KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT INGENIEURSWETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK
Kasteelpark Arenberg 10, 3001 Leuven (Heverlee)

QUANTUM ENTANGLEMENT DISTILLATION IN THE STABILIZER FORMALISM

Promotoren:
Prof. dr. ir. B. De Moor
Dr. ir. J. Dehaene

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de ingenieurswetenschappen

door

Erik HOSTENS

October 2007

KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT INGENIEURSWETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK
Kasteelpark Arenberg 10, 3001 Leuven (Heverlee)

QUANTUM ENTANGLEMENT DISTILLATION IN THE STABILIZER FORMALISM

Jury:
Prof. dr. ir. J. Berlamont, voorzitter
Prof. dr. ir. B. De Moor, promotor
Dr. ir. J. Dehaene, copromotor
Prof. dr. M. Fannes
Prof. dr. ir. J. Vandewalle
Prof. dr. ir. F. Verstraete (Universität Wien)
Prof. dr. ir. K. Audenaert (University of London)

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de ingenieurswetenschappen

door

Erik HOSTENS

U.D.C. 530.145

October 2007

and He shall purify

© Katholieke Universiteit Leuven – Faculteit Ingenieurswetenschappen
Arenbergkasteel, B-3001 Heverlee (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotocopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm or any other means without written permission from the publisher.

D/2007/7515/100

ISBN 978-90-5682-866-0

Voorwoord

Vooreerst zou ik promotor Bart De Moor willen bedanken voor de unieke kans om dit onderzoek te voeren. Als academisch omnivoor heeft hij indertijd de QIT-groep op SISTA in het leven geroepen, wat geen evidentie is, gezien het toch wel exotische karakter van dit onderzoek binnen een ingenieursdepartement (op de vraag waarom luidde zijn antwoord: “It’s poetry!”). Het is nochtans gebleken dat alternatieve invalshoeken voor de ontginning van een dergelijk interdisciplinair niemandsland geen windeieren oplevert (eerder Columbus-eieren). Op dat vlak staat Barts aanpak haaks op de toenemende fragmentarisering van wetenschappelijk onderzoek.

Ik ben copromotor Jeroen Dehaene zeer dankbaar voor zijn bijstand in raad en daad, vooral op onderzoeks- maar ook op menselijk vlak. Begiftigd met een merkwaardig geometrisch inzicht in de meest abstracte zaken en de gave die ook voor anderen te verhelderen, is hij een uitstekende begeleider geweest. Ik moet zijn geduld dikwijls op de proef gesteld hebben door mijn chaotische inzichten onsamenhangend en met molenwiekende armen uit de doeken te doen, maar hij bleek beter in het ontwarren van mijn gedachtenkronkels dan ikzelf. Daarnaast is zijn kritische ingesteldheid (niet in het minst voor zijn eigen werk) een belangrijke maatstaf voor kwaliteitsvol onderzoek. Deze thesis vormt de kroon op een vruchtbare samenwerking.

Verder wil ik ook de andere voormalige leden van de QIT-groep danken, Maarten Van den Nest, Frank Verstraete en Koenraad Audenaert. Hoewel het QIT-verhaal een beetje eindigt zoals het liedje van de kleine negers (Frank ben ik nog net op de drempel tegengekomen op zijn weg naar buiten, Koenraad was al weg, en Jeroen en Maarten zijn uiteindelijk ook vertrokken naar de andere kant van respectievelijk het Arenbergkasteel en de Alpen), is het me een hele eer als hekkensluiter deel te mogen uitgemaakt hebben van deze zeer productieve groep en Frank en Koenraad onder deze omstandigheden opnieuw te mogen begroeten. Ik dank Maarten voor de collegialiteit en de menige diverterende discussies omtrent wetenschap, literatuur en filosofie, maar ook over het leven van alledag en uiteraard The Simpsons.

Dank ook aan de andere leden van de jury, Mark Fannes en Joos Vandewalle, voor het nalezen van dit proefschrift, en prof. Berlamont voor het willen voorzitten (tot tweemaal toe) van de doctoraatsverdediging.

Ik betuig mijn erkenning aan het Instituut voor de Aanmoediging van Innovatie door Wetenschap en Technologie in Vlaanderen (IWT-Vlaanderen), voor de financiering van mijn onderzoek.

I would like to acknowledge my Scottish sister-in-law Lorna Hutchinson, who, despite being pregnant as hell (and beyond), was so kind as to subject this text to thorough subbing. It must have been hard as a professional British linguist trying to ignore the numerous instances of hideous American spelling. Anyway, you did a splendid job, for which many, many thanks!

I am grateful to prof. Hans Briegel, Marc Hein and Wolfgang Dür for giving me the opportunity to visit the IQI research group in Innsbruck twice over the past few years. These were interesting stays during which I experienced the fruitful interaction in this group and had many instructive discussions.

Ik ben Ida Tassens, Ilse Pardon, Péla Noé, Bart Motmans en Lut Vander Bracht zeer erkentelijk voor hun ondersteunende rol binnen SISTA en ESAT. De KULeuven-administratie is vaak een gevecht tegen de bierkaai, en hun werk, onmisbaar voor ons administratieve nitwits, verdient een bloemetje.

Een doctoraat is soms een eenzame strijd. Daarentegen hing er een bijzonder fijne sfeer tussen de collega's op SISTA. Het was me een genoegen de werkuren (en ook daarbuiten) te delen met zulke waardevolle en interessante mensen, in de eerste plaats Tom en Ilse en mijn eilandgenoten Steven, Bart en Jeroen. I would also like to mention the many pleasant walks to and lunches in alma (in the company of a wide variety of nationalities), during which it is safe to say we discussed pretty much everything.

Ook buiten SISTA hebben een aantal mensen onrechtstreeks een gunstige invloed gehad, zij het in de vorm van ontspanning in de natuur, leuke babbels op café, muzikale intermezzi of louterende zuipfestijnen. Ik denk aan de zatselbijeenkomsten en citytrips met Bruno, Daan, Karel en Annelies, de fietsreisjes en andere uitstappen met Raf, Bram, Gerry en Peter, de alcoholische uitspattingen met de D&A-veteranen, de concerten met Soratea, de middeleeuwse anachronismen met Il Nostromo del Sogno en uiteraard de muzikale én andere evenementen met Octopus (Marieke en Kirsten, ik ben het niet vergeten). Ludo, merci voor het perfecte evenwicht tussen cultuur en anti-cultuur en het af en toe mogen crashen op de sofa.

Hierbij mag ik zeker mijn familie niet nalaten te bedanken: Jan en Ivo, voor het hart onder de riem en de vele mountainbiketochten waar we Dafke kapotreden, Dirk, voor het weerbericht en de entertainment op tijd en stond en het sporadisch bewusteloos-drinken, Hilde, voor de muzikale en spirituele noot en nonkel Jan, voor het onbegrensd altruïsme en de stilte van Staleikerheide.

Tot slot wil ik mijn ouders danken voor hun onvoorwaardelijke steun. Eigenlijk is dankbaarheid voor je ouders een soort absurde evidentie (wat betekent voor een boom de grond waarin hij staat?). Maar toch: bedankt, lieve ouders, voor alles.

Erik Hostens,
Leuven, september 2007

Nederlandse samenvatting

Distillatie van kwantumverstrengeling in het stabilisatorformalisme

Het onderwerp van dit proefschrift is de ontwikkeling van verstrengelingsdistillatieprotocols binnen het stabilisatorformalisme. Verstrengelingsdistillatieprotocols zijn methoden om, uitsluitend door middel van lokale operaties en klassieke communicatie, de kwantumverstrengeling die aanwezig is in kopieën van een gegeven kwantumtoestand te concentreren. Ze zijn zowel praktisch als theoretisch van belang. Enerzijds zijn ze een manier om de verstrengeling te zuiveren die nodig is voor bepaalde praktische doeleinden, wanneer enkel lokale operaties en klassieke communicatie (LOKC) toegelaten zijn voor de beoogde toepassing. Anderzijds kunnen ze een fundamenteler inzicht verschaffen in de eigenschappen van verstrengeling en de fysische grenzen van de manipulatie van verstrengeling.

Het stabilisatorformalisme is een wiskundig kader bestaande uit stabilisatortoestanden en Cliffordoperaties, belangrijke specifieke klassen van kwantumtoestanden en -operaties. We leggen de nadruk op de equivalente voorstelling van deze toestanden en operaties, naast de gebruikelijke groeptheoretische, in termen van binaire matrixalgebra. Dit ‘binaire beeld’ maakt een transparante en efficiënte beschrijving van verstrengelingsdistillatieprotocols mogelijk en geeft ons de gelegenheid om bestaande resultaten aanzienlijk te verbeteren. We geven een volledig overzicht van de eigenschappen van stabilisatortoestanden en Cliffordoperaties in het binaire beeld die relevant zijn met het oog op de ontwikkeling van verstrengelingsdistillatieprotocols.

We behandelen de distillatie van zowel twee-partijen- als meer-partijenverstrengeling. Distillatie komt meestal neer op de extractie van informatie door middel van LOKC zodat de entropie van de ingangstoestand afneemt en

het eindresultaat een zuivere toestand is. Bij twee-partijendistillatieprotocols onderscheiden we naast deze informatie-extractie een bijkomende afname van de entropie die voortkomt uit de lokale metingen in het protocol. De beschrijving in het binaire beeld stelt ons in staat om het onderliggend principe dat deze entropie-reductie veroorzaakt te verklaren en dit inzicht te gebruiken om betere protocols te ontwikkelen.

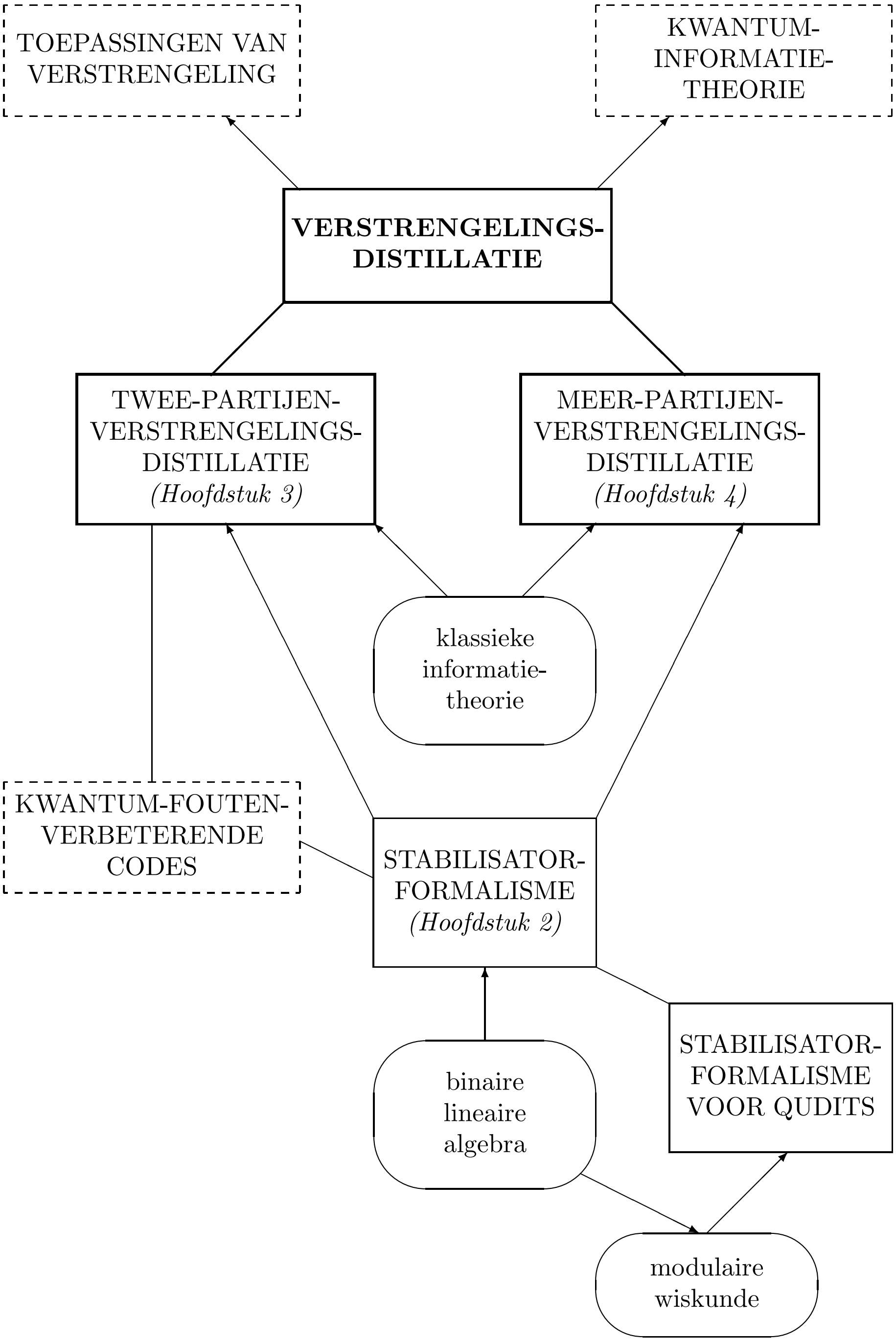
Voor meer-partijendistillatie verkrijgen we verbeteringen op twee niveaus. Ten eerste leiden we, door uitgebreid gebruik te maken van het binaire beeld, de meest algemene structuur af van de lokale Cliffordoperaties die worden gebruikt in het protocol, voor verschillende klassen stabilisatortoestanden. Ten tweede formuleren en exploiteren we bepaalde eigenschappen van de sterk-typische verzameling, een concept uit klassieke informatietheorie.

De structuur van onze resultaten [51, 52, 53, 54, 55] is schematisch weergegeven in figuur 0.1. Verstrengelingsdistillatie is belangrijk zowel voor praktische doeleinden (toepassingen in kwantumcryptografie, kwantumcomputing and kwantumcommunicatie) als voor een beter begrip van fundamentele eigenschappen van verstrengeling (zoals verstrengelingsmaten) in kwantuminformatietheorie. We onderscheiden distillatieprotocols voor twee of meerdere partijen.

De belangrijkste wiskundige technieken die hiervoor worden aangewend zijn klassieke informatietheorie en binaire lineaire algebra. De concepten informatiewinst, entropie-reductie en typische verzameling in ons werk op twee-partijenprotocols [53] zijn directe toepassingen van klassieke informatietheorie. Naast de beschrijving van twee-partijenprotocols in termen van stabilisatorcodes [51] is er een sterk verband tussen entropie-reductie en het bestaan van ontaarde codes, een belangrijk aspect van kwantumcodes zonder klassiek equivalent. We gebruiken het concept sterk-typische verzameling in het meer-partijengeval om de informatiewinst uit de metingen te maximaliseren [54, 55].

We beschrijven het wiskundig kader van het stabilisatorformalisme, dat oorspronkelijk werd ontwikkeld in de context van kwantum-foutencorrectie, volledig in termen van binaire-matrixoperaties. Dit laat ons toe om de entropie-reductie in twee-partijenprotocols uit te buiten en om de meest algemene structuur te vinden van de Cliffordoperaties in meer-partijenprotocols. Zo hebben we bestaande protocols aanzienlijk verbeterd en de winsten op die manier verkregen zijn tot op vandaag niet meer overtroffen.

Het artikel [52] maakt geen deel uit van dit proefschrift. In deze paper beschrijven we veralgemeningen van het stabilisatorformalisme voor willekeurige dimensies (qudits). We onderzoeken een verband met modulaire wiskunde, waarbij we ons in grote mate hebben gebaseerd op het werk voor qubits [24]. We hebben dit werk niet opgenomen in het proefschrift omdat het te ver afligt van verstrengelingsdistillatie en de tekst onnodig zou verzwaren.


TOEPASSINGEN VAN VERSTRENGELING
KWANTUM-INFORMATIE-THEORIE
VERSTRENGELINGS-DISTILLATIE
TWEE-PARTIJEN-VERSTRENGELINGS-DISTILLATIE (Hoofdstuk 3)
MEER-PARTIJEN-VERSTRENGELINGS-DISTILLATIE (Hoofdstuk 4)
klassieke informatie-theorie
KWANTUM-FOUTEN-VERBETERENDE CODES
STABILISATOR-FORMALISME (Hoofdstuk 2)
STABILISATOR-FORMALISME VOOR QUDITS
binaire lineaire algebra
modulaire wiskunde

Figuur 0.1: Structuur en samenhang van de resultaten.

Hoofdstuk 1: Inleiding

Kwantuminformatietheorie is een tamelijk nieuw onderzoeksgebied dat elementen bevat van verschillende takken, waaronder kwantummechanica, informatietheorie en computerwetenschappen. Centraal staat de studie van specifieke eigenschappen van kwantumsystemen, en toepassingen zoals kwantumcommunicatie, kwantumcryptografie en kwantumcomputers. In tegenstelling tot een klassiek systeem, schaalt de beschrijving van de toestand van een kwantumsysteem exponentieel op met de systeemgrootte. Daarom is het simuleren van de evolutie van een groot kwantumsysteem op een klassieke computer, zoals die vandaag bestaat, uiterst inefficiënt en in de praktijk onbegonnen werk. Die vaststelling bracht de bekende natuurkundige Richard Feynman op het idee om hiervoor ook kwantumsystemen te gebruiken [36]. Dat leidde tot het concept van de *kwantumcomputer* [26], en het simuleren van andere kwantumsystemen kan worden beschouwd als de eerste toepassing ervan. Een doorbraak kwam er in 1994, toen Peter Shor een *kwantumalgoritme* ontdekte om de priemfactoren of de discrete logaritme van een geheel getal te vinden met een tijdscomplexiteit die slechts polynomiaal opschaalt ten opzichte van de ingangsgrootte [78], terwijl nog geen polynomiaal klassiek algoritme is gevonden tot op heden. Twee jaar later ontwikkelde Lov Grover een kwantumalgoritme om een ongesorteerde databank te doorzoeken met een snelheid die in grootte-orde het kwadraat is van de snelheid van het beste klassieke algoritme [44]. De *qubit* is de kwantumtegenhanger van de klassieke bit en de standaard bouwsteen van een kwantumcomputer. Het is een kwantumsysteem met twee niveaus. We beschouwen hier geen praktische realisaties van qubits. Typische voorbeelden zijn de polarisatie van een foton en de spin van een elektron.

De *kwantumtoestand* bevat de informatie die we hebben over een bepaald kwantumsysteem. De enige manier waarop we aan die informatie kunnen geraken is door de waarneming van herhaalbare experimenten en de daarmee gepaard gaande kansen. Hoewel kansen ook bestaan in klassieke fysica, waar ze het gevolg zijn van onvolledige kennis van een systeem, zijn ze inherent aan de schijnbaar indeterministische aard van kwantummetingen. Een kwantumtoestand is ofwel *zuiver* ofwel *gemengd*. De laatstgenoemde kan worden gezien als een klassiek statistisch ensemble van zuivere toestanden: a.h.w. een mengsel van kwantumtoestand-'amplitudes' en klassieke kansen. Een zuivere toestand wordt wiskundig geïdentificeerd met een éénheidsvector in een complexe Hilbertruimte,[1] en een gemengde toestand met een positief semi-definiete Hermitische operator op dezelfde ruimte met spoor gelijk aan 1: de *dichtheidsoperator* ρ. De toestand bepaalt de uitkomstwaarschijnlijkheden van welke meting ook op het systeem. Terzelfdertijd veroorzaakt de meting een schijnbaar plotse verandering in de toestand van het systeem. Dit vreemde fenomeen wordt ook wel het *inklappen* van de toestandsvector of de golffunctie genoemd. De interpretatie ervan was en is nog steeds het onderwerp van menig controverse [76], maar we gaan hier niet dieper op in. Op een manier 'vernietigd'

[1] Voor een qubit is deze ruimte tweedimensionaal.

een meting de informatie die in de kwantumtoestand zit. Dit gedrag is nuttig voor encryptie, want een afluisteraar kan geen informatie winnen uit een verzonden kwantumboodschap noch ermee knoeien zonder dat het kan worden opgemerkt. In tegenstelling tot klassieke cryptografie, die gefundeerd is op de rekencomplexiteit van de oplossing van bepaalde wiskundige problemen, biedt *kwantumcryptografie* de mogelijkheid van willekeurige veiligheid [11, 37].

Een andere eigenschap die in klassieke fysica tot dusver nog niet was tegengekomen en die een centrale plaats in dit proefschrift inneemt, is *kwantumverstrengeling.* Het manifesteert zich in het gedrag van een samengesteld systeem als correlaties van meetuitkomsten die onverklaarbaar zijn in elke klassieke theorie. Een typisch voorbeeld is het experiment waarin een atoom vervalt via een intermediaire toestand naar de grondtoestand en zo twee verstrengelde fotonen uitstraalt: het meten van de polarisatie van één foton legt ogenblikkelijk de polarisatie van het andere foton vast. Albert Einstein deed verstrengeling af als een “spookachtige werking op afstand” en gebruikte het in de beruchte EPR-paper om de onvolledigheid van kwantummechanica aan de kaak te stellen [33]. Hij dacht dat de schijnbare niet-lokaliteit in de natuur kon verklaard worden door middel van klassieke correlaties tussen *lokale verborgen variabelen.* Ofschoon het een erg tegenintuïtief fenomeen is, haalde kwantumverstrengeling het uiteindelijk van de lokale-verborgen-variabelen-theorieën: in 1964 –bijna drie decennia nadat de EPR-paper verscheen–, leidde John Bell bovengrenzen af op de correlaties die lokale verborgen variabelen vertonen, de zogenaamde *Bell-ongelijkheden* [9], die wel overschreden worden in kwantummechanica. Dat gaf de mogelijkheid om de onenigheid door middel van experiment op te lossen. Het duurde wel nog eens twee decennia om de experimenten te realiseren, maar ze gaven de kwantummechanica gelijk [4].

Hoewel verstrengeling geen extra parameter is, maar een intrinsieke eigenschap van de kwantumtoestand, is het onderzocht als een grootheid, zoals energie of informatie. Naast klassieke correlaties[2] is verstrengeling de mate van kwantum-niet-lokaliteit in de toestand die het kwantumsysteem, bestaande uit meerdere deelsystemen, beschrijft. Dit leidt tot de vraag hoe het kan onderscheiden worden van klassieke correlaties, en bovendien, hoe het kan gekwantificeerd worden [7]. Meestal wordt verstrengeling gedefinieerd als de negatie van scheidbaarheid: een *scheidbare* toestand vertoont alleen klassieke correlaties [93]. Een belangrijke eigenschap van verstrengeling is dat ergens in het verleden een kwantummechanische wisselwerking moet plaatsgevonden hebben tussen de deelsystemen. Als de partijen die de deelsystemen sturen enkel *lokale operaties* kunnen uitvoeren en enkel via *klassieke* kanalen mogen *communiceren* (LOKC), dan kan er geen verstrengeling tot stand komen uit een scheidbare toestand. Anderzijds kan verstrengeling gemakkelijk vernietigd worden, door lokaal metingen uit te voeren die de toestand van het deelsysteem doen inklappen naar een toestand die zuiver is en dus scheidbaar van de rest van het systeem.

[2] correlaties die evengoed kunnen verklaard worden door een lokale-verborgen-variabelen-theorie

De LOKC-voorwaarde is een beperking op de toestanden waarin een gegeven toestand kan worden getransformeerd, wat aanleiding geeft tot een gedeeltelijke ordening van toestanden. Dit houdt de notie in van een bepaalde hoeveelheid verstrengeling: één toestand heeft minstens zoveel verstrengeling als een ander als het in die andere toestand kan worden getransformeerd door LOKC. Voor *twee partijen* bestaan er toestanden met *maximale verstrengeling*: van deze toestanden spelen de vier *Bell-toestanden* een centrale rol in kwantuminformatietheorie. Elke andere toestand van twee qubits –waaronder de Bell-toestanden zelf– kan worden gerealiseerd door LOKC vanuit een Bell-toestand [69, 71]. Men kan de verstrengeling in zo'n Bell-toestand beschouwen als eenheid van verstrengeling, ook wel *ebit* genoemd, wat aanleiding geeft tot natuurlijke definities van *verstrengelingsmaten* voor twee partijen [71]. Twee belangrijke verstrengelingsmaten verdienen speciale aandacht, met name de *vormingsverstrengeling* en de *distillatieverstrengeling* [14].

De vormingsverstrengeling van een gegeven twee-qubitstoestand, is de verhouding $E_F = \lim_{N\to\infty} \frac{M}{N}$, met M het minimale aantal ebits dat nodig is om N qubitparen in de gegeven toestand te verkrijgen door LOKC. De distillatieverstrengeling is net het omgekeerde: $E_D = \lim_{N\to\infty} \frac{m}{N}$, met m het maximale aantal ebits dat door LOKC kan gewonnen worden uit N qubitparen in de gegeven toestand. Een fundamentele eigenschap van deze maten is

$$E_D \leq E_F,$$

met de gelijkheid voor zuivere toestanden. Dit volgt dadelijk uit het feit dat verstrengeling niet kan toenemen onder LOKC en als E_D groter was dan E_F, dan zou men meer ebits en dus verstrengeling kunnen halen uit N qubitparen in de gegeven toestand dan er nodig waren om ze te creëren.

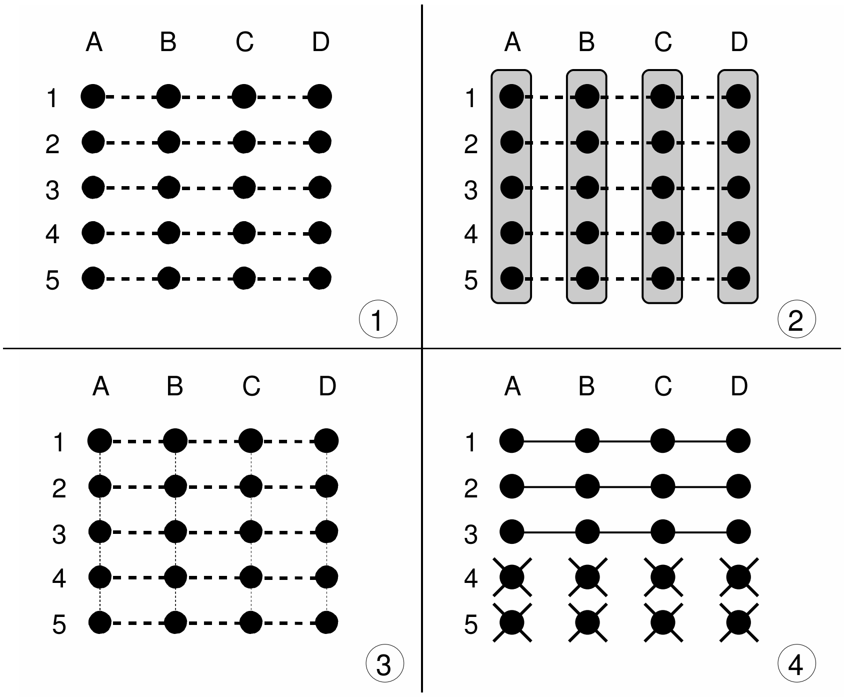
De definitie van de distillatieverstrengeling impliceert de betekenis van *verstrengelingsdistillatie* voor twee partijen [13]: uit qubitparen in een gegeven toestand willen we door LOKC zoveel mogelijk ebits halen. Het leeuwendeel van dit proefschrift handelt over de ontwikkeling van procedures om dit optimaal te doen: *verstrengelingsdistillatieprotocols*. Er zijn twee belangrijke motivaties om distillatieprotocols te bestuderen. De eerste motivatie is praktisch: ze zijn een manier om toestanden te bekomen die zuivere en maximaal verstrengelde toestanden benaderen, wat vereist is voor veel toepassingen, waaronder de bekende voorbeelden teleportatie [12], kwantum-sleuteldistributie [35] en superdichte codering [15]. De tweede motivatie is er één van meer fundamentele aard: de *winst*, oftewel de fractie gewonnen ebits, van elk distillatieprotocol is per definitie een ondergrens voor de distillatieverstrengeling. Daarom brengt het significant verbeteren van distillatieprotocols ons dichter bij een beter begrip van de onomkeerbare aard van de manipulatie van verstrengeling.

In vergelijking met twee partijen is de karakterisatie van *meer-partijen-verstrengeling* veel gecompliceerder, laat staan de kwantificatie ervan. Er bestaat nu geen éénduidige definitie meer van verstrengelingsmaten, gezien er geen maximaal verstrengelde toestanden meer zijn [8, 34, 50, 61, 87]. We kunnen echter nog steeds de winst van een distillatieprotocol definiëren als de fractie gewenste toestanden dat kan worden gedistilleerd uit een aantal kopieën van een gegeven toestand, want er zijn veel toepassingen van verstrengelde meer-partijentoestanden [19, 21, 23, 31, 45, 47, 57]. Er is een voor de hand liggende reden om enkel protocols te beschouwen die vertrekken van kopieën van *dezelfde* toestand: in een praktische opstelling maakt één van de partijen kopieën van de beoogde meer-partijentoestand aan en verdeelt die over de andere partijen via *kwantumkanalen.* We beschouwen geen specifieke implementaties van zulke kanalen: één voorbeeld is een optische vezel, met fotonen als dragers van kwantuminformatie. In de realiteit zijn zulke kwantumkanalen niet onfeilbaar en de qubits die erdoor worden gestuurd zijn onderhevig aan ruis en *decoherentie*, het verlies aan kwantuminformatie door spontane interactie met de oncontroleerbare en onwaarneembare omgeving. Het is natuurlijk om statische en tijdsonafhankelijke kanalen te veronderstellen, zodat na verdeling de ruisige kopieën statistisch onafhankelijk zijn en in dezelfde gemengde toestand.

In een typisch protocol voeren alle partijen lokale operaties en lokale metingen uit, elk op hun deel van meerdere ruisige kopieën. De lokale operaties resulteren in klassieke statistische afhankelijkheid tussen de kopieën. Door lokaal een deel van de kopieën te meten en de uitkomsten via klassieke communicatie uit te wisselen en te vergelijken, wordt informatie gewonnen over de toestand van de kopieën na verdeling. Hierdoor verandert de toestand van de overblijvende kopieën in een meer verstrengelde toestand. De gemeten kopieën zijn scheidbaar en kunnen worden weggegooid. Dit is schematisch weergegeven in figuur 0.2. De volledige procedure vermeerdert de totale verstrengeling niet, maar *concentreert* ze. Vandaar de naam 'distillatie'.

Naast twee- versus meer-partijen, kunnen we protocols indelen volgens asymptotisch versus eindig. *Asymptotische* protocols betrekken een oneindig aantal ruisige kopieën en moeten gezien worden als theoretische limieten. In de praktijk kan hun winst benaderd worden door een groot aantal initiële kopieën te nemen, gelijkaardig aan de mogelijkheid om data te comprimeren tot dicht tegen de Shannon-limiet in klassieke informatietheorie [22]. In de limiet genereert een asymptotisch protocol kopieën van een zuivere toestand, die rechtstreeks kunnen gebruikt worden voor de beoogde toepassing, en we definiëren de *asymptotische winst* als de limiet van de verhouding van het aantal gegenereerde kopieën tot het aantal initiële kopieën. *Eindige* protocols daarentegen betrekken slechts een eindig aantal kopieën en leveren kopieën met meer verstrengeling, maar niet voldoende zodat ze moeten herhaald worden in een iteratieve procedure tot het resultaat bevredigend is of kan gebruikt worden als ingang voor een ander protocol. We noemen zo'n protocol *adaptief* als intermediaire meetuitkomsten volgende acties van het protocol bepalen. In de literatuur worden adaptieve protocols vaak aangeduid als protocols die *tweewegscommunicatie* gebruiken, omdat alle partijen moeten overeenkomen wat die toekomstige acties zijn, en daarom moet klassieke informatie uitgewisseld worden in twee richtingen.

Twee belangrijke distillatieprotocols voor twee-partijenverstrengeling zijn het asymptotische *hashing-protocol* en het eindige *recurrence-protocol* [14]. Bei-

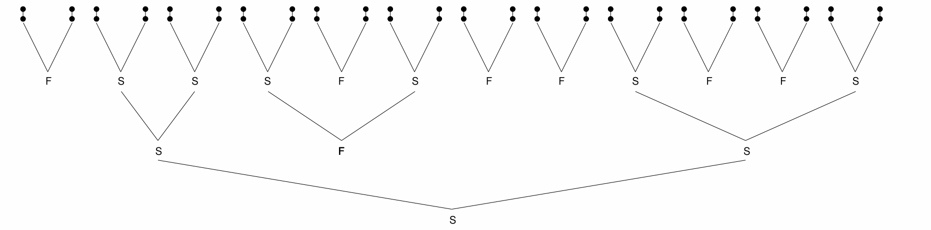


Figuur 0.2: De algemene opstelling van een distillatieprotocol (hier starten vier partijen met vijf kopieën van een verstrengelde toestand van vier qubits). Qubits worden weergegeven door stippen; (ruisige) verstrengeling door (streepjes)lijnen; klassieke correlaties door stippellijnen; lokale operaties door rechthoeken rond de betrokken qubits en gemeten qubits door aangekruiste stippen.
1) De partijen A, B, C and D starten met vijf kopieën van een ruisige verstrengelde toestand.
2) Ze voeren lokale operaties uit, elk op zijn deel van de kopieën.
3) Dit heeft klassieke correlaties tussen de kopieën tot gevolg.
4) Lokale metingen op een aantal kopieën (in dit voorbeeld 4 en 5) verschaft informatie over de ganse toestand, wat resulteert in meer verstrengeling tussen de overblijvende kopieën (1-3). De gemeten kopieën zijn nu scheidbaar en kunnen worden weggegooid.

den starten met kopieën van een toestand die een statistisch ensemble is van Bell-toestanden, een zogenaamde *Bell-diagonale* toestand. Zo'n toestand kan worden beschouwd als één van de vier Bell-toestanden, elk met een bepaalde kans. Via deze interpretatie in termen van klassieke waarschijnlijkheden kunnen we gebruik maken van de *asymptotische-equipartitie-eigenschap* [22], volgens hetwelk een rij van κ discrete toevalsvariabelen met kans $1-\epsilon$ behoort tot de *typische verzameling* $\mathcal{A}_\epsilon^{(\kappa)}$, met $\epsilon \to 0$ voor $\kappa \to \infty$. Er wordt in het hashing-protocol verondersteld dat de toestand van alle kopieën tesamen behoort tot $\mathcal{A}_\epsilon^{(\kappa)}$ met een verwaarloosbare foutkans ϵ. De uitkomst van elke meting komt overeen met gemiddeld de helft van alle mogelijkheden. De andere helft kan worden verwijderd uit $\mathcal{A}_\epsilon^{(\kappa)}$. Het protocol wordt op die manier verdergezet totdat er nog slechts één kandidaat overschiet. Deze is door de veronderstelling de initiële zuivere toestand van de kopieën. Uiteindelijk zijn er zoveel kopieën gemeten als de totale initiële entropie van de kopieën. Elke meting gaat ten koste van één kopie, dat daarna scheidbaar is en kan worden weggegooid. Bijgevolg is de hashing-winst gelijk aan nul als de entropie per kopie groter is dan één, ook al waren de kopieën verstrengeld.

Om tegemoet te komen aan de slechte performantie van hashing op te ruisige kopieën kan het worden voorafgegaan door meerdere iteraties van recurrence. Deze gecombineerde procedure heeft altijd een winst groter dan nul voor verstrengelde Bell-diagonale toestanden, zelfs als de initiële entropie hoog is. Recurrence vertrekt van slechts twee kopieën van een Bell-diagonale toestand, waarvan één wordt gemeten na de lokale operaties. Afhankelijk van de uitkomst van die meting is de overblijvende kopie ofwel in een meer verstrengelde Bell-diagonale toestand, ofwel in een scheidbare toestand. In het laatste geval zeggen we dat het protocol 'mislukt' is en de overblijvende kopie wordt weggegooid. Het protocol wordt per twee toegepast op verschillende kopieën. Zo kan de gemiddelde verstrengeling per overgebleven kopie stapsgewijs vermeerderd worden door de procedure te herhalen op alle overgebleven kopieën van 'geslaagde' protocols (een voorbeeld hiervan is weergegeven in figuur 0.3). In tegenstelling tot hashing is deze procedure adaptief, gezien enkel kopieën die het resultaat zijn van geslaagde protocols opnieuw gecombineerd worden in een volgende stap, en mislukken of slagen hangt van de meetuitkomsten af.

Op een gegeven ogenblik in de zojuist beschreven procedure is het niet langer voordelig om nog een recurrence-iteratie uit te voeren vooraleer op hashing over te stappen. Als recurrence altijd maar wordt verdergezet, is de asymptotische winst zelfs nul, want in elke iteratie wordt het aantal niet-gemeten kopieën gehalveerd [14]. Op één of andere manier vullen iteratieve recurrence en hashing mekaar aan. Een vraag die daarop volgt, is of een protocol kan worden ontwikkeld dat de voordelen van beiden combineert, dat adaptief is en asymptotisch, zonder die abrupte overgang van recurrence naar hashing. Een eerste stap in die richting werd gedaan in [90], waar de hashing-winst werd verbeterd voor alle initiële entropieën. We konden het principe achterhalen dat aan de basis ligt van deze verbeteringen: telkens wanneer een meting wordt



Figuur 0.3: Drie iteraties van recurrence, resulterend in één kopie met meer verstrengeling uit 24 initiële ruisige kopieën. In elke stap worden kopieën paarsgewijs samengenomen in het recurrence-protocol, met als resultaat ofwel kopieën met meer verstrengeling ofwel scheidbare kopieën. In het eerste geval noemen we het protocol geslaagd (S), in het tweede geval mislukt (F). Men blijft dit herhalen met kopieën uit geslaagde protocols totdat een gewenst criterium is bereikt of men overschakelt op hashing.

uitgevoerd, wordt de toestand geprojecteerd op de eigenruimte van de gemeten observabele, en daardoor wordt de meting van observabelen die niet commuteren met de eerste onmogelijk gemaakt.[3] Men zou denken dat dit hetzelfde is als 'vergeten' of het 'weggooien' van informatie, wat altijd resulteert in hogere entropie omdat entropie groter is dan voorwaardelijke entropie [22]. Maar hier *elimineert* de projectie als het ware de entropie die overeenkomt met de observabelen die niet commuteren met de eerste. Dit is sterk verwant aan het concept van *ontaarde kwantumcodes* [39] in kwantum-foutencorrectie. Door dit principe uit te buiten, gecombineerd met andere ideeën, konden we een protocol vinden dat alle bestaande protocols voor Bell-diagonale toestanden achter zich laat, en tot op heden nog niet is verbeterd [53].

Voor onze zoektocht naar betere distillatieprotocols maakten we veelvuldig gebruik van het *stabilisatorformalisme* [68, 72]. Dit tranparant wiskundig kader werd oorspronkelijk ontwikkeld met het oog op het vinden van goede *kwantumfoutverbeterende codes* [18, 40], technieken om de kwetsbare kwantuminformatie te beschermen tegen decoherentie, die absoluut noodzakelijk zijn voor het succesvol realiseren van toepassingen zoals de kwantumcomputer [42]. Het stabilisatorformalisme omvat *stabilisatortoestanden* en *Cliffordoperaties*, belangrijke klassen van (verstrengelde) kwantumtoestanden en operaties. Naast kwantumfoutenverbetering heeft het stabilisatorformalisme veel toepassingen [45]. Ook het concept van de *éénwegs-kwantumcomputer*, een veelbelovende alternatieve organisatie van een kwantumcomputer, bestaat uit één-qubitmetingen op een stabilisatortoestand (clustertoestand) [75].[4]

[3] Deze observabelen kunnen wel degelijk gemeten worden, maar de uitkomsten hiervan geven geen enkele informatie over de initiële toestand.

[4] Bell-toestanden, grafetoestanden en clustertoestanden zijn speciale gevallen van stabilisatortoestanden.

Het stabilisatorformalisme wordt meestal uitgelegd in een groepentheoretisch kader. De beschouwde groepen zijn homomorf met vectorruimten over het veld GF(2). Dat laat een beschrijving toe in termen van binaire lineaire algebra [24], dat meermaals zijn nut bewees voor de ontwikkeling van distillatieprotocols, zowel voor twee als voor meer partijen [25, 51, 53, 54, 55]. Het voordeel van dit 'binaire beeld' is dat het ganse mechanisme van initiële kopieën, lokale operaties en metingen volledig kan worden beschreven door middel van matrixbewerkingen over GF(2), die slechts kwadratisch opschalen in complexiteit ten opzichte van de ingangsgrootte. Op die manier ontwijken we de intrinsieke complexiteit van algemene kwantumtoestanden en -operaties. Bovendien worden interpretaties in termen van klassieke informatietheorie, zoals voor hashing, veel transparanter in het binaire beeld.

Bestaande meer-partijenvarianten van hashing en recurrence [1, 3, 19, 29, 38, 43, 58, 59, 60, 63, 67] hebben twee nadelen. Ten eerste, door informatietheorie niet volledig uit te buiten, resulteren de hashing-varianten in teveel metingen: om alle elementen van de typische verzameling te elimineren, wordt geëist dat het aantal metingen bepaalde marginale entropieën overschrijdt, terwijl de typische verzameling veel sneller kan worden gereduceerd als de informatie die uit de metingen wordt gewonnen, efficiënter gebruikt wordt. Hier werd voor een stuk aan tegemoet gekomen in [19], maar we konden een methode opstellen die het aantal metingen minimaliseert [54, 55]. Daarvoor hebben we een iets strengere versie nodig van de typische verzameling: de *sterk-typische verzameling* $\mathcal{T}_\epsilon^{(\kappa)}$. Ten tweede leidden we voor *CSS-toestanden*, een belangrijke grote klasse van stabilisatortoestanden, de meest algemene lokale operaties af gebruikt in het distillatieprotocol. Zo kunnen niet alleen hashing, maar ook recurrence-achtige protocols met een hogere winst ontwikkeld worden [54]. Op die manier is dit resultaat een veralgemening naar meer partijen van [25] voor twee-partijenprotocols.

Hoofdstuk 2: Het stabilisatorformalisme

In dit preliminaire hoofdstuk behandelen we eigenschappen van de binaire beschrijving van de Pauligroep, stabilisatortoestanden en Cliffordoperaties die relevant zijn in de context van verstrengelingsdistillatieprotocols. We beginnen met de definitie van de Pauligroep op n qubits $\mathcal{G}_n$, een groep voor de vermenigvuldiging die bestaat uit Kroneckerproducten van Paulimatrices met een bijkomende complexe fase (1, i, -1 of $-i$). We laten zien hoe deze groep isomorf is met de groep $\mathbb{Z}_2^{2n}$ voor de optelling als de fases niet in rekening worden gebracht. Dan worden Cliffordoperaties geïntroduceerd als unitaire operaties die de Pauligroep op zichzelf afbeelden onder conjugatie. De werking van een Cliffordoperatie wordt naar het binaire beeld vertaald in een linkse vermenigvuldiging van een binaire symplectische matrix op de vector die de Pauli-operatie identificeert. We tonen hoe een willekeurige Cliffordoperatie kan worden opgebouwd uit een reeks elementaire één- en twee-qubit-Cliffordoperaties, wat van belang kan zijn voor de praktische realisatie van Cliffordoperaties.

We definiëren een stabilisatortoestand als de gemeenschappelijke eigenvector van de stabilisator, een maximale commutatieve deelgroep van de Pauligroep. De stabilisatortoestand is volledig gedetermineerd door de binaire voorstelling van een verzameling Pauli-operaties die de stabilisator voortbrengt. Deze binaire voorstellingen kunnen bijeengevoegd worden als de kolommen van een binaire matrix. Een Cliffordoperatie die op een stabilisatortoestand inwerkt wordt in het binaire beeld de vermenigvuldiging van de respectievelijke voorstellingsmatrices. We wijden een apart stuk aan specifieke eigenschappen van distillatieprotocols in het binaire beeld. Stabilisatorcodes worden kort toegelicht. Het is een belangrijke klasse van kwantum-foutverbeterende codes die eveneens efficiënt in het binaire beeld worden beschreven en op die manier sterk gelinkt zijn aan onze distillatieprotocols

Hoofdstuk 3: Distillatie van twee-partijenverstrengeling

We behandelen de ontwikkeling van distillatieprotocols voor twee-partijenverstrengeling door uitvoering gebruik te maken van het binaire beeld van het stabilisatorformalisme. We leiden kort het concept van twee-partijenverstrengeling en Bell-toestanden in, en we overlopen iets uitvoeriger de verschillende elementen van distillatieprotocols in het stabilisatorformalisme voor het specifieke twee-partijengeval, geïllustreerd met het bestaande recurrence-protocol. We lichten het werkingsprincipe van asymptotische protocols (waaronder hashing) toe. Wegens de slechte prestaties van hashing voor te ruisige ingangskopieën is het nodig dat dergelijke protocols worden voorafgegaan door een iteratie van eindige protocols, zoals recurrence.

Hoewel het operatie-stuk van zowel eindige als asymptotische protocols op dezelfde manier worden beschreven in het binaire beeld, blijken ze op vlak van interpretatie nogal te verschillen. Ten eerste wordt de werking van asymptotische protocols volledig beschreven in termen van het informatie-theoretische concept typische verzameling, terwijl eindige protocols leiden tot een verandering van de waarschijnlijkheden die de gemengde toestand definiëren. Ten tweede laat het hashing-protocol geen incorporatie van adaptiviteit toe, waardoor het slecht presteert voor ruisige ingangskopieën. Eindige protocols zijn intrinsiek adaptief omdat tussenliggende meetuitkomsten het vervolg van het protocol bepalen.

Het doel is manieren vinden om de voordelen van eindige protocols met die van hashing te combineren. Een eerste succesvolle stap in die richting was gedaan in [90], zij het eerder ad hoc. Door de onderliggende principes die aan de basis liggen van de verbeteringen te analyseren, konden we protocols construeren die deze ideeën uitbuiten, en op die manier alle bestaande protocols overtreffen [53]. Het belangrijkste principe is de entropie-afname die gepaard gaat met het lokale inklappen van de toestandsvector door de metingen in het protocol.

Tenslotte bestuderen we het limietgedrag van eindige protocols voor bijna scheidbare Bell-diagonale toestanden.

Hoofdstuk 4: Distillatie van meer-partijen-verstrengeling

We beschrijven asymptotische protocols voor meer-partijenverstrengelingsdistillatie. We tonen hoe bestaande resultaten significant kunnen worden overtroffen door het binaire beeld en klassieke informatietheorie uit te buiten. Ons werk op meer-partijendistillatie verschilt van bestaande resultaten op twee niveaus. Ten eerste kunnen we in het binaire beeld de meest algemene structuur van de lokale Cliffordoperaties, die in het protocol worden aangewend, afleiden, voor verschillende klassen stabilisatortoestanden. Op die manier wordt een grotere statistische afhankelijkheid van de kopieën bewerkstelligd zodat lokale metingen meer informatie uit de initiële toestand kunnen halen. Bijgevolg zijn er minder metingen nodig om de toestand te zuiveren en dus een grotere winst.

Ten tweede maken we optimaal gebruik van bepaalde eigenschappen van de sterk-typische verzameling. In tegenstelling tot het twee-partijengeval levert de lokale meting van een kopie de eigenwaarde van meer dan één stabiliserende Pauli-operatie op. Het doel van het protocol is de totale entropie van de ingangskopieën tot nul te brengen. In bestaande protocols vertaalt dit zich in de vereiste dat het aantal gemeten kopieën bepaalde marginale entropieën (of in het beste geval voorwaardelijke entropieën) moet overstijgen. Door gebruik te maken van de afgeleide eigenschappen van de sterk-typische verzameling kunnen we het exact aantal gemeten kopieën berekenen dat nodig is om de initiële totale entropie tot nul te brengen. Op die manier kunnen we de winst van hashing aanzienlijk vergroten.

Abstract

The topic of this thesis is the development of entanglement distillation protocols within the stabilizer formalism. Entanglement distillation protocols are methods for concentrating the quantum entanglement that is present in copies of a given quantum state by means of local operations and classical communication only. They are both of practical and theoretical importance. On the one hand, they are a means of purifying the entanglement necessary for practical purposes when only local operations and classical communication (LOCC) are allowed for the application in mind. On the other hand, they serve to give a more fundamental insight into the properties of entanglement and the physical boundaries of entanglement manipulation.

The stabilizer formalism is a mathematical framework comprising stabilizer states and Clifford operations, specific important classes of quantum states and operations. Usually formulated in a group theoretical setting, we put special emphasis on the equivalent representation of these states and operations in terms of binary matrix algebra. This 'binary picture' allows for a transparent and efficient description of entanglement distillation protocols and gives us the opportunity to improve existing results significantly. We give a complete overview of the properties of stabilizer states and Clifford operations in the binary picture that are relevant for the purpose of developing entanglement distillation protocols.

We focus on the distillation of both bipartite as multipartite entanglement, i.e. involving two or more than two parties respectively. Distillation mostly boils down to the extraction of information by means of LOCC so that the entropy of the target state decreases and the end result is a pure state. For bipartite distillation protocols, next to this information extraction, an extra reduction in the entropy is distinguished resulting from the local measurements in the protocol. The description in the binary picture enables us to recognize the underlying principle causing this entropy reduction, and to use this insight to devise protocols that outperform existing ones.

For the multipartite case, improvements are obtained on two levels. Firstly, by making extensive use of the binary picture, we derive, for different classes of stabilizer states, the most general structure of the local Clifford operations used in the protocol. Secondly, we derive and exploit particular properties of the strongly typical set, a concept borrowed from classical information theory.

Glossary

Mathematical notation

$\|\psi\rangle$	pure state vector
$\langle\psi\|$	dual of $\|\psi\rangle$
ρ	density matrix
$\mathrm{Tr}\left\{\rho\right\}$	trace of ρ
$S(\rho)$	Von Neumann entropy of the state ρ
$p(x)$ or p_x	probability of x
$H(X)$	Shannon entropy of the random variable X
$\|\mathcal{A}\|$	cardinality of the set $\mathcal{A}$
A_j	j-th column of matrix A
A^T	transpose of A
$A^\dagger$	conjugate transpose of A
A^{-1}	inverse of A
A^{-T}	short for $(A^T)^{-1}$
$\otimes$	Kronecker or tensor product
$\oplus$	direct sum
$\odot$	elementwise product
$\sim$	equality up to a phase factor $e^{i\varphi}$
$\mathrm{col}\,(A)$	column space of A
$\mathcal{J}^\perp$	orthogonal complement of vector space $\mathcal{J}$
$\mathrm{diag}(A)$	vector equal to the diagonal of A
$\mathrm{lows}\,(A)$	strict lower triangular part of A
δ_{ij}	Kronecker delta
$\mathcal{O}$	Landau symbol "big O"

Fixed symbols

$\mathbb{Z}_2$	set of integer numbers modulo 2
$\mathbb{C}$	set of complex numbers
$\mathbb{Z}_2^n$	set of n-dimensional vectors over $\mathbb{Z}_2$
$\mathbb{Z}_2^{m\times n}$	set of $m \times n$ matrices over $\mathbb{Z}_2$
$\mathcal{H}_n$	n-qubit Hilbert space, $\mathcal{H}_n \cong \mathbb{C}^{2^n}$

$\mathcal{G}_n$	Pauli group on n qubits
$\sigma_x, \sigma_y, \sigma_z$	Pauli matrices
I_n	$n \times n$ identity matrix (when n is omitted, dimensions match context)
e	vector containing all ones (dimension matches context)
e_i	vector containing 1 on the i-th entry and zeros elsewhere
$\mathcal{S}$	stabilizer
$\lvert\psi_{\mathcal{S}}\rangle$	stabilizer state with stabilizer $\mathcal{S}$
$\mathcal{H}_{\mathcal{S}}$	stabilizer code space with stabilizer $\mathcal{S}$
$\lvert\psi_{S,b}\rangle$	stabilizer state represented by S and b
$\mathcal{A}_\epsilon^{(\kappa)}$	typical set
$\mathcal{T}_\epsilon^{(\kappa)}$	strongly typical set

Acronyms

LOCC	local operations and classical communication
CNOT	controlled-NOT gate
LHS (RHS)	left-hand (right-hand) side
BPM	bilateral Pauli measurement
AEM	appended ebit measurement
PB	partial breeding
LP	linear programming

Conventions

For notational convenience, we will often denote a binary vector by a string (e.g., 1010 stands for $[1\ 0\ 1\ 0]^T$).

As we are mostly working in finite-dimensional vector spaces with a fixed basis, we will often identify vectors with their coordinate representation and operators with their matrix representation.

Contents

Chapter 1

Introduction

Quantum information theory is a fairly new branch of research covering elements from various areas, including quantum mechanics, information theory and computer science. It is the study of particular properties of quantum systems and applications thereof, such as quantum communication, quantum cryptography and quantum computing. Describing the state of a quantum system, contrary to a classical system, scales exponentially with the system size. Therefore, simulating the evolution of a large quantum system on a classical computer, as we know it today, is highly inefficient and in practice, it quickly becomes an infeasible task. This observation led the famous physicist Richard Feynman to the idea of using quantum devices for this purpose instead [36]. As such, the concept of a *quantum computer* was born [26], and simulating other quantum systems can be regarded as its first conceivable application. A breakthrough came in 1994 when Peter Shor discovered a *quantum algorithm* for finding the prime factors or the discrete logarithm of an integer in polynomial time with respect to the input size [78], whereas no polynomial classical algorithm has been found at the time of writing. Two years later, Lov Grover developed a quantum algorithm for searching an unsorted database with quadratic speedup compared to classical algorithms [44]. The *qubit* is the quantum counterpart of the classical bit and the standard building block of quantum computation. It is a two-level quantum system. We do not consider practical implementations of qubits here. Common examples are the polarization of a photon and the spin of an electron.

The *quantum state* contains the information we have on a particular quantum system. The only way it is accessible to us is by the observation of repeatable experiments and the concomitant probabilities. Though already existing in classical physics, where they express lack of information, probabilities are inherent in the apparent indeterministic nature of quantum measurements. A quantum state is either *pure* or *mixed*. The latter can be regarded as a classical statistical ensemble of pure states, and, therefore, a mixture of quantum state 'amplitudes' and classical probabilities. Mathematically, a pure state is identified with a unit vector $|\psi\rangle$ in a complex Hilbert space,[1] a mixed state with a positive semi-definite Hermitian operator on the same space with trace one: the *density operator* ρ. The state completely determines the outcome probabilities of any measurement on the system. At the same time, the measurement causes an apparently abrupt change in the state of the measured system. This strange phenomenon is referred to as the *collapse* of the state vector or wave function. Its interpretation has been –and still is– the subject of much debate [76], but we will not delve further into that issue. In a sense, measuring 'destroys' the information contained in the quantum state. This behavior is exploited for encryption, as an eavesdropper cannot extract information from a communicated quantum message nor tamper with it without being noticed. Unlike classical cryptography, which is based on the computational complexity of solving particular mathematical problems, *quantum cryptography* guarantees unconditional security [11, 37].

Another feature thus far not encountered in classical physics, and taking a central position in this thesis, is *quantum entanglement.* It manifests itself in the behavior of a composed system as correlations of measurement outcomes that cannot be explained in a classical theory. A typical example is the experiment in which an atom decays via an intermediate state to the ground state, thereby emitting two entangled photons: measuring the polarization of one photon immediately pins down the polarization of the other. Albert Einstein derided entanglement as a "spooky action at a distance" and used it in the famous EPR paper to expose quantum mechanics' incompleteness [33]. He thought that the apparent non-locality in nature could be explained by classical correlations between *local hidden variables.* Although being a highly counterintuitive phenomenon, quantum entanglement eventually triumphed over the local hidden variable theories: in 1964 –almost three decades after publication of the EPR paper–, John Bell derived upper bounds on the correlations exhibited by local hidden variables, the so-called *Bell inequalities* [9], that can be violated in quantum mechanics. This gave rise to the possibility of settling the dispute by experiment. Yet, it took nearly another two decades to put these experiments into practice, in favor of quantum mechanics [4].[2]

Although entanglement is not an extra defining parameter, but an intrinsic property of the quantum state, it has been investigated as a resource, like energy or information. Coexisting with classical correlations,[3] entanglement is the amount of quantum non-locality in the state describing a quantum system, consisting of multiple subsystems. This directly leads to the question of how to distinguish it from classical correlations, and more importantly, how to quantify it [7]. Usually, entanglement is defined as the negation of separability: a *separable* state only displays classical correlations [93]. A key property of entanglement is that it requires the subsystems involved to have interacted quantum

[1] For a qubit, this space is two-dimensional.

[2] Although some loopholes are present in the original experiments, and are hard to circumvent. Nonetheless, most people do not doubt the validity of the experiments.

[3] correlations that can equally be explained by a local hidden variable theory

mechanically somewhere in the past. If the parties controlling the subsystems are only allowed to perform *local operations* and *communicate classically* (LOCC), an entangled state cannot be created out of a separable state. On the other hand, entanglement can easily be destroyed, by locally performing measurements that cause the state of the subsystem to collapse into a state that is pure and therefore separable from the rest of the system.

In general, the LOCC constraint imposes limits on what states a given state can be transformed into, giving rise to a partial ordering of states. This implies the notion of a certain quantity of entanglement: one state is said to contain at least as much entanglement as another state if it can be transformed into the other by LOCC. In the *bipartite* case, states of two qubits exist with *maximal entanglement*: of these, the four *Bell states* play a central role in quantum information theory. Any other state of two qubits –including the Bell states themselves– can be realized by LOCC starting from a Bell state [69, 71]. Using the entanglement present in a Bell state as standard unit of entanglement, often named *ebit*, natural definitions of bipartite *entanglement measures* appear [71]. Two important measures of entanglement deserve special attention here, namely the *entanglement of formation* and the *entanglement of distillation* [14].

The entanglement of formation[4] of a given two-qubit state, is the fraction $E_F = \lim_{N\to\infty} \frac{M}{N}$, where M is the minimal number of ebits needed to create N qubit pairs in the given state by LOCC. The entanglement of distillation is exactly the opposite: $E_D = \lim_{N\to\infty} \frac{m}{N}$, where m is the maximal number of ebits that can be extracted by LOCC out of N qubit pairs in the given state. A fundamental property of these measures is

$$E_D \leq E_F,$$

with equality for pure states. This is a direct consequence of the fact that entanglement cannot increase under LOCC. Indeed, if E_D were larger than E_F, then one would be able to extract more ebits and thus more entanglement out of N qubit pairs in the given state than were needed to create them. Reminiscent of the second law in thermodynamics, this somehow reflects a fundamental irreversibility of entanglement manipulation.

The definition of the entanglement of distillation already implies the meaning of bipartite *entanglement distillation* [13]: out of qubit pairs in a given state, we want to extract as many ebits as possible by means of LOCC. The bulk of this thesis deals with the development of procedures to do this in an optimal way: *entanglement distillation protocols*.[5] There are two main motivations for studying distillation protocols. The first motivation is practical: they

[4] To be precise, what is defined here is the *entanglement cost*, which equals the *asymptotic entanglement of formation* [71]. One believes it also equals the entanglement of formation, although the additivity of the entanglement of formation is still an unproven conjecture.

[5] In the literature, these are sometimes called *entanglement purification protocols*, as they mostly boil down to rendering a mixed state more pure. We prefer not to use the term 'purification', as it might be confused with the interpretation of a mixed state as part of a larger pure state [56].

are a means of obtaining states that approach pure and maximally-entangled states, required for many applications, of which well-known examples are teleportation [12], quantum key distribution [35] and superdense coding [15]. The second motivation is of a more fundamental nature: the *yield*, i.e. the fraction of extracted ebits, of each distillation protocol is a lower bound for the entanglement of distillation, by definition of the latter. Therefore, significantly improving distillation protocols brings us closer to a better understanding of the irreversible nature of entanglement manipulation.

Compared to the bipartite case, the characterization of *multipartite* entanglement, let alone its quantification, is much more complex. No straightforward definition of an entanglement measure exists, as there is no notion of maximally-entangled states any more [8, 34, 50, 61, 87]. Yet, we can still define the yield of a distillation protocol as the fraction of desired states that can be distilled out of a number of copies of a given state, as there are many applications using multipartite entangled states [19, 21, 23, 31, 45, 47, 57]. There is an obvious reason for considering the protocols starting from copies of the *same* state: in a practical setting, one of the parties locally creates copies of the multipartite pure state required for the application in mind and distributes it to the other parties via *quantum channels*. We do not consider specific implementations of such channels: one example is an optic fiber, if photons are the carriers of quantum information. In reality, these quantum channels are not perfect, and the qubits sent through are subject to noise and *decoherence*, which is the loss of quantum information due to spontaneous interaction with the uncontrollable and unobservable environment. It is natural to assume memory-less and time-invariant channels, such that after distribution, the noisy copies are independent and in the same mixed state.

In a typical protocol, all parties perform local operations and local measurements, each on their share of the multiple noisy copies. The local operations result in classical statistical dependence of the copies. By locally measuring a fraction of the copies and classically exchanging and comparing the outcomes, information is gained on the state of the copies after distribution. This information transforms the remaining copies into a state that is more entangled. The measured copies are separable and can be discarded. This is schematically depicted in figure 1.1. The overall procedure does not increase, but *concentrates* the total entanglement. Hence the name ‘distillation’.

Next to bipartite versus multipartite, we categorize protocols as asymptotic versus finite. *Asymptotic* protocols involve an infinite number of noisy copies and are to be seen as theoretical limits. In practice, their yield can be approached by considering a large amount of initial copies, analogous to the possibility of data compression close to the Shannon limit in classical information theory [22]. In the limit, an asymptotic protocol outputs pure state copies, ready to use for the application in mind, and we define the *asymptotic yield* as the limit of the ratio of the number of output copies to the number of input copies. *Finite* protocols on the other hand involve only a finite number of copies and deliver copies with more entanglement, but not enough so that

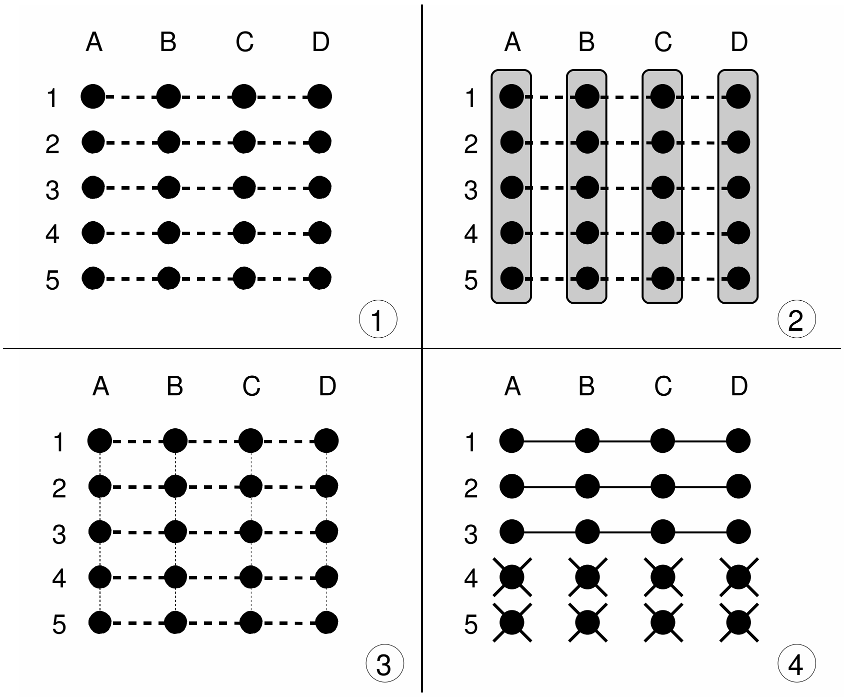


Figure 1.1: A distillation protocol in general (here with four parties starting with five copies of a four-qubit state). Qubits are depicted as dots; (noisy) entanglement as (dashed) lines; classical correlations as dotted lines; local operations by rectangles enclosing the involved qubits and measured qubits as crossed dots.
1) The parties A, B, C and D start with five copies of a noisy entangled state.
2) They apply local operations, each on their share of the copies.
3) This results in classical correlations between the copies.
4) Local measurements on a part of all copies (in this example 4 and 5) then yield information on the overall state, on average resulting in more entanglement on the remaining copies (1-3). The measured copies have become separable and can be discarded.

they have to be repeated in an iterative procedure until the result is satisfying or used as input of another protocol.[6] We call a protocol *adaptive* when intermediate measurement outcomes determine future actions of the protocol. In the literature, these are sometimes referred to as protocols using *two-way communication*, as all parties must agree on the future actions and therefore, classical information needs to be sent in both directions.

Two important bipartite distillation protocols are the asymptotic *hashing protocol* and the finite *recurrence protocol* [14]. Both start, possibly after some pre-processing step, from copies of a state that is an ensemble of Bell states, a so-called *Bell-diagonal* state. Such a state can be regarded as one of the four Bell states, each with a given probability. This interpretation in terms of classical probabilities allows for the use of the *asymptotic equipartition property* [22], which states that a sequence of κ discrete random variables is, with probability $1 - \epsilon$, an element of the *typical set* $\mathcal{A}_\epsilon^{(\kappa)}$, where $\epsilon \rightarrow 0$ for $\kappa \rightarrow \infty$. In the hashing protocol, it is assumed that the state of all copies is contained in $\mathcal{A}_\epsilon^{(\kappa)}$, with vanishing error probability ϵ. On average, the outcome of each measurement is incompatible with half of all possibilities, which can then be eliminated. The protocol is continued in this way until only one candidate remains, which, by assumption, is the initially unknown pure state of the copies. To achieve this, the number of measurements needed equals the total entropy of the initial state of all copies. Every measurement is at the cost of one copy, which is separable afterwards and can be discarded. Therefore, if the entropy per copy exceeds one, the yield of hashing is zero, even when the copies were entangled initially.

To overcome the poor performance of hashing for too noisy copies, it can be preceded by several iterations of recurrence. This combined procedure will always have a nonzero yield for Bell-diagonal states with entanglement, even if the initial entropy is high. One starts with only two copies of a Bell-diagonal state, of which one is measured after the local operations. Depending on the outcome of the measurement, the remaining copy is either in a more entangled Bell-diagonal state, or in a separable state. In the latter case, we say the protocol has ‘failed’ and the remaining copy is discarded. The protocol is applied to several pairs of copies, and the average entanglement per kept copy can be gradually increased by repeating this procedure on all remaining copies of ‘successful’ protocols, an example of which is illustrated in figure 1.2. Contrary to hashing, this procedure is adaptive, as only copies of successful protocols are combined in a next step, and failure or success depends on the measurement outcomes.

At a certain stage in the previously-explained procedure, it is no longer advantageous to do another recurrence iteration before switching to hashing. In fact, if recurrence is continued, the asymptotic yield becomes zero, because in every iteration, the number of unmeasured copies is halved [14]. Somehow,

[6] As such, finite protocols applied simultaneously on an infinite number of copies can also be regarded as an asymptotic protocol.

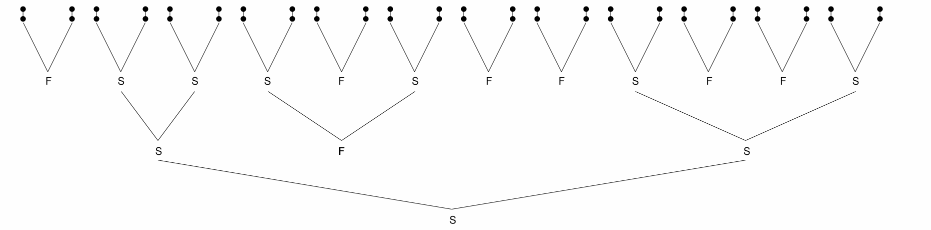


Figure 1.2: Three iteration steps of recurrence, producing one copy with more entanglement out of 24 initial noisy copies. In each step, copies are paired in the recurrence protocol, producing either copies with more entanglement if successful (S) or separable copies otherwise (F). Resulting copies of successful protocols are gathered and the same procedure repeated, until some criterion is reached or the copies can be used as input of hashing.

iterative recurrence and hashing complement each other. A natural question to ask therefore, is whether a protocol can be developed that combines the benefits of both, which is adaptive and asymptotic, without this abrupt transition of recurrence to hashing. A first step in that direction was made in [90], where the hashing yield was beaten over the entire range of initial entropies. We were able to recognize the principle that lies at the basis of these improvements. Whenever a measurement is performed, the state is projected onto an eigenspace of the measured observable –the notorious collapse–, thereby prohibiting the measurement of observables that do not commute with the first.[7] One would expect that this is equivalent to 'forgetting' or 'throwing away' information, which always results in higher entropy, as entropy exceeds conditional entropy [22]. But in this case, the projection actually *eliminates* the entropy associated with the observables not commuting with the first (what we mean by that will become clear when explained in more detail in chapter 3). We believe this is strongly related to the concept of *degenerate quantum codes* [39] in quantum error correction. By exploiting this principle along with other ideas, we created a protocol that significantly outperformed existing protocols for Bell-diagonal states, and has not been surpassed at the time of writing [53].

In our search for better distillation protocols, we made extensive use of the *stabilizer formalism* [68, 72]. This transparent mathematical framework was originally developed for the purpose of finding good *quantum error-correcting codes* [18, 40], techniques for protecting the fragile quantum information against decoherence, that are absolutely necessary for the effective realization of applications such as quantum computation [42]. The stabilizer formalism comprises *stabilizer states* and *Clifford operations*, important sets of (entangled) quan-

[7] To be precise, these observables *can* be measured, but the outcomes will not give any information on the original state.

tum states and operations. Next to quantum error correction, the stabilizer formalism has many applications (for an overview, we refer to [45] and references therein). Also the concept of the *measurement-based quantum computer*, a promising alternative setup for quantum computation, is based on the preparation of a stabilizer (cluster) state and one-qubit measurements [75].[8]

The stabilizer formalism is mostly formulated in a group theoretical setting. The considered groups are homomorphic to vector spaces over the field GF(2). This allows for an efficient description in terms of binary, linear algebra [24], which proved useful for the development of distillation protocols, both bipartite and multipartite [25, 51, 53, 54, 55]. The advantage of this framework for distillation, which we like to refer to as the '*binary picture*', is that the mechanism of initial copies, local operations and measurements is entirely described in terms of matrix operations over the field GF(2), that scale only quadratically over the input size. Doing so, we have evaded the inherent complexity of general quantum states and operations. Furthermore, interpretations in terms of classical information theory, such as for hashing explained above, become much more transparent in the binary picture.

Many multipartite variants of the hashing protocol and the recurrence protocol have been constructed [1, 3, 19, 29, 38, 43, 58, 59, 60, 63, 67]. Two drawbacks exist for those protocols. Firstly, by not exploiting information theory to a full extent, the hashing variants result in overkill: to eliminate all elements of the typical set save one, they demand that the number of measurements exceeds particular marginal entropies, whereas the typical set can be reduced much faster if the information gained by the measurements is used more efficiently. This was partially met in [19] by relaxing to conditional entropies, but we devised a method that minimizes the number of measurements [54, 55]. To this end, we need a slightly stricter version of the typical set, namely the *strongly typical set* $\mathcal{T}_\epsilon^{(\kappa)}$. Secondly, for *CSS states*, an important large class of stabilizer states, we derived the most general local operations suitable for distillation protocols in the binary language. Doing so, not only hashing, but also recurrence-like protocols with a higher yield can be developed [54]. As such, this result is a multipartite generalization of [25] for bipartite protocols.

[8] Bell states, graph states and cluster states are all special cases of stabilizer states.

Outline

In **chapter 2**, we elaborate on properties of the binary description of the Pauli group, Clifford operations and stabilizer states that are relevant in the context of entanglement distillation protocols. We define the Pauli group consisting of Kronecker products of Pauli matrices, and show how they are linked with binary vectors. Clifford operations are introduced as unitary operations mapping the Pauli group to itself under conjugation. This is translated into binary terms as the left multiplication of a binary matrix on the vector representing the Pauli operation. We explain how an arbitrary Clifford operation can be decomposed in elementary one- and two-qubit Clifford operations. We define a stabilizer state as the simultaneous eigenvector of a subgroup of the Pauli group and show how it can be represented by a binary matrix. We focus on some particular general properties in the binary picture of the distillation protocols for stabilizer states that will be the topic of the following chapters. We end this chapter by touching on the theory of stabilizer codes, which is important because of its close relationship with distillation protocols.

In **chapter 3**, we discuss the development of bipartite distillation protocols by making extensive use of the stabilizer formalism in the binary picture. After briefly introducing the concept of bipartite entanglement and Bell states, we recapitulate in more detail the various elements of distillation protocols in the stabilizer formalism for the particular bipartite case, illustrated with the existing recurrence protocol. Our goal is to find ways of combining the benefits of finite protocols and asymptotic protocols, and develop adaptive asymptotic protocols. We give the main lines of thought of asymptotic protocols and elaborate on the basic principles on which adaptiveness can be introduced to improve their performance. More precisely, we show how the description in binary matrix algebra enables us to recognize the local collapse of the state vector and the concomitant entropy reduction as the key principle for the improvements. Finally, we also discuss the search for optimal finite protocols for noisy input states.

In **chapter 4**, we describe asymptotic protocols for multipartite entanglement distillation. We show how existing results can be significantly outperformed by exploiting the binary picture and classical information theory. More precisely, we calculate, for different classes of stabilizer states, the most general structure of the local Clifford operations used for the protocol, such that they effect a larger statistical dependence of multiple noisy copies of the input state and, as such, effect a higher yield. On the level of classical information theory, we derive the particular properties of the strongly typical set that are relevant for the purpose of minimizing the number of measurements necessary to purify the state of the input copies.

In **chapter 5**, we summarize our main results in a general conclusion, and we outline some possible roads for future research.

Results

The structure and cohesion of our work [51, 52, 53, 54, 55] is schematically depicted in figure 1.3. Entanglement distillation is important for both practical purposes (i.e. applications in quantum cryptography, quantum computation and quantum communication) and for the understanding of fundamental properties of entanglement (e.g. entanglement measures) in quantum information theory. We distinguish the distillation of bipartite entanglement and multipartite entanglement.

The main mathematical techniques that are applied for this purpose are classical information theory and binary linear algebra. The concepts information gain, entropy reduction and typical set in our work on bipartite protocols [53] are direct applications of classical information theory. Furthermore, next to the description of bipartite protocols in terms of stabilizer codes [51], there is a strong connection between entropy reduction and the existence of degenerate codes, which is an important feature of quantum codes with no classical analogue. We use the concept strongly typical set in the multipartite setting for calculating the exact minimal number of measurements needed to purify the entanglement [54, 55].

We describe the framework of the stabilizer formalism, originally developed in the context of quantum error-correction, entirely in terms of binary matrix operations. This helps us to recognize and to exploit the entropy reduction in bipartite protocols and to find the most general structure for the Clifford operations used in multipartite protocols. Doing so, we outperform existing protocols significantly and the yields obtained in this way have not been surpassed at the time of writing.

The article [52] is not included in this thesis because it falls outside the scope of entanglement distillation. In this paper, we describe generalizations of the stabilizer formalism concepts Pauli group, Clifford group and stabilizer states for quantum systems of arbitrary dimensions. Such d-level systems, where d is an arbitrary natural number greater than two, are often named *qudits.* We examine a link with modular arithmetic, which yields a transparent way of representing stabilizer states and Clifford operations with matrices over $\mathbb{Z}_d$. As such, this is a generalization of the qubit stabilizer formalism [24], from binary to modular.

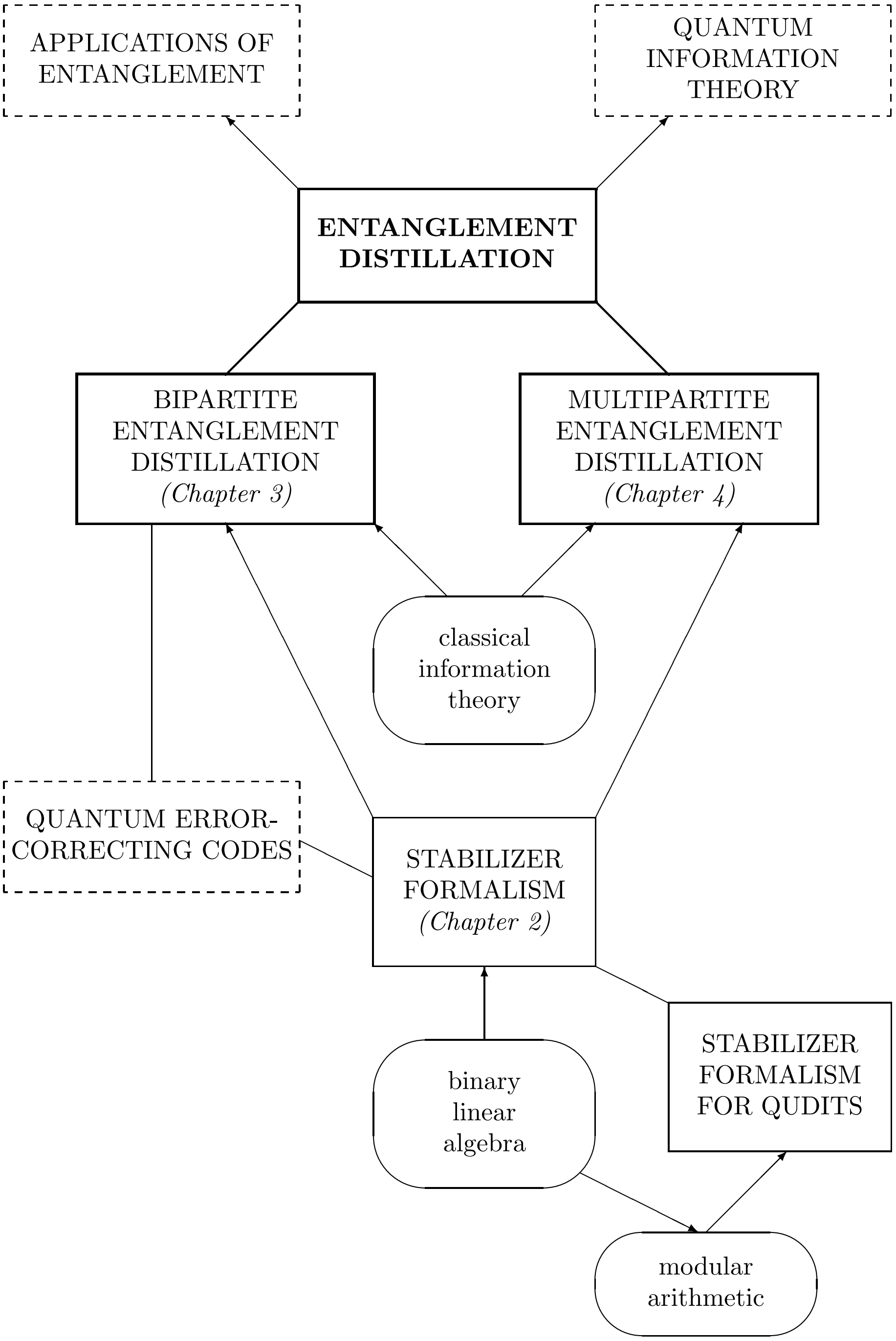

APPLICATIONS OF ENTANGLEMENT
QUANTUM INFORMATION THEORY
ENTANGLEMENT DISTILLATION
BIPARTITE ENTANGLEMENT DISTILLATION (Chapter 3)
MULTIPARTITE ENTANGLEMENT DISTILLATION (Chapter 4)
classical information theory
QUANTUM ERROR-CORRECTING CODES
STABILIZER FORMALISM (Chapter 2)
binary linear algebra
STABILIZER FORMALISM FOR QUDITS
modular arithmetic

Figure 1.3: Structure and cohesion of our results.

Chapter 2

Stabilizer formalism

2.1 Introduction

The stabilizer formalism is a mathematical framework describing *stabilizer states* and *Clifford operations*, an important class of quantum states and operations, in a group theoretic setting. This allows for a complete and efficient description in terms of linear operations on binary matrices. The main topic of this thesis is the distillation of stabilizer states using Clifford operations in this 'binary picture'. We believe it is this binary algebraic view that gave us the opportunity to develop entanglement distillation protocols that significantly improved previous protocols.

In this preliminary chapter, we elaborate on properties of the binary description of stabilizer states and Clifford operations in the context of entanglement distillation protocols. We start in section 2.2 with the definition the Pauli group on n qubits $\mathcal{G}_n$, a multiplicative group consisting of Kronecker products of Pauli matrices, and show that, without taking overall phase factors into account, it is isomorphic with the vector space $\mathbb{Z}_2^{2n}$. Next, in section 2.3, Clifford operations are introduced as unitary operations mapping the Pauli group to itself under conjugation. Again neglecting the action on the phase factors, a Clifford operation on a Pauli operation is translated into binary terms as the left multiplication of a binary matrix on the vector representing the Pauli operation. We show how an arbitrary Clifford operation can be composed as a sequence of elementary one- and two-qubit Clifford operations, which might also be of interest for the practical realization of a Clifford operation.

In section 2.4, we define a stabilizer state as the simultaneous eigenvector of the stabilizer, a maximal Abelian subgroup of the Pauli group. It is completely determined by the binary representations of a set of Pauli operations generating the stabilizer, which can be assembled as the columns of a binary matrix. We show how Clifford operations and Pauli measurements on a stabilizer state, the only quantum operations our distillation protocols will consist of, are described in the binary matrix language.

At that point, we have gathered enough material to focus on some particular general properties in the binary picture of the distillation protocols for stabilizer states that will be the topic of the following chapters. This is the content of section 2.5. The action on the phase factors can be circumvented in the context of distillation protocols. With this, we are able to simplify some intricate formulas of the preceding sections. When describing distillation protocols in later chapters, we will rely upon this section only. This allows the reader not to spend too much attention to the details of the preceding sections.

Finally, in section 2.6, we give a short introduction on stabilizer codes, an important class of quantum error-correcting codes. This serves as an illustration of the application of the stabilizer formalism, but is also relevant for the development of good distillation protocols, as quantum error correction and entanglement distillation are closely related.

The bulk of the binary linear algebra in sections 2.2 to 2.4 is mainly based on the paper of Dehaene et al. [24]. The basic theory of the stabilizer formalism, and more specifically of stabilizer codes, has been explained in various excellent standard texts, such as Nielsen and Chuang [68] and Preskill [72]. Important articles in the development of the stabilizer formalism are by Gottesman and Preskill (extensive material can be found in [40]) and by Calderbank, Rains, Shor and Sloane [17, 18].

2.2 Pauli group

In this section, we describe the Pauli group and its properties that are relevant to our purposes.

We begin with the definition of the *Pauli matrices*

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \ \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \ \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \tag{2.1}$$

that, in quantum mechanics, are well known for representing the observables describing the spin of a spin $\frac{1}{2}$ particle in the three spatial dimensions. By matrix multiplication, the Pauli matrices generate the *Pauli group on one qubit*, denoted by $\mathcal{G}_1$. This group is closely related to the algebra of *quaternions*. The relations

$$\begin{aligned} \sigma_x\sigma_y &= i\sigma_z = -\sigma_y\sigma_x, \\ \sigma_z\sigma_x &= i\sigma_y = -\sigma_x\sigma_z, \\ \sigma_y\sigma_z &= i\sigma_x = -\sigma_z\sigma_y, \\ \sigma_x^2 &= \sigma_y^2 = \sigma_z^2 = I_2, \end{aligned} \tag{2.2}$$

can easily be verified and lead to

$$\mathcal{G}_1 = \{1, i, -1, -i\} \times \{I_2, \sigma_x, \sigma_y, \sigma_z\}. \tag{2.3}$$

The *Pauli group on n qubits*, denoted by $\mathcal{G}_n$, is the n-fold Kronecker product of $\mathcal{G}_1$ with itself. Elements of $\mathcal{G}_n$, which we will refer to as *Pauli operations*, consist of Kronecker products of I_2 and the Pauli matrices, with an overall complex phase factor in $\{1, i, -1, -i\}$. The number of factors in the Kronecker product that differ from the identity, is called the *weight* of the Pauli operation. All Pauli operations are unitary.

The link with binary linear algebra is laid out as follows. We denote the identity and the Pauli matrices by

$$\sigma_{00} = I_2,\ \sigma_{01} = \sigma_x,\ \sigma_{11} = \sigma_y,\ \sigma_{10} = \sigma_z. \tag{2.4}$$

For the sake of simplicity in the formulas below, we use the alternative with real τ matrices, where $\tau_{00} = \sigma_{00},\ \tau_{01} = \sigma_{01},\ \tau_{10} = \sigma_{10}$, and

$$\tau_{11} = i\sigma_{11} = \left[\begin{array}{cc} 0 & 1 \\ -1 & 0 \end{array}\right]. \tag{2.5}$$

We use vector indices to indicate Kronecker products: let $v, w \in \mathbb{Z}_2^n$ and $a = \left[\begin{array}{c} v \\ w \end{array}\right]$, then we denote

$$\tau_a = \tau_{v_1 w_1} \otimes \ldots \otimes \tau_{v_n w_n}. \tag{2.6}$$

Any Pauli operation can be represented as $i^\delta(-1)^\epsilon \tau_a$, where $\delta, \epsilon \in \mathbb{Z}_2$ and $a \in \mathbb{Z}_2^{2n}$.

Remark. In the literature, a Pauli operation, e.g. $\sigma_x \otimes \sigma_z \otimes \sigma_y \otimes I \otimes \sigma_z$, is often denoted in a shorter form as $XZYIZ$. ◇

We have the following multiplication and commutation relations for Pauli operations:

Lemma 2.1 *Let $a, b \in \mathbb{Z}_2^{2n}$, then*

(i) $\tau_a \tau_b = (-1)^{b^T U a} \tau_{a+b}$,

(ii) $\tau_a \tau_b = (-1)^{b^T P a} \tau_b \tau_a$,

$$\textit{with } U = \left[\begin{array}{cc} 0_n & I_n \\ 0_n & 0_n \end{array}\right] \textit{ and } P = U + U^T = \left[\begin{array}{cc} 0_n & I_n \\ I_n & 0_n \end{array}\right].$$

These relations can easily be verified for $n = 1$ and then generalized for $n > 1$. The matrix P defines a *symplectic inner product* $b^T P a = a^T P b$ on the vector space $\mathbb{Z}_2^{2n}$. Note that $a^T P a = 0$, for any $a \in \mathbb{Z}_2^{2n}$.

Remark. The addition of binary objects is performed modulo 2, even in the exponent of i. We then use the following rule for multiplying powers of i:

$$i^s i^t = i^{s+t}(-1)^{st},$$

where $s, t \in \mathbb{Z}_2$. ◇

Remark. τ_{11} is the only non-Hermitian (actually skew Hermitian) of the four τ matrices, and multiplication with i makes it Hermitian. Therefore, $i^{\delta}(-1)^{\epsilon}\tau_a$ is Hermitian if and only if $\delta = a^T U a$. Indeed, $a^T U a$ counts (modulo 2) the number of τ_{11} occurrences in the Kronecker product. ◇

Sometimes, it is insightful to expand quantum states and operations in the canonical basis $|x\rangle$, for $x \in \mathbb{Z}_2^n$, of $\mathcal{H}_n$, also referred to as the *computational basis states.*

Lemma 2.2 *The expansion of τ_a, with $a = \begin{bmatrix} v \\ w \end{bmatrix} \in \mathbb{Z}_2^{2n}$, in the computational basis, is*

$$\tau_a = \sum_{x \in \mathbb{Z}_2^n} (-1)^{v^T x} |x\rangle \langle x + w| .$$

The following lemma will frequently be of use in calculations with these expansions:

Lemma 2.3 *Let $x \in \mathbb{Z}_2^n$, then*

$$\sum_{v \in \mathbb{Z}_2^n} (-1)^{v^T x} = 2^n \delta_{x,0}.$$

Again, these lemmas can easily be verified for $n = 1$ and then generalized for larger n. Application of both lemmas and lemma 2.1 leads directly to

Corollary 2.4 *Let $a \in \mathbb{Z}_2^{2n}$, then* $\mathrm{Tr}\left\{\tau_a\right\} = 2^n \delta_{a,0}$.

Corollary 2.5 *Defining the* matrix inner product $<M, N>= \mathrm{Tr}\left\{M^{\dagger} N\right\}$, *the Pauli operations τ_a, for $a \in \mathbb{Z}_2^{2n}$, form an orthogonal basis for $\mathbb{C}^{2^n \times 2^n}$.*

2.3 Clifford group

2.3.1 Basic properties

A *Clifford operation Q*, by definition, is a unitary operation that maps the Pauli group to itself under conjugation:

$$Q \mathcal{G}_n Q^{\dagger} = \mathcal{G}_n. \tag{2.7}$$

We call the group of all Clifford operations the *Clifford group on n qubits.* Because $Q\tau_a\tau_b Q^{\dagger} = Q\tau_a Q^{\dagger} Q \tau_b Q^{\dagger}$, it suffices to know the image of a generating set of the Pauli group in order to know the image of every Pauli operation. This defines Q up to an overall phase factor. In the binary picture, determining Q boils down to determining the images of τ_{b_k}, for $k = 1, \ldots, 2n$, where b_k, for $k = 1, \ldots, 2n$, form a basis of $\mathbb{Z}_2^{2n}$. Usually one takes the canonical basis e_k, for $k = 1, \ldots, 2n$. These operations τ_{e_k} correspond to single-qubit operations σ_x and

σ_z. Let $Q\tau_{e_k}Q^\dagger = i^{d_k}(-1)^{h_k}\tau_{C_k}$ and assemble the vectors C_k as columns in the matrix $C \in \mathbb{Z}_2^{2n\times 2n}$ and the scalars d_k, h_k in the vectors $d, h \in \mathbb{Z}_2^{2n}$.

Remark. $Q\tau_a Q^\dagger$ is (skew) Hermitian if and only if τ_a is (skew) Hermitian. Since all τ_{e_k} are Hermitian, so are their images. Therefore, $d_k = C_k^T U C_k$, or equivalently $d = \operatorname{diag}(C^T U C)$. In the following, we will always implicitly assume d defined in this way. ◇

The image of τ_a can be found by multiplying those operations $i^{d_k}(-1)^{h_k}\tau_{c_k}$ for which $a_k = 1$. By repeated application of lemma 2.1, this yields

Theorem 2.6 *Given $C \in \mathbb{Z}_2^{2n\times 2n}$ and $h \in \mathbb{Z}_2^{2n}$, representing the Clifford operation Q, we have*

$$\begin{aligned} Q\tau_a Q^\dagger &= i^\delta(-1)^\epsilon \tau_b, \\ \textit{with} \quad b &= Ca, \\ \delta &= d^T a, \\ \epsilon &= h^T a + a^T \operatorname{lows}\left(C^T U C + d d^T\right) a, \end{aligned}$$

where lows (A) *is the strict lower triangular part of the matrix A, i.e. A with all entries in the diagonal and above set to zero.*

Not all $C \in \mathbb{Z}_2^{2n\times 2n}$ represent Clifford operations. Indeed, using lemma 2.1 and theorem 2.6, we have

$$\begin{aligned} \tau_a\tau_b &= (-1)^{a^T P b}\tau_b\tau_a \\ \Leftrightarrow \quad Q\tau_a Q^\dagger Q\tau_b Q^\dagger &= (-1)^{a^T P b} Q\tau_b Q^\dagger Q\tau_a Q^\dagger \\ \Leftrightarrow \quad \tau_{Ca}\tau_{Cb} &= (-1)^{a^T P b}\tau_{Cb}\tau_{Ca}, \\ \text{and} \quad \tau_{Ca}\tau_{Cb} &= (-1)^{a^T C^T P C b}\tau_{Cb}\tau_{Ca}, \end{aligned}$$

where we omitted overall phase factors on the LHS and the RHS in the third equation, as they cancel each other out. We see that Clifford operations preserve the commutation relations between Pauli operations. The third and fourth equation must hold simultaneously for all $a, b \in \mathbb{Z}_2^{2n}$. Therefore, we have the necessary condition

$$C^T P C = P. \tag{2.8}$$

We call a matrix C satisfying (2.8) a *symplectic* matrix. It is also a sufficient condition for representing a Clifford operation. We will show this below.

To find the representation of the composition of two Clifford operations, we apply theorem 2.6 to find the images under the second operation of the images under the first operation of the canonical basis vectors, yielding

Corollary 2.7 *Given $C_{(1)}, C_{(2)} \in \mathbb{Z}_2^{2n\times 2n}$ and $h_{(1)}, h_{(2)} \in \mathbb{Z}_2^{2n}$, representing two Clifford operations $Q_{(1)}$ and $Q_{(2)}$, the product $Q_{(21)} = Q_{(2)}Q_{(1)}$ is represented by $C_{(21)} \in \mathbb{Z}_2^{2n\times 2n}$ and $h_{(21)} \in \mathbb{Z}_2^{2n}$ given by*

$$C_{(21)} = C_{(2)}C_{(1)},$$

$$\begin{aligned} h_{(21)} &= h_{(1)} + C_{(1)}^T h_{(2)} + \text{diag}\left[C_{(1)}\text{lows}\left(C_{(2)}^T U C_{(2)} + d_{(2)}d_{(2)}^T\right) C_{(1)}\right. , \\ &\qquad \left. + d_{(1)}d_{(2)}^T C_{(1)}\right] \end{aligned}$$

where $\text{diag}(A)$ *is the vector containing the diagonal of the matrix* A.

It is useful to know the representation of the inverse $Q^\dagger$ of a Clifford operation Q. With corollary 2.7 and (2.8), the following proposition can be verified.

Proposition 2.8 *Given* $C \in \mathbb{Z}_2^{2n\times 2n}$ *and* $h \in \mathbb{Z}_2^{2n}$, *representing a Clifford operation* Q, *the inverse* $Q^\dagger$ *is represented by*

$$\begin{aligned} C' &= C^{-1} = PC^TP, \\ h' &= C^{-T}(h+d) + \text{diag}[C^{-T}\text{lows}\left(C^TUC + dd^T\right)C^{-1}]. \end{aligned}$$

Finally, using (2.6), the binary representation of the Kronecker product of Clifford operations is straightforward:

Proposition 2.9 *Let* Q *and* Q' *be two Clifford operations represented by* C, h *and* C', h' *respectively, where*

$$C = \begin{bmatrix} C_{(11)} & C_{(12)} \\ C_{(21)} & C_{(22)} \end{bmatrix},\ h = \begin{bmatrix} h_{(1)} \\ h_{(2)} \end{bmatrix},$$
$$\textit{and}\quad C' = \begin{bmatrix} C'_{(11)} & C'_{(12)} \\ C'_{(21)} & C'_{(22)} \end{bmatrix},\ h' = \begin{bmatrix} h'_{(1)} \\ h'_{(2)} \end{bmatrix},$$

then $Q \otimes Q'$ *is represented by*

$$\begin{bmatrix} C_{(11)} & 0 & C_{(12)} & 0 \\ 0 & C'_{(11)} & 0 & C'_{(12)} \\ C_{(21)} & 0 & C_{(22)} & 0 \\ 0 & C'_{(21)} & 0 & C'_{(22)} \end{bmatrix} \textit{ and } h = \begin{bmatrix} h_{(1)} \\ h'_{(1)} \\ h_{(2)} \\ h'_{(2)} \end{bmatrix}.$$

(All $C_{(ij)}, C'_{(ij)}$ *are* $\in \mathbb{Z}_2^{n\times n}$ *and all* $h_{(i)}, h'_{(i)}$ *are* $\in \mathbb{Z}_2^n$.*)*

Corollary 2.10 *A Clifford operation represented by* C, h *acting on a subset* $A \subset \{1, \ldots, n\}$ *of* n *qubits is represented by* C *on the rows and columns with indices in* $A \cup (A+n)$, *embedded in the* $2n \times 2n$ *identity matrix. In the same way,* h *is embedded in the* $2n$ *all zeros vector.*

2.3.2 Special Clifford operations

We present a selected set of special Clifford operations, of particular interest for the decomposition of an arbitrary Clifford operation into one- and two-qubit operations. Often, such operations are called *gates*, in analogy with elementary gates in classical computational devices.

2.3.2.1 Pauli operations

We find the representing C, h of a Pauli operation τ_a by considering the images of τ_{e_k}, for $k = 1, \ldots, 2n$. From the commutation relations of Pauli operations (lemma 2.1), it follows that τ_a is represented by

$$\begin{array}{rcl} C & = & I_{2n}, \\ h & = & Pa. \end{array} \tag{2.9}$$

Note that, as for any Clifford operation, the overall phase factor of a Pauli operation cannot be represented.

2.3.2.2 Configuration space transformations

An *invertible linear transformation of the configuration space* is a unitary operation defined by

$$|x\rangle \to |Tx\rangle, \tag{2.10}$$

where $|x\rangle$, with $x \in \mathbb{Z}_2^n$, are the computational basis states and $T \in \mathbb{Z}_2^{n \times n}$ is an invertible matrix.

Proposition 2.11 *The operation defined by (2.10) is a Clifford operation, represented by*

$$\begin{array}{rcl} C & = & \left[\begin{array}{cc} T^{-T} & 0 \\ 0 & T \end{array} \right], \\ h & = & 0. \end{array}$$

Proof: We calculate the image of τ_a, for an arbitrary $a = \left[\begin{array}{c} v \\ w \end{array} \right] \in \mathbb{Z}_2^{2n}$, expanded in the computational basis using lemma 2.2:

$$\begin{array}{rcl} Q\tau_a Q^\dagger & = & \left(\sum\limits_{x \in \mathbb{Z}_2^n} |Tx\rangle \langle x| \right) \left(\sum\limits_{y \in \mathbb{Z}_2^n} (-1)^{v^T y} |y\rangle \langle y + w| \right) \left(\sum\limits_{z \in \mathbb{Z}_2^n} |z\rangle \langle Tz| \right) \\ & = & \sum\limits_{y \in \mathbb{Z}_2^n} (-1)^{v^T y} |Ty\rangle \langle T(y+w)| \\ & = & \sum\limits_{y' \in \mathbb{Z}_2^n} (-1)^{v^T T^{-1} y'} |y'\rangle \langle y' + Tw| \\ & = & \tau_{Ca}. \end{array}$$

It follows from theorem 2.6 that $h^T a = 0$, for all a, as $C^T U C = U$. Therefore, $h = 0$. □

Example. Familiar examples of configuration space transformations are all permutations of the n qubits, with $T = \Pi$, the permutation matrix, and the two-qubit *CNOT gate*, defined by

$$|x\rangle\,|y\rangle \to |x\rangle\,|x+y\rangle \quad \Rightarrow \quad C = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}. \tag{2.11}$$

The first qubit involved in this operation, is called the *source*, and the second the *target*. We agree to refer to this operation as: a CNOT from source to target. ◇

2.3.2.3 Hadamard gate

The single-qubit *Hadamard gate* is defined as the unitary operation

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{2.12}$$

Proposition 2.12 *The Hadamard gate is represented by*

$$\begin{aligned} C &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \\ h &= 0. \end{aligned}$$

Proof: Expanding H in the computational basis, giving

$$H = \frac{1}{\sqrt{2}} \sum_{x,y\in\mathbb{Z}_2} (-1)^{xy}\,|x\rangle\,\langle y|\,,$$

we calculate the image of τ_a, for arbitrary $a = \begin{bmatrix} v \\ w \end{bmatrix} \in \mathbb{Z}_2^2$, in the same way as for proposition 2.11:

$$\begin{aligned} Q\tau_a Q^\dagger &= \left(\frac{1}{\sqrt{2}} \sum_{x,y\in\mathbb{Z}_2} (-1)^{xy}\,|x\rangle\,\langle y|\right) \left(\sum_{z\in\mathbb{Z}_2} (-1)^{vz}\,|z\rangle\,\langle z+w|\right) \\ &\quad \times \left(\frac{1}{\sqrt{2}} \sum_{s,t\in\mathbb{Z}_2} (-1)^{st}\,|s\rangle\,\langle t|\right) \\ &= \frac{1}{2} \sum_{x,z,t\in\mathbb{Z}_2} (-1)^{xz+vz+zt+wt}\,|x\rangle\,\langle t| \\ &= \sum_{x,t\in\mathbb{Z}_2} \left(\frac{1}{2} \sum_{z\in\mathbb{Z}_2} (-1)^{(x+v+t)z}\right) (-1)^{wt}\,|x\rangle\,\langle t| \end{aligned}$$

$$\begin{aligned} &= \sum_{x\in\mathbb{Z}_2} (-1)^{w(x+v)} \left|x\right\rangle \left\langle x+v\right| \\ &= (-1)^{wv}\tau_{Ca}. \end{aligned}$$

We used lemma 2.3 in the third step. By theorem 2.6 and $C^TUC = U^T$, we have $wv = h^Ta + wv$, for all a. Consequently, $h = 0$. □

2.3.2.4 Phase gate

The single-qubit *phase gate* is the unitary operation

$$F = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix} = \sum_{x\in\mathbb{Z}_2} i^x \left|x\right\rangle \left\langle x\right|. \tag{2.13}$$

Proposition 2.13 *The phase gate is represented by*

$$\begin{aligned} C &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \\ h &= \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \end{aligned}$$

Proof: Given $a = \begin{bmatrix} v \\ w \end{bmatrix} \in \mathbb{Z}_2^2$, we have

$$\begin{aligned} Q\tau_aQ^\dagger &= \left(\sum_{x\in\mathbb{Z}_2} i^x \left|x\right\rangle \left\langle x\right|\right)\left(\sum_{y\in\mathbb{Z}_2} (-1)^{vy} \left|y\right\rangle \left\langle y+w\right|\right)\left(\sum_{z\in\mathbb{Z}_2} (-1)^z i^z \left|z\right\rangle \left\langle z\right|\right) \\ &= \sum_{y\in\mathbb{Z}_2} i^{y+(y+w)}(-1)^{y(y+w)+vy+(y+w)} \left|y\right\rangle \left\langle y+w\right| \\ &= \sum_{y\in\mathbb{Z}_2} i^w(-1)^{(v+w)y+w} \left|y\right\rangle \left\langle y+w\right| \\ &= (-i)^w\tau_{Ca}. \end{aligned}$$

Note that one has to take care while dealing with binary objects in the exponent of i (cf. remark on page 15). We find h with theorem 2.6 and

$$C^TUC = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \quad \Rightarrow \quad d = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

□

2.3.3 Decomposing a Clifford operation into one- and two-qubit operations

We show how, armed with Pauli operations, the elementary two-qubit CNOT gate and the single-qubit hadamard and phase gate, arbitrary symplectic C and

h can be realized. Firstly, this serves as a constructive proof for the sufficiency of (2.8) for representing a Clifford operation. Secondly, it might be of use in the practical realization of a Clifford operation.

We observe that the main problem is realizing C, not h. Indeed, let Q be represented by C, h, then $Q\tau_{P(h+h')} = \tau_{CP(h+h')}Q$ is represented by C, h', for arbitrary h'. This can be verified using (2.9) and corollary 2.7. Therefore, we may neglect h in the following and focus on C.

By corollary 2.10, CNOT, hadamard gate and phase gate translate on a binary level to the following elementary row operations:

- CNOT from qubit k to qubit l results in the adding of the row l to row k and of row $n+k$ to row $n+l$;
- H on qubit k results in swapping rows k and $n+k$;
- F on qubit k results in adding row $n+k$ to row k.

We show how these are applied to transform C into the identity. As I_{2n} is formed by left multiplication of such elementary row operations on C, a decomposition of C then consists of the inverses of these operations in reverse order (the inverses of CNOT, H and F are CNOT, H and $\tau_{10}F$ respectively).

Firstly, we go through a number of steps, transforming C_1 into e_1, schematically depicted as follows:[1]

$$C_1 \xrightarrow{1} \left[\begin{array}{c} 1 \\ . \\ . \\ . \\ \hline . \\ . \\ . \\ . \end{array}\right] \xrightarrow{2} \left[\begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \\ \hline . \\ . \\ . \\ . \end{array}\right] \xrightarrow{3} \left[\begin{array}{c} . \\ 0 \\ 0 \\ 0 \\ \hline 1 \\ . \\ . \\ . \end{array}\right] \xrightarrow{4} \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ \hline 1 \\ . \\ . \\ . \end{array}\right] \xrightarrow{5} \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ \hline 1 \\ 0 \\ 0 \\ 0 \end{array}\right] \xrightarrow{6} \left[\begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \\ \hline 0 \\ 0 \\ 0 \\ 0 \end{array}\right]$$

1. We make sure that $C_{11} = 1$. Note that C_1 must contain ones, otherwise C is not invertible. If the upper half of C_1 is zero, we apply H on qubit k, for some k for which $C_{n+k,1} = 1$. Then, if $C_{11} = 0$, it can be made 1 by applying CNOT from the first qubit to qubit k, where $C_{k1} = 1$.
2. We apply CNOTs from qubits k to the first, for all k for which $C_{k1} = 1$. This turns the upper half of C_1 into e_1.
3. We apply H on the first qubit.
4. We apply F on the first qubit if $C_{11} = 1$.
5. We apply CNOTs from the first qubit to qubits k, for all k for which $C_{n+k,1} = 1$.

[1] We keep on referring to intermediate stages as C_1.

6. Finally, we apply H on the first qubit again.

Secondly, we transform C_{n+1} into e_{n+1} by a very similar procedure. We already have that $C_1^T P C_{n+1} = 1$ because C is symplectic. As $C_1 = e_1$, this means $C_{n+1,n+1} = 1$. Observe that we now also need to take the action on C_1 into account, as we do not want to alter it:

$$C_{n+1} = \left[\begin{array}{c} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \hline 1 \\ \cdot \\ \cdot \\ \cdot \end{array}\right] \xrightarrow{1} \left[\begin{array}{c} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \hline 1 \\ 0 \\ 0 \\ 0 \end{array}\right] \xrightarrow{2} \left[\begin{array}{c} 0 \\ \cdot \\ \cdot \\ \cdot \\ \hline 1 \\ 0 \\ 0 \\ 0 \end{array}\right] \xrightarrow{3} \left[\begin{array}{c} 1 \\ \cdot \\ \cdot \\ \cdot \\ \hline 0 \\ 0 \\ 0 \\ 0 \end{array}\right] \xrightarrow{4} \left[\begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \\ \hline 0 \\ 0 \\ 0 \\ 0 \end{array}\right] \xrightarrow{5} \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ \hline 1 \\ 0 \\ 0 \\ 0 \end{array}\right]$$

$$C_1 = \left[\begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \\ \hline 0 \\ 0 \\ 0 \\ 0 \end{array}\right] \xrightarrow{1} \left[\begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \\ \hline 0 \\ 0 \\ 0 \\ 0 \end{array}\right] \xrightarrow{2} \left[\begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \\ \hline 0 \\ 0 \\ 0 \\ 0 \end{array}\right] \xrightarrow{3} \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ \hline 1 \\ 0 \\ 0 \\ 0 \end{array}\right] \xrightarrow{4} \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ \hline 1 \\ 0 \\ 0 \\ 0 \end{array}\right] \xrightarrow{5} \left[\begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \\ \hline 0 \\ 0 \\ 0 \\ 0 \end{array}\right]$$

1. We apply CNOTs from the first qubit to qubits k, for all k for which $C_{n+k,n+1} = 1$.
2. We apply F on the first qubit if $C_{1,n+1} = 1$.
3. We apply H on the first qubit.
4. We apply CNOTs from qubits k to the first, for all k for which $C_{k,n+1} = 1$.
5. Finally, we apply H on the first qubit again.

Now columns 1 and $n+1$ of C are the corresponding columns of I_{2n}. Because C is symplectic, we find that the rows 1 and $n+1$ of C are also the corresponding rows of I_{2n}:

$$C = \left[\begin{array}{c|ccc|c|ccc} 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \hline 0 & & & & 0 & & & \\ \vdots & & C_{(11)} & & \vdots & & C_{(12)} & \\ 0 & & & & 0 & & & \\ \hline 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ \hline 0 & & & & 0 & & & \\ \vdots & & C_{(21)} & & \vdots & & C_{(22)} & \\ 0 & & & & 0 & & & \end{array}\right].$$

Leaving the first qubit out, we can repeat the same procedure for

$$C' = \left[\begin{array}{cc} C_{(11)} & C_{(12)} \\ C_{(21)} & C_{(22)} \end{array}\right],$$

and so on. This recursively leads to the identity.

2.4 Stabilizer states

At this point, we have gathered enough material to define the special class of quantum states which we will focus on in the next chapters.

A *stabilizer state* $|\psi_{\mathcal{S}}\rangle$ on n qubits is the simultaneous eigenstate, with eigenvalues 1, of n commuting Hermitian Pauli operations $M_j = i^{f_j}(-1)^{b_j}\tau_{S_j}$, for $j = 1,\dots,n$, where $S_j \in \mathbb{Z}_2^{2n}$ are linearly independent and $b_j, f_j \in \mathbb{Z}_2$. These Pauli operations generate an Abelian subgroup of $\mathcal{G}_n$, called the *stabilizer* $\mathcal{S}$.

Example. The computational basis states $|x\rangle$, for $x \in \mathbb{Z}_2^n$, are stabilized by $(-1)^{x_j}\tau_{e_j}$, for $j = 1,\dots,n$. ◇

Proposition 2.14 *Given a stabilizer state $|\psi_{\mathcal{S}}\rangle$, and define $\rho_{\mathcal{S}} = |\psi_{\mathcal{S}}\rangle\langle\psi_{\mathcal{S}}|$, then*

(i) $M|\psi_{\mathcal{S}}\rangle = |\psi_{\mathcal{S}}\rangle, \ \forall M \in \mathcal{S}$;

(ii) $|\mathcal{S}| = 2^n$;

(iii) $\rho_{\mathcal{S}} = \frac{1}{2^n}\sum\limits_{M\in\mathcal{S}} M$.

As a consequence, $\mathcal{S}$ defines $|\psi_{\mathcal{S}}\rangle$ up to an overall phase factor.

Proof:

(i) From $M|\psi_{\mathcal{S}}\rangle = |\psi_{\mathcal{S}}\rangle$ and $N|\psi_{\mathcal{S}}\rangle = |\psi_{\mathcal{S}}\rangle$ it follows that $MN|\psi_{\mathcal{S}}\rangle = |\psi_{\mathcal{S}}\rangle$.

(ii) Since all M_j commute and are Hermitian, one can show that the products $\prod\limits_{j\in A} M_j$ are different for different subsets A of $\{1,\dots,n\}$.

(iii) Let ρ be the RHS of the equation. Firstly, we prove that ρ is a pure state. This is equivalent to $\mathrm{Tr}\left\{\rho^2\right\} = \mathrm{Tr}\{\rho\} = 1$ (see appendix A). We have

$$\rho^2 = \frac{1}{2^{2n}}\sum_{M\in\mathcal{S}}\sum_{N\in\mathcal{S}} MN = \frac{1}{2^{2n}}\sum_{M\in\mathcal{S}}\sum_{N'\in\mathcal{S}} N' = \frac{1}{2^n}\sum_{N'\in\mathcal{S}} N' = \rho.$$

Since all $M \in \mathcal{S}$ are Hermitian, $M^2 = M^\dagger M = I \in \mathcal{S}$. All other $e^{i\varphi}I$ cannot be in $\mathcal{S}$, as $e^{i\varphi}I|\psi_{\mathcal{S}}\rangle \neq |\psi_{\mathcal{S}}\rangle$. It follows from corollary 2.4 that

$$\mathrm{Tr}\{\rho\} = \frac{1}{2^n}\sum_{M\in\mathcal{S}}\mathrm{Tr}\{M\} = 1.$$

Secondly, we prove that $\mathrm{Tr}\,\{\rho\rho_{\mathcal{S}}\} = 1$.

$$\mathrm{Tr}\,\{\rho\rho_{\mathcal{S}}\} = \langle\psi_{\mathcal{S}}| \frac{1}{2^n} \sum_{M\in\mathcal{S}} M\,|\psi_{\mathcal{S}}\rangle = \langle\psi_{\mathcal{S}}|\psi_{\mathcal{S}}\rangle = 1.$$

□

Assembling the vectors S_j as the columns of a full rank matrix $S \in \mathbb{Z}_2^{2n\times n}$ and the scalars b_j, f_j in the vectors $b, f \in \mathbb{Z}_2^n$, Hermiticity and commutativity of $\mathcal{S}$ is reflected by

$$\begin{aligned} f &= \mathrm{diag}(S^T U S), &(2.14)\\ S^T P S &= 0. &(2.15)\end{aligned}$$

We will often denote $|\psi_{\mathcal{S}}\rangle$ by $|\psi_{S,b}\rangle$, or by $|\psi_b\rangle$ if S is known from the context. We refer to S as the *generator matrix* and to b as the *phase vector*.

The representation of $\mathcal{S}$ by S, b is not unique, as every other generating set of $\mathcal{S}$ yields an equivalent description. Going from one generating set to another is accomplished by multiplying an invertible $R \in \mathbb{Z}_2^{n\times n}$ on the right on S. By repeated application of lemma 2.1, we arrive at a new generating set, defined by

$$\begin{aligned} S' &= SR, \\ b' &= R^T b + \mathrm{diag}[R^T \mathrm{lows}\left(S^T U S + f f^T\right) R]. \end{aligned} \qquad (2.16)$$

We will refer to this as a *stabilizer basis change.*

If a stabilizer state $|\psi_{\mathcal{S}}\rangle$ is operated on by a Clifford operation Q, $Q\,|\psi_{\mathcal{S}}\rangle$ is a new stabilizer state with stabilizer $Q\mathcal{S}Q^\dagger$. Using theorem 2.6, we arrive at the following theorem:

Theorem 2.15 *$Q\,|\psi_{S,b}\rangle = |\psi_{S',b'}\rangle$, where*

$$\begin{aligned} S' &= CS, \\ b' &= b + S^T h + \mathrm{diag}[S^T \mathrm{lows}\left(C^T U C + d d^T\right) S + f d^T S]. \end{aligned}$$

In analogy with proposition 2.9, we use (2.6) for the representation of a tensor product of stabilizer states:

Proposition 2.16 *The tensor product $|\psi_{S,b}\rangle \otimes |\psi_{S',b'}\rangle$, where*

$$S = \begin{bmatrix} S_z \\ S_x \end{bmatrix} \quad \textit{and} \quad S' = \begin{bmatrix} S'_z \\ S'_x \end{bmatrix},$$

is a stabilizer state represented by

$$\begin{bmatrix} S_z & 0 \\ 0 & S'_z \\ S_x & 0 \\ 0 & S'_x \end{bmatrix}, \begin{bmatrix} b \\ b' \end{bmatrix}.$$

Finally, we arrive at the following lemma and theorem, describing the measurement of a Hermitian Pauli operation on a stabilizer state,[2] which is a crucial element in the distillation protocols we will describe in the next chapters.

Lemma 2.17 *Given a stabilizer $\mathcal{S}$ and $N \in \mathcal{G}_n$. Then either*

- *N commutes with every element of $\mathcal{S}$, in which case N or $-N \in \mathcal{S}$; or*
- *$\mathcal{S}$ can be split into the subgroup $\mathcal{S}_0$ and coset $\mathcal{S}_1 = M_1\mathcal{S}_0$, containing the elements that respectively commute and anticommute with N.*

Proof: Let $S \in \mathbb{Z}_2^{2n\times n}$ and $a \in \mathbb{Z}_2^{2n}$ represent $\mathcal{S}$ and N respectively. N commutes with every element of $\mathcal{S} \Leftrightarrow S^TPa = 0$. Since S^TP is a full rank $n \times 2n$ matrix, a is contained by an n-dimensional space. But $\text{col}\,(S)$ is an n-dimensional subspace of the same space, as S^TPS. Consequently, $a \in \text{col}\,(S)$, which means that $\pm N \in \mathcal{S}$.

Conversely, if $\pm N \notin \mathcal{S}$, then there exist at least one $M_1 \in \mathcal{S}$ that anticommutes with N. We split $\mathcal{S}$ into two disjoint non-empty subsets $\mathcal{S}_0$ and $\mathcal{S}_1$ as defined above. The product of every two operations commuting with N also commutes with N. Furthermore, M_1M and N anticommute, for each $M \in \mathcal{S}_0$. So $\mathcal{S}_0$ is a subgroup of $\mathcal{S}$ and $\mathcal{S}_1$ is its coset. □

Theorem 2.18 *Given a stabilizer state $|\psi_\mathcal{S}\rangle \in \mathcal{H}_n$ and non-trivial $N \in \mathcal{G}_n$ (i.e. $N \neq \pm I$), then measuring the observable N on $|\psi_\mathcal{S}\rangle$ has either outcome $+1$ or -1. We distinguish two cases:*

(i) if N or $-N \in \mathcal{S}$, then, with certainty, the outcome is respectively $+1$ or -1 and $|\psi_\mathcal{S}\rangle$ is left unchanged;

(ii) if $\pm N \notin \mathcal{S}$, then, both with probability $\frac{1}{2}$, the outcome is $+1$ or -1 [$= (-1)^u$, with $u \in \mathbb{Z}_2$] and $|\psi_\mathcal{S}\rangle$ is transformed into $|\psi_{\mathcal{S}'}\rangle$, where $\mathcal{S}' = \mathcal{S}_0 \cup (-1)^uN\mathcal{S}_0$ and $\mathcal{S}_0$ is the subgroup of $\mathcal{S}$ that commutes with N.

Proof: Measuring N can only yield outcomes $+1$ or -1. Indeed, as $N^2 = I$, the eigenvalues of N must square to 1. Let P_u be the projector onto the eigenspace of N with eigenvalue $(-1)^u$, then

$$P_u = \frac{1}{2}[I + (-1)^uN].$$

Indeed, $P_u^2 = P_u$, $NP_u = (-1)^uP_u$ and $P_0 + P_1 = I$. The measurement of N on the state $\rho_\mathcal{S}$ has outcome $(-1)^u$ with probability $p_u = \text{Tr}\,\{P_u\rho_\mathcal{S}P_u\}$ and, by the measurement, the state is transformed into $P_u\rho_\mathcal{S}P_u/p_u$ (see appendix A). We calculate this for both cases (i) and (ii).

(i) Assume $N \in \mathcal{S}$ (the proof is analogous for $-N \in \mathcal{S}$). It is obvious that $P_0\,|\psi_\mathcal{S}\rangle = |\psi_\mathcal{S}\rangle$ and $P_1\,|\psi_\mathcal{S}\rangle = 0$. Therefore, we have outcome $+1$ with certainty and the state is projected onto itself.

[2] When measuring a Pauli operation, we will always assume it Hermitian.

(ii) We know from lemma 2.17 that $\mathcal{S}$ can be split into respectively commuting and anticommuting $\mathcal{S}_0$ and $\mathcal{S}_1 = M_1\mathcal{S}_0$. Now, using proposition 2.14 (iii) and the given commutation relations, we calculate

$$\begin{aligned}
P_u \rho_{\mathcal{S}} P_u &= \frac{1}{2}[I + (-1)^u N] \left(\frac{1}{2^n} \sum_{M\in\mathcal{S}} M \right) \frac{1}{2}[I + (-1)^u N] \\
&= \frac{1}{2^{n+2}}[I + (-1)^u N] \left((I + M_1) \sum_{M\in\mathcal{S}_0} M \right) [I + (-1)^u N] \\
&= \frac{1}{2^{n+2}}[I + (-1)^u N](I + M_1)[I + (-1)^u N] \sum_{M\in\mathcal{S}_0} M \\
&= \frac{1}{2^{n+1}}[I + (-1)^u N] \sum_{M\in\mathcal{S}_0} M \\
&= \frac{1}{2^{n+1}} \sum_{M\in\mathcal{S}'} M.
\end{aligned}$$

From corollary 2.4, it follows that $\operatorname{Tr}\{P_u \rho_{\mathcal{S}} P_u\} = \frac{1}{2}$. □

In the binary picture, theorem 2.18 reads as follows:

Corollary 2.19 *Measuring the non-trivial observable $i^{a^TUa}\tau_a$ on $|\psi_{S,b}\rangle$ yields outcome $(-1)^u$ and transforms the state into $|\psi_{S'',b''}\rangle$, where,*

(i) if $S^TPa = 0$, then $\exists v : a = Sv$ and $u = v^Tb + v^T\operatorname{lows}\left(S^TUS + ff^T\right)v$ and the state is left unchanged by the measurement: $|\psi_{S'',b''}\rangle = |\psi_{S,b}\rangle$;

(ii) if $S^TPa \neq 0$, then $u = 0$ or 1 with equal probability. Let S', b' be defined by (2.16), where the matrix R is invertible and satisfies $(R^{-T})_1 = S^TPa$. Then $S''_1 = a$ and $b''_1 = u$, and $S''_j = S'_j$ and $b''_j = b'_j$, $\forall j \neq 1$.

Proof: Case (i) follows directly from lemma 2.17 and update formulas (2.16).

Case (ii): $(R^{-T})_1 = R^{-T}e_1 = S^TPa \Rightarrow (SR)^TPa = e_1$. Therefore, S' and b' without the first column and entry respectively, represent a generating set of $\mathcal{S}_0$. The measurement has projected the state onto the eigenspace of $i^{a^TUa}(-1)^u\tau_a$. Consequently, S'' and b'' as defined above represent the state after the measurement. □

2.5 Distillation in the stabilizer formalism

In this section, we explain how distillation protocols in the stabilizer formalism work. Recalling the explanation of the general procedure in chapter 1, we have n parties sharing κ copies of an entangled n-qubit mixed state ρ, each party having control over the same qubit of all copies. In a realistic setting,

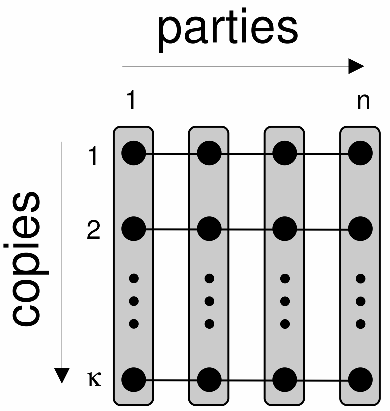


Figure 2.1: Each of n parties performs a local Clifford operation (here depicted as rectangles) on the corresponding qubits (here depicted as dots) of all κ copies, each entangled over the parties.

it is natural to assume the copies in the same state, as a result of the same preparation by one party and the same subsequent distribution via imperfect but time-invariant and memory-less quantum channels. Each party applies a local Clifford operation on the qubits he is in charge of. These local operations result in statistical dependence of the copies. As such, measurements on a part of all copies will provide information on the overall initial state. We observe that there are two kinds of 'locality': firstly, the overall initial state is a tensor product $\rho^{\otimes\kappa}$, describing κ independent states, but entangled over the n parties; secondly, these parties perform Clifford operations that are local with respect to parties, but non-local with respect to copies. This is illustrated in figure 2.1.

This section is organized as follows. In section 2.5.1, we show how a general mixed state can be reduced to a state with a density matrix that is diagonal in the basis of all stabilizer states represented by S (cf. Bell-diagonal states), which gives the possibility of describing the entire protocol in the binary picture. Then, in section 2.5.2, we simplify two formulas of the preceding sections that play a central role in the binary description of distillation protocols and in section 2.5.3 we explain how the two kinds of locality in the protocol are reflected in the binary matrices describing the states and operations. In section 2.5.4, we investigate the limits of information extraction from a stabilizer state by local measurements. Finally, in section 2.5.5, we focus on the properties of a number of important particular kinds of stabilizer states, for which specialized protocols will be dealt with in the following chapters.

2.5.1 Depolarization of mixed states

The distillation protocol starts with multiple copies of a general mixed state ρ, distributed over all parties. In order to have a state that we can fully describe in the binary picture, we need the pre-processing technique *depolarization* [3]. This reduces ρ to a state that can be interpreted as a classical ensemble of stabilizer states, all represented by the same generator matrix S, or, equivalently, as a pure stabilizer state, represented by S and some unknown b.

Lemma 2.20 *The states $|\psi_{S,b}\rangle$, $\forall b \in \mathbb{Z}_2^n$, form a basis for $\mathcal{H}_n$. We will refer to this basis as the S-*basis.

Proof: Since eigenspaces with different eigenvalues of a Hermitian operation are orthogonal, it follows that $|\psi_{S,b}\rangle$ and $|\psi_{S,b'}\rangle$ are orthogonal if $b \neq b'$. □

Example. The computational basis states are $|x\rangle = \left|\psi_{[I\ 0]^T,x}\right\rangle$. ◇

Theorem 2.21 *Given an arbitrary n-qubit mixed state ρ, expanded in the S-basis as*

$$\rho = \sum_{b,b' \in \mathbb{Z}_2^n} \rho_{b,b'} \left|\psi_{S,b}\right\rangle \left\langle\psi_{S,b'}\right| .$$

By subsequently performing the operation τ_{S_j} on the state with probability $\frac{1}{2}$, for $j = 1, \ldots, n$, the initial state ρ is transformed into

$$\rho' = \sum_{b \in \mathbb{Z}_2^n} p_b \left|\psi_{S,b}\right\rangle \left\langle\psi_{S,b}\right| ,$$

where $p_b = \rho_{b,b}$. Note that this operation can be performed locally (as Pauli operations are entirely separable), but requires the use of classical communication between the parties.

Proof: The above operation transforms ρ as follows:

$$\begin{aligned} \rho \quad &\rightarrow \quad \frac{1}{2}(\rho + \tau_{S_j} \rho \tau_{S_j}^\dagger) \\ &= \sum_{b,b' \in \mathbb{Z}_2^n} \rho_{b,b'} \frac{1}{2}[1 + (-1)^{b_j + b'_j}] \left|\psi_{S,b}\right\rangle \left\langle\psi_{S,b'}\right| \\ &= \sum_{b,b' \in \mathbb{Z}_2^n} \rho_{b,b'} \delta_{b_j,b'_j} \left|\psi_{S,b}\right\rangle \left\langle\psi_{S,b'}\right| \end{aligned}$$

Consequently, $\rho' = \sum_{b,b' \in \mathbb{Z}_2^n} \rho_{b,b'} \delta_{b,b'} \left|\psi_{S,b}\right\rangle \left\langle\psi_{S,b'}\right| = \sum_{b \in \mathbb{Z}_2^n} p_b \left|\psi_{S,b}\right\rangle \left\langle\psi_{S,b}\right|$. □

This mixed state ρ' is equivalent to a classical ensemble of pure stabilizer states $|\psi_{S,b}\rangle$, each with probability p_b. We can also regard this state as an unknown pure stabilizer state, i.e. we *assume* the state is a pure stabilizer state, and with probability p_b it is equal to $|\psi_{S,b}\rangle$.

Remark. One might wonder whether the depolarization operation defined in theorem 2.21 really needs to be carried out. Indeed, with probability $\frac{1}{2}$, we perform the operation τ_{S_j}, for $j = 1, \ldots, n$, but afterwards we *do not know* whether it was carried out or not. In a sense, we throw away information about what we have done, which could equally have been nothing at all. In [14], it is argued that revealing what particular Pauli operation was carried out *after* the protocol, cannot change the resulting pure output state. Therefore, the protocol equally works if depolarization is omitted.

However, we believe this reasoning is not correct, since the output state is not exactly pure. We investigate this more rigorously. Let ρ_0 be the initial overall state of all κ n-qubit copies, ρ the state after depolarization, $\mathcal{P}(\rho)$ the result of the protocol and $|D\rangle$ the desired output state. All protocols we will encounter, deliver a state that approach this state:

$$\mathcal{P}(\rho) = (1-\delta) |D\rangle \langle D| + \delta\rho', \tag{2.17}$$

where ρ' is some density operator for which $\langle D| \rho' |D\rangle = 0$, and δ is the (vanishing) failure probability of the protocol. By theorem 2.21, the state after depolarization equals

$$\rho = \sum_i 2^{-n\kappa} \rho_i,$$

where ρ_i is the result of applying a particular Pauli operation on ρ as prescripted by the depolarization procedure. Without loss of generality, we can write

$$\mathcal{P}(\rho) = 2^{-n\kappa} \sum_i \mathcal{P}(\rho_i) = 2^{-n\kappa} \sum_i (1-\delta_i) |D\rangle \langle D| + \delta_i \rho_i'. \tag{2.18}$$

The first equality follows from the fact that the protocol is, like any evolution, a *linear* operation (see appendix A).

Combining (2.17) and (2.18), it follows that

$$2^{-n\kappa} \sum_i \delta_i = \delta, \quad \text{or} \quad \delta_0 \leq 2^{n\kappa}\delta.$$

Note that δ_0 is the failure probability of the protocol without depolarization. Clearly, unless $\delta = \mathcal{O}(2^{-n\kappa})$, for $\kappa \to \infty$, nothing can be said about the asymptotic behavior of δ_0. In fact, we will see later that $\delta = \mathcal{O}(\kappa^{-1})$. ◇

2.5.2 Getting rid of redundant phase information

Following the previous section, the state of each copy after depolarization is an ensemble of pure stabilizer states with fixed generator matrix, but the phase vector is a random variable. Recalling the representation (2.9) of a Pauli operation, used as Clifford operation, one finds $\tau_a |\psi_{S,b}\rangle = |\psi_{S,b'}\rangle$, where

$$b' = b + S^T Pa. \tag{2.19}$$

Since $S^T P$ is full rank, $S^T Pa$ can be any element of $\mathbb{Z}_2^n$ by varying a. Furthermore, it is independent of b. Therefore, we can get rid of the intricate second term in the phase update equations in (2.16) and theorem 2.15, which are respectively function of S and R, or S, C and h, but independent of b, by applying the appropriate extra Pauli operation to the state. Recall that this can be done locally, as Pauli operations are entirely separable. We arrive at the following simplified modifications of (2.16) and theorem 2.15:

- A stabilizer basis change for the stabilizer state $|\psi_{S,b}\rangle$ is described by

$$\begin{array}{rcl} S' & = & SR, \\ b' & = & R^T b. \end{array} \tag{2.20}$$

- The action of a Clifford operation, represented by C, on a stabilizer state $|\psi_{S,b}\rangle$, is described by

$$\begin{array}{rcl} S' & = & CS, \\ b' & = & b. \end{array} \tag{2.21}$$

  Note that h no longer needs to be taken into account.

Both (2.20) and (2.21) hold modulo some extra Pauli operation.

2.5.3 Two kinds of locality

Since ρ can be regarded as a classical ensemble of pure stabilizer states, the same does hold for the overall state $\rho^{\otimes\kappa}$. Consequently, the entire operation without the measurements can be regarded as the action of n local Clifford operations on a tensor product of κ stabilizer states, each represented by the same generator matrix $S \in \mathbb{Z}_2^{2n\times n}$ but independent $b_k \in \mathbb{Z}_2^n$, for $k = 1, \ldots, \kappa$. There are as many parties n as there are qubits in the state: each party has control over the corresponding qubits of all copies.

We observed two kinds of 'locality' in the general setting of a distillation protocol: on the one hand, we have a tensor product of κ stabilizer states on n qubits, entangled over the n parties, and on the other hand, these parties perform Clifford operations that are local with respect to parties, but non-local with respect to copies (cf. figure 2.1).[3] We already saw propositions 2.9 and 2.16 for representing Kronecker products of Clifford operations and tensor products of stabilizer states. However, we need to take into account these different kinds of locality. Following proposition 2.9, we agree that we firstly order the qubits

[3] To be precise, the only true locality is that of the Clifford operations, because the parties are spatially separated; the copies on the other hand are separable, but not spatially separated.

per party firstly, and then per copy. The overall Clifford operation of the protocol is then represented by

$$C = \left[\begin{array}{cc} \tilde{A} & \tilde{B} \\ \tilde{C} & \tilde{D} \end{array} \right] = \left[\begin{array}{ccc|ccc} A_1 & & & B_1 & & \\ & \ddots & & & \ddots & \\ & & A_n & & & B_n \\ \hline C_1 & & & D_1 & & \\ & \ddots & & & \ddots & \\ & & C_n & & & D_n \end{array} \right], \tag{2.22}$$

where $\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D}$ are block diagonal matrices of $\kappa \times \kappa$ blocks A_j, B_j, C_j, D_j. The representations of the local Clifford operations must satisfy the symplecticity condition (2.8), which translates into

$$\left. \begin{array}{rcl} A_j^T C_j + C_j^T A_j & = & 0, \\ B_j^T D_j + D_j^T B_j & = & 0, \\ A_j^T D_j + C_j^T B_j & = & I_\kappa, \end{array} \right\} \quad \text{for } j = 1, \ldots, n. \tag{2.23}$$

The overall stabilizer state is represented by

$$\tilde{S} = S \otimes I_\kappa \quad \text{and} \quad \tilde{b}, \tag{2.24}$$

where $\tilde{b}_{(j-1)n+k} = (b_k)_j$, for $k = 1, \ldots, \kappa$, and for $j = 1, \ldots, n$. Note the difference with proposition 2.16, where the qubits are ordered per copy firstly, and then per party.

Following the line of thought of section 2.5.1, the overall state is a pure stabilizer state $|\psi_{\tilde{S},\tilde{b}}\rangle$ with probability

$$p_{\tilde{b}} = \prod_{k=1}^{\kappa} p_{b_k}. \tag{2.25}$$

Indeed, as the state of each copy is an independent identically-distributed ensemble of pure stabilizer states, the overall state of all copies is an ensemble of tensor products of these stabilizer states, with probability distribution equal to the product of the identical distributions of the single copies.

2.5.4 Extracting information from a stabilizer state

The purpose of local Clifford operations is to spread the information on the unknown random variable $\tilde{b}$ over all copies. Information on $\tilde{b}$ can then be extracted by measuring a part of the copies. We investigate what information can be revealed from a stabilizer state $|\psi_{S,b}\rangle$, where S is given but b is unknown. This is accomplished by performing Pauli measurements on the state. However, as the parties are separate, only local measurements are at their disposal. The value of a non-local observable

$$M = M_1 \otimes \ldots \otimes M_n$$

is determined by locally measuring M_1 on the first qubit, M_2 on the second qubit, and so on, and taking the product of the outcomes.

After a Pauli measurement on a single qubit has been carried out, the outcome of any other Pauli measurement anticommuting with the first will be entirely random and contains no information on the original state. Indeed, after the first measurement, the second measurement falls under case (ii) of theorem 2.18. It follows that, once all qubits have been measured, the state is projected onto eigenstates of these local measurements and no more information can be retrieved from it. We then arrive at the main theorem concerning information extraction on a stabilizer state:

Theorem 2.22 *Given $a \in \mathbb{Z}_2^{2n}$, we define the sets*

$$\begin{aligned} X(a) &= \{j \mid a_{\{j,n+j\}} = 01\}, \\ Y(a) &= \{j \mid a_{\{j,n+j\}} = 11\}, \\ Z(a) &= \{j \mid a_{\{j,n+j\}} = 10\}. \end{aligned}$$

Now let $|\psi_{S,b}\rangle \in \mathcal{H}_n$, where b is unknown, and let $\mathcal{M}$ be a partition of $\{1,\ldots,n\}$ into three subsets X, Y and Z and perform τ_{01}, $i\tau_{11}$ or τ_{10} measurements on qubits $j \in X$, Y or Z respectively. Then from the outcomes of these measurements we can calculate $v^T b$ for all v that satisfy

$$\begin{aligned} X(Sv) &\subset X, \\ Y(Sv) &\subset Y, \\ Z(Sv) &\subset Z. \end{aligned} \tag{2.26}$$

All v satisfying (2.26) constitute a subspace $\mathcal{V}(\mathcal{M})$ of $\mathbb{Z}_2^n$. After the measurements, all other information on b is no longer accessible.

Proof: The value of observable $i^{(Sv)^T USv}\tau_{Sv}$ is equal to the product of the outcomes of its non-trivial factors. This can be done for all v that satisfy (2.26). Using corollary 2.19 (i), we then calculate $v^T b$.

Next, we show that $v_1, v_2 \in \mathcal{V}(\mathcal{M}) \Rightarrow v_1 + v_2 \in \mathcal{V}(\mathcal{M})$. Let $j \in X$. Then $(Sv_1)_{\{j,n+j\}} = (Sv_2)_{\{j,n+j\}} = 00$ or 01. Therefore, $(S[v_1 + v_2])_{\{j,n+j\}} = 00$ or $01 \Rightarrow X(S[v_1 + v_2]) \subset X$. Analogous arguments can be made for Y and Z. □

Example. By measuring $\tau_{10} = \sigma_z$ on the second qubit and $\tau_{01} = \sigma_x$ on the first and third qubit of the stabilizer state $|\psi_{S,b}\rangle$, with

$$S = \left[\begin{array}{ccc} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right],$$

we find b_1 and b_3 by taking the product of the outcomes of the measurements respectively on the first and last two qubits. ◇

2.5.5 Special stabilizer states

2.5.5.1 Bell states

The four Bell states are the simplest and best known example of bipartite quantum entanglement. They are two-qubit pure states and defined as:

$$\begin{array}{rclcl} |B_{00}\rangle &=& |\Phi^+\rangle &=& \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \\ |B_{01}\rangle &=& |\Psi^+\rangle &=& \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \\ |B_{10}\rangle &=& |\Phi^-\rangle &=& \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), \\ |B_{11}\rangle &=& |\Psi^-\rangle &=& \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \end{array} \tag{2.27}$$

With lemma 2.2, one finds that

$$\begin{array}{rclrcl} \tau_{0011}\,|B_{00}\rangle &=& |B_{00}\rangle\,, & \tau_{1100}\,|B_{00}\rangle &=& |B_{00}\rangle\,, \\ \tau_{0011}\,|B_{01}\rangle &=& |B_{01}\rangle\,, & \tau_{1100}\,|B_{01}\rangle &=& -\,|B_{01}\rangle\,, \\ \tau_{0011}\,|B_{10}\rangle &=& -\,|B_{10}\rangle\,, & \tau_{1100}\,|B_{10}\rangle &=& |B_{10}\rangle\,, \\ \tau_{0011}\,|B_{11}\rangle &=& -\,|B_{11}\rangle\,, & \tau_{1100}\,|B_{11}\rangle &=& -\,|B_{11}\rangle\,. \end{array} \tag{2.28}$$

It follows that

$$|B_b\rangle = |\psi_{S_B,b}\rangle\,,\ \text{where } S_B = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}. \tag{2.29}$$

Remark. The *singlet state* $|\Psi^-\rangle = |B_{11}\rangle$ has the peculiar property of always having anticorrelating outcomes when performing the same local measurement with outcomes ± 1 on each qubit. Equivalently, $|B_{11}\rangle$ is an eigenstate of $M \otimes M$ with eigenvalue -1, where M is an arbitrary single-qubit observable with eigenvalues ± 1. Analogous results exist for the other Bell states [68]. ◇

From theorem 2.22, it follows that performing the same measurement on both sides of the Bell state $|B_b\rangle$, we learn the value of:

$$\begin{array}{lrcl} \text{if the measurements are} & \sigma_x & \to & b_1, \\ & \sigma_z & \to & b_2, \\ & \sigma_y & \to & e^T b. \end{array} \tag{2.30}$$

After that, the state is separable and no more information can be retrieved from it.

Using (2.24), a tensor product of κ Bell states is given by

$$|B_{\tilde{b}}\rangle = \left|\psi_{S_B \otimes I_\kappa, \tilde{b}}\right\rangle. \tag{2.31}$$

Example. $|B_{010\,011}\rangle = |B_{00}\rangle \otimes |B_{11}\rangle \otimes |B_{01}\rangle$. ◇

2.5.5.2 Graph states

Graph states are a special kind of stabilizer states, which in recent years have received much attention in the literature (we refer to [45] and references therein). Every n-qubit graph state is identified with an undirected graph on n vertices. An edge connecting two qubits symbolizes a two-qubit interaction that has taken place between these qubits. Graph states are mainly interesting for two reasons. Firstly, any stabilizer state is *local Clifford equivalent* to some graph state.[4] Therefore, the study of entanglement in general stabilizer states, including distillation, can be restricted to graph states. Secondly, many operations on graph states can be translated into graph transformations. This yields a very insightful description of quantum states, and interesting links between the mathematics of graphs and important problems in quantum information and computation theory have already been laid [84, 85, 86]. For most proofs in this section, we will refer to specialized literature, as they are beyond the scope of this thesis.

A graph state is a stabilizer state with generator matrix

$$S = \begin{bmatrix} \Gamma \\ I \end{bmatrix}, \tag{2.32}$$

where $\Gamma \in \mathbb{Z}_2^{n \times n}$ is a zero-diagonal symmetric matrix. This graph state is identified with the graph with adjacency matrix Γ.

Remark. A graph state is said to be *k-colorable* if one is able to label the vertices of the corresponding graph in minimally k different colors such that no vertices with the same color are connected. ◇

Example. The 5-qubit *ring* state is identified with the graph in figure 2.2. This graph has adjacency matrix

$$\Gamma = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}.$$

It is a 3-colorable graph state. ◇

The next lemma and proposition show that an edge in the identifying graph symbolizes a real physical interaction between the qubits involved.

Lemma 2.23 *Analogous to the CNOT gate, the* controlled-Z gate *is the symmetric Clifford operation $|x\rangle |y\rangle \to (-1)^{xy} |x\rangle |y\rangle$ and is represented by*

$$C = \begin{bmatrix} I_2 & P_2 \\ 0 & I_2 \end{bmatrix}.$$

[4] Local Clifford equivalence of two states means that by one-qubit Clifford operations, they can be transformed into one another.

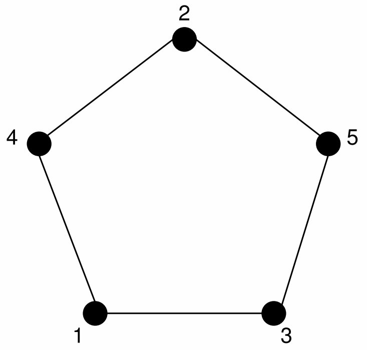


Figure 2.2: Graph of the 5-qubit ring state.

Proof: We calculate the image of τ_a, where $a = \begin{bmatrix} v \\ w \end{bmatrix} \in \mathbb{Z}_2^4$, expanded in the computational basis using lemma 2.2:

$$\begin{aligned}
\sum_{x\in\mathbb{Z}_2^2} (-1)^{v^T x} |x\rangle \langle x+w| \quad &\to \quad \sum_{x\in\mathbb{Z}_2^2} (-1)^{v^T x + x^T U x + (x+w)^T U (x+w)} |x\rangle \langle x+w| \\
&= \quad \sum_{x\in\mathbb{Z}_2^2} (-1)^{v^T x + w^T P x + w^T U w} |x\rangle \langle x+w| \\
&\sim \quad \sum_{x\in\mathbb{Z}_2^2} (-1)^{(v+Pw)^T x} |x\rangle \langle x+w| \\
&= \quad \tau_{Ca}.
\end{aligned}$$

□

Proposition 2.24 *Let* $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, *and perform a controlled-Z on every two qubits of* $|+\rangle^{\otimes n}$ *connected by an edge in a given graph with n vertices. Then the result is a graph state identified with that graph.*

Proof: It is easily verified that $|+\rangle^{\otimes n}$ is a stabilizer state with generator matrix $\begin{bmatrix} 0 \\ I_n \end{bmatrix}$. By induction, the rest then follows from: $C \begin{bmatrix} 0 \\ I_2 \end{bmatrix} = \begin{bmatrix} P_2 \\ I_2 \end{bmatrix}$. □

Theorem 2.25 *Any stabilizer state is local Clifford equivalent to some graph state.*

This theorem is of major importance for stabilizer state distillation, since this allows us to focus on the distillation of graph states only. For proof, we refer to [80, 82].

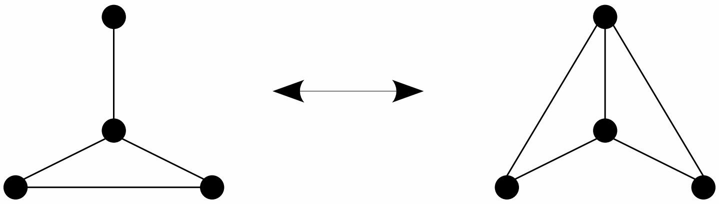


Figure 2.3: Local complementation at the central vertex.

It is conjectured that local unitary equivalence implies local Clifford equivalence for stabilizer states. This means that if two stabilizer states can be transformed into one another by local unitary operations, there exist local Clifford operations that achieve the same goal. This conjecture is already proven for a large class of graph states [83, 95], but until now, no general proof has been found. Local equivalence of stabilizer states and graph states is studied in detail in [80].

Finally, the following theorem gives a criterion to determine whether two given graph states are local Clifford equivalent. A proof is given in [80, 81]. This is important for distillation, since this allows us to always take the best protocol for an entire equivalence class. We only need to transform one state into the other by local Clifford operations before and after the protocol. Firstly, we define the graph transformation of *local complementation*. The *neighborhood* of a vertex j in a given graph is defined as the set of vertices that are connected to vertex j. Then local complementation at vertex j, by definition, yields the same graph, but with complemented neighborhood of vertex j, i.e. in the subgraph on the neighborhood of vertex j, all edges are removed and edges are drawn where there were none before. This is illustrated with an example in figure 2.3.

Theorem 2.26 *Two graph states are local Clifford equivalent if and only if their identifying graphs can be transformed into one another by the (possibly multiple) application of local complementation.*

Example. By performing a local Hadamard on every qubit, the computational basis states are transformed into graph states identified with a graph with only vertices and no edges. ◇

Example. By performing a local Hadamard on one of the qubits of a Bell state, we get a graph state, identified with two connected vertices. ◇

Example. It can be shown (by exhaustive application of local complementation) that no k-colorable graph state, where $k \leq 2$, is local Clifford equivalent to the 5-qubit ring state of figure 2.2. ◇

2.5.5.3 CSS states

CSS states, short for *Calderbank-Shor-Steane states*, are stabilizer states that have a generator matrix, up to a linear transformation on the right (2.20), of the form:

$$S = \begin{bmatrix} S_z & 0 \\ 0 & S_x \end{bmatrix}, \tag{2.33}$$

where $S_z \in \mathbb{Z}_2^{n \times n_z}$ and $S_x \in \mathbb{Z}_2^{n \times n_x}$. Since the generator matrix is full rank, so are S_z and S_x, and $n_z + n_x = n$. The constraint of commutativity (2.15) translates into $S_z^T S_x = 0$. Therefore, once S_z (or S_x) is known, $\mathrm{col}\,(S)$ is entirely defined.

Example. The Bell states are CSS states with $S_z = S_x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. ◇

Example. The so-called *cat state*, often referred to as *GHZ state*, is defined as $\frac{1}{\sqrt{2}}(|00\ldots 0\rangle + |11\ldots 1\rangle)$. It is a CSS state with

$$S_x = e \quad \text{and} \quad S_z = \begin{bmatrix} I \\ e^T \end{bmatrix}. \tag{2.34}$$

◇

Proposition 2.27 *Any two-colorable graph state is local Hadamard equivalent to a CSS state and vice versa.*

Proof: A two-colorable graph state is identified with a graph with adjacency matrix of the form

$$\Gamma = \begin{bmatrix} 0 & \theta^T \\ \theta & 0 \end{bmatrix}.$$

With proposition 2.12, local Hadamards on qubits $j = 1, \ldots, n_z$, yield the transformation

$$S = \begin{bmatrix} 0 & \theta^T \\ \theta & 0 \\ I_{n_z} & 0 \\ 0 & I_{n_x} \end{bmatrix} \quad \rightarrow \quad \begin{bmatrix} I_{n_z} & 0 \\ \theta & 0 \\ 0 & \theta^T \\ 0 & I_{n_x} \end{bmatrix}. \tag{2.35}$$

Applying the same local Hadamards once again yields the inverse transformation. We now prove that any CSS state generator matrix S can be brought in the form of the RHS of (2.35) by multiplying an invertible R on the right of S and possibly permuting some qubits. Since S_z is full rank, there exists a permutation $\Pi \in \mathbb{Z}_2^{n\times n}$ and invertible $R_z \in \mathbb{Z}_2^{n_z\times n_z}$ such that

$$\Pi S_z R_z = \left[\begin{array}{c} I_{n_z} \\ \theta \end{array}\right].$$

The permutation has transformed S_x into

$$\Pi S_x = \left[\begin{array}{c} A \\ B \end{array}\right],$$

where $A \in \mathbb{Z}_2^{n_z\times n_x}$ and $B \in \mathbb{Z}_2^{n_x\times n_x}$. Since $S_z^T S_x = 0$, we have $A = \theta^T B$. It follows that B is invertible. Indeed, if B would not be full rank, then there exists an invertible $R_x \in \mathbb{Z}_2^{n_x\times n_x}$ such that $BR_x = [B'\ 0]$, but then $A = \theta^T B = [A'\ 0]$, yielding a contradiction, since S_x is full rank. Let R_x be the inverse of B. Then S is transformed as follows:

$$S \to \left[\begin{array}{cc} \Pi & 0 \\ 0 & \Pi \end{array}\right]\left[\begin{array}{cc} S_z & 0 \\ 0 & S_x \end{array}\right]\left[\begin{array}{cc} R_z & 0 \\ 0 & R_x \end{array}\right] = \left[\begin{array}{cc} I_{n_z} & 0 \\ \theta & 0 \\ 0 & \theta^T \\ 0 & I_{n_x} \end{array}\right].$$

□

2.6 Stabilizer codes

Fighting decoherence –the uncontrollable influence of the environment– plays a central role in quantum information and computation theory. Since decoherence is extremely hard to prevent, we have to find ways for protecting the fragile quantum state. Therefore, quantum error-correcting codes are an indispensable tool in the road towards effective applications such as quantum computation, quantum communication and quantum cryptography. Similarly as in classical coding, the quantum information of k qubits is *spread out* over n qubits, adding structured redundancy which makes it more robust to noise and decoherence. However, a major difference with classical coding is that the error correction process should not reveal any information on the state, as this would equally destroy its coherence. For the same reasons, the quantum information in qubits, contrary to classical bits, cannot be copied as a consequence of the linearity of quantum mechanical evolution (the *no-cloning theorem* [28, 94]).

Because of its transparent structure, the stabilizer formalism lends itself very well to the task of finding good quantum error-correcting codes. Moreover, stabilizer codes are equivalent to classical linear codes over GF(4) with the additional constraint that the codes must be self-dual [18]. As a consequence, many techniques have been borrowed from an area that was already known and thoroughly investigated long before quantum applications were suggested.

The general evolution (including decoherence) of a quantum state ρ is mathematically described by (see appendix A)

$$\rho \quad \rightarrow \quad \mathcal{E}(\rho) = \sum_i E_i \rho E_i^\dagger, \tag{2.36}$$

where the operations E_i satisfy $\sum_i E_i^\dagger E_i = I$ in order to have a trace-preserving map. A common interpretation of (2.36) is: ρ is mapped to $\rho_i = E_i \rho E_i^\dagger$ with probability $\mathrm{Tr}\{\rho_i\}$. Recovery is possible, if we are able to find a map $\mathcal{R}$ such that $\mathcal{R}[\mathcal{E}(\rho)] = \rho$.

In complete analogy with stabilizer states, a *stabilizer code* $\mathcal{H}_\mathcal{S}$ on n qubits is the simultaneous eigenspace, with eigenvalues 1, of the stabilizer $\mathcal{S}$ generated by $n-k$ commuting Hermitian Pauli operations $M_j = i^{f_j}(-1)^{b_j}\tau_{S_j}$, for $j = 1,\ldots,n-k$, where $S_j \in \mathbb{Z}_2^{2n}$ are linearly independent. The stabilizer code is completely determined by $S \in \mathbb{Z}_2^{2n\times(n-k)}$ and $b \in \mathbb{Z}_2^{n-k}$. We fix the following notation:

- $\mathcal{C} = \mathrm{col}\,(S)$, the binary counterpart of $\mathcal{S}$, and
- $\mathcal{N} = (P\mathcal{C})^\perp$, the binary counterpart of the *normalizer* of $\mathcal{S}$, as $\tau_a^\dagger \mathcal{S} \tau_a = \mathcal{S}$, $\forall a \in \mathcal{N}$. Note that $\mathcal{C} \subseteq \mathcal{N}$.

Proposition 2.28 *The stabilizer code $\mathcal{H}_\mathcal{S}$ as defined above is a subspace of $\mathcal{H}_n$ of dimension 2^k.*

Proof: We can always extend S to a full rank matrix $\bar{S}$ that satisfies $\bar{S}^T P \bar{S} = 0$, where $\bar{S}_{k+j} = S_j$, for $j = 1,\ldots,n-k$, . Indeed, since S is full rank, the subspace $\mathcal{N}$ of $\mathbb{Z}_2^{2n}$ containing all x satisfying $S^T P x = 0$, has dimension $2n - (n-k) = n+k$. The set $\mathcal{N}\backslash\mathcal{C}$ contains $2^{n+k} - 2^{n-k}$ vectors. We pick one and add it as a column to the left of S. Next, we can do the same for this extended matrix, and so on. Then, analogously to $\bar{S}$, we define

$$\bar{b} = \begin{bmatrix} b' \\ b \end{bmatrix} \in \mathbb{Z}_2^n.$$

By lemma 2.20, the set of states $|\psi_{\bar{S},\bar{b}}\rangle$, with fixed b but variable $b' \in \mathbb{Z}_2^k$, are a basis of the space $\mathcal{H}_\mathcal{S}$. It follows that the dimension of $\mathcal{H}_\mathcal{S}$ is 2^k. □

Fixing a basis $\{|\psi_{\bar{S},\bar{b}}\rangle\}_{b'\in\mathbb{Z}_2^k}$ for the code space $\mathcal{H}_\mathcal{S}$, we can encode k *logical qubits* by associating each computational basis state $|x\rangle \in \mathcal{H}_k$ to a stabilizer code basis state $|\psi_{\bar{S},\bar{b}}\rangle$. The unitary operation mapping $|\bar{x}\rangle$, where $\bar{x}^T = [x\ 0] \in \mathbb{Z}_2^n$, to $|\psi_{\bar{S},\bar{b}}\rangle$ is called an *encoding operation* for this stabilizer code. Its inverse is called a *decoding operation.*

Proposition 2.29 *Any Clifford operation represented by $C \in \mathbb{Z}_2^{2n\times 2n}$ and $h \in \mathbb{Z}_2^{2n}$, where*

$$C_{k+1,\ldots,n} = S \quad \textit{and} \quad h_{k+1,\ldots,n} = b,$$

is an encoding operation for the stabilizer code represented by S and b.

Proof: Using theorem 2.15, one can verify that the Clifford operation represented by C and h, where

$$C_{1,\ldots,n} = \bar{S} \quad \text{and} \quad h_{1,\ldots,n} = \bar{b} + \bar{x},$$

maps $|\bar{x}\rangle = \left|\psi_{[I\ 0]^T,\bar{x}}\right\rangle$ to $\left|\psi_{\bar{S},\bar{b}}\right\rangle$. □

The stabilizer code basis, determined by S' and b', should be chosen in such a way that realizing the encoding or decoding operation is most efficient [20].

The next question is what errors (2.36) can be corrected if they have occurred on an encoded state $|\psi\rangle \in \mathcal{H}_\mathcal{S}$. We saw that, with a certain probability, $|\psi\rangle$ is transformed into $E_i\,|\psi\rangle$. Firstly, we focus on the case where E_i is a Pauli operation.

Theorem 2.30 *All errors $E_i \sim \tau_a$ on $|\psi\rangle \in \mathcal{H}_\mathcal{S}$, where $a \in E_\mathcal{S} \subset \mathbb{Z}_2^{2n}$, can be corrected if all $a, b \in E_\mathcal{S}$ satisfy*

$$a + b \notin \mathcal{N}\backslash\mathcal{C}.$$

Equivalently, either $S^T P(a+b) \neq 0$ or $a + b \in \mathcal{C}$ must hold, $\forall a, b \in E_\mathcal{S}$.

Proof: Since two Pauli operations either commute or anticommute, we have $E_i M_j = (-1)^{m_j} M_j E_i$. The scalars m_j are assembled in vector $m \in \mathbb{Z}_2^{n-k}$, which is called the *syndrome* of E_i. It is determined by measuring M_j, for $j = 1, \ldots, n-k$ on the afflicted state $E_i\,|\psi\rangle$, with outcomes $(-1)^{m_j}$. Indeed,

$$M_j E_i\,|\psi\rangle = (-1)^{m_j} E_i M_j\,|\psi\rangle = (-1)^{m_j} E_i\,|\psi\rangle\,.$$

Let $E_i \sim \tau_a$. Then the measurements reveal $m = S^T P a$. Applying τ_b to $E_i\,|\psi\rangle$, where $b \in E_\mathcal{S}$ and $S^T P b = m$, yields $|\psi\rangle$ up to a phase factor, so the error is corrected. Indeed, since $a \in E_\mathcal{S}$ by assumption, and $S^T P(a+b) = m+m = 0$, it must hold that $a+b \in \mathcal{C}$, or equivalently $\tau_b\tau_a \sim M$ for some $M \in \mathcal{S}$. Therefore, $\tau_b E_i\,|\psi\rangle \sim |\psi\rangle$. □

When E_i is not a Pauli operation, by corollary 2.5, we can always expand it as a weighted sum of Pauli operations:

$$E_i = \sum_a \omega_a \tau_a, \text{ where } \omega_a \in \mathbb{C}. \tag{2.37}$$

Corollary 2.31 *All errors E_i, expanded as in (2.37), on $|\psi\rangle \in \mathcal{H}_\mathcal{S}$, can be corrected if all a in the expansion are in $E_\mathcal{S}$.*

Proof: Since the measurements project onto the eigenspace corresponding to the syndrome m, all $\tau_a\,|\psi\rangle$ with a different syndrome are projected onto zero. Therefore, the state after the measurements is expanded as (2.37), where all $a \in E_\mathcal{S}$ and $S^T P a = m$. Again, we apply τ_b, where $b \in E_\mathcal{S}$ and $S^T P b = m$, and we get: $\tau_b E_i\,|\psi\rangle = \sum\limits_a \omega_a \tau_b \tau_a\,|\psi\rangle = \sum\limits_a \omega_a e^{i\varphi_a}\,|\psi\rangle \sim |\psi\rangle$. □

Analogously to classical codes, we define the *distance* of a stabilizer code $\mathcal{H}_\mathcal{S}$ to be the minimal weight Pauli operation τ_a for which $a \in \mathcal{N}\backslash\mathcal{C}$. From the above, it is clear that all Pauli errors of weight $\leq d-1$ can be detected, and all Pauli errors of weight $\leq \left\lfloor \frac{d-1}{2} \right\rfloor$ can be corrected. In general, a quantum error-correcting code, using n qubits to encode k logical qubits and with distance d, is referred to as an $[[n,k,d]]$ code.

Remark. An n-qubit stabilizer state is an $[[n,0,d]]$ stabilizer code. ◇

Example. The *5-qubit code* is a $[[5,1,3]]$ stabilizer code with stabilizer generated by the Pauli operations:

$$XZZXI, IXZZX, XIXZZ, ZXIXZ.$$

The Pauli operations that correspond to $a \in \mathcal{N}\backslash\mathcal{C}$ are, up to phase factors:

$$\begin{array}{cccccc}
IIXYX & IIZXZ & IIYZY & IXIYY & IXXIZ & IXYXI \\
IZIXX & IZXZI & IZZIY & IYIZZ & IYZYI & IYYIX \\
XIIXY & XIZIX & XIYYI & XXIZI & XXXXX & XXZYZ \\
XZIIZ & XZXYY & XZYZX & XYXII & XYZZY & XYYXZ \\
ZIIZX & ZIXXI & ZIYIZ & ZXXZY & ZXZII & ZXYYX \\
ZZIYI & ZZZZZ & ZZYXY & ZYIIY & ZYXYZ & ZYZXX \\
YIIYZ & YIXIY & YIZZI & YXIIX & YXZXY & YXYZZ \\
YZXXZ & YZZYX & YZYII & YYIXI & YYXZX & YYYYY
\end{array}$$

This code detects 2 errors and corrects 1 error. ◇

2.7 Conclusion

In this chapter, we have considered the stabilizer formalism, and the way it is efficiently described in terms of binary linear algebra. We have defined the elementary concepts of Pauli and Clifford operations and stabilizer states. A separate section was devoted to a number of particular properties that are relevant in the domain of entanglement distillation protocols. In the following chapters, we will mainly rely upon this part of this chapter. Finally, we briefly touched on the theory of stabilizer codes, a major application of the stabilizer formalism.

Chapter 3

Bipartite entanglement distillation

3.1 Introduction

In this chapter, we describe protocols for distilling bipartite entanglement in the binary picture. We discern two major families of protocols. On the one hand, there are the asymptotic protocols, that involve an infinite number of qubits pairs and output an infinite number of pure Bell state pairs. Their performance is expressed by the *asymptotic yield*, which is the proportion of distilled Bell pairs to the number of input qubit pairs. The best known of this family is the hashing protocol, which is equivalent to the breeding protocol [13, 14]. On the other hand, there are the recurrence protocol and variants thereof, which we call finite protocols, that involve only a finite number of qubit pairs [14, 25, 27, 63, 64, 65]. These protocols are mostly used in an iterative procedure for gradually increasing the available entanglement and are usually followed by the hashing protocol, since the asymptotic yield of iteratively applying finite protocols can be shown to be zero.

Although the operational part of both finite and asymptotic protocols have the same description in the binary picture, they turn out to be quite different when it comes to interpretation. Firstly, the action of asymptotic protocols is entirely described in terms of the information-theoretical concept typical set, whereas finite protocols result in a change of the probabilities defining the mixed state. Secondly, the hashing protocol as it is leaves no room for incorporating *adaptiveness*, causing it to perform rather poorly on noisy states. An iteration of finite protocols is intrinsically adaptive, as intermediate outcomes influence future actions, albeit that for most procedures, the adaptiveness is restricted to merely discarding the result for certain outcomes. The protocols of [2] take a step further, but despite violating all kinds of one-way communication quantum error correction bounds, they still have an asymptotic yield not exceeding that of hashing.

Our goal is to find ways of introducing the benefits of finite protocols into the hashing protocol. A first successful attempt in that direction was made in [90]. Yet, the strategy that is used there is rather ad hoc, in the sense that suggested generalizations consist of exhaustive searches over a rather unclear decision space. By recognizing the underlying principles that are at the basis of the improvements, we were able to create protocols that, by exploiting these ideas, outperform all existing schemes significantly [53].

This chapter is organized as follows. In section 3.2, we give a few basic characterizations of bipartite entanglement, particularly those that are relevant in the context of distillation. We briefly explain the concepts entanglement versus separability, the pure state entanglement measure entropy of entanglement, the mixed state entanglement measures entanglement of formation and entanglement of distillation, and the process of teleportation. We also argue why we only consider distillation of states that have a diagonal density matrix in the Bell state basis. This section is rather concise, as there are numerous excellent standard texts on entanglement, of which we only cite [50, 71, 87], but this is far from an exhaustive list.

In section 3.3, we recapitulate the contents of section 2.5 on techniques of distillation in the stabilizer formalism, this time applied to the particular case of Bell states. Firstly, we give two main theorems describing the actions of the protocols and illustrate them with the recurrence protocol. Secondly, we show that two different interpretations of the protocols are, in fact, equivalent [51]. This equivalence allows us to combine the advantages of both approaches, which will prove useful in the following sections. Thirdly, we elaborate on two ways of extracting information by measurements on the initial state. These two methods turn out to be equivalent for measurements involving an infinite number of qubit pairs, which is the reason why hashing and breeding are equivalent. Yet, this no longer holds for measurements on a finite number of qubit pairs, which led to the recognition of the major principle giving rise to the improvements of [53, 90].

In section 3.4, we briefly explain the hashing/breeding protocol and the asymptotic equipartition property on which it is based. Then, we are ready to explain in detail the key principles of our adaptive variants of breeding in section 3.5. This section is the main part of this chapter. Most of it was the content of [53]. Although the best variant that we can come up with in that section is fit for distilling even very noisy Bell-diagonal states, in the neighborhood of separable states, we argue that it is, in fact, equivalent to a number of recurrence iterations, producing states that are less noisy, followed by this variant. Therefore, it is useful to analyze finite protocols for very noisy entangled states, in the limit approaching separable states, which is done in the final section 3.6.

3.2 Basic characterizations of bipartite entanglement

Usually, entanglement is defined as the negation of separability. In the bipartite case, a state ρ_{AB} is said to be separable with respect to parties A and B if it can be written in the form [93]:

$$\rho_{AB} = \sum_i p_i \, \rho_A^{(i)} \otimes \rho_B^{(i)}, \tag{3.1}$$

where p_i is a probability distribution and all $\rho_A^{(i)}, \rho_B^{(i)}$ are density matrices for the systems A, B respectively. All correlations between measurement outcomes on the A-side and the B-side of (3.1) can be explained in a local hidden variable model. From the positivity of density matrices, the necessary *positive partial transpose* (PPT) condition for separability immediately follows [70]:

$$\rho_{AB} \text{ is a separable state} \quad \Rightarrow \quad \rho_{AB}^{T_A} > 0,$$

where the LHS of the inequality is the partial transpose of ρ_{AB} with respect to subsystem A. This condition is not sufficient for separability,[1] but it is sufficient for *non-distillability*, giving rise to the concept of *bound entanglement*, which cannot be distilled [49].

The entanglement of a pure state $|\psi\rangle$ is quantified by its *entropy of entanglement* [10], defined as:

$$E(|\psi\rangle) = S(\rho_A) = S(\rho_B), \tag{3.2}$$

where $S(\rho) = -\mathrm{Tr}\,\{\rho \log_2 \rho\}$ is the *Von Neumann entropy* and $\rho_A = \mathrm{Tr}_B\{|\psi\rangle\langle\psi|\}$ the reduced density matrix of subsystem A, obtained by taking the partial trace with respect to subsystem B (vice versa for ρ_B). Without any reference to subsystem B, the state of subsystem A is given by ρ_A. As the Von Neumann entropy is a measure for mixedness, (3.2) quantitatively characterizes pure state entanglement as the degree of coherence of two entangled quantum systems, i.e. the extent in which their states cannot be accurately described separately. The entropy of entanglement cannot increase under the action of LOCC, and it is invariant under local unitary operations. It is zero for a separable state and one for Bell states (ebits). An asymptotic number κ of copies of a pure state with entanglement equal to E can be transformed reversibly in the limit into κE Bell states (with some separable ancilla) by local operations, a process known as *entanglement concentration* [10].

For mixed states, the distinction between various levels of entanglement becomes less clear, let alone its quantification. Considering transformations of states by local operations and classical communication, the definition of two important entanglement measures arises, which we introduced in chapter 1:

[1] Although it is sufficient for two-qubit states.

the *entanglement of formation* E_F and the *entanglement of distillation* E_D. Both measures are defined in terms of LOCC equivalence to Bell states. This is natural, because Bell states are maximally-entangled and therefore their entanglement has been given the role of standard unit of entanglement. But their main importance lies in the fact that if one is able to establish qubit pairs in the Bell state between remote parties, any non-local state can be prepared, or equally, any non-local operation can be carried out. Indeed, by the process of teleportation [12], using only local operations and classical communication, any single-qubit state (possibly part of some larger state) can be reliably transmitted to a remote party if both parties already shared a Bell state pair. This also emphasizes the importance of Bell state distillation: arbitrary qubit states can be perfectly communicated by means of a noisy quantum channel and (faultless) classical communication by distributing noisy Bell states via the channel firstly, distilling them into perfect ebits and using them for teleporting the states under consideration.

In this thesis, we focus on two-qubit states that are diagonal in the Bell state basis:

$$\rho = p_{00} \left|B_{00}\right\rangle \left\langle B_{00}\right| + p_{01} \left|B_{01}\right\rangle \left\langle B_{01}\right| + p_{10} \left|B_{10}\right\rangle \left\langle B_{10}\right| + p_{11} \left|B_{11}\right\rangle \left\langle B_{11}\right| , \quad (3.3)$$

for the following reasons. Firstly, as explained in the previous sections, a Bell-diagonal state can be interpreted as a classical ensemble of the four Bell states. Since Bell states are stabilizer states, and the protocols we consider use Clifford operations and Pauli measurements only, we can describe the protocols entirely in the transparent binary picture. Secondly, we showed in chapter 2 that, by performing depolarization, any density matrix can be reduced to a state that is diagonal in the Bell state basis, while maintaining the values of its diagonal entries. Yet, depolarization in a sense throws away information on the initial state, so one could expect that this is accompanied with some loss of entanglement as well. In [88], an optimal two-qubit LOCC *filtering* method is devised, which, if successful, returns a state, which is Bell-diagonal, with the highest entanglement of formation achievable in this way,[2] thereby providing an alternative for depolarization and an extra argument in favor of exclusively considering Bell-diagonal states.

In our search for good distillation protocols, it is useful to have some clue on their performance by comparing the yields with bounds for E_D for Bell-diagonal states. From the fundamental property that entanglement cannot increase on average under LOCC, we have that E_D is upper bounded by E_F. For a Bell-diagonal state ρ, E_F is a one-to-one function (plotted in figure 3.1) of the *fully entangled fraction* or *fidelity* F, which is the largest overlap $\left\langle e\right| \rho \left|e\right\rangle$ with a maximally-entangled two-qubit state $\left|e\right\rangle$, i.e. a state that is local unitary equivalent to the Bell states [14]. For Bell-diagonal states, F is simply the largest diagonal entry. We observe in figure 3.1 that all Bell-diagonal states

[2] However, this method might not be optimal when the success probability is taken into account [87].

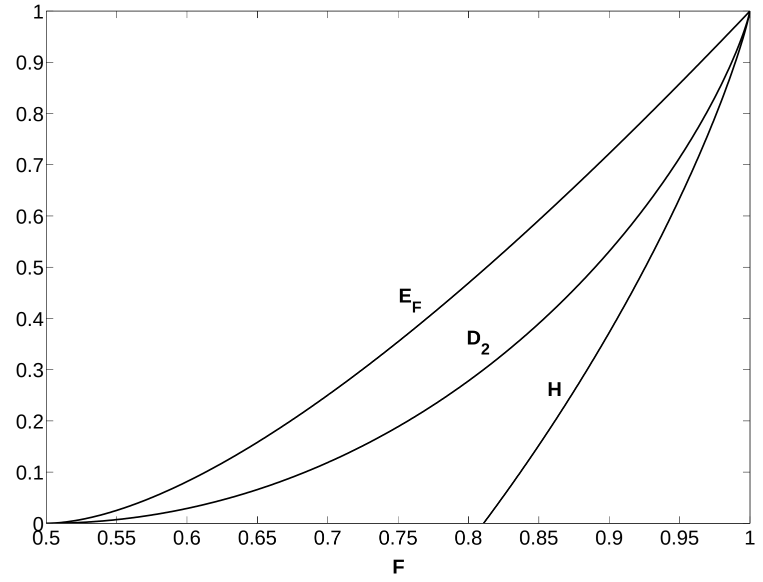


Figure 3.1: The entanglement of formation E_F of a Bell-diagonal state and bounds for E_D of a Werner state, as a function of the fidelity F. The lower bound H is the yield of the hashing protocol and the upper bound D_2 is the entanglement of distillation for rank two Bell-diagonal states and the asymptotic relative entropy of entanglement.

with $F \leq \frac{1}{2}$ are separable. States for which $F > \frac{1}{2}$ are known to be entangled and distillable [48].

Interesting particular cases of Bell-diagonal states are on the one hand *rank two Bell-diagonal states*, with two zeros on the diagonal of ρ, and, on the other hand, *Werner states* or *isotropic states*, with $F = p_{00}$ and $p_{01} = p_{10} = p_{11} = \frac{1-F}{3}$. They are the result of sending one qubit of a perfect ebit through a *bit flip, phase flip or bit-phase flip channel* and a *depolarizing* channel [68] respectively. In a sense, Werner states can be considered most distant to rank two states.[3] In [5, 74], the following lower and upper bound for the entanglement of distillation of Werner states, with fidelity F, were calculated:

$$1 + F \log_2 F + (1-F) \log_2 \frac{1-F}{3} \leq E_D \leq 1 + F \log_2 F + (1-F) \log_2 (1-F).$$

Both bounds are also plotted in figure 3.1. Note that the upper bound is already significantly below E_F, which is an expression of the irreversible nature

[3] In fact, within the set of Bell-diagonal states, the minimal *trace distance* [68] or the minimal *quantum Chernoff bound* [6] to any rank two state, *is* maximal for Werner states.

of entanglement manipulation of mixed states. We will show in section 3.4 that the LHS is the yield of the hashing protocol, and the RHS is the hashing yield for the rank two Bell-diagonal state with the same fidelity, which is proven to be its entanglement of distillation [73]. It is also the *asymptotic relative entropy of entanglement* for Werner states [5]. Since the problem of distillation is completely solved for rank two Bell-diagonal states, we will primarily compare distillation protocols on their performance for Werner states.

3.3 Protocols for Bell-diagonal states

The protocol starts with multiple copies of a Bell-diagonal state (3.3) shared by two parties, Alice and Bob. Recalling the content of section 2.5.1, we regard the overall state of the κ copies as a pure tensor product of Bell states, binary identified with $S_B \otimes I_\kappa$ and $\tilde{b}$ according to (2.31), where $\tilde{b}$ is a random variable with probability distribution $p_{\tilde{b}}$. Both Alice and Bob locally apply a Clifford operation in order that the information contained in $\tilde{b}$ is permuted. How this is accomplished is the content of the following theorem [25]:

Theorem 3.1 *When Alice and Bob respectively perform the same Clifford operation, represented by a symplectic $C \in \mathbb{Z}_2^{2\kappa\times 2\kappa}$, then $\tilde{b}$ is transformed as:*

$$\tilde{b} \to C\tilde{b}.$$

As C is invertible, this transformation is a permutation of $\mathbb{Z}_2^{2\kappa}$.

Proof: Let

$$C = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix},$$

then, with (2.22) and (2.21), $S_B \otimes I_\kappa$ is transformed into

$$\begin{bmatrix} A_1 & & B_1 & \\ & A_1 & & B_1 \\ C_1 & & D_1 & \\ & C_1 & & D_1 \end{bmatrix} \begin{bmatrix} 0 & I \\ 0 & I \\ I & 0 \\ I & 0 \end{bmatrix} = \begin{bmatrix} B_1 & A_1 \\ B_1 & A_1 \\ D_1 & C_1 \\ D_1 & C_1 \end{bmatrix}.$$

As C satisfies (2.23), one can verify that multiplication on the right by C^T yields $S_B \otimes I_\kappa$ again. By (2.20), it follows that $\tilde{b}$ is transformed into $C\tilde{b}$. □

Next, by local measurements, Alice and Bob extract information on $\tilde{b}$: they both measure σ_z on the last $m\kappa$ copies, yielding $u = C_{(4)}^T \tilde{b}$ according to (2.30), where we define

$$C = \begin{bmatrix} C_{(1)}^T \\ C_{(2)}^T \\ C_{(3)}^T \\ C_{(4)}^T \end{bmatrix}, \text{ where } \begin{cases} C_{(1)}, C_{(3)} \in \mathbb{Z}_2^{2\kappa\times(1-m)\kappa} \\ C_{(2)}, C_{(4)} \in \mathbb{Z}_2^{2\kappa\times m\kappa} \end{cases}, \text{ and } \bar{C} = \begin{bmatrix} C_{(1)}^T \\ C_{(3)}^T \end{bmatrix}.$$

Afterwards, the measured copies are discarded. The state of the remaining copies is given by the following theorem [25, 51]:

Theorem 3.2 *Defining $\mathcal{C} = \mathrm{col}\left(PC_{(4)}\right)$ and $\mathcal{N} = (P\mathcal{C})^{\perp}$, then the remaining copies after the protocol are an ensemble of states $|B_w\rangle$, each with new probability*

$$p'_w = \frac{p(\mathcal{C}+v)}{p(\mathcal{N}+v)},$$

where v satisfies $\bar{C}v = w$ and $C^T_{(4)}v = u$.

Proof: Firstly, the numerator equals the total probability of all elements $\tilde{b}$ that are mapped to w and correspond to the observed measurement results: $C^T_{(4)}\tilde{b} = u$. Indeed, because C^T is symplectic [from $C^TP(CPC^T) = (C^TPC)PC^T = PPC^T = C^T$ follows $CPC^T = P$] and full rank, $\mathcal{C}$ is the space of all t satisfying $\bar{C}t = 0$ and $C^T_{(4)}t = 0$.

Secondly, the denominator is a normalizing factor equal to the total probability of all elements that correspond to the observed measurement results. Similarly as for the numerator, $\mathcal{N}$ is the space of all t satisfying $C^T_{(4)}t = 0$. The condition on v selects the right cosets of $\mathcal{C}$ and $\mathcal{N}$. □

The choice of notation for the spaces $\mathcal{C}$ and $\mathcal{N}$ will become clear in section 3.3.2.

3.3.1 Recurrence

We illustrate this procedure with the simplest bipartite distillation protocol, that takes two copies of a Bell-diagonal qubit pair as an input and, if successful, outputs one Bell-diagonal pair with more entanglement. The recurrence protocol was originally presented in [13], and further improved in [27, 65], where it was also named the 'IBM protocol' or the 'Oxford protocol'.

We choose one qubit pair as source, and the other one as target. Both parties apply a CNOT on their qubits from source to target. Afterwards, both measure σ_z on the target qubit. The protocol has succeeded when the outcomes are the same ($C^T_{(4)}\tilde{b} = 0$), and failed otherwise. In the latter case, the remaining qubit pair is separable and can be discarded. By theorem 3.1, $\tilde{b}$ is transformed into $C\tilde{b}$, where C is given by (2.11). From theorem 3.2, we can calculate the probability of success

$$\begin{aligned}
p_{\mathrm{S}} &= p(\mathcal{N}) \\
&= p_{0000} + p_{0011} + p_{0100} + p_{0111} + p_{1000} + p_{1011} + p_{1100} + p_{1111} \\
&= p_{00}^2 + p_{01}^2 + p_{00}p_{10} + p_{01}p_{11} + p_{10}p_{00} + p_{11}p_{01} + p_{10}^2 + p_{11}^2 \\
&= (p_{00} + p_{10})^2 + (p_{01} + p_{11})^2,
\end{aligned}$$

and the resulting probabilities of the remaining state

$$p'_{00} = (p_{00}^2 + p_{10}^2)/p_{\mathrm{S}},$$

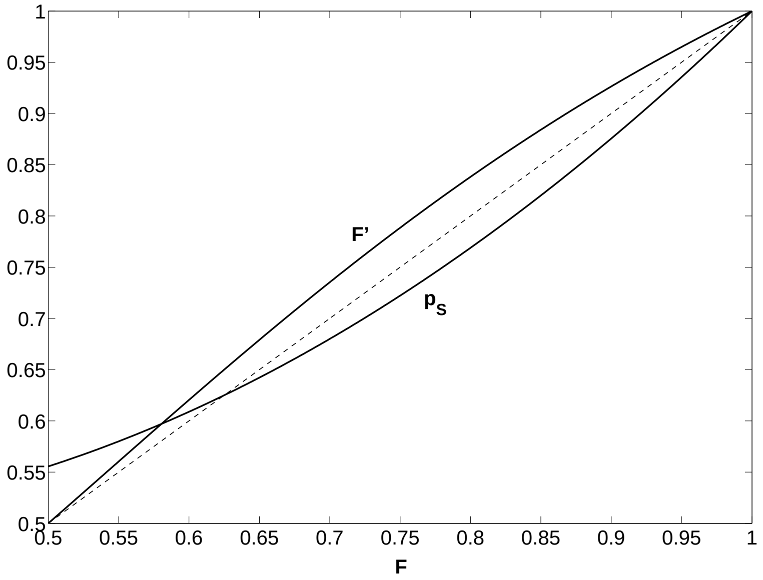


Figure 3.2: The success probability p_S and resulting fidelity F' as a function of the initial fidelity F for the recurrence protocol applied to two copies of a Werner state. Note that for $F > \frac{1}{2}$, we have $F' > F$. For $F \leq \frac{1}{2}$, the initial state is separable, resulting in $F' \leq \frac{1}{2}$ as well.

$$\begin{aligned}
p'_{01} &= (p_{01}^2 + p_{11}^2)/p_\mathrm{S}, \\
p'_{10} &= 2p_{00}p_{10}/p_\mathrm{S}, \\
p'_{11} &= 2p_{01}p_{11}/p_\mathrm{S}.
\end{aligned}$$

In the event of failure, it can be verified that the resulting probabilities satisfy $p''_{00} = p''_{01}$ and $p''_{10} = p''_{11}$. Thus, the remaining state has fidelity $\leq \frac{1}{2}$ and is therefore separable.

The success probability p_S and resulting fidelity F' are plotted in figure 3.2 for an entangled Werner state with fidelity F. In the original scheme of [13], one considers multiple iterations of recurrence on successful pairs, with a *random bilateral rotation* or '*twirl*' between each two iterations. This operation, similar to depolarization, next to setting all off-diagonal entries in the density matrix (with respect to the Bell state basis) to zero, leaves the fidelity invariant but equalizes the other diagonal entries (maintaining their sum). From figure 3.2, it is clear that this iterative procedure results in always increasing fidelity, in the limit going to the pure Bell state $|B_{00}\rangle$.

However, like depolarization, twirling in a sense throws away information, and therefore diminishes the available entanglement. Yet, doing nothing at all causes the procedure to converge to a separable state. In [27], the twirl is replaced by a fixed reordering of the diagonal entries, which causes no loss of information and significantly improves the $F \to F'$ update. In [65], further improvement is achieved by an adaptive reordering of the diagonal entries. Note that, by theorem 3.1, such reordering can always be done by applying the same one-qubit Clifford operation on both qubits.

When operations of the protocol are noisy, it is interesting to consider also assembling non-identical pairs in the protocol, but we do not go deeper into that issue. For a detailed overview of such investigations, we refer to [30].

Remark. Note that the overall procedure approaches zero yield in the limit for high output fidelity, as in each step on average only $p_S/2$ pairs are kept, and an infinite number of iterations is needed to deliver copies in a state that approaches a pure Bell state. ◇

3.3.2 Permutation-based versus code-based approach

We have introduced distillation protocols as starting from multiple copies of a Bell-diagonal state, performing local Clifford operations that, according to theorem 3.1, permute all possible $\tilde{b} \in \mathbb{Z}_2^{2\kappa}$, and local measurements. This we call the *permutation-based approach* [25]. However, the state of these copies can equivalently be interpreted as being initially pure Bell states $|B_{00}\rangle$ prepared by Alice and of which the second qubits are transmitted to Bob via a quantum channel. Leaving the specific physical implementation aside, this quantum channel is generally described by an evolution (2.36), possibly introducing errors. For one copy, a single-qubit Pauli error (on the second qubit only) transforms $|B_{00}\rangle$ as follows:

$$\begin{aligned}
(I \otimes X) |B_{00}\rangle &= |B_{01}\rangle , \\
(I \otimes Y) |B_{00}\rangle &= |B_{11}\rangle , \\
(I \otimes Z) |B_{00}\rangle &= |B_{10}\rangle ,
\end{aligned}$$

which is trivial from (2.1) and (2.27). This gives rise to a state that is diagonal in the Bell state basis.

Given a stabilizer code represented by $S\ (f)$ and b. Referring back to section 2.6, the qubits sent to Bob have undergone some error E_i. By performing the same stabilizer code measurements $M_j = i^{f_j}(-1)^{b_j}\tau_{S_j}$ on Alice's side and on Bob's side, and performing the error correction on Bob's qubits corresponding to the 'syndrome' $m = m_A + m_B + f$, one is able to correct all errors E_i that satisfy theorem 2.30. Indeed, following (2.29), (2.31) and (2.16), the state $|B_{00}\rangle^{\otimes\kappa} = |B_0\rangle$ is stabilized by

$$\tau_a \otimes \tau_a, \quad \forall a \in \mathbb{Z}_2^{2\kappa}.$$

Now $M_j \otimes M_j$ (anti)commutes with $I \otimes E_i$ if and only if M_j (anti)commutes with E_i. After the error correction step, both Alice and Bob apply a decoding

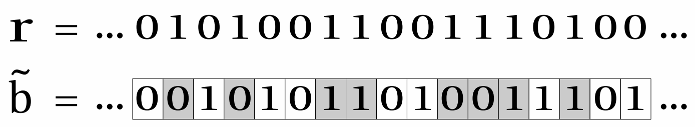

r = ... 0 1 0 1 0 0 1 1 0 0 1 1 1 0 1 0 0 ...
b̃ = ... 0 0 1 0 1 0 1 1 0 1 0 0 1 1 1 0 1 ...

Figure 3.3: Information on $\tilde{b}$ is extracted by determination of the inner product $r^T\tilde{b}$. This is the same as determining the parity (binary sum) of the bits of $\tilde{b}$ on positions i where $r_i = 1$ (shaded in this example).

operation for the stabilizer code. Note that, with theorem 3.1, encoding before transmission leaves the initial state $|B_0\rangle$ invariant and can therefore be omitted. We call this the *code-based approach* [2, 64].

We show that both approaches are equivalent [51]. The local Pauli measurements M_j for determining the syndrome can be realized as follows. Firstly, one applies a decoding operation for the stabilizer code, which transforms the stabilizer code basis states to the computational basis states. By proposition 2.29, the decoding operation is a Clifford operation. Interpreting this action in terms of the stabilizer [41], the stabilizer of the code is transformed to the stabilizer of the computational basis states, which is generated by σ_z on qubits $(1-m)\kappa, \dots, \kappa$. Therefore, measuring M_j is measuring σ_z after the decoding operation on qubits $(1-m)\kappa, \dots, \kappa$. The encoding operation, needed to transform computational basis states back to stabilizer code basis states, is canceled by the final decoding operation in the code-based approach.

Remark. One can verify that the spaces $\mathcal{C}$ and $\mathcal{N}$ defined in theorem 3.2 of the permutation-based approach are the same as the spaces $\mathcal{C}$ and $\mathcal{N}$ defined by the stabilizer code (cf. section 2.6) of the code-based approach [51]. ◇

Example. The stabilizer of the code corresponding to recurrence is generated by ZZ. ◇

3.3.3 Two kinds of parity checks

Information on the initial overall stabilizer state $\left|B_{\tilde{b}}\right\rangle$ is extracted under the form of an inner product $r^T\tilde{b}$, where r is an arbitrary nonzero 2κ-bit vector. This is illustrated in figure 3.3. We call this action a *parity check*, and its value $r^T\tilde{b}$ the *parity*. This can be done in two ways:

1) As explained above, local Clifford operations transform $\left|B_{\tilde{b}}\right\rangle$ into $\left|B_{C\tilde{b}}\right\rangle$. Measuring σ_z on both qubits of one of the pairs yields $r^T\tilde{b}$, where r^T is the corresponding row of the lower half of C. Afterwards, the state of the remaining pairs is $\left|B_{\bar{C}\tilde{b}}\right\rangle$, where $\bar{C}$ is C without this row and the

corresponding row in the upper half. As this procedure is equivalent to performing the same Pauli measurement on both sides (cf. the previous section), we call this a *bilateral Pauli measurement* (BPM).

2) Let $\tilde{b} = \begin{bmatrix} \tilde{b}_1 \\ \tilde{b}_2 \end{bmatrix}$ and $r = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}$. Appending an ebit (out of some pool of predistilled qubit pairs) in the state $|B_{00}\rangle$ yields the state $\left|B_{\tilde{b}_1 0 \tilde{b}_2 0}\right\rangle$. Applying on both sides the Clifford operation represented by

$$\left[\begin{array}{c|c|c|c} & & & 0 \\ I_\kappa & r_2 & 0 & \vdots \\ & & & 0 \\ \hline 0\cdots0 & 1 & 0\cdots0 & 0 \\ \hline & & & 0 \\ 0 & r_1 & I_\kappa & \vdots \\ & & & 0 \\ \hline r_1^T & 0 & r_2^T & 1 \end{array}\right]$$

leaves $\left|B_{\tilde{b}}\right\rangle$ unchanged but transforms the ebit into $|B_{0,r^T\tilde{b}}\rangle$. Measuring σ_z on both sides of the ebit then yields $r^T\tilde{b}$. We call this an *appended ebit measurement* (AEM).

It is clear that after an AEM, any other parity check yielding $s^T\tilde{b}$ can be done, as the AEM only affects the ebit. However, this does not hold for a BPM. After a BPM, the only information on $\left|B_{\tilde{b}}\right\rangle$ left for us to extract is the information we can extract from the resulting state $\left|B_{\bar{C}\tilde{b}}\right\rangle$. Clearly, we can perform parity checks yielding $t^T\bar{C}\tilde{b}$, $\forall t \in \mathbb{Z}_2^{2(\kappa-1)}$. This is equivalent to the possibility of determining $s^T\tilde{b}$ for all $s \in \mathbb{Z}_2^{2\kappa}$ that satisfy $s^T P r = 0$. Indeed, as C is symplectic, all such s are in the column space of $\left[\bar{C}^T \ r\right] \Rightarrow s = \bar{C}^T t + \alpha r$, for some $t \in \mathbb{Z}_2^{2(\kappa-1)}$ and $\alpha \in \mathbb{Z}_2$. Since $r^T\tilde{b}$ was already determined, we know $s^T\tilde{b} = t^T\bar{C}\tilde{b} + \alpha r^T\tilde{b}$ by determining $t^T\bar{C}\tilde{b}$ from $\left|B_{\bar{C}\tilde{b}}\right\rangle$. In general, every time we determine $r^T\tilde{b}$ of $\left|B_{\tilde{b}}\right\rangle$ by a BPM, afterwards we have only access to $s^T\tilde{b}$ where $s^T P r = 0$, whatever method we use. This should not come as a surprise, because when $s^T P r = 1$, the corresponding Pauli observables anticommute, so their values cannot both be determined.

In practice, after a BPM, we should continue working with the transformed state represented by $\bar{C}\tilde{b}$. But this requires knowledge of the entire matrix C, while the parity check is specified only by r. As explained in the previous paragraph, we can describe all future actions in terms of $\tilde{b}$: we only need to know which BPM have been done. This yields a much more transparent description of the protocol.

3.4 Asymptotic protocols

3.4.1 Asymptotic equipartition property

Let $(X_1, \dots, X_\kappa)$ be a sequence of independent identically-distributed discrete random variables, each having event set Ω with probability distribution

$$p : \Omega \to [0,1] : x \to p(x).$$

In the following, we assume that κ is a large number, in the limit approaching infinity, and that ϵ is a small number, in the limit approaching zero. The *typical set* $\mathcal{A}_\epsilon^{(\kappa)}$ with respect to $p(x)$ is defined as the set of sequences $(x_1, \dots, x_\kappa) \in \Omega^\kappa$ with the following property:

$$\left| \frac{-\log_2 p(x_1, \dots, x_\kappa)}{\kappa} - H(X) \right| \leq \epsilon, \tag{3.4}$$

where

$$H(X) = -E[\log_2 p(X)] = -\sum_{x \in \Omega} p(x) \log_2 p(x)$$

is the entropy of the random variable X.

Proposition 3.3 *$p(\mathcal{A}_\epsilon^{(\kappa)}) \geq 1 - \delta$, where $\delta = \mathcal{O}(\kappa^{-1}\epsilon^{-2})$.*

Proof: The random variable $-\sum_{i=1}^{\kappa} \log_2 p(X_i)/\kappa$ has mean $H(X)$ and variance $\mathrm{Var}[-\log_2 p(X)]/\kappa$. By Chebyshev's inequality [91], we then have

$$P(\left| \frac{-\log_2 p(X_1, \dots, X_\kappa)}{\kappa} - H(X) \right| \geq \epsilon) \leq \frac{\mathrm{Var}[-\log_2 p(X)]}{\kappa\epsilon^2}.$$

It follows that $p(\mathcal{A}_\epsilon^{(\kappa)}) \geq 1 - \delta$, where $\delta = \mathcal{O}(\kappa^{-1}\epsilon^{-2})$. □

This proposition states that, when considering a vast number of independent identically-distributed discrete random variables, the joint random variable will almost certainly have a *typical* probability. This property is referred to as the *asymptotic equipartition property*, which is often stated informally as: *almost all events are almost equally surprising.*

We can calculate the cardinality of $\mathcal{A}_\epsilon^{(\kappa)}$ dividing $p(\mathcal{A}_\epsilon^{(\kappa)})$ by $p(x_1, \dots, x_\kappa)$, which, by (3.4), is close to $2^{-\kappa H(X)}$:

Corollary 3.4 $|\mathcal{A}_\epsilon^{(\kappa)}| = 2^{\kappa[H(X) + \mathcal{O}(\epsilon)]}$.

3.4.2 Hashing and breeding

The hashing protocol takes κ copies of a Bell-diagonal state (3.3) as an input. As explained in chapter 2, we interpret the state of the copies as an unknown pure tensor product of κ Bell states, $|B_{\tilde{u}}\rangle$, where $\tilde{u} \in \mathbb{Z}_2^{2\kappa}$. The goal now is to determine $\tilde{u}$. According to proposition 3.3, $\tilde{u}$ is an element of $\mathcal{A}_\epsilon^{(\kappa)}$ with probability $1 - p_1$, where $p_1 = \mathcal{O}(\kappa^{-1}\epsilon^{-2})$, for $\kappa \to \infty$ and for $\epsilon \to 0$.

Then, the protocol consists of BPM unveiling $r^T\tilde{u}$, where the r defining each BPM is randomly chosen from $\mathbb{Z}_2^{2\kappa}$ with uniform probability. After each BPM, we rule out every $\tilde{b} \in \mathcal{A}_\epsilon^{(\kappa)}$ that is not compatible with the outcome: $r^T\tilde{b} \neq r^T\tilde{u} \Rightarrow r^T\Delta\tilde{b} = 1$, where $\Delta\tilde{b} = \tilde{b} + \tilde{u}$. We calculate an upper bound for the probability p_2 that any $\tilde{b} \neq \tilde{u}$ is not eliminated after $m\kappa$ BPM. Firstly, it is clear that, for fixed $\Delta\tilde{b} \neq 0$, the linear map $\mathbb{Z}_2^{2\kappa} \to \mathbb{Z}_2 :\ r \to r^T\Delta\tilde{b}$ splits $\mathbb{Z}_2^{2\kappa}$ in half. Since r is randomly chosen with uniform probability, the probability of $r^T\Delta\tilde{b} = 1$ equals $\frac{1}{2}$. After $m\kappa$ BPM, the probability of not being eliminated is $2^{-m\kappa}$. Consequently, the probability that any $\tilde{b} \in \mathcal{A}_\epsilon^{(\kappa)}$ different from $\tilde{u}$ is not eliminated after all BPM, equals $p_2 \leq |\mathcal{A}_\epsilon^{(\kappa)}|2^{-m\kappa} = 2^{\kappa[S(\rho)-m+\mathcal{O}(\epsilon)]}$ following corollary 3.4, where

$$S(\rho) = -p_{00}\log_2 p_{00} - p_{01}\log_2 p_{01} - p_{10}\log_2 p_{10} - p_{11}\log_2 p_{11}$$

is the Von Neumann entropy of ρ, which is also the entropy associated with the ensemble of Bell states $|B_b\rangle$.

The total failure probability of the protocol is the probability $p_1 + p_2$ that the initial assumption $\tilde{u} \in \mathcal{A}_\epsilon^{(\kappa)}$ is wrong or that not every $\tilde{b} \neq \tilde{u}$ is eliminated after all BPM. If we choose for instance $m = S(\rho) + \mathcal{O}(\sqrt{\epsilon})$, and $\epsilon = \mathcal{O}(\kappa^{-1/4})$, it follows that the failure probability $p_1 + p_2$ vanishes for $\kappa \to \infty$. After all BPM, we are left with $m\kappa$ measured separable pairs and $(1-m)\kappa$ pure ebits. The yield $1 - m$ of the protocol approaches $1 - S(\rho)$. We already plotted the hashing yield for rank two Bell diagonal states (D_2) and Werner states (H) as a function of the fidelity F in figure 3.1 on page 47.

Remark. To be precise, the r, s defining different BPM in the sequence should satisfy $r^TPs = 0$, as we showed in section 3.3.3. Consequently, subsequent r cannot be chosen from $\mathbb{Z}_2^{2\kappa}$ with a uniform probability. However, this does not corrupt the validity of the given calculation of the elimination probability. Indeed, after each BPM, we can equivalently define the following BPM on the resulting $\bar{C}\tilde{b}$ being uniformly chosen from $\mathbb{Z}_2^{2(\kappa-1)}$, and so on, also resulting in $\frac{1}{2}$ elimination probability for each BPM.

We have calculated an upper bound for p_2. One could wonder whether the elimination of all $\tilde{b} \neq \tilde{u}$ does not, in fact, proceed *faster* than predicted. Indeed, a BPM unveils the value of $r^T\tilde{u}$, i.e. *one* bit, and destroys one copy, leaving the remaining copies identified by $\bar{C}\tilde{u}$, which has *two* bits less than $\tilde{u}$. So we learn one bit, but at the same time another is lost and is of no further importance. Because of that, it is possible that different $\tilde{b}, \tilde{b}'$ are both compatible with the outcome, which means $r^T\tilde{b} = r^T\tilde{b}'$, and are mapped to the same $\bar{C}\tilde{b} = \bar{C}\tilde{b}'$. This could cause the number of candidates to drop faster than halving with each BPM. We will show later that this is not the case for hashing, but we will actually exploit this phenomenon to improve the hashing protocol. ◇

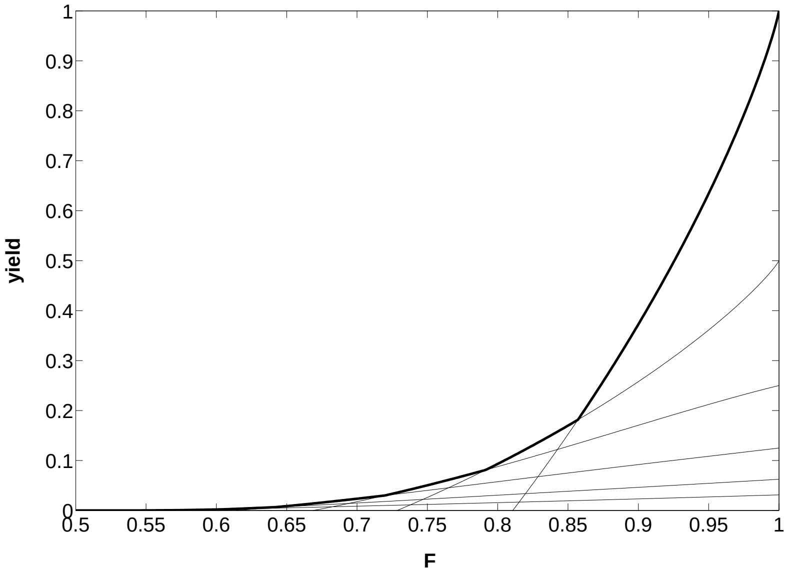


Figure 3.4: The yield of the combined iterative recurrence and hashing for Werner states as a function of the fidelity F. We applied $0, 1, 2, \ldots$ recurrence iterations (with optimal reordering in between [65]) before switching to hashing, which gives rise to the multiple lines. The overall yield is the maximum of these. We observe that more recurrence iterations are needed for low initial F.

The breeding protocol works in the same way as the hashing protocol, except for using AEM instead of BPM. Although this requires an initial pool of *predistilled* ebits, the net output of ebits is the same as for hashing. Similarly as for hashing, we find that we need to apply approximately $\kappa S(\rho)$ AEM, at the cost of as many ebits. After these, the initial κ pairs are transformed to pure ebits, since no measurements were performed on them. The yield is therefore equal to

$$\frac{\#(\text{output ebits}) - \#(\text{input ebits})}{\#(\text{input pairs})} = \frac{\kappa - \kappa S(\rho)}{\kappa} = 1 - S(\rho).$$

As one can see in figure 3.1, the hashing yield is zero for Werner states with fidelity $F < 0.8107$, although E_D is nonzero as soon as $F > \frac{1}{2}$. Since the hashing protocol performs so poorly for noisy states, a common solution is to have hashing preceded by a number of recurrence steps. Using the optimal reordering strategy of [65], and doing another recurrence iteration if this increases the yield, we arrive at the yield plotted in figure 3.4.

Remark. The hashing/breeding protocol can be carried out with the use of only *one-way* classical communication. Indeed, Alice and Bob need to agree on the random parity checks they apply, and their outcomes should be compared in order to know $\tilde{u}$ at the end of the protocol. But since intermediate outcomes do not alter future actions, it suffices that Alice solely decides which random parity checks (of which the number is fixed) will be applied and transmits this information together with the outcomes she observed to Bob. By applying the same operations as Alice, and comparing his results with Alice's, he can determine $\tilde{u}$. Bob then transforms the state $|B_{\tilde{u}}\rangle$ to $|B_0\rangle$ by applying the correct Pauli operation (cf. section 2.5.2). $\diamond$

3.5 Adaptive asymptotic protocols

The hashing/breeding protocol has a nice information-theoretical interpretation: for every measurement, we gain one bit of information, which is the maximal amount we can extract since the measurement outcomes either correlate or anticorrelate.[4] This information gain causes an equal decrease of one bit of the entropy of the state. When this entropy is reduced to zero, the state has become a pure tensor product of Bell states. In the following, we choose to start from breeding instead of hashing, for convenience. All derivations equally hold if we start from hashing.

In section 3.5.1, we show that breeding can be divided into successive stages of partial information extraction, yielding an equivalent protocol. In a first approach, at every stage we replace measurements on ebits by measurements on a finite number of copies, whenever correlating and anticorrelating outcomes are equiprobable. It can be verified that the entropy of the global state is then reduced by more than one bit. This is because whenever an observable is measured, the state is projected onto the eigenspace of the observable, thereby eliminating the entropy associated with observables not commuting with the one measured. This is the content of section 3.5.2. We will explain in sections 3.5.3 and 3.5.4 how our protocol is organized as to have as many replacements as possible. This is illustrated with the explicit calculation of the yield for Werner states, in section 3.5.5. Then, two variants of this first scheme are discussed, in sections 3.5.6 and 3.5.7. The results of this last section were not included in the original paper [53].

3.5.1 Partial breeding

We show how the breeding protocol can be divided into successive stages of partial information extraction. Partial information on $\tilde{u}$ is extracted by restricting

[4] In fact, there are *four* possible outcomes, i.e. two on each side, but only by *comparing* these, can we obtain information on the state. This is because we are restricted to *local* measurements, whereas the state itself is *non-local*.

to parity checks $r^T\tilde{u}$, where r is of the form

$$r = \begin{bmatrix} r' \otimes s_{(1)} \\ r' \otimes s_{(2)} \end{bmatrix}, \text{ where } s = \begin{bmatrix} s_{(1)} \\ s_{(2)} \end{bmatrix} \in \mathbb{Z}_2^{2\lambda}$$

is fixed (with finite λ) and random $r' \in \mathbb{Z}_2^{2\frac{\kappa}{\lambda}}$ take over the role of r. We will call this technique *partial breeding.* Note that it is completely specified by fixing s. Therefore we will refer to it by PB s. We illustrate how partial breeding works with an example. Let $s = 1100$, and divide $\tilde{u}$ into vectors of $2\lambda = 4$ bits (i.e. $\lambda = 2$ pairs). Every such 2λ-bit vector g is either an element of $0^{(s)}$, if $s^Tg = 0$, or of $1^{(s)}$, if $s^Tg = 1$. For this example, we have

$$\begin{aligned} 0^{(s)} &= \{0000, 0001, 0010, 0011, 1100, 1101, 1110, 1111\}, \\ 1^{(s)} &= \{0100, 0101, 0110, 0111, 1000, 1001, 1010, 1011\}. \end{aligned}$$

We have for instance

$$\begin{array}{cccccccc} \tilde{u} & = & \multicolumn{6}{l}{10\,11\,\underline{01}\,01\,00\,10\,11 \quad 10\,10\,\underline{10}\,11\,01\,01\,00} \\ \Rightarrow & g: & 1010 & 1110 & \underline{0110} & 0111 & 0001 & 1001 \quad 1100 \\ & \in & 1^{(s)} & 0^{(s)} & 1^{(s)} & 1^{(s)} & 0^{(s)} & 1^{(s)} \quad 0^{(s)} \quad . \end{array}$$

In the same way as for breeding, a typical set can be associated with the distribution of $0^{(s)}$ and $1^{(s)}$. This set has $\approx 2^{\frac{\kappa}{\lambda}h(0^{(s)},1^{(s)})}$ elements, where

$$h(p, 1-p) = -p\log_2 p - (1-p)\log_2(1-p) \tag{3.5}$$

is the *binary entropy function.* Therefore, we need $\approx \frac{\kappa}{\lambda}h(0^{(s)}, 1^{(s)})$ AEM to determine s^Tg for all 2λ-bit vectors g constituting $\tilde{u}$, with probability close to 1. For this example, we have

$$\begin{aligned} p_{0^{(s)}} &= p_{0000} + p_{0001} + \ldots + p_{1111}, \\ p_{1^{(s)}} &= p_{0100} + p_{0101} + \ldots + p_{1011}. \end{aligned}$$

We have considered partial information extraction on a sequence of independent identically-distributed random variables over the set $\{00, 01, 10, 11\}$. But the same idea can also be applied to the sets $0^{(s)}$ and $1^{(s)}$. Once we have carried out the previous PB step, we know for every 4-bit vector g (determining 2 pairs) whether it is in $0^{(s)}$ or in $1^{(s)}$. If we bring all $g \in 0^{(s)}$ together, again we have i.i.d. random variables over $0^{(s)}$, and again we could perform partial breeding, this time, for instance, PB $t_0 = 0011$. Combining this with, for instance, PB $t_1 = 1000$ for $1^{(s)}$, we get to know for every g in which of the following sets it is:

$$\begin{aligned} S_1 &= 0^{(s)} \cap 0^{(t_0)} &&= \{0000, 0011, 1100, 1111\}, \\ S_2 &= 0^{(s)} \cap 1^{(t_0)} &&= \{0001, 0010, 1101, 1110\}, \end{aligned}$$

$$\begin{aligned} S_3 &= 1^{(s)} \cap 0^{(t_1)} = \{0100, 0101, 0110, 0111\}, \\ S_4 &= 1^{(s)} \cap 1^{(t_1)} = \{1000, 1001, 1010, 1011\}. \end{aligned}$$

It can be verified that the total number of AEM needed in the first and second PB step of this example is equal to

$$-\frac{\kappa}{2}\left(p_{S_1}\log_2 p_{S_1} + p_{S_2}\log_2 p_{S_2} + p_{S_3}\log_2 p_{S_3} + p_{S_4}\log_2 p_{S_4}\right),$$

which is exactly the entropy that is associated with the partition into S_1, S_2, S_3, S_4 times the number of 4-bit vectors g in $\tilde{u}$. This is a consequence of the fact that

$$H(C) = (-p_A \log_2 p_A - p_B \log_2 p_B) + p_A H(A) + p_B H(B),$$

where $\{A, B\}$ is a partition of the set C [22]. So no matter how a certain situation is attained, the number of AEM (= the cost in ebits) always equals the total information gain. We can continue performing PB steps in this way until all sets considered are singletons. We then have determined $\tilde{u}$ completely, at the cost of $\kappa S(\rho)$ ebits.

Of course, there is no point in dividing the breeding protocol in successive stages of partial breeding. In [90], $0^{(s)}$ states are further purified by breeding, but the $1^{(s)}$ states are treated differently: on the first pair of every $1^{(s)}$ state, a BPM 10 is performed, yielding the value of parity check 10 of this pair. As the pair is measured, it is lost, but the measurement also provides information on the second pair. This one is in $\{10, 11\}$ if the parity was 0 and in $\{00, 01\}$ if the parity was 1. So in both cases, we end up with a rank two Bell-diagonal state, for which it has been proved that hashing/breeding is optimal [73]. The yield of this protocol was calculated in [90], and turns out to be larger than that of breeding. But the reason why this should necessarily be the case, remained unclear. We shed light on this issue in the next section.

3.5.2 Entropy reduction

The reason why the protocol of [90] outperforms the breeding protocol, lies in the difference between an AEM and a BPM. If a parity check is performed on a finite number λ of pairs, represented by an ensemble of vectors $g \in \mathbb{Z}_2^{2\lambda}$, the resulting state will have lower entropy by a BPM than by an AEM. Next to extracting information under the form of the parity, a BPM results in the mapping of different vectors to the same new vector, resulting in an extra entropy reduction.

To see this, we recall from section 3.3.3 that a BPM for the parity check $s^T g$ results in a new state (with one pair less) represented by $\bar{C}g$, where the last row of C is s^T. By the measurement, we learn $s^T g$, but we also lose $t^T g$, where t^T is the last row of the upper half of C. This loss causes all g with the same result $\bar{C}g$ and parity $s^T g$ to be mapped to the same vector $\bar{C}g$. Note that the parities should be equal as well, otherwise one of the two is ruled out. From the symplecticity of C, it follows that g and $g + Ps$ are mapped to the same $\bar{C}g$. Indeed, $\bar{C}Ps = 0$ and $s^T Ps = 0$. Consequently, the new state is represented by the ensemble of vectors $\bar{C}g$, with probabilities $p_g + p_{g+Ps}$. This addition of probabilities results in the extra entropy reduction. We see a similarity with *degenerate quantum codes* [39], where different errors yield the same syndrome and have the same effect on the encoded state. It is a feature of quantum codes with no classical equivalent.

Let us illustrate this with an example. We have two pairs represented by an ensemble of 4-bit vectors and we perform a BPM 1111. We are left with only one pair represented by an ensemble of 2-bit vectors. The probabilities are

$$\tfrac{p_{0000}+p_{1111}}{p_{0(s)}}, \quad \tfrac{p_{0011}+p_{1100}}{p_{0(s)}}, \quad \tfrac{p_{0101}+p_{1010}}{p_{0(s)}}, \quad \tfrac{p_{0110}+p_{1001}}{p_{0(s)}},$$

if the parity is 0, and

$$\tfrac{p_{0001}+p_{1110}}{p_{1(s)}}, \quad \tfrac{p_{0010}+p_{1101}}{p_{1(s)}}, \quad \tfrac{p_{0100}+p_{1011}}{p_{1(s)}}, \quad \tfrac{p_{0111}+p_{1000}}{p_{1(s)}},$$

if the parity is 1. Note that we do not identify these probabilities with the two-bit vectors $\bar{C}g$: all future actions are described entirely in terms of the original vectors g, as explained in section 3.3.3. If we had used an AEM, then we would still have two pairs, but represented only by 8 vectors instead of 16, with probabilities

$$\tfrac{p_{0000}}{p_{0(s)}}, \quad \tfrac{p_{1111}}{p_{0(s)}}, \quad \tfrac{p_{0011}}{p_{0(s)}}, \quad \tfrac{p_{1100}}{p_{0(s)}}, \quad \tfrac{p_{0101}}{p_{0(s)}}, \quad \tfrac{p_{1010}}{p_{0(s)}}, \quad \tfrac{p_{0110}}{p_{0(s)}}, \quad \tfrac{p_{1001}}{p_{0(s)}},$$

if the parity is 0, and

$$\tfrac{p_{0001}}{p_{1(s)}}, \quad \tfrac{p_{1110}}{p_{1(s)}}, \quad \tfrac{p_{0010}}{p_{1(s)}}, \quad \tfrac{p_{1101}}{p_{1(s)}}, \quad \tfrac{p_{0100}}{p_{1(s)}}, \quad \tfrac{p_{1011}}{p_{1(s)}}, \quad \tfrac{p_{0111}}{p_{1(s)}}, \quad \tfrac{p_{1000}}{p_{1(s)}},$$

if the parity is 1. The average difference in entropy is equal to

$$\begin{aligned} &[-p_{0000}\log_2 p_{0000} - p_{1111}\log_2 p_{1111} - \ldots - p_{0111}\log_2 p_{0111} - p_{1000}\log_2 p_{1000}] \\ &+ [(p_{0000}+p_{1111})\log_2(p_{0000}+p_{1111}) + \ldots + (p_{0111}+p_{1000})\log_2(p_{0111}+p_{1000})] \end{aligned}$$

and is always positive. Indeed, for all $x, y \geq 0$, we have:

$$[-x\log_2 x - y\log_2 y] + [(x+y)\log_2(x+y)] = (x+y)h(\frac{x}{x+y}, \frac{y}{x+y}), \quad (3.6)$$

where $h(p, 1-p)$ was defined in (3.5), and plotted in figure 3.5.

This plot shows that the entropy reduction, given by the RHS of (3.6), grows larger in line with the equiprobability of the colliding vectors g and $g + Ps$, averaged over all g. If one probability relative to the other becomes small, the entropy reduction vanishes. That is the reason why the hashing protocol, where parity checks are BPM instead of AEM, has the same yield as the breeding protocol:[5] again, we use the fact that almost all weight comes from vectors

[5] This gives an answer to the question posed in the remark on page 55.

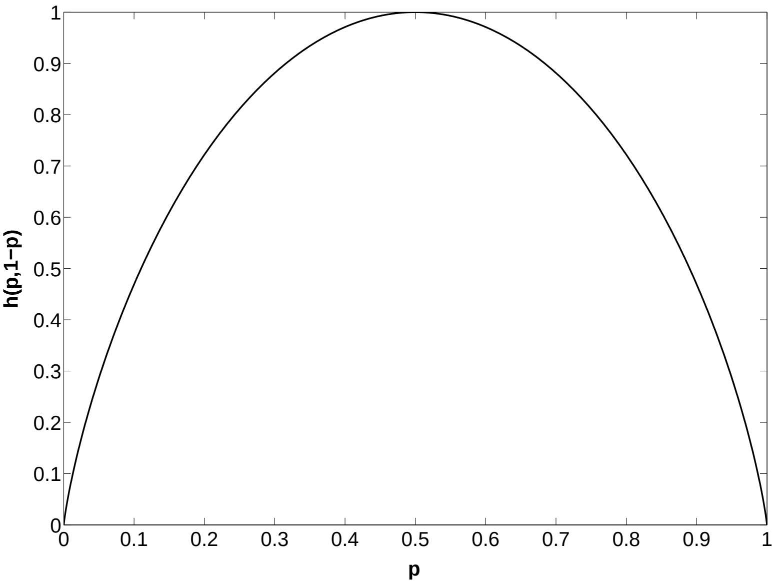


Figure 3.5: The binary entropy function $h(p, 1-p)$.

$\tilde{b} \in \mathcal{A}_\epsilon^{(\kappa)}$. Since the r are uniformly distributed, so are $\tilde{b} + Pr$. Therefore, the probabilities $\approx (p_{00}p_{01}p_{10}p_{11})^{\kappa/4}$ of $\tilde{b} + Pr$ are vanishing (as κ is large) compared to the probabilities $\approx 2^{-\kappa S(\rho)}$ of $\tilde{b}$. A variant of hashing presented in [79], where some of the BPM are on a finite number of copies resulting in a nonzero entropy reduction, performs slightly better than hashing.

It is clear that we should focus on BPM on small numbers of copies, because there lies the benefit of the entropy reduction. However, up until now, we have only spoken of the information gain, but we also have to take the cost into account. PB requires AEM, each at the cost of one ebit, whereas a BPM is at the cost of one of the copies. But as all unmeasured copies will ultimately be pure Bell states, this will not make any difference. By construction, every AEM in PB has equiprobable values, and therefore yields one bit of information. The same does hold for a BPM if r has infinite length and is random. Indeed, hashing is equivalent to breeding. But if we are to perform small, non-random parity checks, their values are not necessarily equiprobable and therefore yield less than one bit of information. If the parities are equiprobable, improvement is guaranteed. Note that the BPM 10 on the first pair of two $1^{(s)}$ pairs does have equiprobable values, which explains the improvement of [90] over breeding. So in some way, we should try to spot as many finite equiprobable parity checks as possible and carry them out by BPM.

3.5.3 Decoupling

Learning the parity of a number of qubit pairs by partial breeding or BPM causes statistical dependence of the pairs involved, which makes the continuation of the protocol very complicated. However, this statistical dependence can be undone in some cases, which we refer to as *decoupling*. The idea of decoupling is best explained by an example. Suppose by PB 1111, we learn for two copies of a Bell-diagonal qubit pair their state $\alpha^{(1111)}$. Where the states of the copies were independent before, this obviously no longer holds afterwards. But if we then perform PB 11 on all first pairs, yielding the state $\beta^{(11)}$ of the first of the two copies under consideration, we now have two independent pairs $\beta^{(11)}$ and $(\alpha+\beta)^{(11)}$. Indeed, we have learned the parities $1111 \to \alpha$ and $1010 \to \beta$, which is equivalent to knowing $1010 \to \beta$ and $0101 \to \alpha + \beta$, thus 11 for both pairs. So where the first PB coupled the ensembles of the two pairs, the second decoupled them again.

The same applies to PB $1111 \to \alpha$ followed by BPM $11 \to \beta$ on the first pair. This is equivalent to BPM $11 \to \beta$ on the first pair and PB $11 \to \alpha+\beta$ on the second pair. And it can be verified that BPM $1111 \to \alpha$ followed by BPM $11 \to \beta$ on the first pair is equivalent to BPM $11 \to \beta$ on the first pair and BPM $11 \to \alpha+\beta$ on the second pair. Indeed, this equivalence is a consequence of choosing another code basis when using the code-based interpretation. The same idea was used in the adaptive stabilizer code formalism of [2].

However, no decoupling rule holds for BPM followed by PB. Once we have carried out a BPM on a number of qubit pairs, we have statistical dependence not only by the knowledge of the overall parity, but also by the mapping together of vectors as explained in section 3.5.2. To highlight this dependence, we will denote the resulting state of BPM s with parity α by $[\alpha^{(s)}]$. Although the knowledge on the parities decouples by PB, the mapping does not. As an example, let BPM 1111 followed by PB 1010 on two particular pairs have parities 0 and 1 respectively. The resulting state of the pairs is $[1^{(11)}1^{(11)}]$ and has probabilities:

$$\tfrac{p_{01}^2+p_{10}^2}{(p_{01}+p_{10})^2} \quad \text{and} \quad \tfrac{2p_{01}p_{10}}{(p_{01}+p_{10})^2}.$$

Therefore, once a BPM is carried out on a number of qubit pairs, we have to take it into account until it is later decoupled by a BPM on some of the qubit pairs.

We summarize all scenarios. Let

$$s = \begin{bmatrix} s_{(1)} \\ s_{(2)} \\ s_{(3)} \\ s_{(4)} \end{bmatrix}, \; t_1 = \begin{bmatrix} s_{(1)} \\ s_{(3)} \end{bmatrix}, \; t_2 = \begin{bmatrix} s_{(2)} \\ s_{(4)} \end{bmatrix},$$

and we perform the parity check t_1 on the corresponding part of state $\alpha^{(s)}$. Then the following update rules apply:

	state	parity for t_1	resulting state
$\boxed{\text{PB}}$	$0^{(s)}$	$\to 0:$	$0^{(t_1)}0^{(t_2)}$
		$\to 1:$	$1^{(t_1)}1^{(t_2)}$
	$1^{(s)}$	$\to 0:$	$0^{(t_1)}1^{(t_2)}$
		$\to 1:$	$1^{(t_1)}0^{(t_2)}$
$\boxed{\text{BPM}}$	$0^{(s)}$	$\to 0:$	$[0^{(t_1)}]0^{(t_2)}$
		$\to 1:$	$[1^{(t_1)}]1^{(t_2)}$
	$1^{(s)}$	$\to 0:$	$[0^{(t_1)}]1^{(t_2)}$
		$\to 1:$	$[1^{(t_1)}]0^{(t_2)}$

(3.7)

If the considered state was connected to others by previous BPM, like in $[x\ \alpha^{(s)}\ y]$, the state transforms as follows:

	state	parity for t_1	resulting state
$\boxed{\text{PB}}$	$[x\ 0^{(s)}\ y]$	$\to 0:$	$[x\ 0^{(t_1)}0^{(t_2)}\ y]$
		$\to 1:$	$[x\ 1^{(t_1)}1^{(t_2)}\ y]$
	$[x\ 1^{(s)}\ y]$	$\to 0:$	$[x\ 0^{(t_1)}1^{(t_2)}\ y]$
		$\to 1:$	$[x\ 1^{(t_1)}0^{(t_2)}\ y]$
$\boxed{\text{BPM}}$	$[x\ 0^{(s)}\ y]$	$\to 0:$	$[0^{(t_1)}][x\ 0^{(t_2)}\ y]$
		$\to 1:$	$[1^{(t_1)}][x\ 1^{(t_2)}\ y]$
	$[x\ 1^{(s)}\ y]$	$\to 0:$	$[0^{(t_1)}][x\ 1^{(t_2)}\ y]$
		$\to 1:$	$[1^{(t_1)}][x\ 0^{(t_2)}\ y]$

(3.8)

Note that decoupling is nothing more than linearity of parity checks. Whenever we have performed a number of parity checks, these generate a space of parity checks. Any generating set of this space is equivalent to the original set of parity checks. For example, $\{0101, 1010\}$ is equivalent to $\{1010, 1111\}$. We will use decoupling parity checks because otherwise, the protocol becomes very complicated and unclear.

3.5.4 Forcing equiprobable parity checks

In section 3.3.3, we showed that once we have performed a BPM, we have to make sure that all following parity checks commute with it. There is a way in which this is automatically achieved. All vectors of the form $11 \otimes x$ commute (we could also have taken 01 or 10). Indeed, for all $x, y \in \mathbb{Z}_2^\kappa$, it holds:

$$\begin{aligned}&\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes x\right)^T P \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes y\right) \\ &= \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes x\right)^T \left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \otimes I_\kappa\right) \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes y\right) \;=\; 0 \otimes x^T y \;=\; 0.\end{aligned}$$

Therefore, if we stick to parity checks of this form, we do not have to care about commutability anymore. In this way, for every qubit pair we eventually find out whether it is $0^{(11)}$ or $1^{(11)}$. For now, let us assume that we go up to this point but not further: we want to find an optimal way of reaching the point where every pair is determined as $0^{(11)}$ or as $1^{(11)}$. In the following, we will denote the all-zeros and all-ones λ-bit vectors by 0_λ and 1_λ respectively, and $1^{(11\otimes 1_\lambda)}$ simply by $1^{(\lambda)}$.

Whenever we spot equiprobable parity checks, we should perform it by BPM. We will now explain how we can force such parity checks. Suppose we have 2λ qubit pairs, determined as $1^{(2\lambda)}$ by a previous parity check. Then the parity check $11 \otimes 1_\lambda 0_\lambda$ has equiprobable values. Indeed, it holds that

$$1^{(2\lambda)} = \begin{cases} 0^{(\lambda)} 1^{(\lambda)} \\ 1^{(\lambda)} 0^{(\lambda)} \end{cases}.$$

Clearly, both possibilities have the same initial probability $p_{0^{(\lambda)}} p_{1^{(\lambda)}}$, which is $\frac{1}{2}$ after normalization. Therefore, performing the parity check $11 \otimes 1_\lambda$ on the left half yields the parities of both halves and this information equals one bit. By performing a BPM, we have the extra entropy reduction. Furthermore, this BPM decouples the two halves of the state.

However, if the 2λ pairs are

$$0^{(2\lambda)} = \begin{cases} 0^{(\lambda)} 0^{(\lambda)} \\ 1^{(\lambda)} 1^{(\lambda)} \end{cases},$$

we do not have equiprobable possibilities. With a little trick, we are still able to force an equiprobable parity check. Two states of this kind can be written as

$$0^{(2\lambda)}0^{(2\lambda)} = \left\{ \begin{array}{c} 0^{(\lambda)}0^{(\lambda)}\ 1^{(\lambda)}1^{(\lambda)} \\ 1^{(\lambda)}1^{(\lambda)}\ 0^{(\lambda)}0^{(\lambda)} \\ \hline 0^{(\lambda)}0^{(\lambda)}\ 0^{(\lambda)}0^{(\lambda)} \\ 1^{(\lambda)}1^{(\lambda)}\ 1^{(\lambda)}1^{(\lambda)} \end{array} \right. .$$

By performing an extra PB $11 \otimes 0_\lambda 1_{2\lambda} 0_\lambda$, we can distinguish the first two possibilities from the last two (as indicated by the line). If its value is 1, again we have two equiprobable possibilities $0^{(\lambda)}0^{(\lambda)}1^{(\lambda)}1^{(\lambda)}$ and $1^{(\lambda)}1^{(\lambda)}0^{(\lambda)}0^{(\lambda)}$, that are separated by a BPM $11 \otimes 1_\lambda$ on one of the four $\alpha^{(\lambda)}$ constituting the state. If its value is 0, the possibilities are not equiprobable, but again we can bring two of these results together, with possibilities

$$\left\{ \begin{array}{c} 0^{(\lambda)}0^{(\lambda)}0^{(\lambda)}0^{(\lambda)}\ 1^{(\lambda)}1^{(\lambda)}1^{(\lambda)}1^{(\lambda)} \\ 1^{(\lambda)}1^{(\lambda)}1^{(\lambda)}1^{(\lambda)}\ 0^{(\lambda)}0^{(\lambda)}0^{(\lambda)}0^{(\lambda)} \\ \hline 0^{(\lambda)}0^{(\lambda)}0^{(\lambda)}0^{(\lambda)}\ 0^{(\lambda)}0^{(\lambda)}0^{(\lambda)}0^{(\lambda)} \\ 1^{(\lambda)}1^{(\lambda)}1^{(\lambda)}1^{(\lambda)}\ 1^{(\lambda)}1^{(\lambda)}1^{(\lambda)}1^{(\lambda)} \end{array} \right. ,$$

and performing PB $11 \otimes 0_{3\lambda} 1_{2\lambda} 0_{3\lambda}$ separating the possibilities as indicated by the line, and so forth. Clearly, this trick can be repeated endlessly.

We calculate the average fraction $\eta(0^{(2\lambda)})$ of $0^{(2\lambda)}$ on half of which a BPM is performed, resulting in $[0^{(\lambda)}]0^{(\lambda)}$ or in $[1^{(\lambda)}]1^{(\lambda)}$ with equal probability [note that $\eta(1^{(2\lambda)}) = 1$]. The procedure explained in the previous paragraph is recursive: at each step, we combine two random variables with two possible values x and y ($p_x + p_y = 1$). The variables of the next step are xx and yy, and so on. Therefore, it is possible to calculate $\eta(0^{(2\lambda)})$ in a recursive way. Let ξ be the probability to reach the situation under consideration and k the total number of $0^{(2\lambda)}$ involved in the present step. Initially, we have

$$\begin{aligned} \xi &= 1, \\ p_x &= \frac{p^2_{0^{(\lambda)}}}{p^2_{0^{(\lambda)}} + p^2_{1^{(\lambda)}}}, \\ p_y &= \frac{p^2_{1^{(\lambda)}}}{p^2_{0^{(\lambda)}} + p^2_{1^{(\lambda)}}}, \\ k &= 2. \end{aligned}$$

From the procedure explained in the previous paragraph, we have the following recursion relation:

$$\begin{aligned} \xi &\leftarrow \xi(p_x^2 + p_y^2), \\ p_x &\leftarrow \frac{p_x^2}{p_x^2 + p_y^2}, \\ p_y &\leftarrow \frac{p_y^2}{p_x^2 + p_y^2}, \end{aligned}$$

$$k \quad \leftarrow \quad 2k.$$

At each step, we have a probability $2p_x p_y$ that half of one of the $0^{(2\lambda)}$ involved is determined by BPM. So each step yields another fraction $2\xi p_x p_y/k$ of $0^{(2\lambda)}$ on half of which a BPM is performed. It can be verified that the total sum of these fractions over all steps is equal to

$$\begin{aligned} \eta(0^{(2\lambda)}) &= \sum_{i=0}^{\infty} \frac{(vw)^{2^i}}{2^i \prod_{j=0}^{i}(v^{2^j}+w^{2^j})}, \\ \text{where} \quad v = \frac{p^2_{0(\lambda)}}{p^2_{0(\lambda)}+p^2_{1(\lambda)}} &= \frac{p^2_{0(\lambda)}}{p_{0(2\lambda)}} \quad \text{and} \quad w = \frac{p^2_{1(\lambda)}}{p^2_{0(\lambda)}+p^2_{1(\lambda)}} = \frac{p^2_{1(\lambda)}}{p_{0(2\lambda)}}. \end{aligned} \tag{3.9}$$

In practice, it suffices to truncate the procedure after a few steps, since the terms in the summation of (3.9) decrease exponentially fast.

3.5.5 Numerical calculation of the yield

The protocol starts with PB $11 \otimes 1_{2^q}$. The next step is an iteration of the procedure explained in the previous section, for $\lambda = 2^{q-1}, 2^{q-2}, \ldots, 1$, where we use the update rules (3.7) and (3.8). For now, we will treat all $0^{(2\lambda)}$ in the same way, i.e. we do not favor particular states being parity checked by BPM. As a consequence, every $0^{(2\lambda)}$ has the same probability $\eta(0^{(2\lambda)})$ of undergoing a BPM $11 \otimes 1_\lambda 0_\lambda$. We find that, from one step to the next, the states transform as follows:

state		transforms to	with probability
$0^{(2\lambda)}$	$\rightarrow$	$[0^{(\lambda)}]0^{(\lambda)}$	$\frac{\eta(0^{(2\lambda)})}{2}$
	$\rightarrow$	$[1^{(\lambda)}]1^{(\lambda)}$	$\frac{\eta(0^{(2\lambda)})}{2}$
	$\rightarrow$	$0^{(\lambda)}0^{(\lambda)}$	$\frac{p^2_{0(\lambda)}}{p_{0(2\lambda)}} - \frac{\eta(0^{(2\lambda)})}{2}$
	$\rightarrow$	$1^{(\lambda)}1^{(\lambda)}$	$\frac{p^2_{1(\lambda)}}{p_{0(2\lambda)}} - \frac{\eta(0^{(2\lambda)})}{2}$
$1^{(2\lambda)}$	$\rightarrow$	$[0^{(\lambda)}]1^{(\lambda)}$	$\frac{1}{2}$
	$\rightarrow$	$[1^{(\lambda)}]0^{(\lambda)}$	$\frac{1}{2}$

(3.10)

With these rules, we are able to calculate the frequencies (i.e. the expected number of occurrences per 2^q initial qubit pairs) of all possibilities from one step to the next. After the last step, we are left only with $0^{(1)}$ and $1^{(1)}$ pairs, in various combinations of BPM (denoted by square brackets). Within square brackets, permutations of pairs yield equivalent states. Therefore, we do not have to calculate the frequencies of all possibilities, but only up to a permutation of the pairs: between square brackets, only the number $n_0^{(\lambda)}$ of $0^{(\lambda)}$ and $n_1^{(\lambda)}$ of $1^{(\lambda)}$ matter. We denote this by $[n_0^{(\lambda)}, n_1^{(\lambda)}]$ and we abbreviate $[n_0^{(1)}, n_1^{(1)}]$

to $[n_0, n_1]$. The possibilities in the end are then:

$$\begin{array}{l} 0^{(1)}, 1^{(1)}, \\ [1,0], [0,1], \\ [2,0], [1,1], [0,2], \\ \quad\vdots \\ [2^{q-1},0], [2^{q-1}-1,1], \ldots, [0,2^{q-1}], \end{array} \tag{3.11}$$

with frequencies $f(0^{(1)}), f(1^{(1)}), f([1,0]), \ldots, f([0,2^{q-1}])$. Note that these must satisfy

$$\sum_A n_0(A) f(A) = 2^q p_{0^{(1)}} \quad \text{and} \quad \sum_A n_1(A) f(A) = 2^q p_{1^{(1)}}, \tag{3.12}$$

where we define $n_0(0^{(1)}) = 1,\ n_1(0^{(1)}) = 0$ and $n_0(1^{(1)}) = 0,\ n_1(1^{(1)}) = 1$. By partial breeding alone, $\kappa h(p_{0^{(1)}}, p_{1^{(1)}})$ ebits would have been sacrificed. Now, for every BPM, we have one ebit less that has been measured. Therefore, the total cost of ebits per qubit pair up to this point equals

$$h(p_{0^{(1)}}, p_{1^{(1)}}) - \frac{1}{2^q} \sum_{[n_0,n_1]} f([n_0,n_1]). \tag{3.13}$$

But the protocol is not finished yet: as mentioned earlier we have purified the state up to the point were the parity of 11 for each pair is determined as 0 or as 1. Now we continue. Breeding is optimal for the pairs that have never been involved in some BPM, as they are independent rank two Bell diagonal states [73]. We show that breeding is optimal for all remaining pairs. Although equiprobable parity checks can still be found, they will no longer result in an entropy reduction if carried out by a BPM. Indeed, all further parity checks a must be of the form $01 \otimes a'$ or of the form $10 \otimes a'$, because for every pair we already know the parity of 11. Therefore, Pa too is of this form. Since every pair is either $0^{(1)} = \{00, 11\}$ or $1^{(1)} = \{01, 10\}$, the mapping of vectors vanishes: one of the two vectors mapped to the same new vector has already been ruled out by the parity checks, because $0^{(1)} + 01 = 0^{(1)} + 10 = 1^{(1)}$. Deprived of the benefit of entropy reduction by BPM, the best thing left is to gain one bit of information for every measurement. The number of ebits needed per qubit pair equals the entropy per pair

$$\frac{1}{2^q} \sum_A f(A) H(A) \tag{3.14}$$

left in the overall state. It can be verified that

$$\begin{aligned} H(0^{(1)}) &= h(q_{00}, q_{11}), \\ H(1^{(1)}) &= h(q_{01}, q_{10}), \end{aligned} \tag{3.15}$$

$$H([n_0,n_1]) \quad = \quad -\frac{1}{2}\sum_{i=0}^{n_0}\sum_{j=0}^{n_1}\binom{n_0}{i}\binom{n_1}{j}Q_{ij}\log_2 Q_{ij},$$

where $q_{00}=\frac{p_{00}}{p_{00}+p_{11}},\quad q_{11}=\frac{p_{11}}{p_{00}+p_{11}},\quad q_{01}=\frac{p_{01}}{p_{01}+p_{10}},\quad q_{10}=\frac{p_{10}}{p_{01}+p_{10}},$

$Q_{ij}=q_{00}^{i}q_{11}^{n_0-i}q_{01}^{j}q_{10}^{n_1-j}+q_{00}^{n_0-i}q_{11}^{i}q_{01}^{n_1-j}q_{10}^{j}.$

Now all unmeasured qubit pairs are pure ebits. The fraction of unmeasured pairs equals

$$1-\frac{1}{2^q}\sum_{[n_0,n_1]}f([n_0,n_1]). \qquad (3.16)$$

If we subtract the total fraction of measured ebits, which is the sum of (3.13) and (3.14), from this value (3.16), we get the yield of the protocol:

$$1-h(p_{0^{(1)}},p_{1^{(1)}})-\frac{1}{2^q}\sum_{A}f(A)H(A). \qquad (3.17)$$

We have numerically calculated (3.17) for Werner states as a function of the fidelity, for $q=1,\dots,6$. This is plotted in figure 3.6. We have truncated the procedure for forcing equiprobable parity checks (see the end of section 3.5.4) after 10 steps. We see that with increasing q, the yields of the protocols increase but converge. This is because incrementing q results only in more BPM on large numbers of pairs, and the entropy reduction of a BPM decreases (and eventually vanishes) with an increasing number of involved pairs. Indeed, since the improvement with respect to hashing only comes from entropy reduction, we calculate the total entropy reduction

$$\frac{1}{2^q}\sum_{A}f(A)R(A),$$

where $R(A)$ is the entropy reduction of A, by subtracting the hashing yield

$$1-S(\rho)=1-h(p_{0^{(1)}},p_{1^{(1)}})-p_{0^{(1)}}h(q_{00},q_{11})-p_{1^{(1)}}h(q_{01},q_{10})$$

from the yield of the adaptive protocol (3.17). Making use of (3.12), if follows that

$$R(A)=n_0(A)h(q_{00},q_{11})+n_1(A)h(q_{01},q_{10})-H(A). \qquad (3.18)$$

Clearly, $R(0^{(1)})=R(1^{(1)})=0$. For Werner states, we have $q_{01}=q_{10}=\frac{1}{2}$, and using (3.15), one can simplify (3.18) to

$$R([n_0,n_1])=n_0h(q_{00},q_{11})+\frac{1}{2}\sum_{i=0}^{n_0}\binom{n_0}{i}Q_i\log_2 Q_i,$$

independent of n_1, where $q_{00}=\frac{3F}{2F+1}$, $q_{11}=\frac{1-F}{2F+1}$ and $Q_i=q_{00}^{i}q_{11}^{n_0-i}+q_{00}^{n_0-i}q_{11}^{i}$. We have plotted $R([n_0,n_1])$ for Werner states in figure 3.7, as a function of F and for increasing n_0.

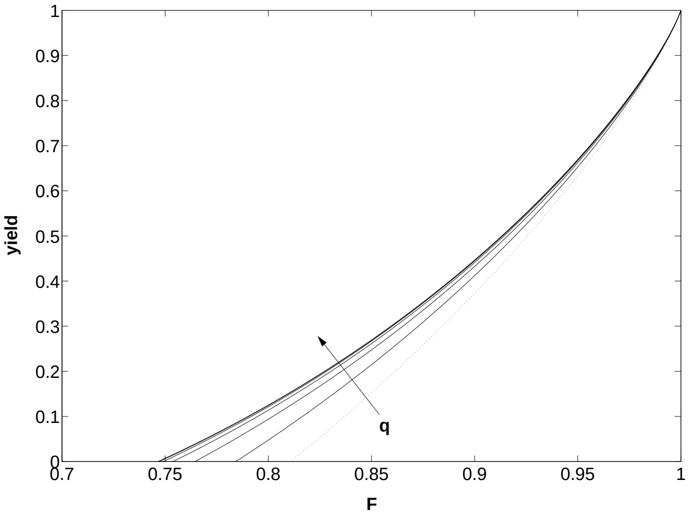


Figure 3.6: The yields of the proposed protocol (solid lines), for $q = 1, 2, 3, 4, 5, 6$ (in the order as indicated by the arrow), compared to the yield of hashing/breeding (dotted line), for Werner states as a function of F. The yield increases with increasing q and converges for large q (note that the yields for $q = 5$ and $q = 6$ almost coincide).

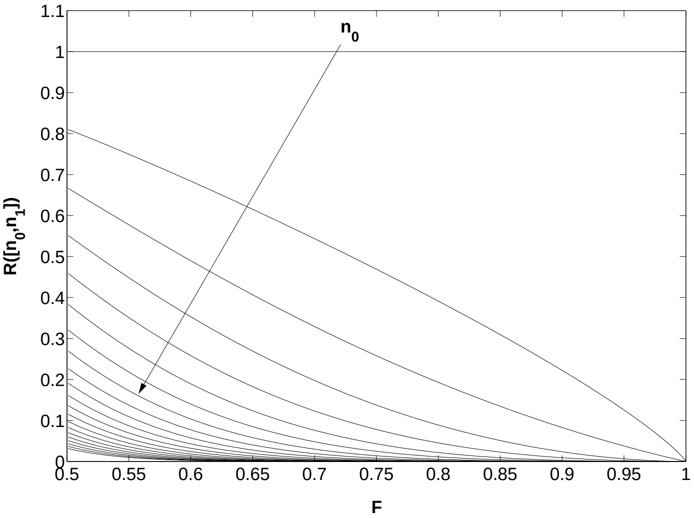


Figure 3.7: The entropy reduction $R([n_0, n_1])$ for Werner states as a function of the fidelity F, for $n_0 = 0, \dots, 20$. We observe that, for large n_0, the entropy reduction vanishes.

3.5.6 Favoring BPM on a small number of pairs

Since the entropy reduction $R[n_0, n_1]$ decreases with increasing n_0, we should try to increase the number of possibilities $[n_0, n_1]$ with small n_0. In the first version of our protocol, we did not make use of this: all $0^{(2\lambda)}$ were treated equally. So there is still room for improvement. A first ad hoc strategy is the following.

At each step, we have $0^{(2\lambda)}$ and $1^{(2\lambda)}$, distributed over all possibilities. We carry out BPM $11 \otimes 1_\lambda 0_\lambda$ on each $1^{(2\lambda)}$, so there the situation remains the same. For the $0^{(2\lambda)}$, first of all we take the ones that are linked by BPM (i.e. in square brackets) to a small number of pairs: this tends to result in more $[n_0^{(\lambda)}, n_1^{(\lambda)}]$ with small $n_0^{(\lambda)}$. Every $0^{(2\lambda)}$ is part of some possibility A, where $n_0^{(2\lambda)}$ is nonzero. We now order all possibilities $[n_0^{(2\lambda)}, n_1^{(2\lambda)}]$ according to increasing $n_0^{(2\lambda)} + n_1^{(2\lambda)}$ and on a second level according to increasing $n_0^{(2\lambda)}$. So for example $[5,3] < [6,2] < [4,5]$. We favor small $n_0^{(2\lambda)}$ on a second level because all $1^{(2\lambda)}$ will certainly be reduced, on average resulting in smaller n_0 in the end. We also define that all $[n_0^{(2\lambda)}, n_1^{(2\lambda)}] < 0^{(2\lambda)}$. Probably better orderings can be found, but this would presumably complicate things further without much benefit. We define

$$L(A) = \frac{\sum\limits_{B<A} n_0^{(2\lambda)}(B) f(B)}{p_{0^{(2\lambda)}} 2^{q-1}/\lambda} \quad \text{and} \quad U(A) = \frac{\sum\limits_{B\leq A} n_0^{(2\lambda)}(B) f(B)}{p_{0^{(2\lambda)}} 2^{q-1}/\lambda}.$$

$L(A)$ and $U(A)$ are the fractions of all $0^{(2\lambda)}$ that are part of some $B < A$ and $\leq A$ respectively. Note that $L([1,0]) = 0$ and $U(0^{(2m)}) = 1$. We combine the $0^{(2\lambda)}$ for the procedure explained in section 3.5.4 as follows: firstly, we partition all $0^{(2\lambda)}$ (in total $p_{0^{(2\lambda)}}\kappa/(2\lambda)$ elements) in two equally large sets: every $0^{(2\lambda)}$ of the first set is part of some $A \leq B$, where all $0^{(2\lambda)}$ that are part of B are in the second set. Now every $0^{(2\lambda)}$ of the first set is combined with one of the second set and PB $11 \otimes 0_\lambda 1_{2\lambda} 0_\lambda$ is performed. Whenever the parity is 1 (the probability of which is calculated in the same way as in section 3.5.4), a BPM $11 \otimes 1_\lambda 0_\lambda$ is performed on the first $0^{(2\lambda)}$. All $0^{(2\lambda)}0^{(2\lambda)}$ with parity 0 are again divided in two halves, according to the ordering of every first $0^{(2\lambda)}$. By continuing in this way, the fraction $\eta(0^{(2\lambda)}|A)$ of $0^{(2\lambda)}$, part of some A, on which a BPM $11 \otimes 1_\lambda 0_\lambda$ is performed, can be calculated, and equals

$$\eta(0^{(2\lambda)}|A) = \sum_{i=1}^{u(A)-1} z_i + \sum_{i=u(A)}^{l(A)} \frac{2^{-i} - L(A)}{U(A) - L(A)} z_i \tag{3.19}$$

$$\begin{array}{rrlrrl} \text{where} & l(A) & = & \lfloor -\log_2 L(A) \rfloor, & v & = & \frac{p^2_{0^{(\lambda)}}}{p_{0^{(2\lambda)}}}, \\ & u(A) & = & \lceil -\log_2 U(A) \rceil, & w & = & \frac{p^2_{1^{(\lambda)}}}{p_{0^{(2\lambda)}}}, \\ & z_i & = & \frac{2(vw)^{2^{(i-1)}}}{\prod\limits_{j=0}^{i-1}(v^{2^j}+w^{2^j})}. & & & \end{array}$$

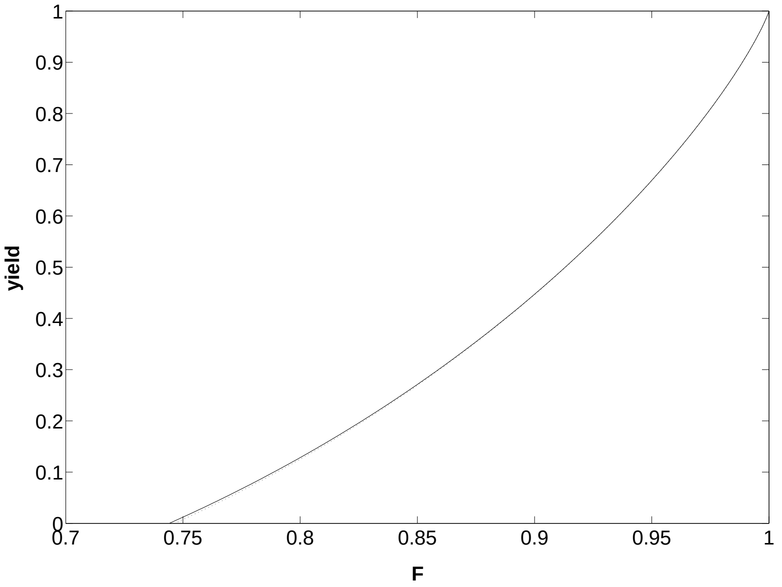


Figure 3.8: The yield of the protocol for Werner states with fidelity F, in the modified version of section 3.5.6 (solid line), compared to the yield of the original (dotted line), with $q = 6$.

As in (3.9), the terms in the second summation in (3.19) decrease exponentially fast. Therefore, when $l(A)$ is large, the procedure may be truncated after a number of steps. In the update rules (3.10), $\eta(0^{(2\lambda)})$ must be replaced by $\eta(0^{(2\lambda)}|A)$. Note that we have different update rules for different possibilities A. With this, we end up with the same possibilities (3.11) but with different frequencies $f(0^{(1)}), f(1^{(1)}), f([1,0]), \ldots, f([0, 2^{q-1}])$. To calculate the yield, we still use (3.15) and (3.17). We have plotted this in figure 3.8 for Werner states as a function of F, for $q = 6$. The difference with the first version is negligible (and even smaller for $q < 6$). Furthermore, this strategy is much more complicated than the original. Therefore, we have not further investigated this path.

3.5.7 Greedy variant

For Werner states, the yield of our best protocol is zero when $F < 0.7424$. This is better than the original breeding protocol (0.8107), but in order to distill states with lower fidelity, we must perform a number of recurrence iterations firstly, until it is no longer advantageous to do another iteration. The result of that is plotted in figure 3.9. One might think that the benefit of recurrence is due to some other phenomenon than information extraction and entropy reduction, because whenever the protocol fails, the remaining separable qubit pair is discarded. Although discarding is not present in any of the asymptotic protocols, one can interpret the discarding of the remaining pair in recurrence as a BPM with equiprobable values. Indeed, recurrence consists of a BPM 1100. If the parity is 1, the parity check 1000 is equiprobable, and should therefore be carried out by a BPM. But then *two* BPM have been carried out on the two qubit pairs involved, so there are no remaining unmeasured pairs.

Therefore, the recurrence iterations before our protocol only improve it by the fact that also *non-equiprobable* parity checks are carried out by BPM. The non-maximal nature of the information gain is more than compensated for by the entropy reduction for low-fidelity states. So we need a more complex criterion for BPM than merely equiprobable parity checks. A natural alternative is a *greedy* approach: we perform a BPM whenever the sum of information gain and entropy reduction is larger than 1. However, this strategy is somewhat shortsighted, since it might be possible that performing a BPM will only become advantageous in future stages of parity checks, even if the information gain and entropy reduction of this particular BPM is smaller than 1. Indeed, performing a BPM will definitely produce more $[n_0, n_1]$ with small n_0, giving rise to a larger overall entropy reduction *in the end*.

A better greedy method is the following: instead of just looking at the *immediate* information gain and entropy reduction, we choose to perform PB or BPM according to what gives us the largest *total* information gain and entropy reduction, or equivalently, the largest expected number of distilled Bell pairs. To this end, we somehow *reverse* the order of the subsequent stages: we do not calculate the frequencies $f(0^{(\lambda)})$, $f(1^{(\lambda)})$ and $f([n_0^{(\lambda)}, n_1^{(\lambda)}])$ from the frequencies $f(0^{(2\lambda)})$, $f(1^{(2\lambda)})$ and $f([n_0^{(2\lambda)}, n_1^{(2\lambda)}])$ of the previous stage, but we calculate the *yields*, i.e. the expected net fraction of ebits that can be distilled, $\gamma(0^{(2\lambda)})$, $\gamma(1^{(2\lambda)})$ and $\gamma([n_0^{(2\lambda)}, n_1^{(2\lambda)}])$, from the yields $\gamma(0^{(\lambda)})$, $\gamma(1^{(\lambda)})$ and $\gamma([n_0^{(\lambda)}, n_1^{(\lambda)}])$.

This is done as follows. Firstly, it is clear that, for the first stage $\lambda = 1$,

$$\begin{aligned}\gamma(0^{(1)}) &= 1 - h(q_{00}, q_{11}),\\ \gamma(1^{(1)}) &= 1 - h(q_{01}, q_{10}),\\ \gamma([n_0, n_1]) &= n_0 + n_1 - 1 - H([n_0, n_1]).\end{aligned}$$

The next question is what to do with $0^{(2\lambda)}$, $1^{(2\lambda)}$ and $[n_0^{(2\lambda)}, n_1^{(2\lambda)}]$. Equiprobable parity checks should still be carried out by BPM, so the situation remains the same for all $1^{(2\lambda)}$. For the $0^{(2\lambda)}$, we can either apply the procedure of section 3.5.4, or directly apply a BPM $11 \otimes 1_\lambda 0_\lambda$. We treat all $0^{(2\lambda)}$ in a single possibility $[n_0^{(2\lambda)}, n_1^{(2\lambda)}]$ in the same way, for the same intuitive reason we gave in section 3.5.6: a BPM becomes more interesting, i.e. a larger entropy reduction, if fewer pairs are involved. Therefore, if one $0^{(2\lambda)}$ in a single possibility is split by BPM, automatically fewer pairs are involved and we have a cascading effect.

So, for all $0^{(2\lambda)}$ in a single possibility, we choose between two actions:

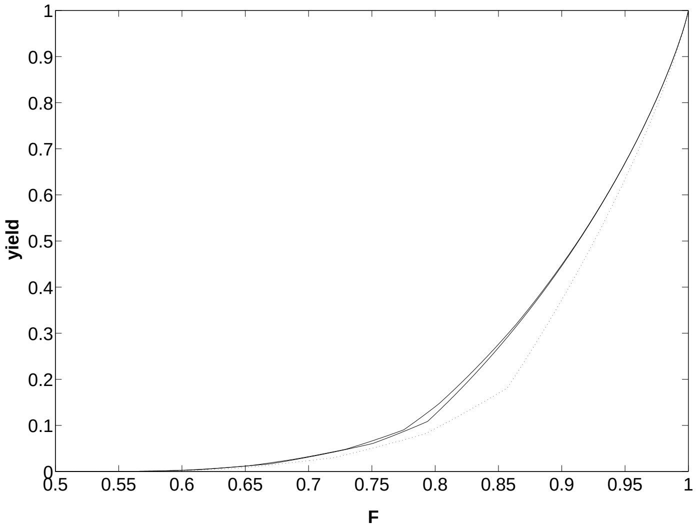


Figure 3.9: The yield of our original adaptive protocol (lower solid line) combined with recurrence iterations and the yield of the greedy variant (upper solid line), for Werner states with fidelity F, in comparison to the yield of hashing with recurrence iterations (dotted line).

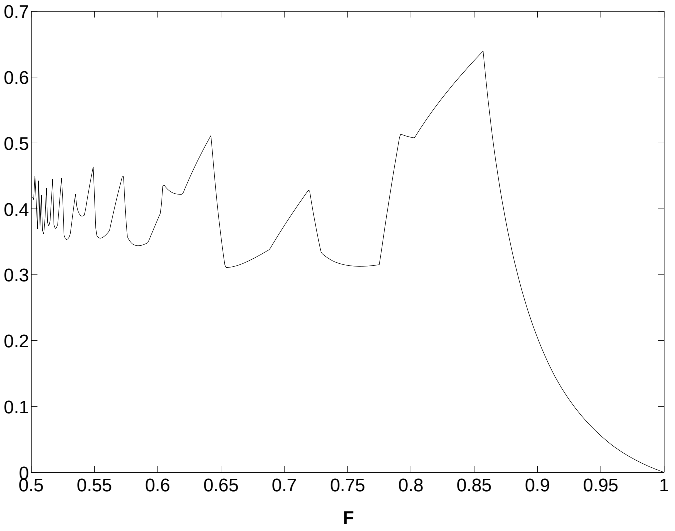


Figure 3.10: The relative difference between the yield of the greedy method and the yield of hashing with recurrence iterations, for Werner states with fidelity F.

1) either we apply the original procedure of section 3.5.4, with update rule given by (3.10),

2) or we perform a BPM $11 \otimes 1_\lambda 0_\lambda$, with the following update rule:

state		transforms to	with probability
$0^{(2\lambda)}$	$\rightarrow$	$[0^{(\lambda)}]0^{(\lambda)}$	$\frac{p^2_{0^{(\lambda)}}}{p_{0^{(2\lambda)}}}$
	$\rightarrow$	$[1^{(\lambda)}]1^{(\lambda)}$	$\frac{p^2_{1^{(\lambda)}}}{p_{0^{(2\lambda)}}}.$

For both cases, we subsequently calculate the resulting fractions of $[n_0^{(\lambda)}, n_1^{(\lambda)}]$ from a single possibility $[n_0^{(2\lambda)}, n_1^{(2\lambda)}]$. The yield is the sum of the products of these fractions with the corresponding yields $\gamma([0^{(\lambda)}, 1^{(\lambda)}])$ calculated in the previous stage, minus the number of used-up ebits. One can verify that this number of used-up ebits is

$$n_0^{(2\lambda)} \left(h(\frac{p^2_{0^{(\lambda)}}}{p_{0^{(2\lambda)}}}, \frac{p^2_{1^{(\lambda)}}}{p_{0^{(2\lambda)}}}) - \eta(0^{2\lambda}) \right)$$

for case (1) and zero for case (2). Now, by choosing the action with the highest yield (greedy), we arrive at the yield $\gamma([0^{(2\lambda)}, 1^{(2\lambda)}])$. The yield of the protocol is then

$$\begin{aligned} \max \ \Big\{ & p_{0^{(2^q)}} \gamma(0^{(2^q)}) + p_{1^{(2^q)}} \gamma(1^{(2^q)}) - h(p_{0^{(2^q)}}, p_{1^{(2^q)}}) \ , \\ & p_{0^{(2^q)}} \gamma([0^{(2^q)}]) + p_{1^{(2^q)}} \gamma([1^{(2^q)}]) \Big\} / 2^q, \end{aligned}$$

since for the parity check $11 \otimes 1_{2^q}$ in the final stage on 2^q pairs, the best is taken of either PB or BPM, although not much difference between the two is expected, as we already pointed out in section 3.5.5.

Remark. Note that, since we reverse the stages and take the best of two paths for every next stage, we do not know in advance what possibilities will eventually be present in the protocol. Therefore, we need to calculate γ for *all* possibilities, causing some overhead in the numerical calculation of the yield. However, this drawback is of no importance, as we are not concerned with the computational complexity of the calculation of the yield. The complexity of the protocol itself does not increase. $\diamond$

This greedy protocol has zero yield for low initial fidelity, so we still need to do recurrence iterations firstly. In order to overcome this weakness, we introduce the following 'back door' in the greedy procedure. Since the state $[0^{(2)}]$ is the state of the resulting qubit pair of a successful recurrence step, we do not submit it to the greedy choice explained above, but use it as an input of the entire protocol, with a possible reordering firstly, and calculate its γ as the yield of the protocol for this (improved) qubit pair. We do not have to repeat this recursive procedure infinitely, as the fraction of such reinserted pairs decreases exponentially fast with respect to the number of recursions. This new procedure now fully covers the initial recurrence iterations. Indeed, at worst we always performed BPM and no PB, leaving us with only $[0^{(2)}]$, which are reinserted, and $[1^{(2)}]$, on which BPM are applied, equivalent to discarding.

Remark. Recall that we judge our protocols based on their performance for Werner states, since of all Bell-diagonal states, they are most differing with rank two Bell-diagonal states, for which the problem of distillation is completely solved. For Werner states, it does not matter whether the parity checks are of the form $01 \otimes 1_\lambda$, $10 \otimes 1_\lambda$ or of the form $11 \otimes 1_\lambda$, as $p_{01} = p_{10} = p_{11}$. But with preceding recurrence iterations or reinserting $[0^{(2)}]$ in the protocol, this no longer holds, and we already mentioned that this requires a reordering like in between multiple recurrence iterations. By numerical simulation, it appears that the same reordering as for multiple recurrence iteration gives the highest yield, but it is hard to verify this analytically. ◇

The yield for Werner states of this final protocol is plotted in figure 3.9, and we see a slight improvement with respect to the combination of the original protocol with recurrence iterations. Although it seems that the greedy protocol does not really outperform the combination of hashing and recurrence for noisy states, since the yield for low fidelity is still very close to zero, we can only appreciate this by looking at the *relative difference* of the yields (i.e. the difference with respect to the yield itself). This is plotted in figure 3.10. We observe that the course of the relative difference, although strongly fluctuating, on average remains more or less constant in the low-fidelity region, which indicates that the improvement is mainly situated in the high-fidelity region. Indeed, the same improvement equally holds in the low-fidelity region because there the greedy protocol boils down to multiple recurrence iterations.

A final suggestion which could be taken into consideration, is the following. Instead of exclusively performing parity checks of the form $11 \otimes x$ and only in the end allowing parity checks $01 \otimes x$ or $10 \otimes x$, we could already introduce these in an earlier stage of the protocol. Commutativity of parity checks could then for instance be guaranteed by taking $1111 \otimes x$ and $1100 \otimes x$. However, we can no longer apply the decoupling rules. As a consequence, to trace all possibilities becomes extremely difficult, let alone deriving a criterion for the choice between BPM and PB. Nevertheless, we have tried this to some extent, resulting in a yield below that of the greedy protocol. Therefore, we have omitted the details here.

3.6 Finite protocols for low-fidelity states

We argued in the previous section that our asymptotic adaptive variants of hashing/breeding only work for high-fidelity states, which is passed on to low-fidelity states by recurrence iterations that boost the fidelity. In this section, we search for variants of recurrence that perform this task in an optimal way. We recall from section 3.3 that the protocols we consider consist of the same local Clifford operation on both sides followed by σ_z measurements on a number of qubit pairs. By theorem 3.2, the posterior probabilities of the unmeasured pairs are

$$\frac{p(\mathcal{C}+v)}{p(\mathcal{N}+v)}, \tag{3.20}$$

where all v satisfy $C_{(4)}^T v = u$. These probabilities are completely determined by u and a generating set of $\mathcal{C}$, as $\mathcal{N}$ is the dual space of $\mathcal{C}$. In the code-based interpretation, $\mathcal{C}$ is the binary representation of the stabilizer of the code. We focus on protocols that output only one qubit pair. The output fidelity F' is the maximum of (3.20) over all v that satisfy the above condition.

The protocol takes low-fidelity Werner states as an input. We now consider $F = \frac{1}{2} + \epsilon$, where ϵ is infinitesimal. We conjecture that in this region,[6] the resulting state is separable, i.e. $F' < \frac{1}{2}$, whenever $u \neq 0$ (depending on the measurement outcome), provided that there was no decoupling.[7] We have numerically verified this for many codes, but we were unable to find a proof. By theorem 3.2, the probability that $u = 0$ is $p(\mathcal{N})$, in which case $F' = p(\mathcal{C})/p(\mathcal{N})$. In first order of ϵ, this output fidelity equals $F' = \frac{1}{2} + \alpha\epsilon$, where $\alpha > 1$. We now prove this and calculate α for given $\mathcal{C}$. The *weight enumerator* of $\mathcal{C}$ is defined as the polynomial

$$W_{\mathcal{C}}(x,y) = \sum_{i=0}^{\kappa} A_i x^{\kappa-i} y^i$$

where A_i is the number of elements of weight i in $\mathcal{C}$. The weight enumerator of $\mathcal{N}$ is then given by the generalized *MacWilliams identity* [18, 62, 77]:

$$W_{\mathcal{N}}(x,y) = \frac{1}{|\mathcal{C}|} W_{\mathcal{C}}(x+3y, x-y) \tag{3.21}$$

For a Werner state with fidelity $F = \frac{1}{2} + \epsilon$, we have:

$$\begin{aligned} F' &= \frac{W_{\mathcal{C}}(F, \frac{1-F}{3})}{W_{\mathcal{N}}(F, \frac{1-F}{3})} = \frac{2^{\kappa-1} W_{\mathcal{C}}(F, \frac{1-F}{3})}{W_{\mathcal{C}}(1, \frac{4F-1}{3})} \\ &= \frac{2^{\kappa-1} \sum_{i=0}^{\kappa} A_i (\frac{1}{2}+\epsilon)^{\kappa-i} (\frac{1}{6} - \frac{1}{3}\epsilon)^i}{\sum_{i=0}^{\kappa} A_i (\frac{1}{3} + \frac{4}{3}\epsilon)^i} \end{aligned}$$

[6] We found counterexamples for higher F, but where $F' < F$, which means there is still entanglement, but it has decreased.

[7] Otherwise, it is possible that the code actually consists of two *separate* codes, for one of which $u = 0$.

$$
\begin{aligned}
&= \frac{2^{\kappa-1}\sum_{i=0}^{\kappa} A_i(-\frac{1}{3})^i \left(\sum_{j=0}^{\kappa-i} \binom{\kappa-i}{j}(\frac{1}{2})^{\kappa-i-j}\epsilon^j\right)\left(\sum_{j=0}^{i}\binom{i}{j}(-\frac{1}{2})^{i-j}\epsilon^j\right)}{\sum_{i=0}^{\kappa} A_i(\frac{4}{3})^i\left(\sum_{j=0}^{i}\binom{i}{j}(\frac{1}{4})^{i-j}\epsilon^j\right)} \\
&= \frac{\left(\frac{1}{2}\sum_{i=0}^{\kappa} A_i(\frac{1}{3})^i\right) + \left(\sum_{i=0}^{\kappa} A_i(\frac{1}{3})^i(\kappa-2i)\right)\epsilon + \mathcal{O}(\epsilon^2)}{\left(\sum_{i=0}^{\kappa} A_i(\frac{1}{3})^i\right) + \left(\sum_{i=0}^{\kappa} A_i(\frac{1}{3})^i 4i\right)\epsilon + \mathcal{O}(\epsilon^2)} \\
&= \frac{1}{2} + \alpha\epsilon + \mathcal{O}(\epsilon^2),
\end{aligned}
$$

where

$$
\alpha = \frac{\sum_{i=0}^{\kappa} A_i(\frac{1}{3})^i(\kappa-4i)}{\sum_{i=0}^{\kappa} A_i(\frac{1}{3})^i}. \tag{3.22}
$$

With q iterations of the protocol, ϵ has scaled by a factor α^q. We find for the recurrence protocol: $A_0 = 1, A_1 = 0, A_2 = 1 \Rightarrow \alpha = 6/5$, which is the best possible value for $\kappa = 2$. We have calculated the α value for numerous random stabilizers, of which the best are given in table 3.1. An efficient algorithm for generating random stabilizers is described in appendix B. The given codes are not unique, as many codes have the same weight distribution. Indeed, a permutation of the κ factors in the stabilizer elements already yields an equivalent code. We observe that applying the best $\kappa = 4$ protocol for Werner states (marked with an asterisk in table 3.1), with $\alpha = 2$, yields a larger fidelity increase than applying recurrence twice, for which $\alpha^2 = 1.44 < 2$. This is to be expected, because two successful recurrence iterations can be regarded as a single iteration of one particular $\kappa = 4$ protocol, as schematically depicted in figure 3.11. Moreover, in between the two recurrence iterations, twirling is applied, causing some loss of entanglement.

Remark. This no longer holds when one of the recurrence executions is not successful ($u \neq 0$), in which case the resulting pair is discarded. We will come back to this later. ◇

This approach has two major drawbacks:

1. Although a code with high α yields a high fidelity increase, this effect could very well be nullified by a corresponding low success probability $p(\mathcal{N})$. We will need a characteristic that somehow combines both.
2. It is implicitly assumed that between each two iterations, twirling is applied, leaving the fidelity invariant but equalizing the other probabilities, since α only gives the fidelity increase for Werner states as an input. However, recalling section 3.3.1, it is better to replace twirling by a reordering of the probabilities, but then α is no longer valid.

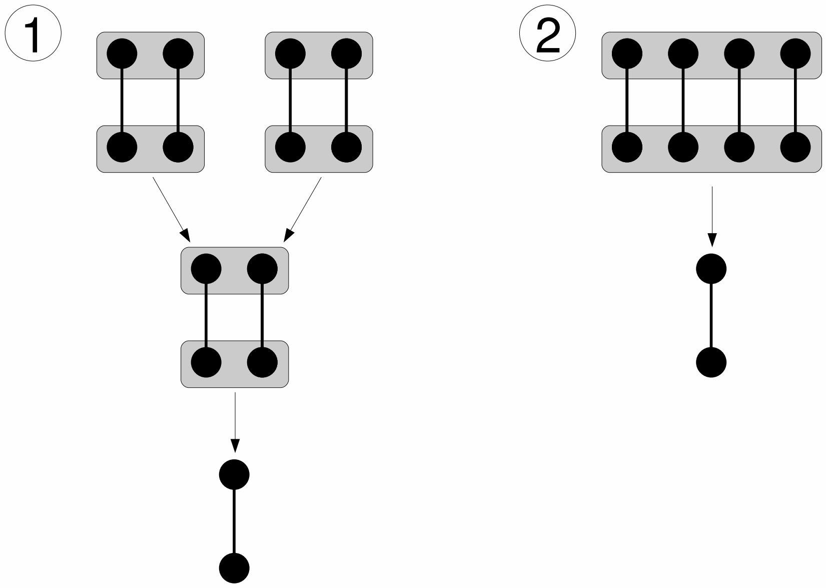


Figure 3.11: Two successful recurrence iterations (1) are a special case of a single successful $\kappa = 4$ protocol (2).

To meet the first drawback, the *joint performance* ξ was defined in [16] as the fraction of remaining pairs after doubling ϵ, or equivalently, let $F' = \frac{1}{2} + 2^t\epsilon$ after a number of iterations, then we are left with ξ^t pairs. It is clear that $2^t = \alpha^q$ and $\xi^t = [p(\mathcal{N})/\kappa]^q$, where q is the number of iterations, necessary for reaching F'. Recall that $p(\mathcal{N}) = 2^{1-\kappa}D + \mathcal{O}(\epsilon)$, where D is the denominator of (3.22). It follows that

$$\log_2 \xi = \frac{\log_2 D + 1 - \kappa - \log_2 \kappa}{\log_2 \alpha}. \tag{3.23}$$

After calculating ξ for multiple random stabilizers, we observe that, fixing κ, a large ξ corresponds to a large α (the optimal values found for ξ and α are given in table 3.1). This means that it is best to boost the fidelity as much as possible in one iteration, and α is, in that respect, a good characteristic after all. However, with increasing κ, initially ξ increases, but then decreases, reaching a maximum in $\kappa = 5$. Apparently, when the number of involved pairs becomes too large, the gain of α is nullified by the exponentially decreasing success probability and the larger loss of pairs in case of failure.

To overcome the second drawback, we should also take other Bell-diagonal states than Werner states into account. Let $p_0 = p_{00}$, $p_x = p_{01}$, $p_y = p_{11}$ and $p_z = p_{10}$. Between each two iterations of the protocol, we perform local single-qubit Clifford operations on the resulting state reordering the probabilities such that $p_x \geq p_y \geq p_z$. Since we are still dealing with low-fidelity states, we define $p_0 = \frac{1}{2} + \epsilon$ and $p_* = (\frac{1}{2} - \epsilon)\beta_*$, where $*$ stands for x, y, or z. It follows that $\sum_* \beta_* = 1$ and $\beta_x \geq \beta_y \geq \beta_z$. We assume that after a number of iterations, the β_* will converge.[8] We now calculate the α_∞ and ξ_∞ that correspond to this asymptotic behavior. The *complete weight enumerator* of $\mathcal{C}$ is defined as the polynomial

$$W_{\mathcal{C}}(w,x,y,z) = \sum_i A_i w^{i_0} x^{i_x} y^{i_y} z^{i_z}$$

where A_i is the number of elements in $\mathcal{C}$ with composition[9] $i = (i_0, i_x, i_y, i_z)$. Note that $i_0 + i_x + i_y + i_z = \kappa$. To calculate the resulting fidelity p_0', we use the MacWilliams identity for complete weight enumerators [62]:

$$W_{\mathcal{N}}(w,x,y,z) = \frac{1}{|\mathcal{C}|} W_{\mathcal{C}}(w+x+y+z, w+x-y-z, w-x+y-z, w-x-y+z). \tag{3.24}$$

In the same way as α, we calculate α_∞:

$$\begin{aligned} p_0' &= \frac{W_{\mathcal{C}}(p_0,p_x,p_y,p_z)}{W_{\mathcal{N}}(p_0,p_x,p_y,p_z)} \\ &= \frac{2^{\kappa-1}\sum_i A_i(\frac{1}{2}+\epsilon)^{i_0}(\frac{1}{2}-\epsilon)^{i_x+i_y+i_z}\beta_x^{i_x}\beta_y^{i_y}\beta_z^{i_z}}{\sum_i A_i[\beta_x + 2(1-\beta_x)\epsilon]^{i_x}[\beta_y + 2(1-\beta_y)\epsilon]^{i_y}[\beta_z + 2(1-\beta_z)\epsilon]^{i_z}} \end{aligned}$$

[8] In rare cases, they do not converge, but for these codes the asymptotic behavior can still be examined and they appear to perform worse than the ones listed below.

[9] the respective number of I, X, Y and Z in the codeword

$$= \frac{1}{2} + \alpha_\infty \epsilon + \mathcal{O}(\epsilon^2),$$

where

$$\alpha_\infty = \frac{\sum_i A_i \beta_x^{i_x} \beta_y^{i_y} \beta_z^{i_z} (i_0 - \frac{i_x}{\beta_x} - \frac{i_y}{\beta_y} - \frac{i_z}{\beta_z})}{\sum_i A_i \beta_x^{i_x} \beta_y^{i_y} \beta_z^{i_z}}. \tag{3.25}$$

Let D_∞ be the denominator of (3.25). Analogous to (3.23), we have

$$\log_2 \xi_\infty = \frac{\log_2 D_\infty + 1 - \kappa - \log_2 \kappa}{\log_2 \alpha_\infty}. \tag{3.26}$$

It remains for us to determine the limits of the converging β_*. The new probabilities p'_x, p'_y and p'_z are given by $p(\mathcal{C} + v)/p(\mathcal{N})$, sorted in descending order, where $\mathcal{C} + v$ are the three cosets of $\mathcal{C}$ in $\mathcal{N}$. There is no concise formula for the weight distribution of the cosets, given that of $\mathcal{C}$. Let A_i^* be the weight distribution corresponding to the coset giving rise to p'_*, then it follows:

$$\beta_* = \frac{\sum_i A_i^* \beta_x^{i_x} \beta_y^{i_y} \beta_z^{i_z}}{N} + \mathcal{O}(\epsilon),$$

where N is a normalizing factor such that $\sum_* \beta_* = 1$. The best ξ_∞ and corresponding other characteristics are given in table 3.1. Apparently, good asymptotic behavior is compatible with good performance for Werner states as well, except for $\kappa = 4$. Surprisingly, the optimal $\kappa = 4$ protocol is equivalent to iterating recurrence twice. One can also verify that the optimal $\kappa = 3$ protocol is equivalent to initially applying recurrence on the first two pairs, and then on the reordered resulting pair and the third pair.

We observe that the best ξ_∞ is reached for the original $\kappa = 2$ recurrence protocol. This is because by considering larger κ, the resulting protocols are in a sense *less adaptive*. Indeed, as we already mentioned in the remark on page 77, two recurrence iterations are no longer a special case of a single $\kappa = 4$ protocol as soon as one of both outcomes in the first iteration yields $u \neq 0$: for recurrence, whenever $u \neq 0$, the resulting (separable) pair is discarded, whereas for $\kappa = 4$, it is always combined with another, for which possibly $u = 0$. This can be circumvented by incorporating the possibility of choosing a different measurement when the outcome does not yield $u = 0$. A number of bounds for codes using only one-way communication are shown to be violated by such adaptive stabilizer codes [2]. However, for $\kappa = 4$, the best possible protocol already consists of two recurrence iterations. Therefore, no fully adaptive $\kappa = 4$ protocol will do better than recurrence. In general, comparing different adaptive protocols becomes difficult, since there is no longer a simple scalar characterizing their behavior. Before, whenever $u \neq 0$ we discarded the result, whereas now, the protocol just takes a different course. This gives rise to an exponentially expanding tree, of which every leaf corresponds to a different resulting state, with a particular probability. Numerically, we could circumvent having to take an exponential growing number of states into account by *mixing* states. A drawback is that this, like twirling, causes loss of entanglement.

κ	ξ	α	ξ_∞	α_∞	β_x	β_y	β_z
2	0.0077	1.2	0.0705	1.414	0.469	0.282	0.249
3	0.0191	1.5	0.0467	1.707	0.469	0.282	0.249
4*	0.0370	2	0.0370	2	1/3	1/3	1/3
4	0.0264	1.846	0.0374	2	0.469	0.282	0.249
5	0.0413	2.5	0.0413	2.5	1/3	1/3	1/3
6	0.0269	2.6	0.0269	2.6	1/3	1/3	1/3
7	0.0230	2.932	0.0233	2.942	0.359	0.325	0.316
8	0.0185	3.176	0.0188	3.190	0.349	0.326	0.325

Table 3.1: The best values of ξ_∞ for $\kappa = 2, \ldots, 8$, and the corresponding other characteristics. The (mostly non-unique) stabilizers are generated by:

$(\kappa = 2)$	ZZ
$(\kappa = 3)$	ZZI, XYZ
$(\kappa = 4)^*$	$ZZZZ, XXXX, IXYZ$
$(\kappa = 4)$	$ZZII, IIZZ, XYXY$
$(\kappa = 5)$	$ZYXYI, XXIYX, XZZXI, YIZYY$
$(\kappa = 6)$	$IZXXIX, YIZXYZ, IXYYYZ, XZYYXI, XIYZYX$
$(\kappa = 7)$	$YXYYIZZ, IZYZIYX, IYXXZIY, IYYYXXZ$
	$ZIYXXZI, XYIZYXI$
$(\kappa = 8)$	$IYIZXYXY, YIIXYIYZ, YXYXZXYX, IIZZIYIY$
	$XZZXZZXY, YYIYIIXZ, XXYIZIZX$

Yet, before much more delving into devising intricate performance criteria, we should ask ourselves firstly whether there is a good *physical* reason for considering infinitesimally small deviations from $F = \frac{1}{2}$, except for being a helpful approximation to simplify the low-fidelity case. In my opinion, the answer to that question is: *no.* The fidelity of the channel used to establish the qubit pairs can really be anything between 0 and 1. For the classical case, $F = \frac{1}{2}$ is the worst possible situation and information can still be transmitted as long as $F \neq \frac{1}{2}$, so considering infinitesimal deviations does make sense. Entanglement distillation, on the other hand, becomes impossible as soon as $F \leq \frac{1}{2}$, and this threshold is passed without any peculiarity.

Finally, we might consider larger values of F. We can still use (3.20) and the MacWilliams identities for calculating the posterior probabilities, but the higher order terms in ϵ can no longer be neglected. This, and the fact that after only a few steps it is already better to switch to a variant of hashing, makes an analytic approach rather difficult. A rule of thumb is: good codes, i.e. with a high distance d, tend to yield good protocols. Indeed, for large fidelity $F = 1 - \epsilon$, it is easy to see that

$$F' = \frac{p(\mathcal{C})}{p(\mathcal{N})} = \frac{1}{1 + p(\mathcal{N}\backslash\mathcal{C})/p(\mathcal{C})} = \frac{1}{1 + \mathcal{O}(\epsilon^d)} = 1 - \mathcal{O}(\epsilon^d).$$

The only thing left for us is numerical simulation, but attempts in that direction were unsuccessful, and we considered it not worthy of putting much more effort into it.

3.7 Conclusion

In this chapter, we have discussed the development of bipartite distillation protocols by making extensive use of the stabilizer formalism. After briefly introducing the concept of bipartite entanglement and Bell states, we have recapitulated in more detail the various elements of distillation protocols in the stabilizer formalism for the particular bipartite case, illustrated with the existing recurrence protocol. We have given the main lines of thought of the asymptotic protocols hashing and breeding and elaborated on the basic principles (partial breeding, entropy reduction, decoupling) on which adaptiveness can be introduced to improve the performance of these protocols. The binary picture allowed us to give a natural explanation of how local measurements give rise to entropy reduction, next to information extraction, which was their initial purpose. This was illustrated by explicitly calculating the yield of a number of variants for Werner states. Finally, we discussed the search for optimal finite protocols for noisy (and almost separable) input states.

Chapter 4

Multipartite entanglement distillation

4.1 Introduction

In this chapter, we describe asymptotic protocols for multipartite entanglement distillation. Compared to the bipartite case, multipartite entanglement is a much more intricate concept. While there is only one form of bipartite pure state entanglement, this definitely no longer applies to the multipartite case. Therefore, the characterization of multipartite entanglement, even when restricted to stabilizer states only, is a complicated matter and we do not delve further into it, but refer to [8, 34, 45, 46, 50, 61, 87, 89] for an extensive account. Multipartite entangled states over distant parties have important applications, especially in quantum cryptography. We cite [19, 21, 23, 31, 47, 57], but this is far from an exhaustive list.

In a sense, for practical purposes, multipartite entanglement distillation is rendered superfluous by bipartite entanglement distillation, because as soon as we are able to establish pure ebit pairs, these can be used as a resource for teleporting multipartite states. However, since we are dealing with distillation, we are concerned with entanglement as a *state characteristic*: given copies of a particular mixed state, how many pure state copies can we extract from these, by LOCC only. It will become clear that multipartite distillation protocols do make a difference when it comes to this task.

Many generalizations of bipartite distillation protocols to a multipartite setting have been found. They can be categorized according to asymptotic (hashing/breeding) [3, 19, 63, 59, 60] versus finite protocols [1, 3, 19, 29, 38, 43, 58, 59, 60, 63, 66, 67], to whether they take noise in the recovering operations into account [3, 29, 58, 59, 60] or to the kind of quantum states they are designed for. The cited references are only suited for CSS states or states that are locally equivalent to CSS states (e.g. two-colorable graph states), except

[66] (for the three-qubit W state[1]) and [38, 58, 60] (for arbitrary stabilizer states). Our work on multipartite distillation differs from the results in these references, and outperforms them, for two reasons. On the one hand, by making extensive use of the binary matrix framework of chapter 2, we are able to derive, for different classes of stabilizer states, the most general structure of the local Clifford operations used for the protocol, such that they effect a larger statistical dependence of multiple noisy copies of the input state, which is crucial for distillation. On the other hand, we derive and exploit particular properties of the so-called *strongly typical set*, a concept borrowed from classical information theory. We explain both ways in more detail.

Firstly, recall from chapter 2 that in order to extract information on the overall state of multiple copies of a given mixed state ρ, the copies that are sacrificed for local measurements should contain information on the overall state. To this end, local unitary operations are applied such that the copies become statistically dependent (cf. figure 1.1 on page 5). In order to stay in the binary picture, we restrict ourselves to local Clifford operations. Since we want the output of the protocol to be copies of a stabilizer state with generator matrix S, we demand that the local Clifford operations map the set of all tensor products of such states onto itself.[2] We derive the most general structure of such local Clifford operations by considering particular classes of states, identified by the structure of S. Intuitively, when S has a more general structure, the constraints on the local Clifford operations become stricter. This was recognized in [38], in a code-based approach, where it was specified what classes of stabilizer codes are fit for the distillation of different classes of stabilizer states: arbitrary stabilizer codes for CSS-H states[3], CSS codes for CSS states and CSS-H codes for arbitrary stabilizer states. There is a large similarity with our own work and this result of [38], except for the fact that for particular CSS states, we find more general local Clifford operations than those giving rise to CSS codes.

Secondly, by local measurements on a stabilizer state we simultaneously learn the eigenvalue of *more than one* stabilizing Pauli operation. For example, by locally measuring σ_z on all qubits of a CSS state, by theorem 2.22 we learn the value of b_i, for $i = 1, \ldots, n_z$. This feature is new with respect to the situation for Bell states. The goal of the protocol is to reduce the total entropy of all copies. In all references given above, the number of measurements according to a particular tripartition $\mathcal{M}$ is required to exceed the marginal entropies of each random variable b_i that is revealed by such a measurement. Yet, in most cases, these random variables are dependent. Therefore, the above requirement is too strict and results in too many measurements, or equally, a lower yield. In [19], this drawback was partially met by relaxing to conditional

[1] For three qubits, there are two local unitary equivalence classes of states [32]. Representatives of both classes are respectively the cat state $\frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$ and the W state $\frac{1}{\sqrt{3}}(|001\rangle + |010\rangle + |100\rangle)$.

[2] Actually, the copies to be measured do not need to be in this set, but this would complicate things too much.

[3] CSS-H states and codes are special subclasses of respectively CSS states and codes.

entropies, but it still gives rise to an overestimation of the necessary number of copies to be measured. The information-theoretical interpretation of multipartite hashing/breeding is very similar to the bipartite case. Again, we regard the overall state as an unknown pure state $|\psi_{\tilde{S},\tilde{u}}\rangle$. It is assumed that $\tilde{u} \in \mathcal{T}_\epsilon^{(\kappa)}$, the strongly typical set, with a vanishing probability that this assumption is false. By local Clifford operations and measurements, parity checks are performed on $\tilde{u}$, revealing its identity. By demanding that at the end of the protocol, all $\tilde{b} \in \mathcal{T}_\epsilon^{(\kappa)}$ that differ from $\tilde{u}$ are eliminated with approximate certainty, we are able to exactly calculate the minimal number of copies to be measured to achieve this goal. This minimum is the solution of a linear programming problem.

This chapter is organized as follows. In section 4.2, we define the information-theoretical concept strongly typical set. It is a stricter version of the typical set we encountered in chapter 3, yet asymptotically, they are almost identical. We will derive a particular property that is crucial for calculating the minimum number of copies that need to be measured in order to eliminate all elements of $\mathcal{T}_\epsilon^{(\kappa)}$ save $\tilde{u}$. In section 4.3, we outline the general procedure of the protocols, since there are many elements in common. In sections 4.4 and 4.5, we apply this procedure to the particular case of CSS states and graph states respectively. Finally, we formulate a conclusion in section 4.6. The results of this chapter were published in [54, 55].

4.2 Strongly typical set

In the context of multipartite distillation protocols, we will need a stricter version of the asymptotic equipartition property. The sequences that are contained by the *strongly* typical set are not only typical concerning their individual probabilities, but all individual *sample frequencies* have to be within 'typical' boundaries [22]. This is formally stated as follows.

Let $\tilde{X} = (X_1, \ldots, X_\kappa)$ be a sequence of independent identically-distributed discrete random variables, each having event set Ω with probability function

$$p : \Omega \to [0,1] : x \to p(x).$$

The strongly typical set $\mathcal{T}_\epsilon^{(\kappa)}$ is defined as the set of sequences $\tilde{x} = (x_1, \ldots, x_\kappa) \in \Omega^\kappa$ for which the sample frequencies

$$f_x(\tilde{x}) = \frac{|\{x_i \mid x_i = x\}|}{\kappa}$$

are close to the true values $p(x)$:

$$\mathcal{T}_\epsilon^{(\kappa)} = \{\tilde{x} \in \Omega^\kappa : \ |f_x(\tilde{x}) - p(x)| < \epsilon, \ \forall x \in \Omega\}. \tag{4.1}$$

In the following of this chapter, we will always assume $\kappa \to \infty$ and $\epsilon \to 0$. The next proposition shows that asymptotically, the strongly typical set is almost equal to the typical set. As a consequence, the main relevant properties of $\mathcal{A}_\epsilon^{(\kappa)}$ also hold for $\mathcal{T}_\epsilon^{(\kappa)}$.

Proposition 4.1 *(i)* $p(\mathcal{T}_\epsilon^{(\kappa)}) \geq 1 - \mathcal{O}(\kappa^{-1}\epsilon^{-2})$*;*

(ii) $\mathcal{T}_\epsilon^{(\kappa)} \subset \mathcal{A}_{\mathcal{O}(\epsilon)}^{(\kappa)}$*.*

Proof:

(i) It can be verified that $f_x(\tilde{X})$ has mean $p(x)$ and variance $p(x)[1-p(x)]/\kappa$. By Chebyshev's inequality [91], we then have

$$P(|f_x(\tilde{X}) - p(x)| \geq \epsilon) \leq \frac{p(x)[1-p(x)]}{\kappa\epsilon^2}.$$

It follows that $p(\mathcal{T}_\epsilon^{(\kappa)}) \geq 1 - \mathcal{O}(\kappa^{-1}\epsilon^{-2})$.

(ii) Let $\tilde{x} \in \mathcal{T}_\epsilon^{(\kappa)}$, then

$$\begin{aligned} \log_2 p(\tilde{x}) &= \log_2 \prod_{x\in\Omega} p(x)^{\kappa f_x(\tilde{x})} \\ &= \log_2 \prod_{x\in\Omega} p(x)^{\kappa[p(x)+\mathcal{O}(\epsilon)]} \\ &= \kappa \sum_{x\in\Omega} [p(x) + \mathcal{O}(\epsilon)] \log_2 p(x) \\ &= -\kappa[H(X) + \mathcal{O}(\epsilon)]. \end{aligned}$$

It follows that $\tilde{x} \in \mathcal{A}_{\mathcal{O}(\epsilon)}^{(\kappa)}$.

□

We define the function $y : \Omega \to \{1, \ldots, t\} : x \to y(x)$. It defines a partition of Ω into subsets Ω_j, for $j = 1, \ldots, t$. The following theorem is of central importance for asymptotic multipartite distillation protocols:

Theorem 4.2 *Given some* $\tilde{u} \in \mathcal{T}_\epsilon^{(\kappa)}$*, and the set*

$$\mathcal{Y}_{\tilde{u}} = \{\tilde{v} \in \mathcal{T}_\epsilon^{(\kappa)} \mid y(v_i) = y(u_i),\ \textit{for } i = 1, \ldots, \kappa\}.$$

Then we have

$$\frac{\log_2 |\mathcal{Y}_{\tilde{u}}|}{\kappa} = H(X) - H(Y) + \mathcal{O}(\epsilon) + \mathcal{O}(\kappa^{-1}\log\kappa),$$

where

$$H(Y) = -\sum_{j=1}^{t} p(\Omega_j) \log_2 p(\Omega_j)$$

is the entropy of the random variable $Y = y(X)$*.*

Proof: We define

$$f_{\Omega_j}(\tilde{x}) = \sum_{x\in\Omega_j} f_x(\tilde{x}).$$

By definition, for all $\tilde{v} \in \mathcal{Y}_{\tilde{u}}$ and for $j = 1,\ldots,t$, it holds

$$f_{\Omega_j}(\tilde{v}) = f_{\Omega_j}(\tilde{u}).$$

For all $x \in \Omega$, let f_x^* be some fixed value, where $\kappa f_x^* \in \mathbb{N}$, that satisfies

$$|f_x^* - p(x)| < \epsilon,\ \forall x \in \Omega \tag{4.2}$$

$$f_{\Omega_j}^* = \sum_{x\in\Omega_j} f_x^* = f_{\Omega_j}(\tilde{u}),\ \text{for } j = 1,\ldots,t. \tag{4.3}$$

We define $\mathcal{N}_{f^*}$ as the set of elements $\tilde{v} \in \mathcal{Y}_{\tilde{u}}$ with these exact sample frequencies f_x^*. Then elementary combinatorics tells us

$$|\mathcal{N}_{f^*}| = \prod_{j=1}^{t} \frac{[f_{\Omega_j}^*\kappa]!}{\prod_{x\in\Omega_j}[f_x^*\kappa]!}.$$

Using Stirling's approximation [92]:

$$\log\kappa! = \kappa\log\kappa - \kappa + \mathcal{O}(\log\kappa),\ \text{for } \kappa\to\infty,$$

and (4.3), we find

$$\log_2|\mathcal{N}_{f^*}| = \kappa\sum_{j=1}^{t}\left[f_{\Omega_j}^*\log_2 f_{\Omega_j}^* - \sum_{x\in\Omega_j} f_x^*\log_2 f_x^*\right] + \mathcal{O}(\log\kappa).$$

As f^* satisfies (4.2), we have $f_x^* = p(x) + \mathcal{O}(\epsilon),\ \forall x \in \Omega.$ Therefore,

$$\log_2|\mathcal{N}_{f^*}| = \kappa[H(X) - H(Y) + \mathcal{O}(\epsilon)] + \mathcal{O}(\log\kappa).$$

It is clear that $|\mathcal{N}_{f^*}| \le |\mathcal{Y}_{\tilde{u}}|$, as $\mathcal{N}_{f^*} \subseteq \mathcal{Y}_{\tilde{u}}$. Since there is a total $\le (2\epsilon\kappa)^t$ of f^* that satisfy (4.2), an upper bound for $|\mathcal{Y}_{\tilde{u}}|$ is

$$(2\epsilon\kappa)^t \max_{f^*}|\mathcal{N}_{f^*}|,$$

where the maximum is taken over all f^* that satisfy (4.2)-(4.3). It follows that

$$\log_2|\mathcal{Y}_{\tilde{u}}| = \kappa[H(X) - H(Y) + \mathcal{O}(\epsilon)] + \mathcal{O}(\log\kappa).$$

□

4.3 General procedure of multipartite stabilizer state hashing/breeding

In this section, we explain the main lines of thought of multipartite stabilizer state hashing or breeding, without assuming particular properties of the states under consideration. For each protocol, we distinguish between an *operation* section and an *information* section. Recalling the general setting of a distillation protocol (bipartite as well as multipartite) in section 2.5, we start with κ copies of a mixed n-qubit state that is diagonal in the S-basis, where S is the generator matrix of the pure stabilizer state we want to distill. The protocol consists of local Clifford operations, represented by (2.22), followed by measurements on a number of copies (or for breeding: on appended predistilled states). Determining which freedom we have in choosing these local Clifford operations constitutes the operation section. Next, taking these limitations into account, we show how to calculate the minimal number of copies that have to be measured in order to purify the remaining copies. This is the information section. Both are explained in detail in sections 4.3.1 and 4.3.2.

4.3.1 Operation section

The local Clifford operations, represented by $C \in \mathbb{Z}_2^{2n\kappa\times 2n\kappa}$ defined in (2.22), transform the overall generator matrix $\tilde{S} \in \mathbb{Z}_2^{2n\kappa\times n\kappa}$, defined in (2.24), into $C\tilde{S}$. We will only allow local Clifford operations that transform the original unknown state $|\psi_{\tilde{S},\tilde{u}}\rangle$ into a state that is represented by the same generator matrix $\tilde{S}$. As such, these local Clifford operations constitute a group and map the set of tensor products of states represented by S onto itself. In order to satisfy this constraint, there must exist some $R \in \mathbb{Z}_2^{n\kappa\times n\kappa}$ such that

$$C\tilde{S}R = \tilde{S}. \tag{4.4}$$

Clearly, the more we restrict the structure of S, the more freedom we have in the structure of C. That is why we separately solve this problem for different types of stabilizer states in the next sections. Applying the local Cliffords to the state $|\psi_{\tilde{S},\tilde{u}}\rangle$, by (2.20) and (2.21), the i-th copy is transformed into $|\psi_{S,\bar{R}^T\tilde{u}}\rangle$, where the columns of $\bar{R} \in \mathbb{Z}_2^{n\kappa\times n}$ are columns $i, \kappa+i, \ldots, (n-1)\kappa+i$ of R. We will show later that all $\bar{R}$ resulting from (4.4) constitute a vector space $\mathcal{R}$.

Remark. To be precise, the set of all possible $\bar{R}$ only *approximates* a vector space. Indeed, since R is invertible, the columns of $\bar{R}$ are independent, so for instance $\bar{R} = 0$ is not allowed. However, we will show that the possibility that an element of $\mathcal{R}$ is not allowed, is negligible.

Nonetheless, a problem might arise because we will see that particular linear constraints linking different columns of $\bar{R}$ (resulting from simplecticity constraints), are not incorporated into the definition of $\mathcal{R}$. However, we will circumvent this problem by showing that either

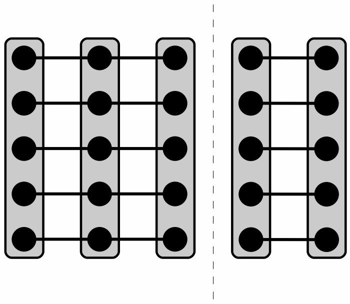


Figure 4.1: Because all operations are local, the protocol applied on multiple copies of a separable state is effectively two distinct protocols on the respective separable parts.

- not all columns of $\bar{R}$ are relevant at the same time (section 4.4); or
- only a fixed part of each column of $\bar{R}$ is relevant (section 4.5).

◇

Two things should be pointed out concerning the solution of (4.4). Firstly, we will demand that the state we want to distill is *fully entangled*. This means that there is no dichotomy of the qubits of the state such that the state is separable with respect to this dichotomy. Otherwise, since all operations are local, no entanglement can be created between the two parts, and the overall protocol then really consists of *two* distinct protocols, as illustrated in figure 4.1. This is an unnecessary complication. Moreover, the yield of these protocols combined is the minimum of the respective yields.

Secondly, in some cases it is useful to consider two equivalent views of the protocol, for reasons that will become clear later. Since the local Clifford operations satisfying (4.4) form a group, subsequently applying such Clifford operations locally on unmeasured qubits followed by measuring one qubit is equivalent to applying a single Clifford operation, followed by all measurements. Both views are illustrated in figure 4.2.

4.3.2 Information section

After the local Clifford operations, we apply measurements on $m\kappa$ copies to extract information on the unknown $\tilde{u}$. For each type of measurement, specified by the tripartition $\mathcal{M}$ as defined in theorem 2.22, we derive lower bounds on the fractions $m(\mathcal{M})$ of each type of measurement $\mathcal{M}$ that are necessary to purify the state. Indeed, next to the possibility of our initial assumption $\tilde{u} \in \mathcal{T}_\epsilon^{(\kappa)}$

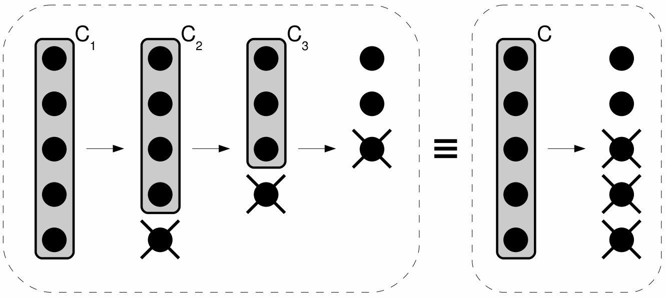


Figure 4.2: Two equivalent views of the protocol. Subsequent Clifford operations (C_1, C_2, C_3) performed on unmeasured qubits, each followed by the measurement of a single qubit, are equivalent to performing a single Clifford operation $(C = C_3C_2C_1)$ and the same measurements.

being wrong, with vanishing probability

$$p_1 = \mathcal{O}(\kappa^{-1}\epsilon^{-2}), \tag{4.5}$$

the protocol fails if some $\tilde{b} \neq \tilde{u}$ in $\mathcal{T}_\epsilon^{(\kappa)}$ survives the elimination process. The lower bounds on the $m(\mathcal{M})$ guarantee that the probability p_2 of this event vanishes. For both hashing and breeding, it is clear that the yield then equals $1 - m$, where m is the minimal $\sum_{\mathcal{M}} m(\mathcal{M})$ over all $m(\mathcal{M})$ satisfying these lower bounds.

To derive these, firstly we calculate the probability that a particular $\tilde{b}$ is not eliminated by the measurement of a single copy according a given tripartition $\mathcal{M}$, which occurs when $\tilde{b}$ is compatible with the observed measurement outcome (which is, by assumption, compatible with $\tilde{u}$). According to theorem 2.22, the measurement of the i-th copy $|\psi_{S,\bar{R}^T\tilde{u}}\rangle$ yields $v^T\bar{R}^T\tilde{u}$, $\forall v \in \mathcal{V}(\mathcal{M})$, which is a subspace of $\mathbb{Z}_2^n$. Therefore, $\tilde{b}$ is not eliminated when $V(\mathcal{M})^T\bar{R}^T\Delta\tilde{b} = 0$, where $\Delta\tilde{b} = \tilde{b} + \tilde{u}$ and col$[V(\mathcal{M})] = \mathcal{V}(\mathcal{M})$. Fixing $\Delta\tilde{b}$ and $\mathcal{M}$ thus yields a linear function

$$\phi : \mathcal{R} \to \mathbb{Z}_2^{n(\mathcal{M})} : \bar{R} \to \phi(\bar{R}) = V(\mathcal{M})^T\bar{R}^T\Delta\tilde{b},$$

where $n(\mathcal{M}) = \dim[\mathcal{V}(\mathcal{M})] \leq n$. We define

$$d(\mathcal{M}, \Delta\tilde{b}) = \dim[\phi(\mathcal{R})],$$

the dimension of the range of ϕ. It follows that $d(\mathcal{M}, \tilde{b}) \leq n(\mathcal{M})$. The best we can do is randomly choose $\bar{R}$ from $\mathcal{R}$ with uniform probability: then $\phi(\mathcal{M})$ is uniformly distributed too, yielding maximal information. Indeed, all cosets in $\mathcal{R}$ of the kernel ϕ have the same number of elements. Consequently, the probability that $\tilde{b}$ is not eliminated by a measurement according to $\mathcal{M}$ equals $2^{-d(\mathcal{M},\Delta\tilde{b})}$, which is the inverse of the number of possible values of $\phi(\bar{R})$.

Secondly, we calculate the probability that a particular $\tilde{b}$ is not eliminated by *all* measurements. It is easily verified that this probability equals

$$\prod_{\mathcal{M}} \left[2^{-d(\mathcal{M},\Delta\tilde{b})}\right]^{m(\mathcal{M})\kappa} = 2^{-\kappa\sum_{\mathcal{M}} m(\mathcal{M})d(\mathcal{M},\Delta\tilde{b})},$$

where the product in the LHS (and sum in the RHS) runs over all possible tripartitions $\mathcal{M}$ of $\{1,\ldots,n\}$.

Thirdly, we calculate an upper bound for the probability p_2 that some $\tilde{b} \neq \tilde{u}$ survives the entire procedure. This probability p_2 is at most

$$\sum_{f\neq 0} N_f^* 2^{-\kappa\sum\limits_{\mathcal{M}} m(\mathcal{M})f(\mathcal{M})}, \tag{4.6}$$

where the sum runs over all functions $f : f(\mathcal{M}) \in \{0,\ldots,n(\mathcal{M})\}$ that are not identical to zero. N_f^* is the number of $\tilde{b} \in \mathcal{T}_\epsilon^{(\kappa)}$ for which $d(\mathcal{M},\Delta\tilde{b}) = f(\mathcal{M})$. Let N_f be the number of $\tilde{b} \in \mathcal{T}_\epsilon^{(\kappa)}$ for which $d(\mathcal{M},\Delta\tilde{b}) \leq f(\mathcal{M})$.

Lemma 4.3 *Given natural numbers N_i that scale exponentially with κ, for $i = 1,\ldots,t$, where t is independent of κ. Then*

$$\frac{\log_2\left(\sum\limits_{i=1}^{t} N_i\right)}{\kappa} = \max_i \frac{\log_2 N_i}{\kappa} + \mathcal{O}(\kappa^{-1}), \text{ for } \kappa \to \infty.$$

Proof: We have

$$\max_i N_i \leq \sum_{i=1}^{t} N_i \leq t \max_i N_i.$$

Then there exists some r such that

$$\sum_{i=1}^{t} N_i = r \max_i N_i,$$

where $1 \leq r \leq t$, therefore $r = \mathcal{O}(1)$, and lemma 4.3 follows. □

Proposition 4.4 *$p_2 = \mathcal{O}(2^{-\kappa^{1-\eta}})$ and therefore vanishes for $\kappa \to \infty$ if the following inequalities hold:*

$$\sum_{\mathcal{M}} m(\mathcal{M})f(\mathcal{M}) \geq \frac{\log_2 N_f}{\kappa} + \mathcal{O}(\kappa^{-\eta}), \ \forall f \not\equiv 0, \tag{4.7}$$

for some $\eta < 1$.

Proof: It can be verified that the upper bound (4.6) is $\mathcal{O}(2^{-\kappa^{1-\eta}})$ if:

$$\sum_{\mathcal{M}} m(\mathcal{M})f(\mathcal{M}) \geq \frac{\log_2 N_f^*}{\kappa} + \mathcal{O}(\kappa^{-\eta}),\ \forall f \not\equiv 0, \tag{4.8}$$

and $\eta < 1$. We now prove that the inequalities (4.7) are equivalent to (4.8). Evidently,

$$N_f = \sum_{f' \leq f} N_{f'}^*,$$

where $f' \leq f$ stands for: $f'(\mathcal{M}) \leq f(\mathcal{M})$, $\forall \mathcal{M}$. With lemma 4.3, we have

$$\frac{\log_2 N_f}{\kappa} = \max_{f' \leq f} \frac{\log_2 N_{f'}^*}{\kappa} + \mathcal{O}(\kappa^{-1}) \geq \frac{\log_2 N_f^*}{\kappa} + \mathcal{O}(\kappa^{-1}),$$

so (4.7) implies (4.8). Let $f'' = \operatorname*{argmax}_{f' \leq f} N_{f'}^*$. Then it follows from (4.8) that

$$\begin{aligned}
\sum_{\mathcal{M}} m(\mathcal{M})f(\mathcal{M}) &\geq \sum_{\mathcal{M}} m(\mathcal{M})f''(\mathcal{M}) \\
&\geq \frac{\log_2 N_{f''}^*}{\kappa} + \mathcal{O}(\kappa^{-\eta}) \\
&= \frac{\log_2 N_f}{\kappa} - \mathcal{O}(\kappa^{-1}) + \mathcal{O}(\kappa^{-\eta}) \\
&= \frac{\log_2 N_f}{\kappa} + \mathcal{O}(\kappa^{-\eta}).
\end{aligned}$$

□

Fourthly, we calculate $\frac{\log_2 N_f}{\kappa}$. To this end, we use the following proposition:

Proposition 4.5 *Given a particular subspace $\mathcal{J}$ of $\mathbb{Z}_2^n$, we define the set*

$$\mathcal{L}(\mathcal{J}) = \{\tilde{b} \in \mathcal{T}_\epsilon^{(\kappa)} \mid \Delta\tilde{b} \in \mathcal{J}^\perp \otimes \mathbb{Z}_2^\kappa\}.$$

Then it follows that

$$\frac{\log_2 |\mathcal{L}(\mathcal{J})|}{\kappa} = G(\mathcal{J}^\perp) + \mathcal{O}(\epsilon) + \mathcal{O}(\kappa^{-1}\log\kappa),$$

where we define the function

$$G(\mathcal{J}^\perp) = -\sum_{b \in \mathbb{Z}_2^n} p(b) \log_2 \frac{p(b)}{p(b+\mathcal{J}^\perp)}.$$

Note that $G(\mathcal{J}^\perp)$ equals the relative entropy $H(p||q)$, *where the probability distribution q is defined by $q(b) = p(b+\mathcal{J}^\perp)$.*

Proof: $\Delta\tilde{b} \in \mathcal{J}^\perp \otimes \mathbb{Z}_2^\kappa \;\Leftrightarrow\; v^T \Delta b_i = 0$, for $i = 1, \ldots, \kappa$ and $\forall v \in \mathcal{J} \;\Leftrightarrow\; b_i$ and u_i are in the same coset of $\mathcal{J}^\perp$, for $i = 1, \ldots, \kappa$. Let Ω_j, for $j = 1, \ldots, t = 2^{\dim \mathcal{J}}$, be the cosets of $\mathcal{J}^\perp$ in $\mathbb{Z}_2^n$. From theorem 4.2, it follows that

$$\begin{aligned}\frac{\log_2 |\mathcal{L}(\mathcal{J})|}{\kappa} &= -\sum_{b \in \mathbb{Z}_2^n} p(b) \log_2 p(b) + \sum_{j=1}^{t} p(\Omega_j) \log_2 p(\Omega_j) \; + \mathcal{O}(\epsilon) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad + \mathcal{O}(\kappa^{-1} \log \kappa) \\ &= -\sum_{b \in \mathbb{Z}_2^n} p(b) \log_2 \frac{p(b)}{p(b + \mathcal{J}^\perp)} + \mathcal{O}(\epsilon) + \mathcal{O}(\kappa^{-1} \log \kappa).\end{aligned}$$

□

Recall that N_f is the number of $\tilde{b} \in \mathcal{T}_\epsilon^{(\kappa)}$ for which $d(\mathcal{M}, \Delta\tilde{b}) \leq f(\mathcal{M})$. Let $\mathcal{W}_f(\mathcal{M})$ be a subspace of dimension $n(\mathcal{M}) - f(\mathcal{M})$ of $\mathcal{V}(\mathcal{M})$. Given $\mathcal{M}$, we see that $d(\mathcal{M}, \Delta\tilde{b}) \leq f(\mathcal{M})$ for all $\Delta\tilde{b} \in \mathbb{Z}_2^{n\kappa}$ that satisfy $(\bar{R}w)^T \Delta\tilde{b} = 0$, $\forall \bar{R} \in \mathcal{R}$ and $\forall w \in \mathcal{W}_f(\mathcal{M})$. We will show in the next sections that these $\Delta\tilde{b}$ constitute a space $\mathcal{J}[\mathcal{W}_f(\mathcal{M})]^\perp \otimes \mathbb{Z}_2^\kappa$, where $\mathcal{J}[\mathcal{W}_f(\mathcal{M})]$ is a subspace of $\mathbb{Z}_2^n$ and function of $\mathcal{W}_f(\mathcal{M})$. It follows that

$$N_f = \left| \bigcup_{\mathcal{W}_f} \mathcal{L}\left(\sum_{\mathcal{M}} \mathcal{J}[\mathcal{W}_f(\mathcal{M})] \right) \right|$$

where the union runs through all combinations of subspaces $\mathcal{W}_f(\mathcal{M})$ of dimension $n(\mathcal{M}) - f(\mathcal{M})$ of $\mathcal{V}(\mathcal{M})$. The number of such subspaces is independent of κ. Therefore, similar to lemma 4.3, we have

$$\frac{\log_2 N_f}{\kappa} = \frac{\log_2 \max\limits_{\mathcal{W}_f} \left| \mathcal{L}\left(\sum_{\mathcal{M}} \mathcal{J}[\mathcal{W}_f(\mathcal{M})] \right) \right|}{\kappa} + \mathcal{O}(\kappa^{-1}).$$

By proposition 4.5, it follows that

$$\frac{\log_2 N_f}{\kappa} = \max_{\mathcal{W}_f} G\left(\left[\sum_{\mathcal{M}} \mathcal{J}[\mathcal{W}_f(\mathcal{M})] \right]^\perp \right) + \mathcal{O}(\epsilon) + \mathcal{O}(\kappa^{-1} \log \kappa). \tag{4.9}$$

Finally, let for instance $\epsilon = \mathcal{O}(\kappa^{-1/3})$ and $\eta = \frac{1}{4}$, then it follows from (4.5), (4.9) and proposition 4.4 that

$$\begin{aligned} p_1 &= \mathcal{O}(\kappa^{-1/3}), \\ p_2 &= \mathcal{O}(2^{-3\kappa/4}), \end{aligned}$$

and the total failure probability $p_1 + p_2$ of the protocol vanishes for $\kappa \to \infty$. The yield $1 - m$ is maximized by minimizing the total number of measured copies $m\kappa$. Combining (4.9) with proposition 4.4 and neglecting vanishing terms, we arrive at the following linear programming (LP) problem for calculating the optimal m:

$$\begin{array}{ll} \text{minimize} & m = \sum_{\mathcal{M}} m(\mathcal{M}) \\ \text{subject to} & \sum_{\mathcal{M}} m(\mathcal{M}) f(\mathcal{M}) \geq \max_{\mathcal{W}_f} G\left(\left[\sum_{\mathcal{M}} \mathcal{J}[\mathcal{W}_f(\mathcal{M})]\right]^{\perp}\right), \text{ for all } f \not\equiv 0, \end{array} \tag{4.10}$$

where, given the function $f : \mathcal{M} \to f(\mathcal{M})$, the maximum in the RHS is taken over all combinations of subspaces $\mathcal{W}_f(\mathcal{M})$ of dimension $n(\mathcal{M}) - f(\mathcal{M})$ of $\mathcal{V}(\mathcal{M})$.

4.4 CSS state hashing

In this section, we focus on the hashing protocol for CSS states. Although CSS state mixtures can also be distilled by protocols for arbitrary stabilizer states, like the stabilizer state breeding protocol of the next section, higher yields can be achieved when exploiting the particular structure of stabilizer states. Indeed, as already explained in the previous section, a more restricted structure of the stabilizer generator matrix S results in more freedom in the local Clifford operations by the constraint (4.4). In section 4.4.1, we will derive the exact constraints on C that result from (4.4) and the corresponding space $\mathcal{R}$. Then, in section 4.4.2, we calculate the space $\mathcal{J}[\mathcal{W}_f(\mathcal{M})] \otimes \mathbb{Z}_2^{\kappa}$ of states $\Delta\tilde{b}$ that satisfy $(\bar{R}w)^T \Delta\tilde{b} = 0$, $\forall \bar{R} \in \mathcal{R}$ and $\forall w \in \mathcal{W}_f(\mathcal{M})$ for the particular case of CSS states, necessary to derive the LP problem to find the yield of the protocol. Finally, in section 4.4.3, we illustrate this with two examples.

4.4.1 Operation section

The stabilizer generator matrix of a CSS state is of the form (2.33). To check whether it represents a fully entangled state, as required from section 4.3.1, we have the following proposition:

Proposition 4.6 *A CSS state with generator matrix given by (2.33) is separable if and only if there exists some permutation matrix $\Pi \in \mathbb{Z}_2^{n\times n}$ and an invertible matrix $R \in \mathbb{Z}_2^{n_z\times n_z}$ such that*

$$\Pi S_z R = \left[\begin{array}{cc} S_z' & 0 \\ 0 & S_z'' \end{array}\right],$$

where $S_z' \in \mathbb{Z}_2^{n'\times n_z'}$, $S_z'' \in \mathbb{Z}_2^{n''\times n_z''}$ and $0 < n_z' < n_z$. This also holds when using S_x instead of S_z.

Proof: Recall that the CSS state is already entirely defined by S_z, as $S_z^T S_x = 0$. Since S_z is full rank, also S_z' and S_z'' are full rank, and it is possible to find

$S'_x \in \mathbb{Z}_2^{n' \times (n'-n'_z)}$ and $S''_x \in \mathbb{Z}_2^{n'' \times (n''-n''_z)}$ such that ${S'_z}^T S'_x = 0$ and ${S''_z}^T S''_x = 0$. The stabilizer that results from the qubit permutation Π is represented by

$$\left[\begin{array}{cccc} S'_z & 0 & 0 & 0 \\ 0 & 0 & S''_z & 0 \\ 0 & S'_x & 0 & 0 \\ 0 & 0 & 0 & S''_x \end{array}\right],$$

and is thus separable by proposition 2.16. □

By proposition 2.27, the CSS state under consideration is represented by

$$S_z = \left[\begin{array}{c} I_{n_z} \\ \theta \end{array}\right] \text{ and } S_x = \left[\begin{array}{c} \theta^T \\ I_{n_x} \end{array}\right], \tag{4.11}$$

up to a permutation of the n qubits. Let

$$\tilde{A}_z = \left[\begin{array}{ccc} A_1 & & \\ & \ddots & \\ & & A_{n_z} \end{array}\right], \ \tilde{A}_x = \left[\begin{array}{ccc} A_{n_z+1} & & \\ & \ddots & \\ & & A_n \end{array}\right].$$

Using analogous definitions for $\tilde{B}_z, \tilde{B}_x, \tilde{C}_z, \tilde{C}_x, \tilde{D}_z$ and $\tilde{D}_x$, we arrive at the following theorem:

Theorem 4.7 *Local Clifford operations represented by C, given by (2.22), satisfy (4.4) when applied on a CSS state, if and only if*

$$A_1 = \cdots = A_n, \quad \textit{i.e.} \quad \tilde{A} = I_n \otimes A_1, \tag{4.12}$$

$$D_1 = \cdots = D_n, \quad \textit{i.e.} \quad \tilde{D} = I_n \otimes D_1, \tag{4.13}$$

$$\left(\left[\begin{array}{cc} \theta & I_{n_x} \\ L_{\theta^T}^T & 0 \end{array}\right] \otimes I_\kappa\right) \left[\begin{array}{c} B_1 \\ \vdots \\ B_n \end{array}\right] = 0, \tag{4.14}$$

$$\left(\left[\begin{array}{cc} I_{n_z} & \theta^T \\ 0 & L_\theta^T \end{array}\right] \otimes I_\kappa\right) \left[\begin{array}{c} C_1 \\ \vdots \\ C_n \end{array}\right] = 0. \tag{4.15}$$

where the n_x-bit columns of L_θ are $\theta_i \odot \theta_j$ $\forall i,j : 1 \le i < j \le n_z$, and an analogous definition holds for L_{θ^T}.

We distinguish two cases:

(i) θ is orthogonal:

Then $\operatorname{col}(S_z) = \operatorname{col}(S_x)$, and (4.14)-(4.15) is equivalent to:

$$B_1 = \cdots = B_n, \quad \textit{i.e.} \quad \tilde{B} = I_n \otimes B_1,$$
$$C_1 = \cdots = C_n, \quad \textit{i.e.} \quad \tilde{C} = I_n \otimes C_1.$$

(ii) <u>θ *is not orthogonal:*</u>

Then, for all solutions to (4.12)-(4.15), $B_i = 0$, $\forall i \in Z_B$, *and* $C_i = 0$, $\forall i \in Z_C$, *where* $Z_B, Z_C \subseteq \{1, \ldots, n\}$ *and* $Z_B \cup Z_C = \{1, \ldots, n\}$. *Furthermore,* $D_1 = A_1^{-T}$, *and* $A_1^T C_i$ *and* $A_1^{-1} B_i$ *are symmetric, for* $i = 1, \ldots, n$.

Proof: With (4.11), the LHS of (4.4) becomes

$$\begin{bmatrix} \tilde{A}_z & 0 & \tilde{B}_z & 0 \\ 0 & \tilde{A}_x & 0 & \tilde{B}_x \\ \tilde{C}_z & 0 & \tilde{D}_z & 0 \\ 0 & \tilde{C}_x & 0 & \tilde{D}_x \end{bmatrix} \begin{bmatrix} I_{n_z} \otimes I_\kappa & 0 \\ \theta \otimes I_\kappa & 0 \\ 0 & \theta^T \otimes I_\kappa \\ 0 & I_{n_x} \otimes I_\kappa \end{bmatrix} R =$$

$$\begin{bmatrix} \tilde{A}_z & \tilde{B}_z(\theta^T \otimes I_\kappa) \\ \tilde{A}_x(\theta \otimes I_\kappa) & \tilde{B}_x \\ \tilde{C}_z & \tilde{D}_z(\theta^T \otimes I_\kappa) \\ \tilde{C}_x(\theta \otimes I_\kappa) & \tilde{D}_x \end{bmatrix} R.$$

We can write this as two separate equations:

$$\begin{aligned} \begin{bmatrix} \tilde{A}_z & \tilde{B}_z(\theta^T \otimes I_\kappa) \\ \tilde{C}_x(\theta \otimes I_\kappa) & \tilde{D}_x \end{bmatrix} R &= I_{n\kappa} \qquad (4.16) \\ \begin{bmatrix} \tilde{C}_z & \tilde{D}_z(\theta^T \otimes I_\kappa) \\ \tilde{A}_x(\theta \otimes I_\kappa) & \tilde{B}_x \end{bmatrix} R &= \begin{bmatrix} 0 & \theta^T \otimes I_\kappa \\ \theta \otimes I_\kappa & 0 \end{bmatrix}. \end{aligned}$$

Eliminating R, we arrive at

$$\begin{bmatrix} 0 & \theta^T \otimes I_\kappa \\ \theta \otimes I_\kappa & 0 \end{bmatrix} \begin{bmatrix} \tilde{A}_z & \tilde{B}_z(\theta^T \otimes I_\kappa) \\ \tilde{C}_x(\theta \otimes I_\kappa) & \tilde{D}_x \end{bmatrix} =$$

$$\begin{bmatrix} \tilde{C}_z & \tilde{D}_z(\theta^T \otimes I_\kappa) \\ \tilde{A}_x(\theta \otimes I_\kappa) & \tilde{B}_x \end{bmatrix},$$

which is a necessary and sufficient condition for the existence of an R such that (4.4) holds. Blockwise comparison of both sides yields the following equations

$$\begin{aligned} (\theta \otimes I_\kappa)\tilde{A}_z &= \tilde{A}_x(\theta \otimes I_\kappa), && (4.17) \\ (\theta^T \otimes I_\kappa)\tilde{D}_x &= \tilde{D}_z(\theta^T \otimes I_\kappa), && (4.18) \\ (\theta \otimes I_\kappa)\tilde{B}_z(\theta^T \otimes I_\kappa) &= \tilde{B}_x, && (4.19) \\ (\theta^T \otimes I_\kappa)\tilde{C}_x(\theta \otimes I_\kappa) &= \tilde{C}_z. && (4.20) \end{aligned}$$

We show that (4.12) follows from (4.17). Comparing each corresponding block on both sides of (4.17) yields:

$$A_j = A_{n_z+i} \quad \text{if } \theta_{ij} = 1, \text{ for } i = 1, \ldots, n_x \text{ and for } j = 1, \ldots, n_z.$$

From this, it is clear that all A_i, for $i = 1, \ldots, n$, must be equal. If not, it is possible to divide $\{1, \ldots, n\}$ into two disjunct non-empty subsets ω_1 and ω_2 for which $\theta_{ij} = 0$ if $n_z + i \in \omega_1$ and $j \in \omega_2$ or vice versa. We could permute rows and columns of θ such that the resulting $\theta' = \Pi_r \theta \Pi_c$ has all rows i_1 for which $n_z + i_1 \in \omega_1$ above rows i_2 for which $n_z + i_2 \in \omega_2$, and all columns j_1 for which $j_1 \in \omega_1$ on the left of columns j_2 for which $j_2 \in \omega_2$. We then have

$$\begin{bmatrix} \Pi_c^T & 0 \\ 0 & \Pi_r \end{bmatrix} \begin{bmatrix} I \\ \theta \end{bmatrix} \Pi_c = \begin{bmatrix} I \\ \theta' \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \\ \cdot & 0 \\ 0 & \cdot \end{bmatrix}.$$

By proposition 4.6, this represents a separable CSS state, which we excluded from the beginning. An analogous proof holds for (4.13).

One can verify that (4.19) is equivalent to

$$(S_x^T \otimes I_\kappa)\tilde{B}(S_x \otimes I_\kappa) = 0.$$

Since $\tilde{B}$ is block diagonal, we can rewrite this as the linear constraints

$$\left(L_x^T \otimes I_\kappa\right) \begin{bmatrix} B_1 \\ \vdots \\ B_n \end{bmatrix} = 0,$$

where the n-bit columns of L_x are $(S_x)_i \odot (S_x)_j$ $\forall i, j : 1 \leq i \leq j \leq n_z$. These constraints are equivalent to (4.14). An analogous proof holds for (4.15). As the constraints (4.19)-(4.20) are independent, all solutions $\tilde{B}$ must be consistent with all solutions $\tilde{C}$. From (4.19)-(4.20), it follows that

$$\begin{aligned} (\theta \otimes I_\kappa)\tilde{B}_z\tilde{C}_z &= (\theta \otimes I_\kappa)\tilde{B}_z(\theta^T \otimes I_\kappa)\tilde{C}_x(\theta \otimes I_\kappa) \\ &= \tilde{B}_x\tilde{C}_x(\theta \otimes I_\kappa). \end{aligned}$$

In the same way as (4.12) follows from (4.17), this implies $B_1C_1 = \cdots = B_nC_n$.

If $B_iC_i = 0$, then $B_i = 0$ or $C_i = 0$. Indeed, given a particular i, suppose $B_i \neq 0$. Then $e_i \notin \mathrm{col}\,(L_x)$. Consequently, there exists some solution $u \in \mathbb{Z}_2^n$ to $L_x^T u = 0$ with $u_i = 1$. Note that $u \otimes I_\kappa$ is a solution to (4.14). It follows that $B_iC_i = I_\kappa C_i = 0$. From the simplecticity constraint (2.23) it then follows that $D_1 = A_1^{-T}$, and $A_1^T C_i$ and $A_1^{-1} B_i$ are symmetric, for $i = 1, \ldots, n$.

On the other hand, if $B_iC_i \neq 0$, for all i, then there exists a solution u to $L_x^T u = 0$ with $u_i = u_j = 1$, for any pair i, j. Indeed, there are solutions $u, v \in \mathbb{Z}_2^n$ with $u_i = 1$ and $v_j = 1$, so if $u_j = v_i = 0$, then $u + v$ is a solution with $(u+v)_i = (u+v)_j = 1$. Thus, for every i and j, we have a solution $\tilde{B}$ to (4.14) with $B_i = B_j = I_\kappa$ that must be consistent with all solutions $\tilde{C}$. Therefore, $C_1 = \cdots = C_n$. The same applies to the B_i.

It remains for us to prove that $B_iC_i \neq 0$ only if $\mathrm{col}\,(S_z) = \mathrm{col}\,(S_x)$. From $B_1 = \cdots = B_n$, it follows that the space $\mathrm{col}\,(L_x)$ consists of all vectors of even weight. It cannot contain any vector of odd weight, otherwise it would be the entire space $\mathbb{Z}_2^n$ and consequently $C_i = 0$. The same holds for $\mathrm{col}\,(L_z)$, and both spaces are equal. So all $(S_x)_i \odot (S_x)_j$ and $(S_z)_i \odot (S_z)_j$ must have even weight. With (4.11), it can be verified that this only holds if $\theta_i^T \theta_j = (\theta^T)_i^T (\theta^T)_j = \delta_{ij}$. This is equivalent to $\theta^T\theta = \theta\theta^T = I$. Consequently, $S_z = S_x\theta$ and $\mathrm{col}\,(S_z) = \mathrm{col}\,(S_x)$. □

In the following, we agree that S_z and S_x are of the form (4.11), except when θ is orthogonal. Then we take $S_x = S_z$ instead. Since such states are invariant under the simultaneous application of a local Hadamard on every qubit, they are occasionally referred to as *CSS-H states* [38].

Corollary 4.8 *(i) For orthogonal θ:*

$$R = \left[\begin{array}{cc} I_{n/2} \otimes D_1^T & I_{n/2} \otimes B_1^T \\ I_{n/2} \otimes C_1^T & I_{n/2} \otimes A_1^T \end{array}\right].$$

(ii) For non-orthogonal θ:

$$R = \left[\begin{array}{cc} I_{n_z} \otimes A_1^{-1} & \tilde{B}_z^T(\theta^T \otimes I_\kappa) \\ \tilde{C}_x^T(\theta \otimes I_\kappa) & I_{n_x} \otimes A_1^T \end{array}\right].$$

Proof: Case (i) follows from (4.4) and the simplecticity constraints (2.23). Case (ii) can be verified by substituting R in (4.16) and using (4.19)-(4.20). □

We show that all possible $\bar{R} \in \mathbb{Z}_2^{n\kappa \times n}$, of which the columns are columns $i, \kappa + i, \ldots, (n-1)\kappa + i$ of R, constitute a vector space $\mathcal{R}$, as claimed in section 4.3.1. For case (i), $\bar{R}$ is of the form

$$\left[\begin{array}{cc} I_{n/2} \otimes d & I_{n/2} \otimes b \\ I_{n/2} \otimes c & I_{n/2} \otimes a \end{array}\right], \tag{4.21}$$

where $a, b, c, d \in \mathbb{Z}_2^\kappa$. Clearly, the sum of matrices of this form is again of this form. For case (ii), one can verify that $\bar{R}$ is of the form

$$\left[\begin{array}{cc} I_{n_z} \otimes d & (\theta^T \otimes e) \odot (be^T) \\ (\theta \otimes e) \odot (ce^T) & I_{n_z} \otimes a \end{array}\right], \tag{4.22}$$

where $a, d \in \mathbb{Z}_2^\kappa$, the all-ones vectors e have the appropriate dimensions, and $b, c \in \mathbb{Z}_2^{n\kappa}$ satisfy:

$$\begin{array}{rcl} (L_{\theta^T}^T \otimes I_\kappa)b & = & 0, \\ (L_\theta^T \otimes I_\kappa)c & = & 0, \end{array}$$

so the b and c respectively constitute a vector space. Therefore, the sum of matrices of this form is again of this form. One could object that a, d are columns of invertible matrices, and therefore cannot be zero, but the effect of this possibility is negligible (cf. the remark in section 4.3.1).

Remark. This reasoning certainly holds when considering the measurement of only *one* copy. One can see that it also applies to *more* measurements, when interpreting the protocol as in the left of figure 4.2 (cf. section 4.3.1). After t measurements, we consider $\bar{R} \in \mathbb{Z}_2^{n(\kappa-t)\times n}$ for the next measurement. This is the same situation as in the beginning, but on $\kappa - t$ copies instead of κ, and the $\bar{R}$ still constitute a vector space. ◇

4.4.2 Information section

For the particular case of a CSS state, by theorem 2.22, measuring σ_z on every qubit of the state $|\psi_{S,b}\rangle$ yields b_i for $i = 1, \ldots, n_z$ and measuring σ_x on every qubit yields b_i for $i = n_z + 1, \ldots, n$. Although there are definitely other tripartitions $\mathcal{M}$ along which information on the state can be extracted, we will restrict ourselves to exclusively σ_z or σ_x measurements, since these are expected to have the highest information gain, i.e. n_z and n_x bits respectively. Note that these correspond to tripartitions $\mathcal{M}$:

$$\begin{aligned} &X = \emptyset, && Y = \emptyset, && Z = \{1, \ldots, n\}, \quad \text{or} \\ &X = \{1, \ldots, n\}, && Y = \emptyset, && Z = \emptyset, \end{aligned}$$

respectively. In the following, we reason in terms of σ_z measurements, but the same derivation also applies to σ_x measurements. We use the subscript 'z' and omit '$\mathcal{M}$'.

4.4.2.1 Non-orthogonal θ

Local σ_z measurements on one copy yield $\bar{R}_z \Delta\tilde{b}$, where $\bar{R}_z$ is the leftmost $n \times n_z$ part of $\bar{R}$. It is easily verified from (4.22) that

$$\bar{R}_z w = \begin{bmatrix} w \otimes d \\ (\theta w \otimes e) \odot c \end{bmatrix}.$$

If $(\bar{R}_z w)^T \Delta\tilde{b} = 0$ for all possible $\bar{R}_z$ and $w \in \mathcal{W}_{f_z}$, an $(n_z - f_z)$-dimensional subspace of $\mathbb{Z}_2^{n_z}$, then $d_z(\Delta\tilde{b}) \leq f_z$. As d and c are independent, this is equivalent to

$$\left(\begin{bmatrix} w & 0 \\ 0 & \theta w \odot v \end{bmatrix}^T \otimes I_\kappa \right) \Delta\tilde{b} = 0, \ \forall v \in \operatorname{col}(L_\theta)^\perp \text{ and } \forall w \in \mathcal{W}_{f_z}.$$

Defining matrices W_{f_z} and M_θ such that $\operatorname{col}(W_{f_z}) = \mathcal{W}_{f_z}$ and $\operatorname{col}(M_\theta) = \operatorname{col}(L_\theta)^\perp$, it follows that $\Delta\tilde{b} \in \mathcal{J}(\mathcal{W}_{f_z})^\perp \otimes \mathbb{Z}_2^\kappa$, where

$$\mathcal{J}(\mathcal{W}_{f_z}) = \operatorname{col}\left(\begin{bmatrix} W_{f_z} & 0 \\ 0 & \left(\theta W_{f_z} \otimes e^T\right) \odot \left(e^T \otimes M_\theta\right) \end{bmatrix} \right). \tag{4.23}$$

Note that $(U \otimes e^T) \odot (e^T \otimes V)$ for matrices U and V is just a tricky way to express the matrix with columns $U_i \odot V_j,\ \forall i,j$. Similarly, we have

$$\mathcal{J}(\mathcal{W}_{f_x}) = \operatorname{col}\left(\left[\begin{array}{cc} 0 & \left(\theta^T W_{f_x} \otimes e^T\right) \odot \left(e^T \otimes M_{\theta^T}\right) \\ W_{f_x} & 0 \end{array} \right] \right). \tag{4.24}$$

Now these spaces are defined, we can solve the LP problem (4.10) to find the yield of the protocol.

4.4.2.2 Orthogonal θ (CSS-H states)

For orthogonal θ, the derivation is much easier, and one can verify from (4.21) that

$$\mathcal{J}(\mathcal{W}_{f_z}) = \mathcal{W}_{f_z} \oplus \mathcal{W}_{f_z} \quad \text{and} \quad \mathcal{J}(\mathcal{W}_{f_x}) = \mathcal{W}_{f_x} \oplus \mathcal{W}_{f_x}. \tag{4.25}$$

Since the RHS of (4.10) is a function of $\mathcal{J}(\mathcal{W}_{f_z}) + \mathcal{J}(\mathcal{W}_{f_x})$, which remains the same if $\mathcal{W}_{f_z}$ and $\mathcal{W}_{f_x}$ are switched, it follows that there is always a solution to the LP problem for which $m_z = m_x$. Indeed, for each constraint there is a constraint with the same RHS and the coefficients of m_z and m_x switched. The constraints of the LP problem become

$$(f_z + f_x)\frac{m}{2} \geq \max_{\mathcal{W}^\perp_{f_z}, \mathcal{W}^\perp_{f_x}} G\left[\left(\mathcal{W}^\perp_{f_z} \cap \mathcal{W}^\perp_{f_x}\right) \oplus \left(\mathcal{W}^\perp_{f_z} \cap \mathcal{W}^\perp_{f_x}\right)\right],$$

where $\mathcal{W}^\perp_{f_z}$ and $\mathcal{W}^\perp_{f_x}$ are respectively f_z- and f_x-dimensional subspaces of $\mathbb{Z}_2^{n/2}$. From the definition of G in proposition 4.5, it is clear that $G(\mathcal{K}_1) \leq G(\mathcal{K}_2)$ if $\mathcal{K}_1 \subseteq \mathcal{K}_2$. Therefore, given f_z and f_x, the maximal $G(\cdot)$ will be reached for $\mathcal{W}^\perp_{f_z} \subseteq \mathcal{W}^\perp_{f_x}$ or $\mathcal{W}^\perp_{f_x} \subseteq \mathcal{W}^\perp_{f_z}$. Since the largest of the two spaces does not influence this maximum, only the constraints for which $f_z = f_x$ can be active. Consequently, the LP problem for CSS-H states becomes a simple equation, and the yield of the protocol is

$$1 - \max_{\mathcal{K} \neq \{0\}} \frac{G\left(\mathcal{K} \oplus \mathcal{K}\right)}{\dim(\mathcal{K})}, \tag{4.26}$$

where the maximum is taken over all nonzero subspaces $\mathcal{K}$ of $\mathbb{Z}_2^{n/2}$.

4.4.2.3 Constant elimination probability

To derive an upper bound for p_2 in section 4.3.2, firstly we calculated the probability that $\tilde{b}$ is not eliminated by the measurement on a single copy, which is $2^{-d(\mathcal{M},\Delta\tilde{b})}$. We then assumed the same probability for all next measurements of the same kind. However, for every next measurement, we cannot choose just any $\bar{R}$ from $\mathcal{R}$ because of the simplecticity constraints (2.23). Although (2.23) are constraints on the *columns* of A_1, B_1, C_i and D_i, for $i = 1, \ldots, n$, one can verify that from the simplecticity of $C^{-1} = PC^TP$, it follows that

$$C_i D_1^T \quad = \quad D_1 C_i^T,$$

$$\begin{aligned} B_i A_1^T &= A_1 B_i^T, \\ D_1 A_1^T + C_i B_i^T &= I_\kappa, \end{aligned} \tag{4.27}$$

which are constraints on the *rows* of A_1, B_1, C_i and D_i, and therefore on $\bar{R}$. This is the problem addressed in the remark in section 4.3.1.

Apparently, this reasoning only holds when taking the second of two equivalent views of the protocol, as explained in section 4.3.1 (i.e. the right of figure 4.2). Yet, the problem remains in the first view. Indeed, the measurement of one copy only partially reveals $\bar{R}\Delta\tilde{u}$ and the other part is lost.[4] It is possible that because of this loss, the resulting ${R'}^T\tilde{b}$, representing the state of the remaining $\kappa - 1$ copies, satisfies $d_z({R'}^T\tilde{b}) < d_z(\tilde{b})$ or $d_x({R'}^T\tilde{b}) < d_x(\tilde{b})$, where $R' \in \mathbb{Z}_2^{n\kappa \times n(\kappa-1)}$ is R without the columns of $\bar{R}$. As a consequence, the probability that $\tilde{b}$ is not eliminated after the protocol has ended, is larger than $2^{-\kappa[m_z d_z(\Delta\tilde{b}) + m_x d_x(\Delta\tilde{b})]}$.

But, when taking a closer look at the constraints (4.27), we can interpret them as restricting corresponding columns of either R_z or R_x. Indeed, we have for instance $(C_i^T)_j^T(D_1^T)_k = (C_i^T)_k^T(D_1^T)_j$, so for fixed $(C_i^T)_j$ and $(D_1^T)_j$, we can always randomly choose $(C_i^T)_k$ and choose $(D_1^T)_k$ such that the constraint holds, and vice versa. $(C_i^T)_k$ and $(D_1^T)_k$ are part of R_z and R_x respectively, and either one of those is relevant for the measurement under consideration. Therefore, the elimination probability *does* remain constant throughout the entire protocol.

4.4.3 Illustration with cat states and a CSS-H state

4.4.3.1 Cat state distillation

In section 2.5.5, we found that the cat state

$$\frac{1}{\sqrt{2}}(|00\ldots0\rangle + |11\ldots1\rangle)$$

is the CSS state that is represented by (2.34), which corresponds to non-orthogonal $\theta = e^T$, and $b = 0$. It is symmetric over all qubits and by applying local Hadamards on all qubits save one, we arrive at the two-colorable graph state of which the graph, for four qubits, is shown in figure 4.3.[5]

We find

$$\mathrm{col}\,(L_\theta) = \mathbb{Z}_2^1 \quad \text{and} \quad \mathrm{col}\,(L_{\theta^T}) = \{0_{n-1}\},$$

[4] This is a manifestation of the 'impossibility' of measuring those observables that do not commute with the ones already measured.

[5] Although this graph is *not* symmetric over all vertices, one can verify that local complementation at the central vertex, which is physically achieved by applying the inverse phase gate on this qubit and the phase gate on the others, results in the fully-connected graph, which is symmetric.

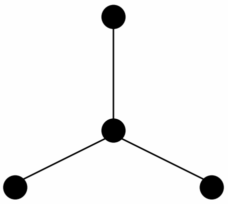


Figure 4.3: This graph represents the graph state we obtain by applying local Hadamards on all qubits of the four-qubit cat state except the one that corresponds to the central vertex.

where 0_{n-1} denotes the all-zeros vector of length $n-1$. Consequently, $M_\theta = 0$ and $M_{\theta^T} = I_{n-1}$. The linear constraints (4.14)-(4.15) become

$$\begin{aligned} B_1 + B_2 + \cdots + B_n = 0, \\ C_1 = C_2 = \cdots = C_n = 0, \end{aligned}$$

so the representation of local Clifford operations that satisfy (4.4) is of the form

$$\left[\begin{array}{cccc|cccc} A_1 & & & & B_1 & & & \\ & \ddots & & & & \ddots & & \\ & & A_1 & & & & B_{n-1} & \\ & & & A_1 & & & & \sum_{i=1}^{n-1} B_i \\ \hline & & & & A_1^{-T} & & & \\ & & & & & \ddots & & \\ & & & & & & A_1^{-T} & \\ & & & & & & & A_1^{-T} \end{array}\right] \tag{4.28}$$

and R is of the form

$$\left[\begin{array}{ccc|c} A_1^{-1} & & & B_1^T \\ & \ddots & & \vdots \\ & & A_1^{-1} & B_{n-1}^T \\ \hline & & & A_1^T \end{array}\right], \tag{4.29}$$

according to corollary 4.8.

We formulate the LP problem to calculate the yield of the protocol. Initially, the four parties share κ copies of the state

$$\rho = \sum_{b \in \mathbb{Z}_2^n} p(b) \left|\psi_b\right\rangle \left\langle\psi_b\right|,$$

where $|\psi_b\rangle = \frac{1}{\sqrt{2}}(|b_1\rangle \dots |b_{n-1}\rangle |0\rangle + (-1)^{b_n} |1+b_1\rangle \dots |1+b_{n-1}\rangle |1\rangle)$. Using (4.23), one can verify that

$$\mathcal{J}(\mathcal{W}_{f_z}) = \mathcal{W}_{f_z} \oplus \{0\}.$$

Since $n_x = 1$, f_x is either 0 or 1, for which the corresponding spaces can be only $\mathcal{W}_{f_x} = \mathbb{Z}_2$ and $\mathcal{W}_{f_x} = \{0\}$ respectively. From (4.24), it follows that $\mathcal{J}(\mathbb{Z}_2) = \mathbb{Z}_2^n$ and $\mathcal{J}(\{0\}) = \{0_n\}$. Consequently, if $f_x = 0$, then $\sum_{\mathcal{M}} \mathcal{J}[\mathcal{W}_f(\mathcal{M})] = \mathbb{Z}_2^n$, for all f_z, and the RHS of the corresponding inequalities in (4.10) is zero, yielding the inequality

$$m_z \geq 0. \tag{4.30}$$

On the other hand, if $f_x = 1$, then $\mathcal{J} = \sum_{\mathcal{M}} \mathcal{J}[\mathcal{W}_f(\mathcal{M})] = \mathcal{W}_{f_z} \oplus \{0\}$, and it follows that $\mathcal{J}^\perp = \mathcal{W}_{f_z}^\perp \oplus \mathbb{Z}_2$. The inequalities of (4.10) become

$$f_z m_z + m_x \;\geq\; \max_{\mathcal{W}_{f_z}^\perp} \left(- \sum_{b \in \mathbb{Z}_2^n} p(b) \log_2 \frac{p(b)}{p(b + \mathcal{W}_{f_z}^\perp \oplus \mathbb{Z}_2)} \right), \tag{4.31}$$

where the maximum is taken over all different f_z-dimensional subspaces $\mathcal{W}_{f_z}^\perp$ of $\mathbb{Z}_2^{n-1}$. For $f_z = 0$ and $f_z = n-1$, these inequalities can be simplified to

$$m_x \;\geq\; - \sum_{b \in \mathbb{Z}_2^n} p(b) \log_2 \frac{p(b)}{p(b) + p(b + e_n)}, \tag{4.32}$$

$$(n-1) m_z + m_x \;\geq\; - \sum_{b \in \mathbb{Z}_2^n} p(b) \log_2 p(b). \tag{4.33}$$

The LP problem is thus minimizing $m_z + m_x$ subject to the constraints (4.30), (4.31) for $f_z \neq 0$ or $n-1$, (4.32) and (4.33), and the yield is $1 - m_z - m_x$.

In comparison, protocols requiring that the fraction of measured copies exceeds the respective marginal entropies [3, 63], have yield

$$1 - \max_{j=1,\dots,n-1} [H(b_j)] - H(b_n). \tag{4.34}$$

When these marginal entropies are relaxed to conditional entropies [19], the yield is

$$\max\Big\{ \; 1 - \max_{j=1,\dots,n-1} [H(b_j)] - H(b_n | b_1, \dots, b_{n-1}), \tag{4.35}$$
$$1 - \max_{j=1,\dots,n-1} [H(b_j | b_n)] - H(b_n) \Big\}.$$

Let us take a typical situation for numerically verifying that our protocol has a higher yield than (4.34) or (4.35). We start with copies of the four-qubit cat state $|\psi_0\rangle$, prepared by one party. It does not matter which party this is, since the state is symmetric. Let us take the first. The second, third and fourth qubit of each copy is sent through identical depolarizing channels with fidelity F to the corresponding parties. The action of each channel is

$$\rho \quad \rightarrow \quad F\rho + \frac{1-F}{3}(\sigma_x \rho \sigma_x + \sigma_y \rho \sigma_y + \sigma_z \rho \sigma_z).$$

It can be verified that this yields a mixture with probabilities

$$\begin{bmatrix} p_{0000} \\ p_{0001} \\ p_{0010} \\ p_{0011} \\ p_{0100} \\ p_{0101} \\ p_{0110} \\ p_{0111} \\ p_{1000} \\ p_{1001} \\ p_{1010} \\ p_{1011} \\ p_{1100} \\ p_{1101} \\ p_{1110} \\ p_{1111} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 3 & 0 \\ 0 & 3 & 0 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 4 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \\ 0 & 1 & 2 & 1 \\ 0 & 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} F^3 \\ F^2 \frac{1-F}{3} \\ F\left(\frac{1-F}{3}\right)^2 \\ \left(\frac{1-F}{3}\right)^3 \end{bmatrix}.$$

Note that for calculating the RHS of constraint (4.31) for $f_z = 1$ or 2, we have to maximize over seven subspaces. The yield of our protocol for this example is plotted as a function of the fidelity F of the channels in figure 4.4, compared with (4.34) and (4.35).

Remark. One can observe that this yield is *lower* than the yield of bipartite hashing for Werner states (H in figure 3.1 on page 47). A just criticism to the given example would be that it is better to use the channels to distribute and distill ebits between the first party and all others, and subsequently teleporting the respective qubits of each copy from the first party to the others. However, in our treatment, we consider the yield to be a *state characteristic*, since we are dealing with *distillation*, i.e. the extraction of pure state copies by LOCC out of copies of a *given* mixed state (e.g. after distribution), and not necessarily the best way to organize the overall procedure of distribution and distillation of these copies. Surely, the latter is also of importance (for practical purposes), but would lead us into a completely different (and much more heuristic) analysis [59], and we decided not to pursue this track.

Still, one might suggest distilling *ebits* out of the copies of the given mixed state, and using those for teleportation. For this cat state example, the best option is measuring σ_x on the third and fourth qubit of a copy, yielding a Bell-diagonal state ρ_{12} between the first and second party. The same can be done between the first and the third party, and between the first and the fourth

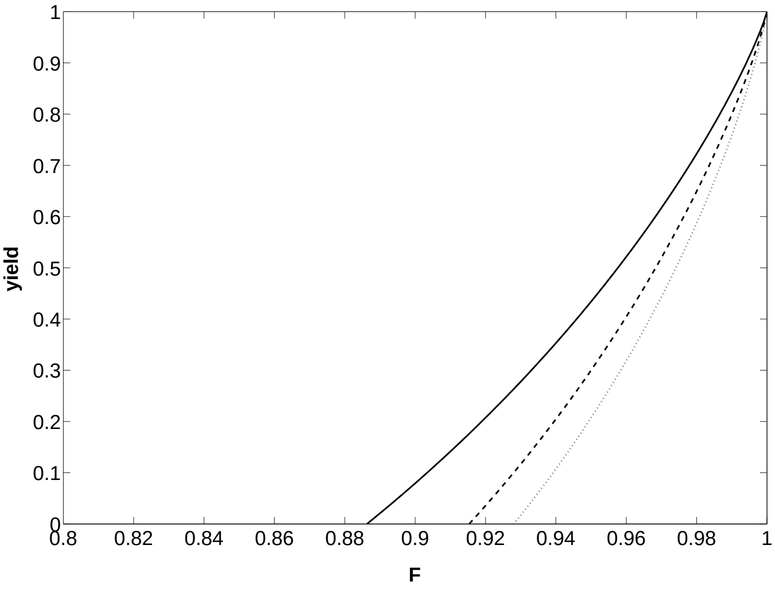


Figure 4.4: The yield of our protocol (solid line) for the given cat state example as a function of the channel fidelity F, compared with (4.34) (dotted line) and (4.35) (dashed line).

party, on other copies. The copies of ρ_{12}, ρ_{13} and ρ_{14} obtained in this way, are subsequently distilled into ebits. Then, the first party creates more pure cat state copies and teleports the respective qubits to the others. Note that the yield of this procedure is at most one third of the best bipartite protocol for ρ_{12}, since three ebits are needed for teleporting a pure cat state copy. Therefore, at least for $F > 0.9$, this suggestion is worse than direct distillation of the mixed state copies. ◇

Remark. One could wonder whether we actually gain anything by using local Clifford operations of the general form (4.28) instead of local Clifford operations built only of CNOTs. For the particular case of cat states (and cat states only), in fact, we *do not.* We prove this for the general n-qubit case. When using only CNOTs, the matrix R is, contrary to (4.29), of the form

$$\left[\begin{array}{ccc|c} A_1^{-1} & & & \\ & \ddots & & \\ & & A_1^{-1} & \\ \hline & & & A_1^T \end{array}\right].$$

It follows that for all f_z, the $\mathcal{J}(\mathcal{W}_{f_z})$ remain the same. For $f_x = 1$, we still have $\mathcal{J}(\mathcal{W}_{f_x}) = \{0_n\}$, but for $f_x = 0$, we now have $\mathcal{J}(\mathcal{W}_{f_x}) = \mathrm{col}\,(e_n)$ instead of $\mathbb{Z}_2^n$. As a consequence, the inequalities (4.30) are to be replaced by

$$f_z m_z \;\geq\; \max_{\mathcal{W}_{f_z}^\perp} \left(- \sum_{b \in \mathbb{Z}_2^n} p(b) \log_2 \frac{p(b)}{p(b + \mathcal{W}_{f_z}^\perp \oplus \{0\})} \right), \tag{4.36}$$

for $f_z = 1, \ldots, n-1$. The inequalities (4.31) remain the same. We now show that the inequalities (4.30)-(4.31) imply (4.36), such that the solution to the LP problem, and consequently the yield, is unaltered. To this end, it is a crucial observation from the geometry of the LP problem (shown in figure 4.5) that there is always a solution for which the inequality (4.32) is an *active* constraint.[6] By subtracting the RHS of (4.32) from the RHS of (4.31), we arrive at

$$f_z m_z \geq \max_{\mathcal{W}_{f_z}^\perp} \left(- \sum_{b \in \mathbb{Z}_2^n} p(b) \log_2 \frac{p(b) + p(b + e_n)}{p(b + \mathcal{W}_{f_z}^\perp \oplus \mathbb{Z}_2)} \right).$$

This inequality implies (4.36) if the argument of the maximum in the RHS is larger than the argument of the maximum in the RHS of (4.36), for all appropriate spaces $\mathcal{W}_{f_z}^\perp$. Subtracting the latter from the former gives

$$- \sum_{b \in \mathbb{Z}_2^n} p(b) \log_2 \frac{p(b + \mathcal{W}_{f_z}^\perp \oplus \{0\})}{p(b + \mathcal{W}_{f_z}^\perp \oplus \mathbb{Z}_2)} + \sum_{b \in \mathbb{Z}_2^n} p(b) \log_2 \frac{p(b)}{p(b) + p(b + e_n)} = T_1 - T_2.$$

[6] This means that the constraint is tight for this solution.

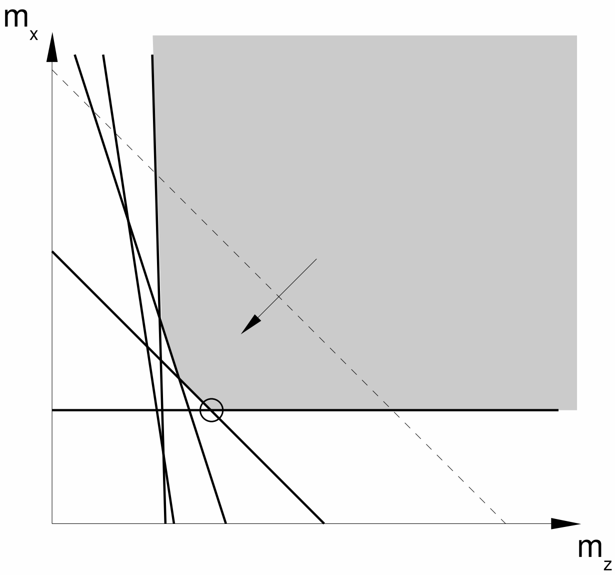


Figure 4.5: All constraints are of the form $f_z m_z + f_x m_x \geq *$, where $f_z \geq 0$ and $0 \leq f_x \leq 1$. Therefore, except (4.32), which has a horizontal boundary, there is no constraint of which the boundary (solid lines) has a slope higher than -1. Since the lines of constant objective function $m_z + m_x$ (dashed line) have slope -1, there must be a solution (circle) that lies on the boundary of (4.32).

We prove that $T_1 \geq T_2$. We define Ω_j, for $j = 1, \ldots, t$, as the cosets of $\mathcal{W}_{f_z}^{\perp} \oplus \{0\}$ in $\mathbb{Z}_2^{n-1} \oplus \{0\}$ and the functions $q(b) = p(b) + p(b + e_n)$ and $p'(b) = p(b)/q(b)$. We then have

$$\begin{aligned}
T_1 &= -\sum_{j=1}^{t} p(\Omega_j) \log_2 \frac{p(\Omega_j)}{q(\Omega_j)} + p(\Omega_j + e_n) \log_2 \frac{p(\Omega_j + e_n)}{q(\Omega_j)} \\
&= \sum_{j=1}^{t} q(\Omega_j)\Big(- p'(\Omega_j) \log_2 p'(\Omega_j) - [1 - p'(\Omega_j)] \log_2 [1 - p'(\Omega_j)]\Big) \\
&= \sum_{j=1}^{t} q(\Omega_j) h\Big(p'(\Omega_j), 1 - p'(\Omega_j)\Big) \\
&= \sum_{j=1}^{t} q(\Omega_j) h\left(\frac{\sum_{b\in\Omega_j} q(b)p'(b)}{\sum_{b\in\Omega_j} q(b)}, 1 - \frac{\sum_{b\in\Omega_j} q(b)p'(b)}{\sum_{b\in\Omega_j} q(b)}\right) \\
&\geq \sum_{j=1}^{t} q(\Omega_j) \frac{\sum_{b\in\Omega_j} q(b) h\Big(p'(b), 1 - p'(b)\Big)}{\sum_{b\in\Omega_j} q(b)} \\
&= \sum_{j=1}^{t} \sum_{b\in\Omega_j} q(b) h\Big(p'(b), 1 - p'(b)\Big) \\
&= \sum_{b\in\mathbb{Z}_2^{n-1}\oplus\{0\}} q(b)\Big(- p'(b) \log_2 p'(b) - [1 - p'(b)] \log_2 [1 - p'(b)]\Big) \\
&= -\sum_{b\in\mathbb{Z}_2^n} p(b) \log_2 p'(b) \; = \; T_2,
\end{aligned}$$

where $h(p, 1-p)$ is the binary entropy function, defined by (3.5). The inequality on the fifth line follows from the concavity[7] of $h(p, 1-p)$, which can be clearly seen in figure 3.5 on page 61.

Note that the above arguing only holds because the coefficient of m_x in the LP constraints can be only 0 or 1. One can verify that the cat state (and all states local Clifford equivalent with it) is the *only* fully entangled CSS state with this property. In the next section, we give an example of a state for which using the most general local Clifford operations does give a higher yield. ◇

4.4.3.2 CSS-H state distillation

Recall from section 4.4.1 that CSS-H states can be written in the form (4.11) with an orthogonal θ. These states only exist on an even number of qubits. The Bell states are an example of CSS-H states, and one can verify that the

[7] A function f is concave if $f(\sum_i p_i x_i) \geq \sum_i p_i f(x_i)$, $\forall x_i \in \mathrm{dom}(f)$ and $\forall p_i \geq 0$ that satisfy $\sum_i p_i = 1$.

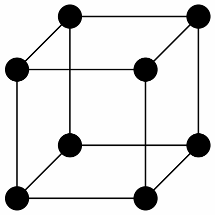


Figure 4.6: The graph corresponding to the cube state that is local Hadamard equivalent to the eight-qubit CSS-H state.

next CSS-H state that is not a tensor product of Bell states (and the only one on eight qubits), is the state corresponding to

$$\theta = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}.$$

This CSS-H state is local Hadamard equivalent to the graph state that corresponds to the graph of figure 4.6. For obvious reasons, we will refer to this graph state as the *cube state*.

We show that for this state it is better that all parties respectively perform the same random Clifford operation, as allowed by theorem 4.7, instead of a local Clifford operation built only of CNOTs. In the former case, the yield is given by (4.26). For the latter case, we have

$$\mathcal{J}(\mathcal{W}_{f_z}) = \mathcal{W}_{f_z} \oplus \{0_{n/2}\} \quad \text{and} \quad \mathcal{J}(\mathcal{W}_{f_x}) = \{0_{n/2}\} \oplus \mathcal{W}_{f_x},$$

instead of (4.25), and the LP problem is

$$\begin{array}{rl} \text{minimize} & m = m_z + m_x \\ \text{subject to} & f_z m_z + f_x m_x \geq \displaystyle\max_{\mathcal{W}^{\perp}_{f_z}, \mathcal{W}^{\perp}_{f_x}} G\left(\mathcal{W}^{\perp}_{f_z} \oplus \mathcal{W}^{\perp}_{f_x}\right), \text{ for all } f \not\equiv 0, \end{array}$$

where the maximum is respectively taken over all f_z- and f_x-dimensional subspaces $\mathcal{W}^{\perp}_{f_z}$ and $\mathcal{W}^{\perp}_{f_x}$ of $\mathbb{Z}_2^{n/2}$.

We consider the following situation. The first party prepares pure copies of the cube state, and sends qubits $2, \ldots, 8$, of each copy to parties $2, \ldots, 8$, respectively. The qubits are sent via identical quantum channels that cause a phase flip with probability $1 - F$. The action of each channel is

$$\rho \quad \rightarrow \quad F\rho + (1-F)\sigma_z \rho \sigma_z.$$

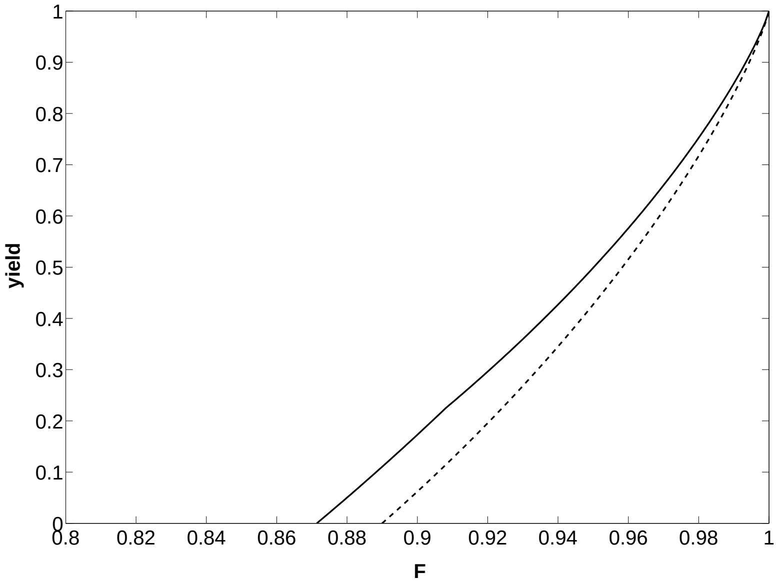


Figure 4.7: The yield of our protocol (solid line) for the given cube state example as a function of the channel fidelity F, compared with the yield when using only CNOTs (dashed line).

Equivalently, the parties share the eight-qubit CSS-H state that has undergone σ_z on the qubits of which the vertices are not connected to the first, and σ_x on the other qubits, each with probability $1 - F$. The yield of the distillation protocol, compared to the yield when using only CNOTs, is plotted in figure 4.7.

4.5 Graph state breeding

In this section, we consider the distillation of arbitrary graph states. Recall from section 2.5.5 that any stabilizer state is local Clifford equivalent to some graph state. Therefore, our protocol is applicable to copies of any stabilizer state. Contrary to CSS states, the general structure of an arbitrary graph state does not allow for much freedom in the choice of local Clifford operations satisfying (4.4). Therefore, in the operation section 4.5.1, we will not so much be concerned with the derivation of their most general structure, as this will be rather simple compared to the CSS state case, but with ways for guaranteeing the randomness of the parity checks, despite the severe constraints on the structure of the local Clifford operations. For this reason, we have chosen a breeding-like approach instead of hashing. The information section 4.5.2, on the other hand, is straightforward. In section 4.5.3, we illustrate this for the five-qubit ring state.

4.5.1 Operation section

We consider κ copies of a graph state, corresponding to the graph with adjacency matrix $\Gamma \in \mathbb{Z}_2^{n\times n}$. According to (2.24) and (2.32), the copies are represented by

$$\begin{bmatrix} \Gamma \\ I_n \end{bmatrix} \otimes I_\kappa.$$

Since arbitrary graph states include CSS states, up to local Hadamards, the local Clifford operations satisfying (4.4) will certainly also satisfy the constraints of theorem 4.7. Therefore, it holds that

$$\begin{array}{ccccccc} A_1 & = & \cdots & = & A_n, & & \\ B_1 & = & \cdots & = & B_n & = & 0, \\ C_1 & = & \cdots & = & C_n & = & 0, \\ D_1 & = & \cdots & = & D_n & = & A_1^{-T}. \end{array}$$

The generator matrix of the copies is transformed into

$$\begin{bmatrix} \Gamma \otimes A_1 \\ I_n \otimes A_1^{-T} \end{bmatrix}.$$

From (4.4), it follows that $(\Gamma \otimes A_1)R = \Gamma \otimes I_\kappa$ and $(I_n \otimes A_1^{-T})R = I_n \otimes I_\kappa$. Consequently, A_1 must be orthogonal and $R = I_n \otimes A_1^T$.

Remark. For particular graph states, more general local Clifford operations may exist that satisfy (4.4). However, even if so, it is unlikely that they give rise to better yields (cf. the cat state case). ◇

Remark. Note that it is no longer necessary to restrict ourselves to graph states. Indeed, for an arbitrary stabilizer state with stabilizer matrix S, we have

$$C(S \otimes I_\kappa)R = \begin{bmatrix} I_n \otimes A_1 & \\ & I_n \otimes A_1 \end{bmatrix} \begin{bmatrix} S_z \otimes I_\kappa \\ S_x \otimes I_\kappa \end{bmatrix} (I_n \otimes A_1^T) = S \otimes I_\kappa.$$

◇

The orthogonality of A_1 puts a severe constraint on its rows and columns, jeopardizing the possibility of creating true randomness in the parity checks that constitute the distillation protocol (cf. the remark in section 4.3.1). This problem can be circumvented by demanding that only a part of the columns of A_1 need to be random. This is the case when we use a breeding-like approach, where no information is to be extracted from appended (predistilled) pure state copies, on which the measurements are performed. Let Q be a random full rank matrix in $\mathbb{Z}_2^{\kappa\times m\kappa}$. We prove that one can always construct an orthogonal $A_1 \in \mathbb{Z}_2^{(1+m)\kappa\times(1+m)\kappa}$ with lower left part ${Q'}^T$, where Q' is either equal to Q or equivalent to Q for the protocol. To this end, we need the following theorems:

Theorem 4.9 *Any symmetric matrix $W \in \mathbb{Z}_2^{n\times n}$ of rank r can be factorized as follows:*

$$W = RDR^T,$$

where R is invertible and

(i) $D = \begin{bmatrix} I_{r/2} \otimes \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & \\ & 0 \end{bmatrix}$ *if* $\operatorname{diag}(W) = 0$,

(ii) $D = \begin{bmatrix} I_r & \\ & 0 \end{bmatrix}$ *if* $\operatorname{diag}(W) \neq 0$.

Proof: We prove that if the theorem is true for all $n \leq N$, it also holds for $n = N+1$. Note that the theorem is trivial for zero matrices, as $0 = R0R^T$, and matrices of zero dimension.

(i) Without loss of generality, we may consider (nonzero) $W \in \mathbb{Z}_2^{(N+1)\times(N+1)}$ of the following form:

$$W = \begin{bmatrix} 0 & 1 & a^T \\ 1 & 0 & b^T \\ a & b & W_2 \end{bmatrix},$$

where a, b, W_2 have appropriate dimensions and $\operatorname{diag}(W_2) = 0$. Indeed, note that identical permutations of rows and columns of W are for free, as they can be absorbed into R as follows:

$$\Pi W \Pi^T = RDR^T \Rightarrow W = (\Pi^T R) D (\Pi^T R)^T.$$

Since $W_2 + ab^T + ba^T$ has zero diagonal and is an $(N-1)\times(N-1)$ matrix, we can write

$$W_2 + ab^T + ba^T = R_2 D_2 R_2^T.$$

It follows that

$$W = RDR^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ b & a & R_2 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & D_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & b^T \\ 0 & 1 & a^T \\ 0 & 0 & R_2^T \end{bmatrix}.$$

By construction, R is invertible because so is R_2.

(ii) Again, without loss of generality, we may consider W of the form:

$$W = \begin{bmatrix} 1 & a^T \\ a & W_2 \end{bmatrix}.$$

We can write

$$W_2 + aa^T = R_2 D_2 R_2^T,$$

where D_2 is either (i) or (ii). It follows that

$$W = RDR^T = \begin{bmatrix} 1 & 0 \\ a & R_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & D_2 \end{bmatrix} \begin{bmatrix} 1 & a^T \\ 0 & R_2^T \end{bmatrix}.$$

If D_2 is of the form (i), it can be brought to (ii), by using the identity

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = VV^T, \quad \text{with } V = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

□

Corollary 4.10 *If and only if a given symmetric matrix $W \in \mathbb{Z}_2^{n\times n}$ is not both full rank and zero-diagonal, we can find square M such that $W = M^T M$.*

Proof: Using theorem 4.9, we have $W = RDR^T$. If D is of the form (ii), we take M equal to R^T with the rightmost $n - r$ columns set to zero. If D is of the form (i) and not full rank, we can find U such that $D = U^T U$, by using the identity

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = V^T V, \quad \text{with } V = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Then take M equal to UR^T with the rightmost $n - r$ columns set to zero.

Finally, we show that if W is full rank and zero-diagonal, there is no M satisfying $M^T M = W$. An equivalent statement is that there exists no square M such that

$$M^T M = D = I \otimes \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Suppose there was. As $x^T D x = 0$ for all x, $M^T M = D$ implies that $y^T y = 0$ for all $y = Mx$. Consequently, M cannot be full rank. But then $M^T M = D$ cannot be true, as D is full rank. □

Theorem 4.11 *A matrix $W \in \mathbb{Z}_2^{n\times r}$ can be extended to an orthogonal matrix $\bar{W} \in \mathbb{Z}_2^{n\times n}$ by adding columns, if and only if*

- $W^T W = I_r$,

- $e \notin \mathrm{col}\,(W)$.

Proof: Define a full rank matrix $Y \in \mathbb{Z}_2^{n\times(n-r)}$ such that $W^TY = 0$. By theorem 4.9, we can find R and D such that $Y^TY = RDR^T$. For now, we assume that Y^TY is full rank. As $e \notin \mathrm{col}\,(W)$, we know that $D = I_{n-r}$. Otherwise $\mathrm{diag}(Y^TY) = 0 \Rightarrow Y_i^TY_i = 0,\ \forall i$. Equivalently, we have $y_i^Te = 0,\ \forall i \Rightarrow Y^Te = 0$, which contradicts $e \notin \mathrm{col}\,(W)$. Let $Z = YR^{-T}$, then $\bar{W} = [W\ Z]$ is orthogonal. Indeed, $Z^TW = R^{-1}Y^TW = 0$ and $Z^TZ = R^{-1}Y^TYR^{-T} = D = I$.

It remains for us to prove that Y^TY is full rank. If not, then there exists some $x \neq 0$ that satisfies $Y^TYx = 0$. By definition of Y, it follows that $Yx \in \mathrm{col}\,(W) \Rightarrow Yx = Wz$ for some $z \neq 0$. But then $W^TWz = W^TYx = 0$, which contradicts $W^TW = I$. □

We now show, for a given full rank $\kappa \times m\kappa$ matrix Q, how to construct an orthogonal $(1+m)\kappa \times (1+m)\kappa$ matrix A_1 with lower left part equal to ${Q'}^T$, where Q' is either equal to Q or equivalent to Q for the protocol. We apply the following procedure:

1) find square matrix M such that $M^TM = I + Q^TQ$;

2) create orthogonal A_1^T by adding columns to the left of
$$W = \begin{bmatrix} Q \\ M \end{bmatrix}.$$

By corollary 4.10, step (1) is possible provided that $I + Q^TQ$ is not both full rank and zero diagonal. In that case, this can be solved by adding just one column to Q, as the resulting matrix Q' then has an odd number of columns. Consequently, we will have one extra copy measured, but as κ is large, this will not influence the yield.

By theorem 4.11, step (2) is possible provided that $e \notin \mathrm{col}\,(W)$. If $e \in \mathrm{col}\,(W)$, then there exists some $x \neq 0$ that satisfies $Qx = e$ and $Mx = e$. Without loss of generality, we may assume that $x_1 = 1$. We define $Q' = Q(I + e_2e_1^T)$, and repeat step (1) yielding M'. Now e will be no longer in $\mathrm{col}\,(W')$. This is shown as follows. Suppose $e \in \mathrm{col}\,(W')$, then there exists some y satisfying $Q'y = e$ and $M'y = e$. From the definition of Q' and the fact that Q is full rank, we have $y = (I + e_2e_1^T)x$. Consequently, $y^Ty = x^Tx + x_1 \neq x^Tx$. However, this is contradicted by

$$\begin{aligned} x^Tx + y^Ty &= x^TIx + y^TIy \\ &= x^T(M^TM + Q^TQ)x + y^T({M'}^TM' + {Q'}^TQ')y \\ &= (x^TM^TMx + y^T{M'}^TM'y) + (x^TQ^TQx + y^T{Q'}^TQ'y) \\ &= (e^Te + e^Te) + (e^Te + e^Te) \ = \ 0. \end{aligned}$$

Therefore, $e \notin \mathrm{col}\,(W')$. Finally, note that $Q' = Q(I + e_2e_1^T)$ is equivalent to Q for the protocol, as $\mathrm{col}\,(Q) = \mathrm{col}\,(Q')$.

4.5.2 Information section

Since we are considering a breeding-like approach, the given copies, in state $|\psi_{\bar{S},\tilde{u}}\rangle$, are appended with copies of the pure state $|\psi_{S,0}\rangle$. Let q be the first column of Q, which is distributed uniformly over $\mathbb{Z}_2^\kappa$. The local Clifford operations, represented by $I_n \otimes A_1$, transform the state of the first appended copy into $\left|\psi_{S,\bar{R}\tilde{u}}\right\rangle$, where $\bar{R} = I_n \otimes q$. Using similar derivations as in section 4.4.2, we find that

$$\mathcal{J}[\mathcal{W}_f(\mathcal{M})] = \mathcal{W}_f(\mathcal{M}).$$

With this, the LP problem (4.10) becomes

$$\begin{aligned} &\text{minimize} && m = \sum_{\mathcal{M}} m(\mathcal{M}) \\ &\text{subject to} && \sum_{\mathcal{M}} m(\mathcal{M}) f(\mathcal{M}) \geq \max_{\mathcal{W}_f} G\left(\left[\sum_{\mathcal{M}} \mathcal{W}_f(\mathcal{M})\right]^{\perp}\right), \text{ for all } f \not\equiv 0, \end{aligned} \tag{4.37}$$

where, given the function $f : \mathcal{M} \to f(\mathcal{M})$, the maximum is taken over all combinations of subspaces $\mathcal{W}_f(\mathcal{M})$ of dimension $n(\mathcal{M}) - f(\mathcal{M})$ of $\mathcal{V}(\mathcal{M})$.

4.5.3 Illustration with the five-qubit ring state

Recall that the five-qubit ring state (cf. page 36) is a three-colorable graph state that is not local Clifford equivalent to some CSS state. As such, it is not covered by the protocols in the previous sections. We calculate the yield for the following mixture:

$$\rho = p_0 \left|\psi_0\right\rangle \left\langle\psi_0\right| + \frac{1-p_0}{2^5-1} \sum_{b \in \mathbb{Z}_2^5 \setminus \{0\}} \left|\psi_b\right\rangle \left\langle\psi_b\right|,$$

where $|\psi_b\rangle$ is the stabilizer state represented by S and b, and

$$S = \begin{bmatrix} \Gamma \\ I_5 \end{bmatrix}, \text{ with } \Gamma = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}.$$

We will only consider the following tripartitions:

$$\begin{aligned} &\mathcal{M}_1: && Z = \{3,4,5\}, && X = \{1,2\}, && Y = \emptyset && \Rightarrow && \mathcal{V}(\mathcal{M}_1) = \operatorname{col}([e_1\ e_2]); \\ &\mathcal{M}_2: && Z = \{1,4,5\}, && X = \{2,3\}, && Y = \emptyset && \Rightarrow && \mathcal{V}(\mathcal{M}_2) = \operatorname{col}([e_2\ e_3]); \\ &\mathcal{M}_3: && Z = \{1,2,5\}, && X = \{3,4\}, && Y = \emptyset && \Rightarrow && \mathcal{V}(\mathcal{M}_3) = \operatorname{col}([e_3\ e_4]); \\ &\mathcal{M}_4: && Z = \{1,2,3\}, && X = \{4,5\}, && Y = \emptyset && \Rightarrow && \mathcal{V}(\mathcal{M}_4) = \operatorname{col}([e_4\ e_5]); \\ &\mathcal{M}_5: && Z = \{2,3,4\}, && X = \{1,5\}, && Y = \emptyset && \Rightarrow && \mathcal{V}(\mathcal{M}_5) = \operatorname{col}([e_5\ e_1]). \end{aligned}$$

By restricting to these tripartitions, we risk finding a suboptimal solution. However, we will show below that the solution found is, in fact, optimal.

Intuitively, $G(\mathcal{J}^{\perp})$ grows larger in line with the size of $\mathcal{J}^{\perp}$. Since for this example all $b \neq 0$ have the same probability, $G(\mathcal{J}^{\perp})$ is a function of $|\mathcal{J}^{\perp}| = 2^{\dim(\mathcal{J}^{\perp})}$ only. Let $d = \dim(\mathcal{J}^{\perp})$, then it can be verified that

$$\begin{aligned} G(\mathcal{J}^{\perp}) \;=\; g(d) \;&=\; -p_0 \log_2 \tfrac{31p_0}{31p_0+(2^d-1)(1-p_0)} + \tfrac{32-2^d}{31}(1-p_0)d \\ &\quad -\tfrac{2^d-1}{31}(1-p_0)\log_2 \tfrac{1-p_0}{31p_0+(2^d-1)(1-p_0)}. \end{aligned}$$

Note that $g(5) = S(\rho) =$ the total entropy of the state. For symmetry reasons, we may assume that the optimal $m(\mathcal{M}_i)$ will be equal for all $\mathcal{M}_i$, thus $m(\mathcal{M}_i) = m/5$, and the inequalities of the LP problem become

$$\left(10 - \sum_{i=1}^{5} \dim[\mathcal{W}_f(\mathcal{M}_i)]\right) \frac{m}{5} \;\geq\; g\left(5 - \dim\left[\sum_{i=1}^{5} \mathcal{W}_f(\mathcal{M}_i)\right]\right).$$

For a fixed LHS, the RHS is maximal and consequently yields an active constraint when the spaces $\mathcal{W}_f(\mathcal{M}_i)$ overlap as much as possible. Therefore, for $f(\mathcal{M}_1) = 1$, we do not consider $\mathcal{W}_f(\mathcal{M}_1) = \mathrm{col}\,(e_1 + e_2)$ as this will definitely decrease the RHS if e_1 or e_2 were not already in $\sum_i \mathcal{W}_f(\mathcal{M}_i)$, and similarly for $i = 2, \ldots, 5$. Thus $\sum_i \mathcal{W}_f(\mathcal{M}_i)$ is generated by e_i, for all i in some subset of $\{1, \ldots, 5\}$. For a fixed RHS, the inequality is an active constraint when the coefficient of m in the LHS is minimal, which is the case if $\dim[\mathcal{W}_f(\mathcal{M}_i)]$ are as large as possible, for all i. One can see that in that case $\sum_i \dim[\mathcal{W}_f(\mathcal{M}_i)] = 2\dim\left[\sum_i \mathcal{W}_f(\mathcal{M}_i)\right]$. It follows that the yield of the protocol equals

$$= 1 - \frac{5}{2} \max_{d} \frac{g(d)}{d}. \tag{4.38}$$

Numerical calculation shows that the maximum is found for $d = 5$. Consequently, (4.38) equals $1 - \frac{S(\rho)}{2}$. We could not have done better. Indeed, it can be verified that there is no tripartition $\mathcal{M}$ for which $n(\mathcal{M}) > 2$, and a measurement according to the tripartition $\mathcal{M}$ yields at most $n(\mathcal{M})$ bits of information. We have plotted the yield of the protocol as a function of p_0 in figure 4.8.

4.6 Conclusion

We have described a generalization of the hashing/breeding protocol from bipartite to multipartite. For particular classes of stabilizer states, i.e. CSS-H states, CSS states and arbitrary graph states respectively, we have derived the most general structure of local Clifford operations that transform the set of tensor products of stabilizer states with generator matrix S onto itself (operation section). This is important as it allows us to increase the randomness in the parity checks performed on the overall state of the copies. Consequently, the information gain of each measurement is increased, such that we need less measurements to purify the overall state. In order to calculate the exact number of copies that need to be measured to this end, we made use of properties of

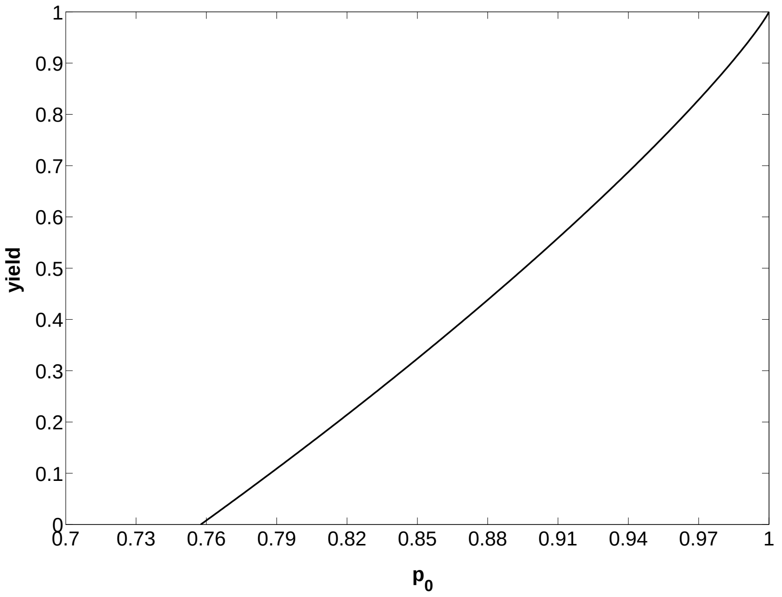


Figure 4.8: The yield of the protocol for the state $\rho = p_0 \left|\psi_0\right\rangle \left\langle\psi_0\right| + \frac{1-p_0}{31} \sum\limits_{b\neq 0} \left|\psi_b\right\rangle \left\langle\psi_b\right|$ as a function of p_0.

the information-theoretical concept strongly typical set (information section). Both operation section and information section enabled us to improve the yields of existing generalizations of hashing/breeding.

Chapter 5

Conclusion

In this thesis, we have treated the development of entanglement distillation protocols within the framework of the stabilizer formalism. After introducing the concept of quantum entanglement and entanglement distillation, we have shown how the entire setup of the distillation protocols under consideration is transparently described in terms of binary matrix algebra (the 'binary picture' of the stabilizer formalism). This description enabled us to improve existing distillation protocols, in a bipartite setting as well as in a multipartite setting.

Our main contribution for the bipartite case was the recognition of the underlying principle that explains why local measurements involving a finite and preferably small number of qubit pairs result in an extra reduction of the state entropy, next to the decrease of entropy that equals the information extracted by the measurements. Using this insight, we devised adaptive variants of existing asymptotic protocols and improved the best yield of all protocols for Bell-diagonal states, which serves as a lower bound for the important entanglement measure entanglement of distillation. This yield has not been surpassed at the time of writing.

For multipartite stabilizer states, we improved existing protocols on two levels. Firstly, for particular classes of stabilizer states (i.e. CSS-H states, CSS states and graph states[1]), we derived the general structure of the binary representation of local Clifford operations that transform the set of tensor products of such pure states onto itself. As such, applying these most general local Clifford operations results in a higher statistical dependence between the input copies of the protocol, increasing the information that can be retrieved from local measurements. Secondly, we generalized the idea of bipartite hashing where the initial state of the input copies is regarded as an unknown pure state that is contained in the typical set. We gave a method to calculate the minimal number of measurements necessary to eliminate all other states from this (strongly) typical set as the solution of a linear programming problem. We have shown, for various examples, that on both levels the yield of the protocols

[1] and all local unitary equivalent states

is increased.

A number of interesting open problems remain. We summarize a few possible directions from where our work has ended, of practical and of more fundamental nature.

- A next and important step towards entanglement distillation in a realistic setting, is taking *noise in the local operations* of the protocol into account. Some attempts in that direction have already been made (for an overview, we refer to [30]), yet many of them are rather ad hoc and encompass a heuristic numerical optimization. Therefore, we have chosen not to deal with it here.

  For theoretical purposes, this step is of minor importance, because when local noise is incorporated, we are no longer dealing with a *state characteristic* and as such, the yield of the resulting protocol is no longer representative for the entanglement that is present in a given state. However, if some fundamental threshold concerning this local noise is found, it could give a clue towards the possibility of realizing practical applications.

- In the multipartite setting, we could look for optimal ways of *obtaining* copies of a desired pure state. Such ways are studied in [59] and compared with one another with respect to the *quantum communication cost*, which is the total number of qubits that have been sent through a quantum channel during the entire procedure. As such, all methods can be compared with one another (e.g. bipartite with multipartite distillation) on an equal footing and the yield of the method is now a *channel characteristic.*

- For our protocols, we only considered local operations that map the set of tensor products of a given stabilizer state to itself. As already mentioned in footnote in chapter 4, the state of the copies to be measured does not necessarily have to be of this kind, but could for instance be a tensor product of different states, or even be entangled over the copies. It remains unclear whether this could increase the yield of the protocol. Because of this and for the sake of manageability, we decided not to delve further into this issue.

- One could generalize the idea of entropy reduction to multipartite distillation. However, protocols that operate in this way will very likely no longer be as clear-cut as for the bipartite case, where the original hashing protocol was much simpler than its multipartite generalizations. We have explored this path to some extent, but were unable to produce any satisfying results.

- Investigating the LP problem for finding the multipartite hashing yield could give a more fundamental insight into multipartite entanglement manipulation. It resembles the problem of calculating the entanglement measure asymptotic relative entropy of entanglement, which is a distance measure to the convex set of all separable (or PPT) states: the LP problem is also convex and involves particular relative entropies.

- For bipartite protocols, we saw that local measurements decrease the overall input state entropy by two mechanisms. On the one hand, information on the state is extracted from the outcomes of the measurements (*information gain* IG), on the other hand, the local collapse of the state vector results in an extra *entropy reduction* ER. Doing so, we unified asymptotic protocols and finite protocols in a single interpretation, and we gave methods to increase IG+ER.

  What could be asked next is whether there exists a *fundamental upper bound* for IG+ER. Clearly, both IG and ER are smaller than 1 for a single parity check. Consequently, an absolute upper bound for the yield of bipartite protocols is $1-S(\rho)/2$, but this is a meaningless upper bound since it exceeds the asymptotic relative entropy of entanglement [5], which in its turn is an upper bound for the entanglement of distillation (cf. page 47). In our treatment of finding optimal protocols, we encountered a sort of trade-off between IG and ER. Indeed, on the one hand, maximal IG is guaranteed for parity checks involving an infinite number of qubit pairs, but then we have zero ER, which gives rise to the hashing yield $1-S(\rho)$. On the other hand, attempts to increase ER are often accompanied by a decrease of IG.

  Somehow, this trade-off has its origin in the *linear nature* of the parity checks. Indeed, for a parity check $r^T\tilde{u}$, only those $\tilde{b}$ that correspond to the observed value are kept (IG), and each vector $\tilde{b}$ 'collides' with $\tilde{b}+Pr$, giving rise to ER. The magnitude of IG and ER are both dependent on the choice of r. Yet, investigating methods for maximizing IG+ER or even finding a useful upper bound is a difficult task, since they both are expectation values of nonlinear entropy-like functions. Furthermore, one actually needs to take *all* parity checks into account and maximize (or upper bound) the *total* IG+ER. But then adaptiveness comes into play and we are left with an optimization problem in a very large decision space.

  Finding better lower bounds for the entanglement of distillation in this way boils down to a mathematical hard problem. Furthermore, since we restrict ourselves to local Clifford operations only, it is not clear in how far the best yield of such protocols is a tight lower bound for the entanglement of distillation (of Bell-diagonal states). Nevertheless, we believe the behavior of IG and ER and their connection *does* give a fundamental insight into the nature of entanglement distillation.

Appendix A

Elementary rules of quantum mechanics

We touch on some basics of quantum mechanics. This appendix is not intended as an introduction to quantum mechanics, but should be seen as a concise summary of notations and conventions in this thesis, as those are not always commonplace in this vast area. We have mainly based ourselves on the excellent standard texts [68, 72] on quantum information theory.

We adopt the *Dirac notation*, using a 'ket' $|\cdot\rangle$ to denote a pure state vector and a 'bra' $\langle\cdot|$ for its dual. The advantage of this notation is the independence of any coordinate basis. We only consider finite-dimensional systems. A *pure state* $|\psi\rangle$ of an n-qubit system is a unit vector in a 2^n-dimensional Hilbert space $\mathcal{H}_n = \mathbb{C}^{2^n}$. For a single qubit, it is mostly agreed that $|0\rangle$ and $|1\rangle$ denote the eigenstates of the Pauli matrix σ_z (defined on page 14). In matrix notation, this reads

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

They form a basis for $\mathcal{H}_1$. Any *superposition* $a\,|0\rangle + b\,|1\rangle$, where $a, b \in \mathbb{C}$ and $|a|^2 + |b|^2 = 1$ is a valid pure state vector. Tensor products of $|0\rangle$ and $|1\rangle$ are often abbreviated, e.g. $|010010\rangle$ is short for $|0\rangle \otimes |1\rangle \otimes |0\rangle \otimes |0\rangle \otimes |1\rangle \otimes |0\rangle$.

The time evolution of a closed quantum system is governed by the *Schrödinger equation*

$$i\hbar \frac{d\,|\psi\rangle}{dt} = H\,|\psi\rangle\,,$$

where $\hbar$ is known as *Planck's constant* and the *Hamiltonian* of the system H is a Hermitian operator: $H = H^\dagger$. Integrated over time, $|\psi\rangle$ is transformed into $|\psi'\rangle$, and the Schrödinger equation takes the form

$$|\psi'\rangle = U\,|\psi\rangle\,,$$

where U is a *unitary* operator: $UU^\dagger = I$.

A *projective measurement* is identified with a Hermitian operator M, called *observable*, with spectral decomposition

$$M = \sum_i \lambda_i P_i,$$

where P_i is the projector onto the eigenspace of M with eigenvalue λ_i. The indices i refer to the possible measurement outcomes. If the state immediately before the measurement is $|\psi\rangle$, then the probability that outcome i occurs equals

$$p_i = \langle\psi| P_i |\psi\rangle \,,$$

and the state immediately after the measurement with this outcome is

$$\frac{P_i |\psi\rangle}{\sqrt{p_i}}.$$

Note that $(\sum_i P_i)(\sum_j P_j) = \sum_i \sum_j \delta_{ij} P_i = \sum_i P_i \Rightarrow \sum_i P_i = I$ such that $\sum_i p_i = 1$ is ensured.

Example. Measuring the observable σ_z on a qubit boils down to projecting its state $|\psi\rangle$ onto $|0\rangle$ or $|1\rangle$, as $\sigma_z = |0\rangle\langle 0| - |1\rangle\langle 1|$. The probabilities of the outcomes are $p_0 = |\langle\psi|\,0\rangle|^2$ and $p_1 = |\langle\psi|\,1\rangle|^2$. ◇

When we do not have maximal knowledge on a system, the concept of a *mixed state* is introduced. It is described by a positive semi-definite Hermitian *density operator* ρ with trace one: $\mathrm{Tr}\,\{\rho\} = 1$. When written in the form

$$\rho = \sum_i p_i |\psi_i\rangle \langle\psi_i| \,,$$

we can regard the mixed state as a *statistical ensemble* of pure states $\{p_i, |\psi_i\rangle\}$. Given ρ, such ensembles are not unique but physically indistinguishable from one another. Note that for a pure state $\rho = |\psi\rangle\langle\psi|$ we have $\mathrm{Tr}\left\{\rho^2\right\} = 1$.

The evolution postulate now reads

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

The probability of outcome i when measuring the observable M defined above now equals

$$p_i = \mathrm{Tr}\,\{P_i\rho\} = \mathrm{Tr}\,\{P_i\rho P_i\} \,,$$

and the state after the measurement with this outcome is

$$\frac{P_i\rho P_i}{p_i}.$$

A mixed state appears naturally as the state of a subsystem of a composite quantum system (that is possibly in a pure state). Let ρ_{AB} be the state of the composite system AB, then the state of subsystem A is given by the *reduced density matrix*

$$\rho_A = \mathrm{Tr}_B \left\{\rho_{AB}\right\},$$

where 'Tr_B' stands for the *partial trace* over subsystem B. This definition ensures that if there is no information on subsystem B (e.g. retrieved by measurements), the behavior of subsystem A is completely determined by ρ_A. In quantum information theory, B often represents the unobservable and uncontrollable *environment* of A. The most general evolution of a system A is given by

$$\rho_A \to \rho'_A = \mathrm{Tr}_B \left\{U\rho_{AB}U^\dagger\right\} = \sum_i M_i \rho M_i^\dagger,$$

where the M_i satisfy $\sum_i M_i^\dagger M_i = I$.

Appendix B

An efficient algorithm for generating random stabilizers

We describe an efficient algorithm for generating a random stabilizer $\mathcal{S}$, which is useful in the search for good stabilizer codes and distillation protocols. Generating the stabilizer of a random n-qubit stabilizer state with this algorithm has a time complexity of $\mathcal{O}(n^3)$. Let n be the number of qubits involved and $n-k$ the number of generating Pauli operations. Firstly, we observe that the difficulty lies only in finding a generator matrix $S \in \mathbb{Z}_2^{2n \times (n-k)}$, since it needs to be full rank and satisfy $S^T P S = 0$. Any $b \in \mathbb{Z}_2^{n-k}$ fixing the phase factors of the generating Pauli operations, will do.

A straightforward strategy is iteratively generating random columns $S_j \in \mathbb{Z}_2^{2n}$, for $j = 1, \ldots, n-k$, and for each next column checking whether the corresponding Pauli operation is independent of and commutes with the Pauli operations represented by the previous columns. However, the constraint of commutability excludes a lot: the probability that a random S_j satisfies this constraint scales roughly as 2^{-j+1}. The requirement of independence does not change this behavior. So, before we arrive at the random stabilizer generator matrix S, the expected total number of randomly-generated columns is $\mathcal{O}(2^{n-k})$, which is of course unfeasible when $n-k$ becomes large.

Instead, we propose the following iterative algorithm. Initially, we pick random nonzero columns $S_1, T_1 \in \mathbb{Z}_2^{2n}$, where $S_1^T P T_1 = 1$. The probability that a random column T_j satisfies the linear constraint $S_j^T P T_j = 1$, where S_j is nonzero, is exactly $\frac{1}{2}$, and $S_j^T P S_j = T_j^T P T_j = 0$ always holds. At the beginning of each step j of the iteration, we have matrices $S, T \in \mathbb{Z}_2^{2n \times (j-1)}$ that satisfy

1) $S^T P T = I$,

2) $S^T P S = 0$,

3) $T^T P T = 0$.

We now add columns S_j and T_j respectively such that the resulting $S', T' \in \mathbb{Z}_2^{2n\times j}$ still satisfy constraints (1), (2) and (3), as follows. Let x be a random element of $\mathbb{Z}_2^{2n}$. We define

$$S_j = (I + TS^T P + ST^T P)x \tag{B.1}$$

and we repeat trying random x until we arrive at a nonzero S_j. We then have: $S^T PS_j = S^T Px + S^T PTS^T Px + S^T PST^T Px = S^T Px + S^T Px = 0$, and similarly, $T^T PS_j = 0$. Furthermore, this S_j is independent of the joined columns of S and T, or equivalently:

$$S_j = 0 \Leftrightarrow S_j \in \mathrm{col}\left([S\ T]\right). \tag{B.2}$$

Indeed, let $S_j \in \mathrm{col}\left([S\ T]\right)$, or $S_j = Ss + Tt$. It follows, using constraints (1), (2) and (3), that $S^T PS_j = t$ and $T^T PS_j = s$, which we already proved to be zero. The implication from left to right is trivial. The probability that a random x gives rise to nonzero S_j is therefore equal to $1 - 2^{2(j-1-n)}$. In fact, this probability also applies to the initial step $j = 1$.

Next, we generate a random $y \in \mathbb{Z}_2^{2n}$ that satisfies $S_j^T Py = 1$, and we define

$$T_j = (I + TS^T P + ST^T P)y. \tag{B.3}$$

It follows that $S_j^T PT_j = S_j^T Py = 1$ and $S^T PT_j = T^T PT_j = 0$. Therefore, the extended matrices S' and T' still satisfy constraints (1), (2) and (3). This algorithm can also be used to generate a random symplectic matrix $C = [S\ T]$, for $k = 0$. One can verify that, for generating $S, T \in \mathbb{Z}_2^{2n\times(n-k)}$ satisfying the constraints, the expected total number of random columns that needs to be generated scales roughly as $3(n-k)$ (i.e. 1 for each column of S and 2 for each column of T). For each generated random column, we only need to perform a fixed number of matrix-column multiplications with time complexity $\mathcal{O}[n(n-k)]$, since independence is guaranteed by construction.

Finally, we show that the resulting matrix S generates a truly random self-dual space $\mathcal{C} = \mathrm{col}\,(S)$. By adding a nonzero column S_j to S, the column space $\mathcal{C}$ is expanded to $\mathcal{C}' = \mathcal{C} + S_j$. This space is self-dual if and only if $S_j \in \mathcal{N}$, the dual space of $\mathcal{C}$. Let $C \in \mathbb{Z}_2^{2n\times 2n}$ be a symplectic matrix of the form $C = [S\ U\ T\ V]$. Such matrix always exists, albeit constructed in the way we have just described. From the simplecticity constraint (2.8), it follows that $\mathcal{N} = \mathrm{col}\left([S\ U\ V]\right)$. So S_j needs to be of the form $S_j = Ss + Uu + Vv$, but this yields the same space $\mathcal{C}'$ as $S_j = Uu + Vv$. Since C is invertible, there exists some $w \in \mathbb{Z}_2^{2n}$ for which $x = Cw$. As this is a one-to-one relation, w acquires the full randomness of x. Equivalently, we have $x = Ss + Uu + Tt + Vv$, where s, u, t, v are random. Plugging this into (B.1), and using (2.8), we arrive at $S_j = Uu + Vv$.

List of Publications

- E. Hostens, J. Dehaene, B. De Moor. The equivalence of two approaches to the design of entanglement distillation protocols.
  Internal Report 04-124, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), 2004. `E-print quant-ph/0406017`.[1]

- E. Hostens, J. Dehaene, B. De Moor. Stabilizer states and Clifford operations for systems of arbitrary dimensions and modular arithmetic.
  Phys. Rev. A **71**, 042315 (2005). `E-print quant-ph/0408190`.

- E. Hostens, J. Dehaene, B. De Moor. Hashing protocol for distilling multipartite Calderbank-Shor-Steane states.
  Phys. Rev. A **73**, 042316 (2006). `E-print quant-ph/0510096`.

- E. Hostens, J. Dehaene, B. De Moor. Asymptotic adaptive bipartite entanglement distillation protocol.
  Phys. Rev. A **73**, 062337 (2006). `E-print quant-ph/0602205`.

- E. Hostens, J. Dehaene, B. De Moor. Hashing protocol for multipartite entanglement distillation.
  Proceedings of the 17th International Symposium of Mathematical Theory of Networks and Systems (MTNS 2006), Kyoto, Japan, 24-28 July 2006.

- E. Hostens, J. Dehaene, B. De Moor. Stabilizer state breeding.
  Phys. Rev. A **74**, 062318 (2006). `E-print quant-ph/0608145`.

[1] Articles for which this number is given can be retrieved from the electronic Quantum Physics Archive. For this particular article, the URL is http://arxiv.org/abs/quant-ph/0406017.

Curriculum Vitae

Erik Hostens was born on September 18th, 1980, in Tienen, Belgium. He went to the Heilige-Drievuldigheidscollege in Leuven, with majors Ancient Greek and mathematics. In July 2003 he received the degree of *Burgerlijk Werktuigkundig-Elektrotechnisch Ingenieur* (Electrical Engineer), option Data Mining and Automation, at the Katholieke Universiteit Leuven. In October 2003, he started pursuing a Ph.D. on the subject of quantum information theory in the SCD research group of the Department of Electrical Engineering (ESAT), under the supervision of promoters Bart De Moor and Jeroen Dehaene. This research is funded by a Ph.D. grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

References

[1] A. Acin, E. Jané, W. Dür, and G. Vidal. Optimal distillation of a Greenberger-Horne-Zeilinger state. *Phys. Rev. Lett.*, 85:4811–4814, 2000. E-print quant-ph/0007042.

[2] A. Ambainis and D. Gottesman. The minimum distance problem for two-way entanglement purification. *IEEE Trans. Info. Theory*, 52(2):748–753, 2006. E-print quant-ph/0310097.

[3] H. Aschauer, W. Dür, and H.-J. Briegel. Multiparticle entanglement purification for two-colorable graph states. *Phys. Rev. A*, 71:012319, 2005. E-print quant-ph/0405045.

[4] A. Aspect, Ph. Grangier, and G. Roger. Experimental realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: a new violation of Bell's inequalities. *Phys. Rev. Lett.*, 49:91–94, 1982.

[5] K. Audenaert, J. Eisert, E. Jané, M. B. Plenio, S. Virmani, and B. De Moor. Asymptotic relative entropy of entanglement. *Phys. Rev. Lett.*, 87:217902, 2001. E-print quant-ph/0103096.

[6] K. M. R. Audenaert, J. Calsamiglia, Ll. Masanes, R. Muñoz-Tapia, A. Acin, E. Bagan, and F. Verstraete. Discriminating states: The quantum Chernoff bound. *Phys. Rev. Lett.*, 98:160501, 2007. E-print quant-ph/0610027.

[7] K. M. R. Audenaert and M. B. Plenio. When are correlations quantum? – verification and quantification of entanglement by simple measurements. *New J. Phys.*, 8:266, 2006. E-print quant-ph/0608067.

[8] H. Barnum and N. Linden. Monotones and invariants for multi-particle quantum states. *J. Phys. A: Math. Gen.*, 34:6787–6805, 2001. E-print quant-ph/0103155.

[9] J. S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200, 1964.

[10] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher. Concentrating partial entanglement by local operations. *Phys. Rev. A*, 53:2046–2052, 1996. E-print quant-ph/9511030.

[11] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin. Experimental quantum cryptography. *Journal of Cryptology*, 5(1):3–28, 1992.

[12] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.*, 70:1895–1899, 1993.

[13] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters. Purification of noisy entanglement and faithful teleportation via noisy channels. *Phys. Rev. Lett.*, 76:722–725, 1996. E-print quant-ph/9511027.

[14] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed state entanglement and quantum error correction. *Phys. Rev. A*, 54:3824–3851, 1996. E-print quant-ph/9604024.

[15] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Phys. Rev. Lett.*, 69:2881–2884, 1992.

[16] H. Bombin and M. A. Martin-Delgado. Entanglement distillation protocols and number theory. *Phys. Rev. A*, 72:032313, 2005. E-print quant-ph/0503013.

[17] A. R. Calderbank, E. M. Rains, P. W. Shor, and N. J. A. Sloane. Quantum error correction and orthogonal geometry. *Phys. Rev. Lett.*, 78(3):405–408, 1997. E-print quant-ph/9605005.

[18] A.R. Calderbank, E.M. Rains, P.W. Shor, and N.J.A. Sloane. Quantum error correction via codes over GF(4). *IEEE trans. Inform. Theory*, 44(4):1369–1387, 1998. E-print quant-ph/9608006.

[19] K. Chen and H.-K. Lo. Multi-partite quantum cryptographic protocols with noisy GHZ states. quant-ph/0404133, 2004.

[20] R. Cleve and D. Gottesman. Efficient computations of encodings for quantum error correction. *Phys. Rev. A*, 56:76–82, 1997. E-print quant-ph/9607030.

[21] R. Cleve, D. Gottesman, and H.-K. Lo. How to share a quantum secret. *Phys. Rev. Lett.*, 83:648–651, 1999. E-print quant-ph/9901025.

[22] T. Cover and J. Thomas. *Elements of information theory.* Wiley, New York, 1991.

[23] C. Crépeau, D. Gottesman, and A. Smith. Secure multi-party quantum computing. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC)*, 2002. E-print quant-ph/0206138.

[24] J. Dehaene and B. De Moor. Clifford group, stabilizer states, and linear and quadratic operations over GF(2). *Phys. Rev. A*, 68:042318, 2003. E-print quant-ph/0304125.

[25] J. Dehaene, M. Van den Nest, B. De Moor, and F. Verstraete. Local permutations of products of Bell states and entanglement distillation. *Phys. Rev. A*, 67:022310, 2003. E-print quant-ph/0207154.

[26] D. Deutsch. Quantum theory, the Church-Turing Principle and the universal quantum computer. *Proc. Roy. Soc. Lond. A*, 400:97–117, 1985.

[27] D. Deutsch, A. Ekert, R. Jozsa, C. Macchiavello, S. Popescu, and A. Sanpera. Quantum privacy amplification and the security of quantum cryptography over noisy channels. *Phys. Rev. Lett.*, 77:2818–2821, 1996. E-print quant-ph/9604039.

[28] D. Dieks. Communication by EPR devices. *Physics Letters A*, 92:271–272, 1982.

[29] W. Dür, H. Aschauer, and H.-J. Briegel. Multiparticle entanglement purification for graph states. *Phys. Rev. Lett.*, 91:107903, 2003. E-print quant-ph/0303087.

[30] W. Dür and H.-J. Briegel. Entanglement purification and quantum error correction. *Rep. Prog. Phys.*, 70:1381, 2007. E-print quant-ph/0705.4165.

[31] W. Dür, J. Calsamiglia, and H.-J. Briegel. Multipartite secure state distribution. *Phys. Rev. A*, 71:042336, 2005. E-print quant-ph/0411209.

[32] W. Dür, G. Vidal, and J. I. Cirac. Three qubits can be entangled in two inequivalent ways. *Phys. Rev. A*, 62:062314, 2000. E-print quant-ph/0005115.

[33] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780, 1935.

[34] J. Eisert and D. Gross. Multi-particle entanglement. In D. Bruss and G. Leuchs, editors, *Lectures on quantum information.* Wiley-VCH, Weinheim, 2006. E-print quant-ph/0505149.

[35] A. K. Ekert. Quantum cryptography based on Bells theorem. *Phys. Rev. Lett.*, 67:661–663, 1991.

[36] R. P. Feynman. Simulating physics with computers. *Int. J. Theor. Phys.*, 21:467, 1982.

[37] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden. Quantum cryptography. *Rev. Mod. Phys.*, 74:145–195, 2002.

[38] S. Glancy, E. Knill, and H. M. Vasconcelos. Entanglement purification of any stabilizer state. *Phys. Rev. A*, 74:032319, 2006. E-print quant-ph/0606125.

[39] D. Gottesman. A class of quantum error-correcting codes saturating the quantum Hamming bound. *Phys. Rev. A*, 54:1862–1868, 1996. E-print quant-ph/9604038.

[40] D. Gottesman. *Stabilizer codes and quantum error correction.* PhD thesis, Caltech, 1997. E-print quant-ph/9705052.

[41] D. Gottesman. The Heisenberg representation of quantum computers. E-print quant-ph/9807006, 1998.

[42] D. Gottesman. Theory of fault-tolerant quantum computation. *Phys. Rev. A*, 57:127–137, 1998. E-print quant-ph/9702029.

[43] K. Goyal, A. McCauley, and R. Raussendorf. Purification of large bicolorable graph states. *Phys. Rev. Lett.*, 74:032318, 2006. E-print quant-ph/0605228.

[44] L. K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the 28th Annual ACM Symposium on the Theory of Computing (STOC)*, pages 212–219, 1996. E-print quant-ph/9605043.

[45] M. Hein, W. Dür, J. Eisert, R. Raussendorf, M. Van den Nest, and H.-J. Briegel. Entanglement in graph states and its applications. In P. Zoller, G. Casati, D. Shepelyansky, and G. Benenti, editors, *Proceedings of the International School of Physics "Enrico Fermi" on "Quantum Computers, Algorithms and Chaos"*, volume 162, Varenna, Italy, 2006. E-print quant-ph/0602096.

[46] M. Hein, J. Eisert, and H.-J. Briegel. Multi-party entanglement in graph states. *Phys. Rev. A*, 69:062311, 2004. E-print quant-ph/0307130.

[47] M. Hillery, V. Bužek, and A. Berthiaume. Quantum secret sharing. *Phys. Rev. A*, 59:1829–1834, 1999. E-print quant-ph/9806063.

[48] M. Horodecki and P. Horodecki. Reduction criterion of separability and limits for a class of distillation protocols. *Phys. Rev. A*, 59:4206–4216, 1999. E-print quant-ph/9708015.

[49] M. Horodecki, P. Horodecki, and R. Horodecki. Mixed-state entanglement and distillation: is there a "bound" entanglement in nature? *Phys. Rev. Lett.*, 80:5239–5242, 1998. E-print quant-ph/9801069.

[50] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 2007. E-print quant-ph/0702225.

[51] E. Hostens, J. Dehaene, and B. De Moor. The equivalence of two approaches to the design of entanglement distillation protocols. E-print quant-ph/0406017, 2004.

[52] E. Hostens, J. Dehaene, and B. De Moor. Stabilizer states and clifford operations for systems of arbitrary dimensions and modular arithmetic. *Phys. Rev. A*, 71:042315, 2005. E-print quant-ph/0408190.

[53] E. Hostens, J. Dehaene, and B. De Moor. Asymptotic adaptive bipartite entanglement-distillation protocol. *Phys. Rev. A*, 73:062337, 2006. E-print quant-ph/0602205.

[54] E. Hostens, J. Dehaene, and B. De Moor. Hashing protocol for distilling multipartite Calderbank-Shor-Steane states. *Phys. Rev. A*, 73:042316, 2006. E-print quant-ph/0510096.

[55] E. Hostens, J. Dehaene, and B. De Moor. Stabilizer state breeding. *Phys. Rev. A*, 74:062318, 2006. E-print quant-ph/0608145.

[56] R. Jozsa. Fidelity for mixed quantum states. *J. Mod. Opt.*, 41:2315–2325, 1994.

[57] A. Karlsson, M. Koashi, and N. Imoto. Quantum entanglement for secret sharing and secret splitting. *Phys. Rev. A*, 59:162–168, 1999.

[58] A. Kay and J. K. Pachos. Multipartite purification protocols: upper and optimal bounds. *Phys. Rev. A*, 75:062307, 2007. E-print quant-ph/0608080.

[59] C. Kruszynska, S. Anders, W. Dür, and H.-J. Briegel. Quantum communication cost of preparing multipartite entanglement. *Phys. Rev. A*, 73:062328, 2006. E-print quant-ph/0512218.

[60] C. Kruszynska, A. Miyake, H.-J. Briegel, and W. Dür. Entanglement purification protocols for all graph states. *Phys. Rev. A*, 74:052316, 2006. E-print quant-ph/0606090.

[61] N. Linden and S. Popescu. On multi-particle entanglement. *Fortsch. Phys.*, 46:567–578, 1998. E-print quant-ph/9711016.

[62] F. J. MacWilliams and N. J. A. Sloane, editors. *The theory of error-correcting codes.* Mathematical Library. North-Holland, 1977.

[63] E. Maneva and J. A. Smolin. Improved two-party and multi-party purification protocols. *AMS Cont. Math.*, 305:203–212, 2002. E-print quant-ph/0003099.

[64] R. Matsumoto. Conversion of a general quantum stabilizer code to an entanglement distillation protocol. *J. Phys. A: Math. Gen.*, 36(29):8113–8127, 2003. E-print quant-ph/0209091.

[65] N. Metwally. More efficient entanglement purification. *Phys. Rev. A*, 66:054302, 2002. E-print quant-ph/0109051.

[66] A. Miyake and H.-J. Briegel. Distillation of multipartite entanglement by complementary stabilizer measurements. *Phys. Rev. Lett.*, 95:220501, 2005. E-print quant-ph/0506092.

[67] M. Murao, M. B. Plenio, S. Popescu, V. Vedral, and P.L. Knight. Multiparticle entanglement purification protocols. *Phys. Rev. A*, 57:R4075–R4078, 1998. E-print quant-ph/9712045.

[68] M. Nielsen and I. Chuang. *Quantum computation and quantum information.* Cambridge University press, Cambridge, 2000.

[69] M. A. Nielsen. Conditions for a class of entanglement transformations. *Phys. Rev. Lett.*, 83:436–439, 1999.

[70] A. Peres. Separability criterion for density matrices. *Phys. Rev. Lett.*, 77:1413–1415, 1996. E-print quant-ph/9604005.

[71] M. B. Plenio and S. Virmani. An introduction to entanglement measures. *Quant. Inf. Comp.*, 7:1–51, 2007. E-print quant-ph/0504163.

[72] J. Preskill. Lecture notes on quantum computation. http://www.theory.caltech.edu/people/preskill/ph219/.

[73] E. M. Rains. Entanglement purification via separable superoperators. E-print quant-ph/9707002, 1997.

[74] E. M. Rains. An improved bound on distillable entanglement. E-print quant-ph/9809082, 1998.

[75] R. Raussendorf, D.E. Browne, and H.-J. Briegel. Measurement-based quantum computation with cluster states. *Phys. Rev. A*, 68:022312, 2003. E-print quant-ph/0301052.

[76] M. Schlosshauer. Decoherence, the measurement problem, and interpretations of quantum mechanics. *Rev. Mod. Phys.*, 76:1267–1305, 2004. E-print quant-ph/0312059.

[77] P. Shor and R. Laflamme. Quantum analog of the MacWilliams identities for classical coding theory. *Phys. Rev. Lett.*, 78(8):1600–1602, 1997. E-print quant-ph/9610040.

[78] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Sci. Statist. Comput.*, 26:1484–1509, 1997. E-print quant-ph/9508027.

[79] P. W. Shor and J. A. Smolin. Quantum error-corrrecting codes need not completely reveal the error syndrome. E-print quant-ph/9604006, 1996.

[80] M. Van den Nest. *Local equivalence of stabilizer states and codes.* PhD thesis, KULeuven, 2005.

[81] M. Van den Nest, J. Dehaene, and B. De Moor. Efficient algorithm to recognize local Clifford equivalence of graph states. *Phys. Rev. A*, 70:034302, 2004. E-print quant-ph/0405023.

[82] M. Van den Nest, J. Dehaene, and B. De Moor. Graphical description of the action of local Clifford transformations on graph states. *Phys. Rev. A*, 69:022316, 2004. E-print quant-ph/0308151.

[83] M. Van den Nest, J. Dehaene, and B. De Moor. Local unitary versus local Clifford equivalence of stabilizer states. *Phys. Rev. A*, 71:062323, 2005. E-print quant-ph/0411115.

[84] M. Van den Nest, W. Dür, A. Miyake, and H.-J. Briegel. Fundamentals of universality in one-way quantum computation. *New J. Phys.*, 9:204, 2007. E-print quant-ph/0702116.

[85] M. Van den Nest, W. Dür, G. Vidal, and H.-J. Briegel. Classical simulation versus universality in measurement-based quantum computation. *Phys. Rev. A*, 75:012337, 2007. E-print quant-ph/0608060.

[86] M. Van den Nest, A. Miyake, W. Dür, and H.-J. Briegel. Universal resources for measurement-based quantum computation. *Phys. Rev. Lett.*, 97:150504, 2006. E-print quant-ph/0604010.

[87] F. Verstraete. *A study of entanglement in quantum information theory.* PhD thesis, K.U.Leuven, 2002.

[88] F. Verstraete, J. Dehaene, and B. De Moor. Local filtering operations on two qubits. *Phys. Rev. A*, 64:010101(R), 2001. E-print quant-ph/0011111.

[89] F. Verstraete, J. Dehaene, and B. De Moor. Normal forms and entanglement measures for multipartite quantum states. *Phys. Rev. A*, 68:012103, 2003. E-print quant-ph/0105090.

[90] K. G. H. Vollbrecht and F. Verstraete. Interpolation of recurrence and hashing entanglement distillation protocols. *Phys. Rev. A*, 71:062325, 2005. E-print quant-ph/0404111.

[91] E. W. Weisstein. Chebyshev inequality. From MathWorld. http://mathworld.wolfram.com/ChebyshevInequality.html.

[92] E. W. Weisstein. Stirling's approximation. From MathWorld. http://mathworld.wolfram.com/StirlingsApproximation.html.

[93] R.F. Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Phys. Rev. A*, 40:4277–4281, 1989.

[94] W. K. Wootters and W. H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982.

[95] B. Zeng, H. Chung, A. W. Cross, and I. L. Chuang. Local unitary versus local Clifford equivalence of stabilizer and graph states. *Phys. Rev. A*, 75:032325, 2007. E-print quant-ph/0611214.