Se um *corte abrupto* existe em, digamos, $n \log n$, então não apenas é verdade que $t_{mix}^{(n)} = \Theta(n \log n)$, mas o coeficiente constante nesta dependência é precisamente 1. Encontrar o ponto exato de *corte abrupto* é portanto, em geral, uma tarefa difícil.

Sabe-se hoje que uma grande variedade de cadeias de Markov apresenta *corte abrupto* como por exemplos passeios aleatórios em grupo e muitos tipos de embaralhamento. Também se conhecem contra-exemplos de cadeias que não exibem o fenômeno. Mesmo assim, ainda não existe um critério geral para caracterizar as cadeias que possuem *corte abrupto*, e continua sendo um tema de pesquisa na comunidade de cadeias de Markov provar e caracterizar a existência de *corte abrupto* em diversas categorias especiais de cadeias.

Existe porém o seguinte critério *necessário* simples para a existência de um *corte abrupto* [59]:

Teorema 5.6.1. *Se uma sequência de cadeias ergódicas exibe* corte abrupto, *então necessariamente seus tempos de mistura e* gaps *espectrais satisfazem*

$$\lim_{n\to\infty} \left(t_{mix}^{(n)} \cdot \Delta^{(n)} \right) \to \infty \quad . \tag{5.33}$$

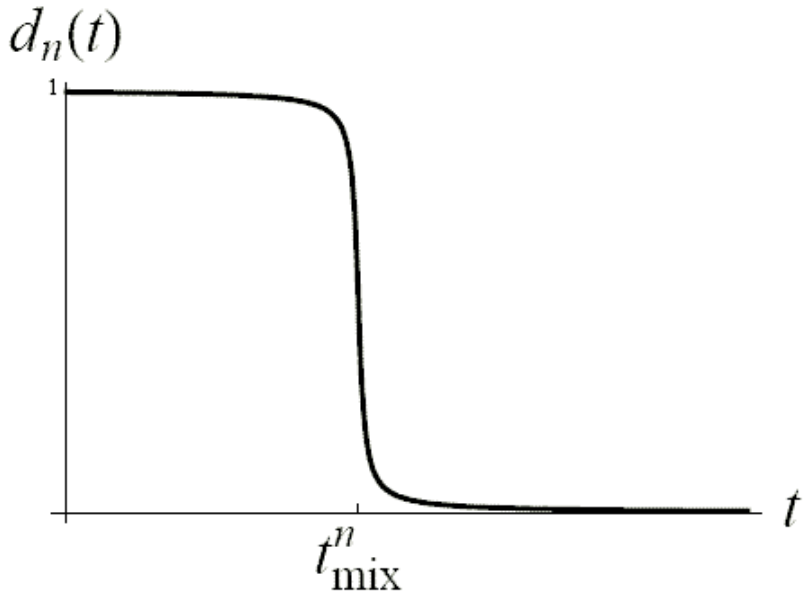


Figura 5.2: Comportamento qualitativo da distância à estacionariedade de uma sequência de cadeias que exibe *corte abrupto*.

5.6.2 Cutoff em cadeias *Birth-and-Death*

Como vimos na Seção 4.6, uma cadeia *Birth-and-Death* (BD) é uma cadeia com espaço de estados $\Omega_{BD} = \mathcal{Z}_n$, e que só faz transições entre primeiros vizinhos. Para esse tipo de cadeia recentemente foi provado que o critério do Teorema 5.6.1 é também suficiente, ou seja:

Teorema 5.6.2. *([59]) Seja $BD^{(n)}$ uma sequência de cadeias BD ergódicas. Então ela exibe* corte abrupto *se e só se a eq. (5.33) for válida. Ainda, a janela de* corte abrupto *é no máximo $O\left(\sqrt{t_{mix}^{(n)}/\Delta^{(n)}}\right)$.*

Este resultado nada diz, porém, sobre o tempo em que o *corte abrupto* ocorre.

5.6.3 Cutoff em CQAs

Em estudos anteriores de CQAs, alguns autores notaram, através de simuações numéricas, a aparente existência de um *corte abrupto* neste processo [18, 28]. Por exemplo, na fig. 5.3 reproduzimos uma figura de [18], no qual nota-se um comportamento semelhante à fig. 5.2 acima.

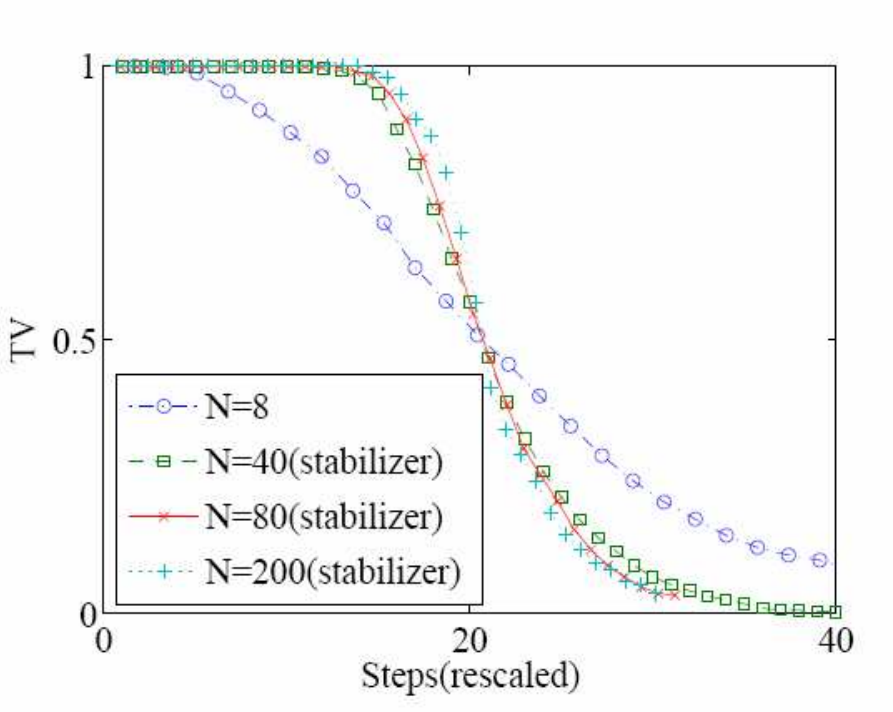


Figura 5.3: Reproduzido de [18]. Evidência numérica do efeito de corte abrupto: a distância TV cai cada vez mais abruptamente a medida que o número n de q-bits cresce.

Nesta Seção apresentamos alguns resultados analíticos que dão apoio a essa possibilidade. Embora não chegamos a conseguir mostrar a existência de um *corte abrupto* para a cadeia P como um todo, podemos mostrar os seguintes resultados parciais

Lema 5.6.1. *A cadeia de peso de Hamming Z exibe* corte abrupto .

Prova: Sabemos, da eq. (4.22) que Z é uma cadeia do tipo BD ergódica. Vimos na Seção 4.6 que $t_{mixZ} = \Theta(n \log n)$ e $\Delta_Z = \Omega(1/n)$. Portanto a cadeia satisfaz às condições do Teorema 5.6.2 e possui *corte abrupto*.

Podemos ainda mostrar que

Lema 5.6.2. *O passeio de grupo G (e sua projeção Z_G) possuem um mesmo* corte abrupto.

Prova: Para a cadeia Z_G, que também é uma cadeia BD ergódica, podemos usar exatamente os mesmos argumentos do lema anterior: primeiro, sabemos pelo Teorema 5.5.1 que $t_{mixZ_G} = O(n \log n)$, e pelos mesmos argumentos que serão apresentados na Seção 5.7 temos que o gap é $\Omega(1/n)$. Portanto essa cadeia também possui *corte abrupto*.

Podemos estender este resultado para a cadeia G como um todo usando a propriedade de que todo passeio aleatório em grupos é uma cadeia transitiva. Dizemos que uma cadeia P é transitiva se para cada par de elementos $(g_1, g_2) \in \Omega$ existe uma bijeção $\Omega \to \Omega$ tal que:

$$f(g_1) = g_2 \quad \text{e} \quad P(g_3, g_4) = P(f(g_3), f(g_4)) \; \forall g_3, g_4 \in \Omega \; .$$

Intuitivamente, uma cadeia ser transitiva significa que ela se comporta de forma semelhante em todos os pontos do espaço de estados. Todo passeio em grupo é transitivo, pois basta escolher $f(g) = g g_1^{-1} g_2$. Ou seja, a cadeia G é a mesma independentemente do rótulo dos seus elementos. Assim o tempo necessário para chegar próximo à estacionaridade é o mesmo para qualquer condição inicial concentrada em um único elemento. Ainda, argumentos apresentados no capítulo 18 de [32] mostram que o tempo de mistura de G é exatamente o mesmo de sua projeção Z_G. Portanto a cadeia G também tem *corte abrupto* de janela $O(n\sqrt{\log n})$. □

O argumento usado acima para G não pode ser usado da mesma forma para P, pois esta última cadeia não é transitiva. Como ela é ainda invariante por permutações, continua sendo verdade que toda condição inicial com essa simetria converge no mesmo instante, o qual será também o instante de *corte abrupto* de Z pelos mesmos argumentos de antes. Porém, ainda é possível em princípio que uma condição inicial assimétrica leve mais tempo para convergir ao estado estacionário. A questão continua em aberto portanto. Um caminho possível para atacar o problema talvez seja tentar adaptar resultados recentes sobre *corte abruptos* em passeios tipo Ising (ver Seção 7.2.1 na Conclusão), dada a semelhança dessas cadeias com a cadeia L.

5.7 Análise da convergência a um 2-desenho através de P

Nessa Seção chegamos aos resultados finais da demonstração de convergência, ligando o tempo de convergência da cadeia de Markov P ao tempo de convergência do ensemble de unitários gerada pelo CQA a um 2-desenho. Esta ligação está feita em [27], e segue igualmente para a nossa demonstração. Enunciamos apenas então os principais resultados:

Teorema 5.7.1. *[30] Se $d(t)$ é a distância à estacionaridade da cadeia P no sentido TV (eq. (2.22)), então o ensemble μ de unitários gerado por t passos do CQA é um 2-desenho* ***de canal*** *$d(t)$-aproximado.*

Portanto para obter um 2-desenho de canal ε-aproximado basta fazer $d(t) \leq \varepsilon$

Teorema 5.7.2. *[27] Se a distância na norma L_2 é, após t passos de P, $d_2(t)$ então o ensemble μ de unitários gerado por t passos do CQA é um 2-desenho $2^{2n}d(t)$-aproximado.*

Portanto para obter um 2-desenho ε-aproximado basta fazer $d_2(t) \leq \varepsilon/2^{2n}$.

Usando então os Teoremas 5.3.1 e 5.7.1, podemos concluir que

Teorema 5.7.3. *Seja $t = O(n\log(n/\varepsilon))$, após t passos do CQA o ensemble de unitários gerado é um 2-desenho* ***de canal*** *ε-aproximado .*

Já para a convergência na definição 2.4.2 precisamos, segundo o Teorema 5.7.2, obter a convergência de Q na norma L_2.

Porém, resultados gerais da literatura de cadeia de Markov ligam o tempo de mistura à convergência na norma L_2, ou seja, é possível obter cotas para d_2 a partir de $d(t)$. Para isso, usamos as eqs. (2.27) e (2.26) para obter uma cota do *gap* espectral e com o gap obteremos t_{2-mix}.

Das eqs. (2.28) e (2.24) e do Teorema 5.3.1 temos que, para alguma constante k

$$kn\log(n/\varepsilon) \geq \frac{1-\Delta}{\Delta}\ln(1/2\varepsilon)\ ,$$

no limite $\varepsilon \to 0$

$$kn \geq \frac{1-\Delta}{\Delta}\ ,$$

o que implica em $\Delta = \Omega(1/n)$. Agora da eq. (2.26) temos:

$$t_{2-mix_Q}(\varepsilon') = O(n\log(1/\varepsilon'))\ ,$$

mas, pelo Teorema 5.7.2, para obter um 2-desenho ε-aproximado devemos ter $d_2 \leq (\varepsilon/2^{2n})$. Pela equação acima temos

$$t_{2-mixQ}(\varepsilon/2^2 n) = O(n \log(2^2 n/\varepsilon)) = O(\, n(n + \log \varepsilon^{-1})) \;,$$

Concluímos assim que

Teorema 5.7.4. *Seja $t = O(\, n(n+\log \varepsilon^{-1}))$, após t passos do CQA o ensemble de unitários gerado é um 2-desenho ε-aproximado .*

Capítulo 6

Paralelização do Algoritmo

Todas as escalas de tempo de convergência obtidas nessa dissertação assumem que cada passo do circuito quântico aleatório é aplicado em sequência. Entretanto, como cada passo afeta apenas um único par de q-bits, é natural questionar se algum ganho de escala pode ser obtido por paralelização do circuito. Nesse capítulo apresentaremos fortes evidências analíticas e numéricas de que isso de fato ocorre: se as portas forem aplicadas em paralelo, a profundidade do circuito pode ser reduzida de um fator $O(n)$. Então, se cada porta de 2 q-bits pode ser realizada em uma unidade de tempo, o tempo necessário para a convergência para um 2-desenho **de canal** será $O(\log n)$ e para um 2-desenho será $O(n)$. Vale notar que uma conjectura similar quanto à paralelização de CQAs foi também feita em [33]. Nesse trabalho os autores consideram um circuito diferente onde várias portas CZ são aplicadas em paralelo, antes do próxima aplicação de portas CZ realiza-se rotações unitárias locais em todos os q-bits. Simulações numéricas para $n \leq 10$ parecem indicar que o *gap* é constante nesse caso, mas os resultados são inconclusivos (e não dão maior informação sobre o tempo de mistura).

Repare que esse tipo de estratégia não reduz o número de portas lógicas que têm que ser implementadas, mas diminui o tempo gasto na computação, também chamado de profundidade do algoritmo. A diminuição da profundidade do algoritmo é especialmente útil para computadores quânticos, pois a maioria das implementações destes sistemas é fortemente afetada pela decoerência gerada pela interação com algum "reservatório". Assim, algoritmos mais longos necessitam de maior correção de erros, o que leva a um aumento do número de q-bits físicos usados para codificar os q-bits lógicos, e requer ainda a implementação de portas lógicas adicionais.

Nesse capítulo vamos examinar uma maneira de paralelizar o CQA proposto.

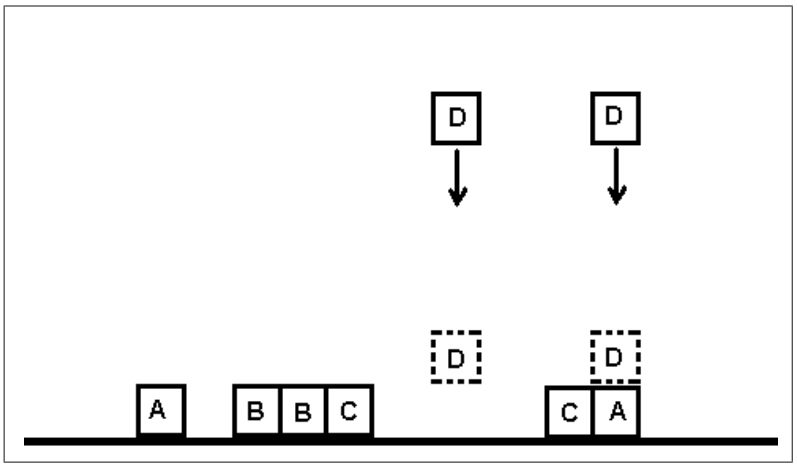


Figura 6.1: Representação gráfica da paralelização. Cada letra representa um par que foi escolhido para interagir. O par D que, acaba de ser escolhido, não pode ir para a primeira linha pois a porta lógica referente à escolha desse par não comuta com a porta lógica referente ao par A, e por isso vai para a segunda linha.

Na figura 6.1 vemos a representação da paralelização proposta. Ela é feita com o auxílio de um computador clássico que sorteia aleatoriamente e independentemente uma sequência de pares[1] e preenche uma tabela como a representada pela figura 6.2. Vê-se claramente que esse processo deixa vazios ou espaços não utilizados em cada linha. Embora não seja possível utilizar todos os espaços em cada linha, só precisamos que a densidade de preenchimento das linhas seja limitada inferiormente por uma constante. Assim poderemos realizar $O(n)$ portas lógicas por unidades de tempo, e o tempo total de execução diminuirá de um fator n.

A figura 6.3 apresenta a evidência numérica de que esse procedimento de paralelização, apesar de sua simplicidade, leva a um resultado que é qualitativamente tão bom quanto possível, já que uma fração constante dos q-bits interage em cada passo. Na próxima Seção veremos que o procedimento de paralelização é um modelo inusitado de crescimento de superfícies, e faremos seu estudo usando algumas ferramentas dessa área.

[1]Se o objetivo é a implementação do algoritmo de geração de um 2-design, a sequência de pares deve conter $O(n \log n)$ ou $O(n^2)$ pares, de acordo com o que foi discutido em capítulos anteriores.

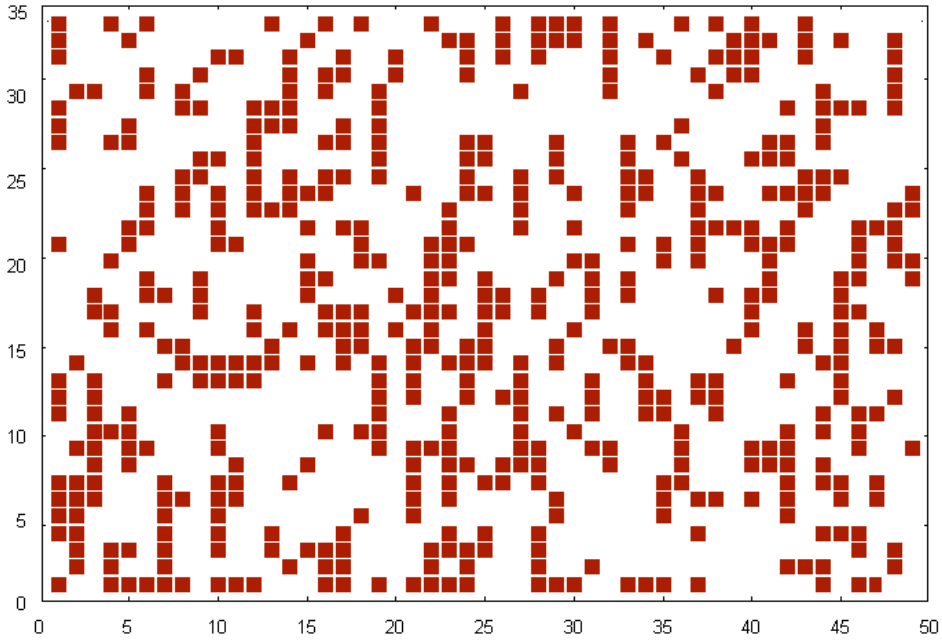


Figura 6.2: Representação de uma instância típica do processo de paralelização. Os quadrados presentes mostram as portas lógicas que seriam realizadas, e todos as portas lógicas da mesma linha são realizadas simultaneamente. Nessa figura, o número de colunas é $n = 50$ e vemos as primeiras 35 linhas.

6.1 Paralelização como um problema de deposição

A situação sugerida pela fig. 6.2 lembra muito a de modelos de deposição de partículas em superfícies, uma área da Física Estatística na qual muitos resultados (em sua maioria, numéricos) têm sido obtido em anos recentes [60]. No nosso caso, trata-se de um modelo inusitado do ponto de vista físico, dado que nossas 'partículas' sempre caem duas a duas, mas a distâncias arbitrárias uma da outra. Ainda por cima 'grudam' em uma mesma altura assim que a primeira colidir com a superfície abaixo, independente se a outra deixar um espaço abaixo e/ou aos lados. Mesmo assim, técnicas desta área podem ser adaptadas.

Começamos notando que a rugosidade da superfície, ou o desvio absoluto da média de altura das colunas, é o que controla a densidade do preenchimento das linhas. Isso acontece pois quando a rugosidade é grande, a chance de escolhermos colunas com alturas muito diferentes aumenta, o que gera mais regiões vazias. Isso acontece pois, ao escolher a coluna i e a coluna j, o número de espaços vazios gerados é exatamente o módulo da diferença entre as alturas das duas colunas.

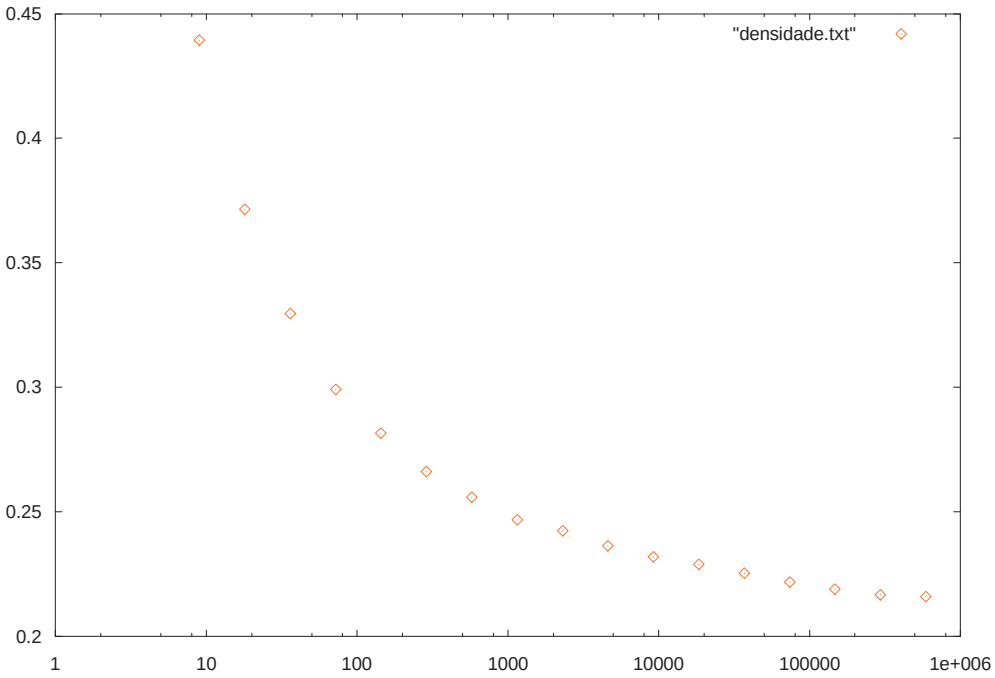


Figura 6.3: Simulação numérica mostrando a densidade de ocupação média das linhas da figura 6.2 em função do número de q-bits. Vê-se que a eficiência da paralelização se mantém mesmo no limite $n \to \infty$. O eixo com o número de q-bits está em escala logarítmica.

Seja $h_i(t)$ a altura (número da maior linha ocupada) da i-ésima coluna da figura no tempo t, e $h(t)$ a média de todas as alturas. Temos:

$$\begin{aligned} \langle h(t+1)\rangle &= \langle h(t)\rangle + \frac{1}{n}\frac{1}{n(n-1)}\sum_{i,j}(|h_i(t)-h_j(t)|+2) \\ &= \langle h(t)\rangle + \frac{1}{n}\left\{2+\frac{1}{n(n-1)}\sum_{i,j}|h_i(t)-h_j(t)|\right\}\ , \end{aligned} \tag{6.1}$$

onde $\langle\rangle$ é a média sobre as escolhas de diferentes pares.

Seja ainda $w_i(t) \equiv h_i(t) - h(t)$ o desvio da altura da coluna i para a média das alturas, e ainda $w(t) = \frac{1}{n}\sum_i |w_i|$. Então:

$$\begin{aligned}\frac{1}{n(n-1)}\sum_{i,j}|h_i(t)-h_j(t)| &= \frac{1}{n(n-1)}\sum_{i,j}|h_i(t)-h(t)-(h_j(t)-h(t))| \\ &\leq \frac{1}{n}\left(\sum_i |w_i| + \sum_j |w_j|\right) \\ &= 2w(t)\end{aligned}$$

De volta a eq. (6.1), temos

$$\langle h(t+1)\rangle - \langle h(t)\rangle \leq \frac{2}{n}(1+w(t)) \tag{6.2}$$

Assim, basta mostrar que a rugosidade é limitada por uma constante (w_0) durante toda a deposição para que tenhamos

$$\langle h(t)\rangle \leq \frac{2t}{n}(1+w_0)). \tag{6.3}$$

Em termos do CQA, isso significa que a profundidade média em cada q-bit fica reduzida por um fator n após a paralelização. Como a rugosidade média é limitada por w_0, não esperamos que haja q-bits com profundidades muito superiores a $\langle h(t)\rangle + w_0$. Isso pode ainda ser visto pelo fato de que são justamente as colunas com maior altura que tem menor crescimento em média, ou seja, se a coluna tiver uma altura maior do que a dispersão seu crescimento será impedido pela ausência de colunas ainda maiores.

Infelizmente, não sabemos como demonstrar isto analiticamente (e como foi mencionado acima, poucos resultados completamente analíticos existem nesta área, mesmo com modelos de deposição mais bem-comportados fisicamente). Ao invés disso, apresentamos a seguir agora uma série de argumentos parcialmente analíticos e parcialmente heurísticos. Fazemos também algumas simulações que confirmam esses argumentos.

6.1.1 Descrição estocástica

Uma estratégia frequentemente utilizada é construir uma equação de crescimento estocástico para cada variável h_i. Essa equação deve descrever o mesmo comportamento médio do modelo original, e deve conter flutuações de maneira a imitar o modelo. Veremos que o mesmo comportamento qualitativo é obtido de maneira largamente independente da intensidade da flutuação, o que justifica o emprego da descrição estocástica.

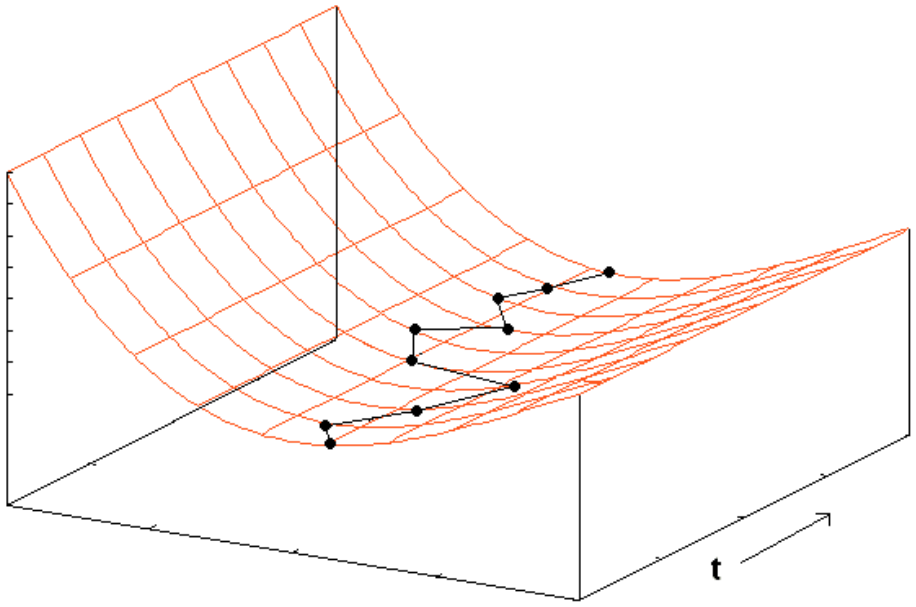


Figura 6.4: Representação de uma instância possível para a dinâmica da variável w_i, que é decorrente da ação de um força estocástica e de um potencial atrator (em vermelho).

Do modelo, temos que a altura de uma coluna i muda quando essa é escolhida junto com a coluna j. Se a coluna j é menor ou igual à coluna i, a coluna i simplesmente aumenta em 1 unidade sua altura, e se a coluna j é mais alta, a coluna i aumenta de $(h_j - h_i + 1)$. Assim, a variação de altura que uma coluna sofre é tanto maior quanto menor for sua altura, o que tende a equalizar as alturas (e portanto reduzir a rugosidade). Veremos que esse comportamento qualitativo é suficiente para mostrarmos que a rugosidade é limitada. Para isso formulamos a equação estocástica

$$h_i(t+1) - h_i(t) = \mathcal{F}(h_i)$$

onde $\mathcal{F}(h_i)$ é uma força estocástica que tem a mesma forma para todas as colunas, devido à simetria do problema, e com a propriedade:

$$\langle \mathcal{F}(h_i) \rangle > \langle \mathcal{F}(h_j) \rangle \text{ , se } h_j > h_i.$$

A equação para a dispersão fica:

$$w_i(t+1) - w_i(t) = \mathcal{F}(h_i) + h(t+1) - h(t) = \mathcal{G}(w_i).$$

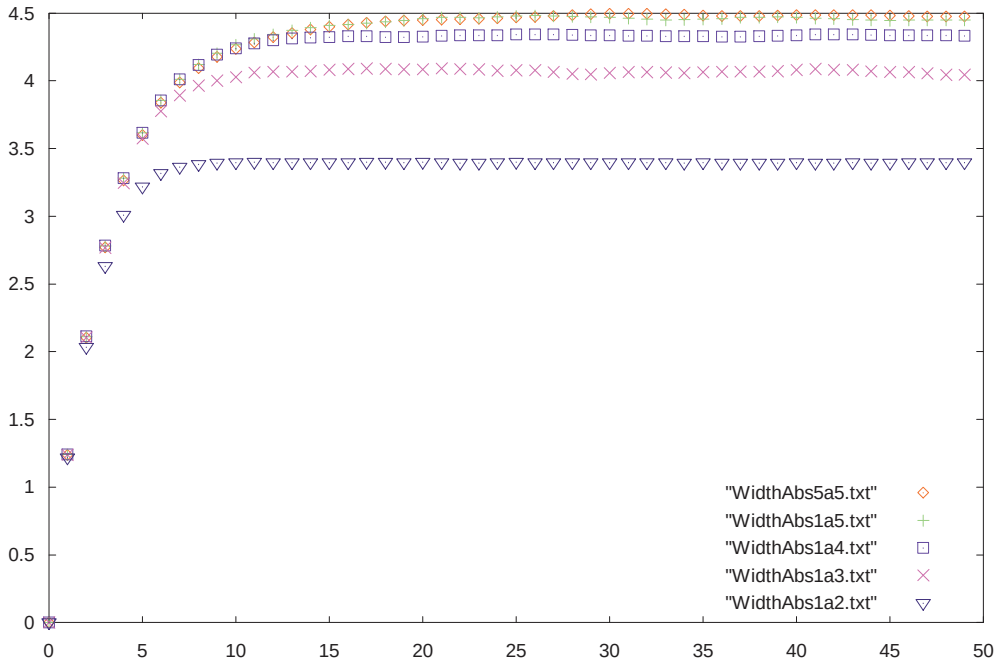


Figura 6.5: Simulação numérica do desvio absoluto médio das alturas das colunas do processo de paralelização representado na fig 6.2 em função do tempo dividido pelo tamanho n.

Como $\langle w_i(t)\rangle = 0$ e $\langle \mathcal{G}(w_i)\rangle$ é também monotonamente decrescente, então $\langle \mathcal{G}(w_i)\rangle$ deve ser nula para algum valor de $w_i = w_{(\mathcal{G}=0)}$. Para valores de w_i maiores do que $w_{(\mathcal{G}=0)}$, a força média $\langle \mathcal{G}(w_i)\rangle$ é negativa, e para valores menores, é positiva. Ou seja, a variável w_i é como uma partícula browniana que é impedida de difundir por um potencial atrator. Esse potencial é o que faz com que a força estocástica tenha média não-nula. Assim, a dispersão média de $w(t)$ é determinada pela forma do potencial e pela intensidade da força estocástica. A origem do potencial atrator efetivo é a correlação entre as alturas das colunas gerada pelo modelo de deposição, que é independente de n. A parte ruidosa de média zero da força estocástica está ligada à possibilidade de escolha de diferentes pares de colunas para interagir[2], e também é independente de n, mas está sujeita a efeitos de tamanho finito como vemos na figura 6.4. Por isso, a dispersão de w_i é essencialmente independente de n e é limitada superiormente.

[2] A rigor, o ruído também depende de w_i. Considere, por exemplo, uma coluna com w_i muito maior que a média. Essa coluna vai, com quase certeza, crescer muito pouco, mas como $w_i(0) = 0$, podemos supor o ruído independente de w_i e mostrar que a dispersão é limitada, justificando essa simplificação.

Capítulo 7

Conclusão

Resumimos agora os principais resultados alcançados deste trabalho, e traçamos algumas linhas para investigação futura.

7.1 Resumo dos resultados

Antes de resumir nossos resultados, convém colocá-los em contexto relembrando mais uma vez os resultados de outros autores, em particular os de Harrow e Low na ref. [27], dos quais nosso trabalho pode ser considerado uma (modesta) correção e extensão.

Recordando, esses autores identificaram o problema de demonstrar a convergência de um CQA a um k-desenho de unitários, e desenvolveram as linhas gerais da análise que apresentamos. Apoiando-se nos resultados anteriores de [18], eles reduziram o problema, no caso $k = 2$ à convergência de uma cadeia de Markov P. Eles demonstraram a existência dessa cadeia para uma ampla família de ensembles (mais larga do que consideramos neste trabalho), e mostraram que, para todos esses casos, o *gap* espectral Δ da cadeia tem a mesma ordem em n.

A partir daí, porém, seus resultados se apóiam na análise detalhada do tempo de mistura t_{mix} da cadeia no caso de um ensemble em particular, a saber o uniforme. Com uma análise extremamente rebuscada, eles conseguiram demonstrar que a cadeia de peso de Hamming (cadeia Z) para esta cadeia converge (no sentido de norma TV) em $O(n \log(n/\varepsilon))$ passos. Em seguida, argumentaram, usando um argumento do tipo *coupon collector*, que isso implicaria a convergência da cadeia P como um todo nesta mesma ordem em n. Sendo este resultado verdade para o ensemble uniforme, eles mostraram que a equivalência em gaps mencionada acima pode ser usada para ver que, para todos os outros ensembles vale o resultado mais fraco de que a convergência (no sentido tanto da norma TV quanto da

quadrática) ocorre em $O(n(n+\log(\varepsilon^{-1}))$ passos.

Finalmente, usando este último resultado HL mostraram que, para todos os ensembles, a convergência para um 2-desenho ε-aproximado ocorre em $O(n(n+\log(\varepsilon^{-1}))$ passos. No sentido de um 2-desenho *de canal*, que depende da norma TV e não da norma quadrática, a convergência ocorreria em $O(n\log(n/\varepsilon))$ passos no caso do ensemble uniforme, devido à cota mais justa obtida neste caso para a cadeia de Markov.

Neste contexto, podemos fazer o seguinte resumo e comentários gerais sobre os nossos resultados

- Começamos mostrando que, para uma larga classe de ensembles que denominamos 'localmente invariantes', a evolução dos segundos momentos é descrita exatamente por uma cadeia de Markov (sem termos que decaem exponencialmente) (Seção 3.1). Esta cadeia tem, para todos os ensembles desta categoria, uma mesma forma geral simples, dependendo de apenas dois parâmetros (Capítulo 4). Estudamos como esses parâmetros se comportam para diversos exemplos de ensembles localmente invariantes, inclusive muitos que são mais convenientes à implementação experimental do que o ensemble uniforme (Seção 4.2).

- Em seguida, apontamos uma falha relevante na demonstração de [27] descrita acima, a saber na passagem da convergência da cadeia Z para a P por um argumento de *coupon collector* simples, que está incorreta. Por implicação, o restante do argumento ficaria comprometido.

- Para contornar este problema, fizemos (Seção 5.2) uma derivação independente de uma cota de ordem $O(n\log(n/\varepsilon))$ para o tempo de mistura da cadeia P. O argumento fez uso de uma decomposição favorável da cadeia original em uma combinação convexa de cadeias comutantes (eq. (4.16)), a qual pôde ser reduzida ainda a uma cadeia mais simples L do tipo 'campo médio' (eq. (4.19)), onde no máximo uma coordenada evolui em cada passo. Neste caso pudemos aplicar uma técnica de acoplamento para encontrar o tempo de mistura (Seção 5.3).

- Este resultado não só confirma aquele alegado em [27] para o ensemble uniforme, mas estende-o para todos os ensembles localmente invariantes (com $b>0$) (melhorando, portanto, a cota para o tempo de mistura nestes casos). Da mesma forma, automaticamente todos os resultados mais fortes com respeito à convergência a 2-desenhos de canal podem igualmente ser estendidos para essa classe mais ampla de ensembles. Ainda, a dependência detalhada da cadeia com os parâmetros a e b permite estabelecermos uma hierarquia parcial dos ensembles localmente invariantes, identificando aqueles para os quais a convergência deve ser mais rápida (Seção 5.4).

- Nossos resultados ainda se apóiam na longa derivação feita em [27] para o tempo de mistura da cadeia Z. Usando técnicas de passeios aleatórios em grupos, apresentamos um argumento, infelizmente não formalizado, que se correto permitiria reduzir parcialmente esta dependência (Seção 5.5).

- Finalmente, mostramos, por argumentos parcialmente analíticos e parcialmente numéricos, que explorando o paralelismo natural do modelo de CQA, é possível reduzir os tempos de convergência obtidos anteriormente por um fator adicional de ordem n. Em outras palavras, com isso pode-se obter para os ensembles localmente invariantes um tempo de convergência de ordem $O(log(n/\varepsilon))$ para um 2-desenho de canal ε-aproximado, ou de ordem $O(n + \log(1/\varepsilon))$ para um 2-desenho ε-aproximado.

7.2 Extensões possíveis

7.2.1 Argumentos para *cutoff* em L a partir de cadeias de Ising?

O formato da cadeia L (eq. (4.19)) permite interpretá-la como uma cadeia de *campo médio*. Em cada passo da cadeia, apenas um sítio, uniformemente escolhido, evolui, e esta evolução é dada por uma cadeia local cujas probabilidades dependem de uma propriedade global do estado, o peso de Hamming H.

Esta situação lembra a que ocorre na chamada cadeia de Ising, uma versão para cadeias de Markov do modelo de Ising clássico (não quântico) da Física Estatística. Essa cadeia pode ser definida de seguinte maneira (para a definição e detalhes, veja Seção 3.3.5 de [32]): o espaço de estados é dado por vetores σ com n componentes, cada uma igual a 1 ou -1 (esses valores correspondem a um 'spin para cima' ou 'para baixo' em cada sítio)[1]. Em cada passo da cadeia, escolhe-se um sítio v uniformemente, e define-se o novo valor ± 1 para aquele spin com probabilidades

$$\mathbb{P}(\pm 1) = \frac{1}{2}\left(1 \mp \tanh \beta S(\sigma)\right)$$

onde β representa a temperatura inversa da cadeia, e

$$S(\sigma, v) = \sum_{w \sim v} \sigma_w$$

é a magnetização do estado σ em todos os sítios adjacentes a v (i.e., cujos spins interagem com o spin em v). Para fazer a analogia apropriada com L (na qual assumimos o grafo

[1] Note que há uma ligeira diferença com respeito à cadeia L, onde usamos valores 1 ou 0

Γ completo), vamos assumir que todos os pares interagem (inclusive cada spin com si próprio). Nesse caso S é a magnetização global do sistema, e traduzindo para o espaço de estados da cadeia L, temos $S = 2H - n$.

Note que, no regime de 'altas temperaturas' da cadeia de Ising ferromagnética, que tomamos como $\beta \ll 1/n$, podemos aproximar $\tanh(x) \sim x \ll 1$ na expressão acima, e as probabilidades de obter 0 e 1 ficam ambas aproximadamente iguais a $1/2$. Nesse caso é natural supor que a cadeia misture rápido, e efetivamente sabe-se que o tempo de mistura é $\Theta(n \log n)$ [37]. Por outro lado, no limite de 'temperatura baixa' $\beta \gg 1/n$, a probabilidade de obter spin para cima é muito baixa ($\mathbb{P}(1) \sim e^{-\beta S} \ll 1$), independente da magnetização. Assim, o tempo necessário para 'virar' um número apreciável de spins de baixo para cima deve ser muito grande. Efetivamente sabe-se que ele fica exponencial em n (há na verdade uma transição de fase de um regime para o outro em $\beta = 1/n$).

No caso da cadeia L, a eq. (4.19) implica que, para condições iniciais peso de Hamming baixo, a probabilidade de obter um 'um' é muito pequena, o que lembra o comportamento da cadeia de Ising a baixa temperatura. No entanto, a analogia não é boa, pois há uma diferença crucial: aqui a probabilidade de obter 1 aumenta rapidamente à medida em que H vai aumentando, o que acaba acelerando a cadeia após o período lento inicial. Comparando-se com a cadeia de Ising , é como se a 'temperatura' da cadeia L fosse aumentando à medida em que H cresce. Nessas condições não é óbvio o que vai acontecer com o tempo de mistura. O que o longo argumento de HL para a cadeia Z faz é essencialmente mostrar que a fase de 'temperatura baixa' dura suficientemente pouco para que o seu tempo de mistura continue de ordem $O(n \log n)$. Como vimos, porém, seu argumento, além de longo não se aplica diretamente à cadeia L, necessitando de mais trabalho.

Existe uma enorme quantidade de trabalho feito sobre cadeias de Ising, incusive recentemente foram obtidas informações detalhadas sobre o *cutoff* nos vários regimes de temperatura [38]. É concebível que algumas das técnicas usadas nesses estudos possa ser adaptada para a cadeia L, fornecendo tanto uma prova mais simples do seu tempo de mistura, como também uma prova de *cutoff* nesta cadeia. Para isto usaria-se um início de demonstração semelhante ao feito por Harrow e Low, até sair da fase de 'temperatura baixa', e então continuaria-se a demonstração usando as técnicas adaptadas.

7.2.2 Convergência a k-desenhos de ordem $k \geq 3$

Como mencionamos na Introdução, foi conjecturado em [27] e há evidências numéricas [31] de que CQAs também convirjam em tempo polinomial para k-desenhos aproximados com valores de $k \geq 3$. Se verdadeira para todo k, esta seria a primeira construção eficiente deste tipo.

A princípio, esta questão pode ser investigada de forma semelhante ao feito para o caso

$k = 2$, ou seja, estudando a evolução dos momentos de ordem k na base de Pauli gerados pelo CQA. Conforme descrito na Seção 3.2, esta evolução depende dos elementos de matriz do super-operador de k-*twirl* $\hat{G}_{\mu}^{(k)}$ definido na eq. (2.11). Nesta Seção, esboçamos como seria este cálculo para o caso $k = 3$, apontando algumas diferenças significativas para o caso $k = 2$ já estudado.

Precisamos calcular os elementos de matriz $\langle\langle \sigma_{\vec{q},\vec{q}\,',\vec{q}\,''} | \hat{G}^{(3)}_{\mu_{ij}} | \sigma_{\vec{p},\vec{p}\,',\vec{p}\,''} \rangle\rangle$. Se μ é o ensemble uniforme $\mathcal{H}$, o Teorema 2.3.1 garante que o super-operador $\hat{G}^{(3)}_{\mathcal{H}_{ij}}$ é o projetor no subespaço $\mathcal{S}_{3,2}$ gerado pelos operadores que permutam três cópias do espaço de dois q-bits. Para simplificar a discussão, vamos nos ater a este caso.

O primeiro problema então é obter estes projetores; para isto, temos de obter uma base ortonormal a partir dos 6 elementos do grupo de permutações de três cópias $S_3 = \{I, (12), (13), (23), (123), (132)\}$. Aplicando o mesmo procedimento de Gram-Schmidt feito na eq. (2.17) às 3 transposições, nós obtemos os operadores mutuamente ortogonais:

$$\bar{S}^{(12)} \equiv 2^n S^{(12)} - I = \sum_{\vec{p}\neq\vec{0}} \sigma_{\vec{p},\vec{p},\vec{0}} \,, \tag{7.1}$$

$$\bar{S}^{(13)} \equiv 2^n S^{(13)} - I = \sum_{\vec{p}\neq\vec{0}} \sigma_{\vec{p},\vec{0},\vec{p}} \,, \tag{7.2}$$

$$\bar{S}^{(23)} \equiv 2^n S^{(23)} - I = \sum_{\vec{p}\neq\vec{0}} \sigma_{\vec{0},\vec{p},\vec{p}} \,. \tag{7.3}$$

O 3-ciclo $(123) = (13)(12)$ é o produto de duas transposições. Usando a eq. (2.15), ele é representado nesse caso por

$$S^{(123)} = S^{(13)} S^{(12)} = \frac{1}{2^{2n}} \sum_{\vec{p},\vec{q}} \sigma_{\vec{p}} \sigma_{\vec{q}} \otimes \sigma_{\vec{q}} \otimes \sigma_{\vec{p}} \,.$$

Removendo da soma os termos com: $\vec{p},\vec{q} = 0; \vec{p} = 0, \vec{q} \neq 0; \vec{p} \neq 0, \vec{q} = 0; \vec{p} = \vec{q} \neq 0$, podemos reescrever isso como

$$S^{(123)} = \frac{1}{2^{2n}} \left(I + \bar{S}^{(12)} + \bar{S}^{(13)} + \bar{S}^{(23)} \right) + \sum_{\substack{\vec{p}\neq\vec{q} \\ \vec{p},\vec{q}\neq\vec{0}}} \sigma_{\vec{p}} \sigma_{\vec{q}} \otimes \sigma_{\vec{q}} \otimes \sigma_{\vec{p}} \,. \tag{7.4}$$

Pode ser facilmente checado então que

$$\bar{S}^{(123)} \equiv S^{(123)} - \frac{1}{2^{2n}} \left(I + \bar{S}^{(12)} + \bar{S}^{(13)} + \bar{S}^{(23)} \right) = \sum_{\substack{\vec{p}\neq\vec{q} \\ \vec{p},\vec{q}\neq\vec{0}}} \sigma_{\vec{p}} \sigma_{\vec{q}} \otimes \sigma_{\vec{q}} \otimes \sigma_{\vec{p}} \tag{7.5}$$

é ortogonal a $I, \bar{S}^{(12)}, \bar{S}^{(13)}, \bar{S}^{(23)}$. Usando eq. (2.4), é imediato ver que sua norma quadrada é

$$\left\langle\!\left\langle \bar{S}^{(123)} | \bar{S}^{(123)} \right\rangle\!\right\rangle = \sum_{\substack{\vec{p}\neq\vec{q} \\ \vec{p},\vec{q}\neq\vec{0}}} \mathrm{Tr}\left[\sigma_{\vec{q}}\sigma_{\vec{p}}\sigma_{\vec{p}}\sigma_{\vec{q}} \otimes I \otimes I\right] = 2^{3n}\left(2^{2n}-1\right)\left(2^{2n}-2\right) \ . \tag{7.6}$$

Considere agora o 3-ciclo restante $(132) = (12)(13)$. Seguindo os mesmos passos, podemos ver que ele é representado por

$$S^{(132)} = \frac{1}{2^{2n}}\left(I + \bar{S}^{(12)} + \bar{S}^{(13)} + \bar{S}^{(23)}\right) + \sum_{\substack{\vec{p}\neq\vec{q} \\ \vec{p},\vec{q}\neq\vec{0}}} \sigma_{\vec{q}}\sigma_{\vec{p}} \otimes \sigma_{\vec{q}} \otimes \sigma_{\vec{p}};\, .$$

Note que a única diferença com respeito a eq. (7.4) é a ordenação do produto no último termo. Portanto, se defirnirmos analogamente

$$\bar{S}^{(132)} \equiv S^{(132)} - \frac{1}{2^{2n}}\left(I + \bar{S}^{(12)} + \bar{S}^{(13)} + \bar{S}^{(23)}\right) = \sum_{\substack{\vec{p}\neq\vec{q} \\ \vec{p},\vec{q}\neq\vec{0}}} \sigma_{\vec{q}}\sigma_{\vec{p}} \otimes \sigma_{\vec{q}} \otimes \sigma_{\vec{p}} \ , \tag{7.7}$$

nós novamente obtemos um oerador ortogonal aos primeiros quatro, e com a mesma norma que $\bar{S}^{(123)}$. Nós podemos então completar nossa base de seis operadores mutuamente ortogonais tomando a soma e a diferença desse dois:

$$\bar{S}^{(+)} \equiv \bar{S}^{(132)} + \bar{S}^{(123)} = \sum_{\substack{\vec{p}\neq\vec{q} \\ \vec{p},\vec{q}\neq\vec{0}}} \left\{\sigma_{\vec{q}}, \sigma_{\vec{p}}\right\} \otimes \sigma_{\vec{q}} \otimes \sigma_{\vec{p}} \ , \tag{7.8}$$

$$\bar{S}^{(-)} \equiv \bar{S}^{(132)} - \bar{S}^{(123)} = \sum_{\substack{\vec{p}\neq\vec{q} \\ \vec{p},\vec{q}\neq\vec{0}}} \left[\sigma_{\vec{q}}, \sigma_{\vec{p}}\right] \otimes \sigma_{\vec{q}} \otimes \sigma_{\vec{p}} \ . \tag{7.9}$$

Um caso especial ocorre para $n = 1$: nesse caso, os operadores de Pauli $\sigma_p \neq \sigma_q$ que são ambos $\neq \sigma_0$ satisfazem as relações de anti-comutação $\sigma_p\sigma_q = -\sigma_q\sigma_p = i\epsilon_{rpq}\sigma_r$, onde ϵ_{rpq} é o símbolo de Levi-Civita anti-simétrico. Portanto o operador $\bar{S}^{(+)}$ some, e

$$\bar{S}^{(-)} = 2\bar{S}^{(132)} = 2i \sum_{\substack{p\neq q \\ p,q\neq 0}} \epsilon_{rpq}\sigma_{r,p,q} \ . \tag{7.10}$$

Nesse caso os operadores de permutação de cópia não são linearmente independentes, gerando um subespaço com dimensão 5 em vez de 6.

Podemos observar agora dois fatos importante: em primeiro lugar, cada um dos operadores da base ortogonal obtida acima é uma soma uniforme sobre um certo conjunto de operadores da base de Pauli. Ainda, pode-se notar que os operadores que aparecem em cada uma destas somas *não* aparecem nas outras. Isto significa então que, na base de Pauli, *a matriz de* $\hat{G}^{(3)}_{\mathcal{H}_{ij}}$ *pode ser escrita na forma bloco-diagonal*, sendo cada um dos seis blocos da forma $\frac{1}{m}F_m$, onde F_m é a matriz $m \times m$ contendo 1 em todas as suas entradas. (Compare com a eq. (3.21) para $\hat{G}^{(2)}_{\mathcal{H}_{ij}}$). Em outras palavras, assim como no caso $k = 2$, a matriz do $\sum_{i \neq j} \hat{G}^{(3)}_{\mathcal{H}_{ij}}$ será novamente uma matriz de Markov bi-estocástica. É razoável conjecturar que o mesmo ocorrerá para qualquer k, algo que deve ser demonstrável utilizando propriedades gerais da dualidade Schur-Weyl [39].

Não é óbvio, porém, que se possa utilizar diretamente técnicas de cadeias de Markov para avaliar a convergência neste caso, por um motivo simples. Ao contrário do caso $k = 2$, onde os únicos momentos não-nulos $\langle \xi_t^2(\vec{p}) \rangle$ formavam automaticamente uma distribuição de probabilidade quando o estado é puro, para $k = 3$ a matriz de Markov atua em vetores cujos elementos, da forma $\langle \xi_t(\vec{p}) \xi_t(\vec{p}\,') \xi_t(\vec{p}\,'') \rangle$ não somam 1, e podem em geral ter componentes negativas. Assim, apesar da matriz de Markov ser bem-comportada, não é evidente, a princípio, se os resultados usuais de convergência de cadeias de Markov sejam adaptáveis a este caso. Mesmo que sejam, fica ainda naturalmente o problema de encontrar cotas para a convergência. Como se viu, já no caso $k = 2$ este problema se revelou não-trivial, então a não ser que se consiga um método mais simples de análise, pode ser difícil generalizar o resultado para valores superiores de k. Vale notar que já seria um resultado interessante apenas obter que a convergência da cadeia é polinomial (ou seja, sem necessariamente obter uma cota justa).

Apêndice A

Cadeias de Markov Preguiçosas

Uma cadeia de Markov P^L é dita a versão preguiçosa de P se $P^L = \frac{\mathbb{I}+P}{2}$. É comum se ver afirmado (e.g., em [41]) que o tempo de mistura de uma cadeia preguiçosa é o dobro da sua versão original. Como comentamos na Seção 5.2, não encontramos uma demonstração rigorosa deste fato em livros-texto de cadeias de Markov. Veremos que o tempo dobrado não é uma regra geral, entretanto poderemos mostrar que esse é um limite superior para o tempo de mistura da cadeia preguiçosa.

Antes de analisar o caso geral vamos tentar entender o que acontece num caso simples.

Considere uma cadeia C cujo espaço de estado é formado pela cara e a coroa de uma moeda. Em cada passo essa cadeia muda a face virada para cima com probabilidade $\delta \leq 1$.

$$C = \begin{pmatrix} 1-\delta & \delta \\ \delta & 1-\delta \end{pmatrix} \qquad \text{autovetores} : \begin{cases} \pi = \frac{1}{2}(1\ ,\ 1) & \leftrightarrow \beta_0 = 1\ , \\ \pi_1 = \frac{1}{2}(1\ ,\ -1) & \leftrightarrow \beta_1 = 1-2\delta\ . \end{cases}$$

A versão preguiçosa de C, que chamaremos C^L só difere pelo autovalor $\beta_1^L = 1-\delta$. Assim a distribuição de probabilidade $(\mu(t))$ após t passos de C é dada por: $\mu(t) = \mu(o)P^t = \pi + c_1\pi_1\beta_1^t$, onde a condição inicial é definida pela constante $|c_1| \leq 1$. Tomaremos $c_1 = 1$ por simplicidade. Assim a distância TV à distribuição estacionária, Seção 2.5.2, será $d(t) = \frac{1}{2}|\beta_1^t|$ e o tempo de mistura será,

$$t_{mixC} = \left\lceil \frac{\ln(1/2)}{\ln(\beta_1)} \right\rceil \quad \text{e} \quad t_{mixC^L} = \left\lceil \frac{\ln(1/2)}{\ln(\beta_1^L)} \right\rceil\ .$$

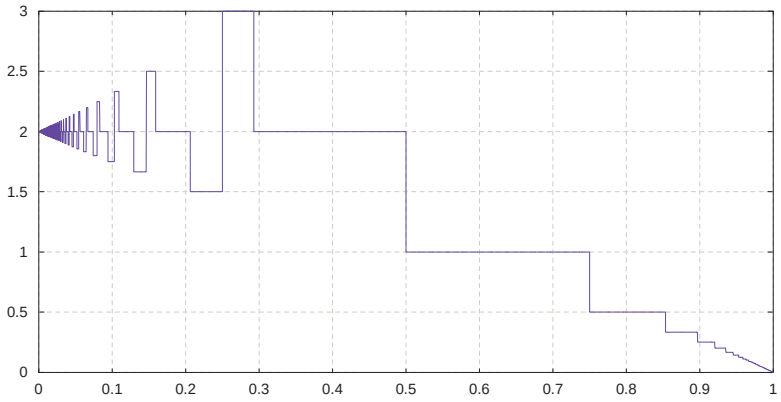


Figura A.1: Razão entre os tempos de mistura (t_{mixC^L}/t_{mixC}) em função do parâmetro δ.

No limite $\delta \to 0$ temos para a razão dos tempos de mistura $\simeq 2$, o que é compatível com a intuição. Por outro lado, no limite $\delta \to 1$, o tempo de mistura da cadeia preguiçosa é muito menor do que a da cadeia original. Isso não é inesperado, pois uma cadeia com gap espectral, veja Seção 2.5.2, $\Delta \to 0$ e $\beta_1 < 0$ é um cadeia que quase sempre se mexe e está tendendo para a aperiodicidade, em outras palavras, se tivermos $\delta = 1$ a moeda sempre muda de face e se ela começa com cara para cima sempre terá cara para cima a cada duas rodadas e nunca chega a uma distribuição estacionária. Quando esse tipo de cadeia é tornada preguiçosa é natural que o tempo de mistura diminua. A interpretação acima se mantém mesmo para cadeias maiores e portanto é necessário que o *gap* venha de autovalores positivos para que o tempo mistura das duas cadeias seja comparável. No caso geral podemos afirmar que:

Teorema A.0.1. *Para qualquer cadeia de Markov P sua versão preguiçosa P^L tem o tempo de mistura limitado por $2t_{mixP}$.*

Prova : Suponha que iniciarmos a cadeia P^L no estado μ. Após t passos, a distância TV para o estado estacionário vale

$$\begin{aligned} d(t) &= \left\|\mu(P^L)^t - \pi\right\|_{TV} \\ &= \left\|\mu \sum_{i=0}^{t} \binom{t}{i} (1/2)^t \, \mathbb{I}^{t-i} \, P^i - \pi\right\|_{TV} \\ &\leq \sum_{i} \binom{t}{i} (1/2)^t \left\|\mu P^i - \pi\right\|_{TV} \end{aligned} \tag{A.1}$$

onde usamos a expansão binomial e a desigualdade triangular. Para colocar uma cota nesta última expressão, recordamos que a distribuição binomial é fortemente concentrada em $i = t/2$, e tem desvio padrão $\sigma = O(\sqrt{t})$. Separemos então a eq. (A.1) em duas somas $d_1(t), d_2(t)$, contendo respectivamente os termos com $i < t/2-\delta$ e $i \geq t/2-\delta$, com $\delta < t/2$.

Como a distância TV com respeito ao estado estacionário de uma cadeia de Markov M é não-crescente em cada passo, todos as distâncias TV em d_2 são limitadas pela do termo para $i = t/2 - \delta$,

o qual por sua vez é no máximo igual ao que seria sem a presença de $\Gamma^{t/2-\delta}$:

$$d_2 \leq \left[\sum_{i \geq t/2-\delta} \binom{t}{i} (1/2)^t \right] \left\| \mu M^i - \pi \right\|_{TV} \leq \left\| \mu M^{t/2-\delta} - \pi \right\|_{TV}$$

onde usamos ainda que a soma em colchetes é ≤ 1.

Por outro lado, para d_1 podemos primeiro usar o fato de que distâncias TV entre distribuições de probabilidade são sempre ≤ 1, de modo que

$$d_1 \leq \sum_{i < t/2-\delta} \binom{t}{i} (1/2)^t.$$

Ainda, podemos explorar o fato de que os termos envolvidos estão na cauda da distribuição binomial, a qual pode ser aproximada usando diversas desigualdades. Por exemplo, usando a desigualdade de Hoeffding [57] podemos escrever

$$d_1 \leq \exp(-2\delta^2)$$

de modo que:

$$d(t) \leq \exp(-2\delta^2) + \left\| \mu M^{t/2-\delta} - \pi \right\|_{TV} \ . \tag{A.2}$$

Se escolhermos o primeiro instante t para o qual $\left\| \mu M^{t/2-\delta} - \pi \right\|_{TV} \leq 1/4 - e^{-2\delta^2}$, temos então, pela definição de tempos de mistura (veja Seção 2.5.2), que

$$t_{mix P^L}(\varepsilon) \leq 2 t_{mix P} \left(\varepsilon - e^{-2\delta^2} \right) + 2\delta \ , \tag{A.3}$$

agora se escolhemos $\delta = \sqrt{log(1/\varepsilon)}$,

$$t_{mix P^L}(\varepsilon) \leq 2 t_{mix P} \left(\varepsilon - \varepsilon^2 \right) + 2\sqrt{log(1/\varepsilon)} \ . \tag{A.4}$$

A diferença entre chegar a uma distância ε ou $\varepsilon(1-\varepsilon)$ é desprezível, assim como a correção $\sqrt{log(1/\varepsilon)}$ já que o termo $t_{mix P}$ cresce com o tamanho do espaço de fase em que a cadeia passeia.

Se ao invés de combinar uma matriz de Markov com a identidade fizermos a combinação com outra matriz de Markov que comute com a primeira e tenha o mesmo estado estacionário os resultados acima podem ser trivialmente generalizados. Se as matrizes A e B tem essa propriedade e são combinadas formando $P = \frac{A+B}{2}$ temos que

$$t_{mixP} \leq 2t_{mixA} \ ,$$

ou,

$$t_{mixP} \leq 2t_{mixB} \ ,$$

sob as mesmas condições discutidas acima, usando o fato de que um passo de A ou B nunca aumenta a distância a estacionariedade.

Finalmente vamos considerar a combinação $P = aA + (1-a)B$, onde $a \in [0,1]$. Com isso os tempos de misturas serão dadas por:

$$t_{mixP} \leq \frac{1}{a} t_{mixA} \ ,$$

ou,

$$t_{mixP} \leq \frac{1}{1-a} t_{mixB} \ .$$

Referências Bibliográficas

[1] K. Binder and D.W. Heermann. *Monte Carlo simulation in statistical physics: an introduction*. Springer Verlag, 2002.

[2] D.W. Heermann. Computer simulation methods: in theoretical physics. 1986.

[3] H. Goult and J. Tobochnik. *An introduction to computer simulation methods: applications to physical systems*. Addison-Wesley, Reading, MA, 1996.

[4] T.J. Stapko. *Practical embedded security: building secure resource-constrained systems*. Newnes, 2008.

[5] D.E. Knuth. The art of computer programming, Vol. 3. *Reading, MA*, 1973.

[6] DJ Berkeland, DA Raymondson, and VM Tassin. Tests for nonrandomness in quantum jumps. *Physical Review A*, 69(5):52103, 2004.

[7] D. Zuckerman. Pseudorandomness and combinatorial constructions. Lecture Notes, 2001. Univ. Texas Austin.

[8] S. Bornholdt and H.G. Schuster. *Handbook of graphs and networks: From the genome to the Internet*. Vch Verlagsgesellschaft Mbh, 2003.

[9] R.M. Karp. Reducibility among combinatorial problems. *Complexity of computer computations*, 43:85–103, 1972.

[10] D.E. Knuth. Big omicron and big omega and big theta. *ACM Sigact News*, 8(2):24, 1976.

[11] MA Nielsen and I.L. Chuang. *Quantum computing and quantum information*. Cambridge University Press, Cambridge, 2000.

[12] A. Ambainis and J. Emerson. Quantum t-designs: t-wise independence in the quantum world. In *Proceedings of the 22nd Annual IEEE Conference on Computational Complexity*, pages 129–140, 2007.

[13] P. Hayden, D. Leung, P.W. Shor, and A. Winter. Randomizing quantum states: Constructions and applications. *Communications in Mathematical Physics*, 250(2):371–391, 2004.

[14] J. Emerson, R. Alicki, and K. Zyczkowski. Scalable noise estimation with random unitary operators. *prevention*, 3:8, 2005.

[15] A. Bendersky, F. Pastawski, and J.P. Paz. Selective and efficient estimation of parameters for quantum process tomography. *Physical Review Letters*, 100(19):190403, 2008.

[16] J.M. Renes, R. Blume-Kohout, AJ Scott, and C.M. Caves. Symmetric informationally complete quantum measurements. *Journal of Mathematical Physics*, 45:2171, 2004.

[17] P. Hayden, D.W. Leung, and A. Winter. Aspects of generic entanglement. *Communications in Mathematical Physics*, 265(1):95–117, 2006.

[18] O.C.O. Dahlsten, R. Oliveira, and M.B. Plenio. The emergence of typical entanglement in two-party random processes. *Journal of Physics A-Mathematical and Theoretical*, 40:8081–8108, 2007.

[19] A. Harrow, P. Hayden, and D. Leung. Superdense coding of quantum states. *Physical review letters*, 92(18):187901, 2004.

[20] C.H. Bennett, D.P. DiVincenzo, J.A. Smolin, and W.K. Wootters. Mixed-state entanglement and quantum error correction. *Physical Review A*, 54(5):3824–3851, 1996.

[21] A.W. Harrow and R.A. Low. Efficient Quantum Tensor Product Expanders and k-Designs. In *Proceedings of the 12th International Workshop and 13th International Workshop on Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, page 561. Springer, 2009.

[22] J. Emerson, Y.S. Weinstein, M. Saraceno, S. Lloyd, and D.G. Cory. Pseudo-random unitary operators for quantum information processing, 2003.

[23] J. Emerson, E. Livine, and S. Lloyd. Convergence conditions for random quantum circuits. *Physical Review A*, 72(6):60302, 2005.

[24] R. Blatt H. Haeffner, C.F. Roos. Quantum computing with trapped ions. *Physics Reports*, 469(4):115–203, December 2008.

[25] D. J. Wineland, M. Barrett, J. Britton, J. Chiaverini, B. DeMarco, W. M. Itano, Jelenkovi B., C. Langer, D. Leibfried, V. Meyer, T. Rosenband, and T. Schtz. Quantum information processing with trapped ions. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 361(1808):1349–1361, July 2003.

[26] J. Calsamiglia, L. Hartmann, W. D
"ur, and H.J. Briegel. Entanglement and decoherence in spin gases. *Arxiv preprint quant-ph/0502017*, 2005.

[27] A.W. Harrow and R.A. Low. Random quantum circuits are approximate 2-designs. *Communications in Mathematical Physics*, 291(1):257–302, 2009.

[28] M. Žnidarič. Optimal two-qubit gate for generation of random bipartite entanglement. *Physical Review A*, 76(1):12318, 2007.

[29] M. Žnidarič. Exact convergence times for generation of random bipartite entanglement. *Physical Review A*, 78(3):32324, 2008.

[30] C. Dnakert. Efficient simulation of random quantum states and operators. Master's thesis, University of Waterloo, 2005.

[31] L. Arnaud and D. Braun. Efficiency of producing random unitary matrices with quantum circuits. *Physical Review A*, 78(6):62329, 2008.

[32] D.A. Levin, Y. Peres, and E.L. Wilmer. *Markov Chains and Mixing Times: With a Chapter on Coupling from the Past by James G. Propp and David B. Wilson*. Amer Mathematical Society, 2008.

[33] Y.S. Weinstein, W.G. Brown, and L. Viola. Parameters of pseudorandom quantum circuits. *Physical Review A*, 78(5):52332, 2008.

[34] P. Diaconis. The cutoff phenomenon in finite Markov chains. *Proc. Nat. Acad. Sci. USA*, 93(4):1659–1664, 1996.

[35] D. Aldous and P. Diaconis. Shuffling cards and stopping times. *The American Mathematical Monthly*, 93(5):333–348, 1986.

[36] P. Diaconis. *Group representations in probability and statistics*. Institute of Mathematical Statistics Hayward, CA, 1988.

[37] J. Ding, E. Lubetzky, and Y. Peres. The mixing time evolution of Glauber dynamics for the mean-field Ising model. *Communications in Mathematical Physics*, 289(2):725–764, 2009.

[38] D.A. Levin, M.J. Luczak, and Y. Peres. Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability. *Probability Theory and Related Fields*, pages 1–43.

[39] W. Fulton and J. Harris. *Representation theory: A first course*. Springer, 2004.

[40] A.Y. Kitaev, A. Shen, and M.N. Vyalyi. *Classical and quantum computation*. Amer Mathematical Society, 2002.

[41] R. Montenegro and P. Tetali. *Mathematical aspects of mixing times in Markov chains*. Now Pub, 2006.

[42] D.P. DiVincenzo, D.W. Leung, and B.M. Terhal. Quantum Physics Title: Quantum Data Hiding. *Journal reference: IEEE Trans. Inf Theory Vol*, 48(3):580–599, 2002. 2-design, algoritmo com n2 passos quant e n8 passos classicos.

[43] R. Bhatia. *Matrix analysis*. Springer Verlag, 1997.

[44] B. Kraus and JI Cirac. Optimal creation of entanglement using a two-qubit gate. *Physical Review A*, 63(6):62309, 2001.

[45] N. Khaneja, R. Brockett, and S.J. Glaser. Time optimal control in spin systems. *Physical Review A*, 63(3):32308, 2001.

[46] Y. Makhlin. Nonlocal properties of two-qubit gates and mixed states, and the optimization of quantum computations. *Quantum Information Processing*, 1(4):243–252, 2002.

[47] J. Zhang, J. Vala, S. Sastry, and K.B. Whaley. Minimum construction of two-qubit quantum operations. *Physical review letters*, 93(2):20502, 2004.

[48] AT Rezakhani. Characterization of two-qubit perfect entanglers. *Physical Review A*, 70(5):52313, 2004.

[49] D. Gottesman. *Stabilizer codes and quantum error correction*. PhD thesis, Caltech Ph. D, 1997.

[50] G. Vidal and CM Dawson. Universal quantum circuit for two-qubit transformations with three controlled-not gates. *Physical Review A*, 69(1):10301, 2004.

[51] J. Zhang, J. Vala, S. Sastry, and K.B. Whaley. Minimum construction of two-qubit quantum operations. *Physical review letters*, 93(2):20502, 2004.

[52] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Probability Theory and Related Fields*, 57(2):159–179, 1981.

[53] D. Aldous and J. Fill. Reversible Markov chains and random walks on graphs. *Book in preparation*, 2001.

[54] P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible Markov chains. *The Annals of Applied Probability*, 3(3):696–730, 1993.

[55] P. Diaconis and J.A. Fill. Strong stationary times via a new form of duality. *The Annals of Probability*, 18(4):1483–1522, 1990.

[56] Joseph Emerson Christoph Dankert, Richard Cleve and Etera Livine. Exact and approximate unitary 2-designs: Constructions and applications. *Arxiv preprint quant-ph/0606161v1*, 2006.

[57] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, pages 13–30, 1963.

[58] H.F. Jones. *Groups, representations, and physics*. Taylor & Francis, 1998.

[59] Jian Ding, Eyal Lubetzky, and Yuval Peres. Total-variation cutoff in birth-and-death chains. 2008.

[60] A.L. Barabási and H.E. Stanley. *Fractal concepts in surface growth*. Cambridge Univ Pr, 1995.