**KATHOLIEKE UNIVERSITEIT LEUVEN**
FACULTEIT TOEGEPASTE WETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK
Kasteelpark Arenberg 10, 3001 Leuven (Heverlee)

# A STUDY OF ENTANGLEMENT IN QUANTUM INFORMATION THEORY

Promotoren:
Prof. dr. ir. B. De Moor
Prof. dr. H. Verschelde (UG)

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de toegepaste wetenschappen

door **Frank Verstraete**

Oktober 2002

**KATHOLIEKE UNIVERSITEIT LEUVEN**
FACULTEIT TOEGEPASTE WETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK
Kasteelpark Arenberg 10, 3001 Leuven (Heverlee)

# A STUDY OF ENTANGLEMENT IN QUANTUM INFORMATION THEORY

Jury:
Prof. dr. ir. J. Berlamont, voorzitter
Prof. dr. ir. B. De Moor, promotor
Prof. dr. H. Verschelde (UG), promotor
Dr. ir. J. Dehaene
Prof. dr. M. Fannes
Prof. dr. S. Massar (ULB)
Prof. dr. M. Plenio (Imperial College)
Prof. dr. ir. J. Vandewalle

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de toegepaste wetenschappen

door **Frank Verstraete**

U.D.C. 681.3*E4

Oktober 2002

D/2002/7515/48

ISBN 90-5682-377-9

"A possible experience or truth is not the same as an actual experience or truth minus its *reality value* but has - according to its partisans, at least - something quite divine about it, a fire, a soaring, a readiness to build and a conscious utopianism that does not shrink from reality but sees it as a project, something yet to be invented." [Robert Musil]

# Voorwoord

Op de eerste dag schiep God de aarde, de natuur en alle materie. Dankbaarheid ben ik verschuldigd voor al het schone op deze wereld, voor het feit dat alles op deze wereld uit zichzelf vooruit gaat en aangetrokken wordt tot het goede[1].

Op de tweede dag schiep God de mens. Dankbaarheid ben ik verschuldigd voor de onwetendheid van de mens, waardoor verwondering hem deelachtig kan worden, waardoor een klein beetje inzicht in de wonderen der natuur hem reeds in een roes kan brengen, waardoor hij het grootste genoegen kan scheppen aan het oplossen van wiskundige spelletjes. Dankbaarheid ook voor de vergankelijkheid van het bestaan, waardoor we verplicht worden te vechten en te genieten.

Op de derde dag werd vriendschap geschapen. Mijn vrienden leerden me dat er geen wezenlijk verschil bestaat tussen de roes van het denken en de roes van een ongebreidelde braspartij, zolang je jezelf maar compleet laat opgaan in hetgeen je doet, zolang je gelooft in schoonheid, zolang je maar 100 procent eerlijk bent ten opzichte van jezelf en de medemens.

Op de vierde dag besliste God dat orde en structuur in het leven geroepen moest worden: hij schiep de promotor. Ik dank Bart De Moor voor zijn uitbundig enthousiasme, voor zijn bezieling van SISTA en voor de luxe die hij me bood te doctoreren op een onderwerp naar mijn keuze. Ik dank Henri Verschelde om mij te laten meegenieten van zijn ongebreidelde kennis, en om mij te inspireren onderzoek te doen in het gebied van de quantum informatie.

Op de vijfde dag ontstonden aldus collega's. Ik had het geluk zowel in Leuven als in Gent terecht te komen in een stimulerende omgeving. Ik dank dan ook al mijn collega's voor de aangename werksfeer en discussies. Speciale dank ben ik verschuldigd aan Ida, Annie en Hugo voor de invulling van mijn praktische verplichtingen, aan Prof. Vandewalle en Prof. Verheest om goede en rechtvaardige departementsvoorzitters te zijn, en aan Prof. Fannes en Prof. Vandewalle voor het kritisch nalezen van deze thesis. Ik denk met plezier terug aan Koenraad Audenaert met zijn recalcitrant gedrag, aan David Dudal en Karel Van

[1] *Maître Pangloss prouvait admirablement qu'il n'y a point d'effet sans cause, et que, dans ce meilleur des monde possible, les choses ne peuvent être autrement: car tout étant fait pour une fin, tout est nécessairement pour la meilleure fin.* [Voltaire]

Acoleyen die me bijbrachten dat het schoner is de ronde van Vlaanderen te winnen dan de Nobelprijs, en bovenal aan Jeroen Dehaene: het is fantastisch met iemand samen te werken die beschikt over een complementaire manier van denken, met iemand die gedreven wordt door een drang naar eenvoud en schoonheid.

The world of physics was disconsolate, subject to unsound beliefs and shrouded in darkness. On the sixth day however, God created the Quantum Information Theory Community, and their devotion to the concept of entanglement resulted in order and wonder. I am very grateful to Andrew Doherty, Steven van Enk, Chris Fuchs, Hideo Mabuchi, John Preskill, Terry Rudolph, Guifré Vidal and Michael Wolf for enlighting me and giving me the opportunity to collaborate with them: I enjoyed their company very much, and thanks to their hospitality in Caltech, Bell Labs and Braunschweig I could fully experience the luxury and privilige to do fundamental research. Special thanks also to T. Brun, N. Cerf, J.I. Cirac, T. De Bie, P. Hayden, M. and P. Horodecki, T. Laustsen, S. Massar, B. Munro, M. Plenio, P. Scudo, L. Vandersijpen, K. Vollbrecht, R. Werner and H. Woerdeman who helped to shape my view of quantum information theory and with whom it was a real pleasure to discuss quantum mechanics.

Op de zevende dag zag God dat alles schoon en goed was, maar dat nog één essentiële schakel in de cirkel der levensloop ontbrak: Hij creëerde liefde, bezieling, ouders, zus, vrouw en kinderen. Ik dank mijn ganse familie van harte voor hun begrip en steun, en mijn ouders voor hun toewijding aan hun kinderen. Ik dank Ludovic en Amaryllis voor hun uitbundige lach en spel, en voor hun verwondering, het schoonste geschenk dat een mens kan koesteren. Boven alles houd ik er aan mijn echtgenote Katrien te danken, voor haar liefde en raad, voor haar volledige toewijding aan ons gezin, voor haar geduld met mijn onmogelijke werkuren, voor al wat een vrouw doet zonder dat haar man het beseft, en voor haar schoonheid en zuiverheid.

Toen de volgende dag aangebroken was, zag de jury dat het goed was.

# Abstract

Although the concept of quantum entanglement has been known for about seventy years, it only recently quit the realms of meta-theoretical discussions when it was discovered how entanglement can be exploited to compute and communicate with an unprecedented power. The primary motivation of the work presented in this thesis has been to contribute to the big effort that has been done during the last decade to understand and quantify quantum entanglement. We have developed advanced techniques of linear and multilinear algebra to investigate and classify entangled pure and mixed quantum states, and discussed some novel applications in the field of quantum information theory.

The results presented in this thesis are mainly of interest from a *fundamental* point a view: entanglement is *the* characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought [**186**]. It is however a real privilege that fundamental research in quantum information theory bears the tools of tomorrow's electrical engineers: the ongoing miniaturization of electronic components will soon reach a scale where quantum mechanical effects play a major role.

The first part of this thesis is devoted to the study of entanglement. Local equivalence classes of multipartite pure and mixed quantum systems are discussed, and different entanglement measures are introduced and compared. The second part is mainly concerned with the problem of transmission and extraction of classical and quantum information through quantum channels. Optimal detection strategies for continuously monitored systems are derived, and we exploit a duality between quantum maps and entangled quantum states to present a unified description of quantum channels.

# Contents

# Glossary

## Mathematical Notation

| | |
|---|---|
| $X^\dagger$ | Hermitean Conjugate of the matrix $X$ (following notation in Peres [**169**]). |
| $X^T$ | Transpose of the matrix $X$ |
| $X^*$ | Complex Conjugate of the matrix $X$ (elementswise) |
| $X^{T_i}$ | Partial Transpose over subsystem $i$ |
| $X^\Gamma$ | shortcut notation for $X^{T_2}$ |
| $\mathrm{Tr}\,(X)$ | trace operation |
| $\mathrm{Tr}_i(X)$ | Partial trace over subsystem $i$ |
| $A \otimes B$ | tensor (or Kronecker) product |
| $A \circ B$ | Hadamard product |
| $A \oplus B$ | Direct sum |
| $A \geq B$ | $A - B$ is positive (semi)-definite |
| $\lvert a \rvert$ | Absolute value |
| $\det(A)$ | determinant |
| $\mathrm{Im}\{X\}$ | Imaginary part of the matrix $X$ |
| $\mathrm{Re}\{X\}$ | Real Part of the matrix $X$ |
| $\{p_i\}$ | the set of elements $\{p_1, p_2, \cdots\}$ |

## Acronyms and Abbreviations

| | |
|---|---|
| CP | Completely Positive |
| EM | Entanglement Monotone/ Entanglement Measure |
| EoF | Entanglement of Formation |
| EPR | Einstein-Podolsky-Rosen |
| GHZ | quantum state named after Greenberger-Horne-Zeilinger |
| LO | Local Operations |
| LOCC | Local Operations assisted by Classical Communication |
| LSVD | Lorentz Singular Value Decomposition |
| LU | Local Unitary operations |
| ME | Maximally Entangled |
| MEMS | Maximally Entangled Mixed States |
| MREGS | Minimal Reversible Entanglement Generating Set |
| POVM | Positive Operator Valued Measure |
| PPT | Positive Partial Transpose |
| RelEnt | Relative Entropy of Entanglement |
| SLOCC | Stochastic Local Operations assisted by Classical Communication or Filtering Operation |
| SVD | Singular Value Decomposition |
| TP | Trace Preserving |
| TPCP | Trace Preserving Completely Positive |

## Fixed Symbols

| | |
|---|---|
| $\beta(\rho)$ | violation of CHSH-inequality |
| $\epsilon_2$ | Completely antisymmetric $2 \times 2$ tensor |
| $\epsilon_{ijk\ldots}$ | Completely antisymmetric tensor |
| $C(\rho)$ | Concurrence |
| $C_A(\rho)$ | Concurrence of Assistance |
| $E_f(\rho)$ | Entanglement of formation |
| $E_N(\rho)$ | Negativity |
| $E_R(\rho)$ | Relative Entropy of Entanglement |
| $f(\rho)$ | Teleportation fidelity |
| $F(\rho)$ | Fidelity or maximal singlet fraction |
| $H(\{p_i\})$ | Shannon entropy $(= -\sum_i p_i \log_2(p_i))$ |
| $N(\rho)$ | Negativity |
| $O(N)$ | Group of real orthogonal $N \times N$ matrices |
| $O(N, \mathcal{C})$ | Group of complex orthogonal $N \times N$ matrices |
| $S(\rho)$ | von-Neumann entropy $(= -\mathrm{Tr}\,(\rho \log_2(\rho)))$ |
| $S(\rho\|\|\sigma)$ | Umegaki Relative Entropy |
| $SO(3,1)$ | Group of Lorentz transformations with determinant $+1$ |
| $SO(N)$ | Group of real orthogonal $N \times N$ matrices with determinant $+1$ |
| $SO(N, \mathcal{C})$ | Group of complex orthogonal $N \times N$ matrices with determinant $+1$ |
| $SL(N, \mathcal{C})$ | Group of complex $N \times N$ matrices with determinant $+1$ |
| $SU(N)$ | Group of complex Unitary $N \times N$ matrices with determinant $+1$ |
| $U(N)$ | Group of complex Unitary $N \times N$ matrices |

# Nederlandse Samenvatting: Een Studie van 'Entanglement' in het licht van de Quantum Informatie Theorie

Een van de grootste wetenschappelijke revoluties ontstond ongeveer 100 jaar geleden toen Max Planck opperde dat energie gequantizeerd was. De genieën Bohr, Einstein, Heisenberg, Schrödinger, Jordan, Born, Pauli en Dirac slaagden er in de jaren twintig in een consistente theorie neer te schrijven die het gedrag van atomen en elementaire deeltjes beschrijft. Dit leidde tot ongeziene en verregaande inzichten in alle takken van de fysica en de chemie. Gedurende zestig jaar werd het onderzoek in de quantummechanica toegespitst op toepassingen en veralgemeningen met spectaculaire successen zoals het uitvinden van de transistor en de ontdekking van de quantum elektrodynamica.

Op het einde van de twintigste eeuw ontstond echter een hernieuwde interesse voor de fundamenten van de quantummechanica, mede gedreven door het feit dat voor het eerst experimenten konden uitgevoerd worden op individueel interagerende elementaire deeltjes. Vrij snel bleek dat het manipuleren van individuele atomen tot ongekende mogelijkheden kon leiden op het vlak van computing, cryptografie en communicatie. Daarenboven werd duidelijk dat er een intrigerende connectie bestond tussen quantummechanica en informatietheorie, wat leidde tot quantuminformatietheorie. Het domein van de quantuminformatietheorie geniet een gepriviligieerde status: het combineert de quantummechanica en de informatietheorie op een zulksdanige manier dat beide onderzoeksgebieden er wel bij varen. Informatietheorie werd in de jaren 1940 uitgevonden door C. Shannon in Bell Labs toen hij onderzocht hoe men optimaal gebruik kan maken van een gegeven communicatienetwerk. Dit leidde tot een fascinerende abstracte mathematische formulering van het begrip informatie.

De term *informatie* in quantuminformatietheorie heeft een dubbele betekenis. Enerzijds slaat deze op het feit dat het gebied onderzoekt hoe informatie verzonden of bewerkt kan worden gebruik makende van quantummechanische systemen. Anderzijds slaat het op het feit dat een quantumtoestand een parameterizatie is van de kennis of informatie die men heeft over een systeem zonder dat er een onderliggende fysische betekenis te hechten is aan de quantumtoestand.

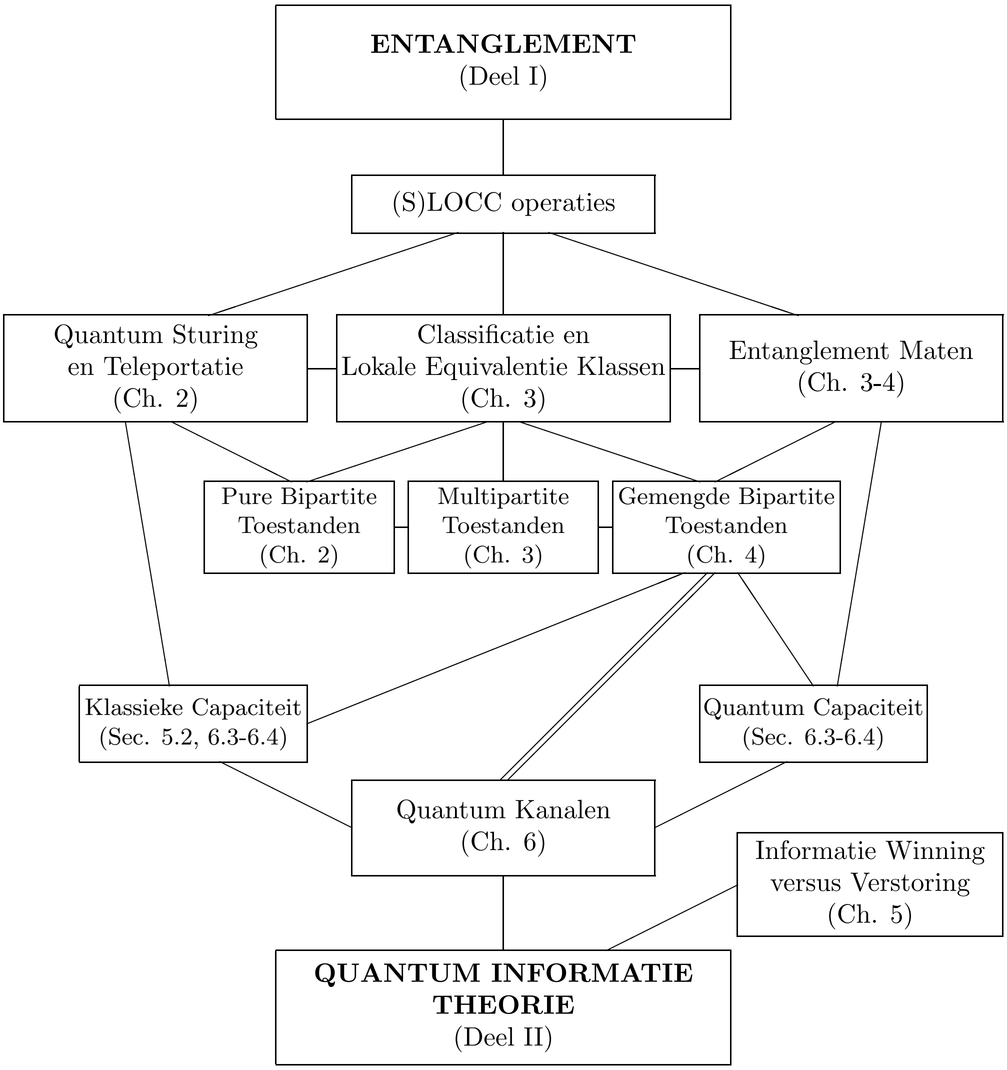
Het centrale concept in quantuminformatietheorie is *entanglement*, wat een direct gevolg is van het feit dat meerdere quantumdeeltjes beschreven worden in een Hilbertruimte die het tensorproduct is van de individuele Hilbertruimten. Entanglement is ervoor verantwoordelijk dat niet-lokale correlaties aanwezig kunnen zijn in quantumsystemen die in schijnbare tegenspraak zijn met het causaliteitsprincipe. Dit werd ontdekt in 1935 door Einstein, Podolsky en Rosen [**82**], en dit verschijnsel werd de naam *entanglement* gegeven door Erwin Schrödinger, die het *de* karakteristieke eigenschap van de quantummechanica noemde. Daarna was het onderwerp entanglement veeleer een onderzoeksobject op de grens van de metafysica, tot men er in de jaren 1980 experimenteel in slaagde entanglement te creëren tussen individuele deeltjes. Sindsdien is entanglement het toverwoord voor elkeen die in het gebied van de quantuminformatietheorie werkt. De belangrijkste uitvindingen in dit gebied werden gedaan na het stellen van de volgende vraag: "Wat kan ik meer doen in deze of gene situatie indien entanglement beschikbaar is?"

Hoewel het overduidelijk is dat entanglement het centrale begrip is in het gebied van quantuminformatietheorie, is het nog helemaal niet duidelijk hoe entanglement beschreven en gequantizeerd moet worden. Dit is het onderwerp van het eerste deel van deze thesis. Het tweede deel handelt over de manier waarop quantumsystemen evolueren, en over de manier waarop (klassieke) informatie uit quantumsystemen geëxtraheerd kan worden. Het zal blijken dat dit tweede deel een soort duale tegenhanger is van het eerste deel.

Nu volgt een summier overzicht van de verschillende hoofdstukken in deze thesis. Een schematisch overzicht van de belangrijkste onderlinge verbanden wordt gegeven in figuur 1.

## Hoofdstuk 1

In het eerste hoofdstuk beogen we de begrippen quantummechanica en informatietheorie in de juiste context te plaatsen. De postulaten van de quantummechanica worden summier behandeld, waarbij de analogie tussen de evolutie van probabiliteitsdistributies en van quantumtoestanden geaccentueerd wordt. Er wordt dieper ingegaan op de essentiele verschillen tussen quantummechanica en klassieke mechanica: volledige kennis over een quantumsysteem impliceert geen determinimisme, en het observeren van een quantumsysteem brengt een irreversibele stochastische verstoring met zich mee. We argumenteren waarom de begrippen informatie en quantummechanica met elkaar verstrengeld zijn, en geven een summier overzicht van de doorbraken in het gebied van quantuminformatietheorie. Tenslotte wordt een overzicht gegeven van deze thesis.




**Figuur 1.** Overzicht van de thesis.

## Hoofdstuk 2

Het tweede hoofdstuk behandelt de eenvoudigste quantumsystemen waarin entanglement aanwezig is: pure bipartite quantumtoestanden (i.e. quantumtoestanden van twee deeltjes in een pure golffunctie). Het *quantum-sturings-theorema* van E. Schrödinger, tot nog toe zo goed als onbekend, wordt geïdentificeerd als het fundamentele theorema dat verklaart hoe entanglement zich manifesteert en hoe het zich laat manipuleren. Het quantum-sturings-theorema geeft een antwoord op de volgende vraag: gegeven een pure bipartite toestand met deeltjes $A$ (Alice) en $B$ (Bob); hoe evolueert de lokale densiteitsoperator van Bob onder invloed van een POVM-meting door Alice? We beschrijven hoe quantum teleportatie aanzien kan worden als een speciaal geval van quantum-sturing, en hoe men nodige en voldoende voorwaarden eruit kan afleiden voor het lokaal transformeren van een toestand in andere toestanden. Samengevat wordt er aangetoond hoe het quantum-sturings-theorema in zijn eenvoud de essentie van de structuur van puur bipartiet entanglement omvat.

## Hoofdstuk 3

Het derde hoofdstuk behandelt het probleem van entanglement in pure toestanden waarbij meer dan twee partijen betrokken zijn. In tegenstelling tot het bipartite geval, waar extensief gebruik kan gemaakt worden van matrixalgebra, wordt de beschrijving van entanglement bemoeilijkt door het feit dat de quantumtoestanden geparameterizeerd worden door hoger dimensionale tensoren. De aandacht wordt vooral toegespitst op hetvolgende probleem: op hoeveel verschillende manieren kunnen multipartite systemen met elkaar entangled zijn? Dit probleem wordt behandeld door het definieren van lokale equivalentieklassen van quantumtoestanden, geïnduceerd door lokaal unitaire of door lokale filtering operaties.

Het definiëren van equivalentieklassen onder de groep van lokaal unitaire operaties wordt gedeeltelijk opgelost door een veralgemening van de singuliere waarden ontbinding (SVD) te introduceren op een constructieve manier. Deze veralgemening blijkt echter niet uniek te zijn (er zijn een discreet aantal oplossingen waarnaar het algoritme kan convergeren), een probleem dat inherent is aan het feit dat met hoger dimensionale tensoren gewerkt wordt. Niettegenstaande dit euvel levert het theorema echter altijd een normaalvorm met het maximaal aantal nullen, en wordt ook een natuurlijke veralgemening van de variationele karakterizering van singuliere waarden bekomen.

De meest algemene lokale transformaties die men fysisch kan implementeren worden beschreven door lokale filtering operaties. Een natuurlijke vraag is dan om equivalentieklassen te gaan definieren van toestanden die in elkaar getransformeerd kunnen worden via deze filtering operaties. Wiskundig gezien komt dit neer op een veralgemening van de singuliere waardenontbinding waarbij de lokale operatoren niet unitair hoeven te zijn ($\in SU(N)$) maar gewoon volle rank ($\in SL(N,\mathcal{C})$). Het formalisme dat daaruit voortspruit is veel krachtiger aangezien het niet-triviale entanglement transformaties mogelijk maakt. Er wordt opnieuw een constructief theorema voorgesteld in het multipartite geval, waarin de normaalvorm een variationele betekenis krijgt. Indicaties over de uniciteit van de normaalvorm worden gegeven, en de continuiteit van de normaalvorm wordt bewezen. Het verkregen formalisme leidt op een natuurlijke manier tot een nieuwe klasse van maten van entanglement. Daarenboven wordt aangetoond dat de normaalvorm van een toestand deze is waarvoor al die maten van entanglement maximaal zijn ten op zichte van alle andere toestanden in zijn equivalentieklasse. Dit impliceert automatisch dat de optimale filtering operaties om maximaal entanglement te creëren uitgaande van een zekere toestand geïdentificeerd zijn. Er wordt daarenboven aangetoond hoe deze resultaten geldig blijven in het geval van gemengde (i.e. niet pure) toestanden.

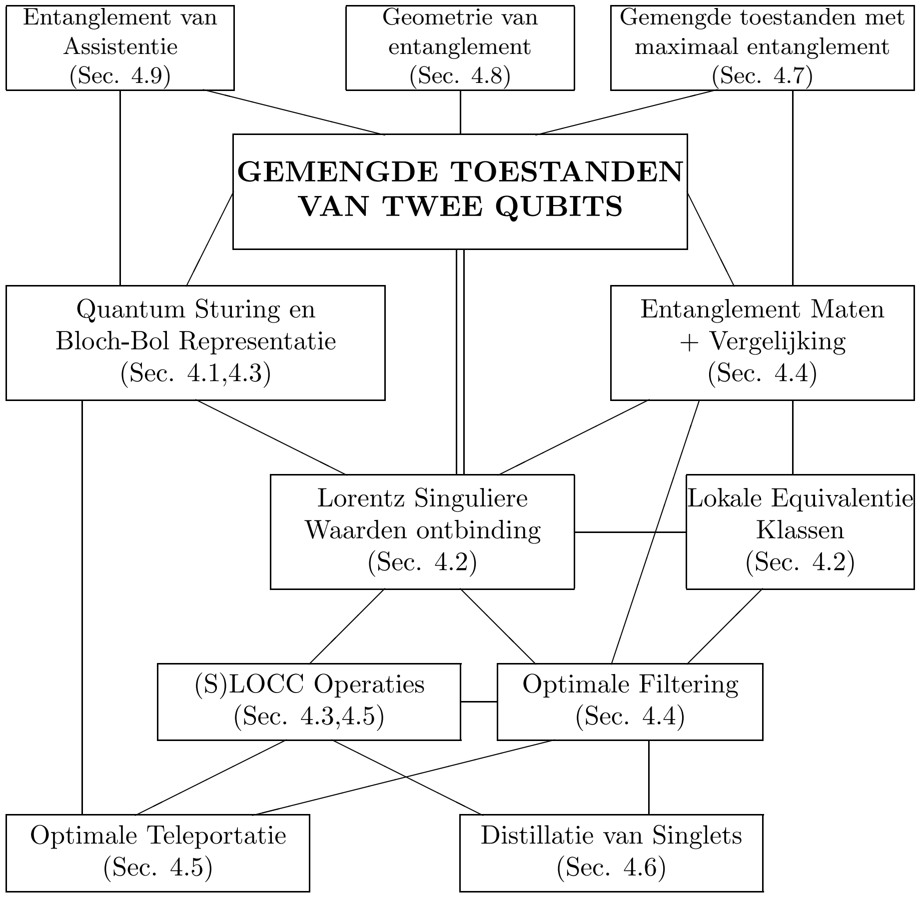
Een volgende paragraaf 3.2 behandelt pure toestanden in het geval van een $2 \times 2 \times N$ dimensionele Hilbertruimte. We bewijzen dat deze toestanden op negen verschillende manieren met elkaar kunnen *entangled* zijn (i.e. er zijn negen verschillende equivalentieklassen onder lokale filtering), en introduceren maten om de hoeveelheid entanglement te quantificeren. We bespreken hoe een quantumtoestand slechts een beperkte susceptibiliteit heeft om *entangled* te zijn: indien een quantumtoestand meer entangled is met één deeltje moet dit ten koste gaan van entanglement met een ander. Het laatste deel van deze paragraaf lost het probleem van optimale 1-copie distillatie van een GHZ-state (genoemd naar Greenberger-Horne-Zeilinger) op.

Paragraaf 3.3 beschrijft alle lokale equivalentieklassen van vier qubits. Dit was mogelijk door het uitbuiten van een toevallige eigenschap van Lie-groepen waarbij geldt dat $SL(2,\mathcal{C}) \otimes SL(2,\mathcal{C}) \simeq SO(4,\mathcal{C})$: het tensorproduct van de groep van complexe matrices met determinant $+1$ met zichzelf is equivalent met de groep van complex orthogonale matrices. Een ingewikkelde veralgemening van de singuliere waardenontbinding naar complex orthogonale equivalentieklassen werd dan ontwikkeld om een volledige classificatie te bekomen, en het bleek dat negen verschillende families met elk een continu aantal equivalentieklassen bestaan. Er werden terug maten voor entanglement voorgesteld, en de verschillende eigenschappen van al die verschillende soorten entanglement worden besproken.

We besluiten het derde hoofdstuk met enkele algemene bemerkingen over equivalentieklassen voor hoger dimensionale systemen.

## Hoofdstuk 4

Het vierde hoofdstuk is het langste hoofdstuk van de thesis en behandelt het probleem van gemengde toestanden van twee qubits. Gemengde toestanden ontstaan doordat pure toestanden *entangled* geraken met de omgeving indien ze niet perfect geïsoleerd zijn. Dit zorgt natuurlijk voor een degradatie van het




**Figuur 2.** Structuur van Hoofdstuk 4

entanglement in de pure toestand, en bemoeilijkt de studie van entanglement in zeer grote mate waardoor zelfs het geval van twee qubits uitermate uitdagend is. Een overzicht van hoofdstuk 4 vindt u in figuur 2.

- In paragraaf 4.2 wordt een natuurlijke ontbinding van alle mogelijke toestanden van twee qubits bekomen in enerzijds lokale en anderzijds globale parameters. Het centrale resultaat is het bewijs van de existentie van de Lorentz singuliere waardenontbinding (LSVD), waarbij in plaats van unitaire equivalentieklassen equivalentieklassen onder Lorentztransformaties beschouwd worden. Dit volgt uit een toevallige eigenschap van Lie-groepen, nl. $SL(2,\mathcal{C}) \simeq SO(3,1)$. Er wordt een reële matrix-parameterizatie voorgesteld van een toestand van twee qubits waarin lokale filteringoperaties overeenkomen met linkse en rechtse vermenigvuldiging met Lorentztransformaties. In analogie met de SVD is de generische normaalform uniek en diagonaal, waarbij de Lorentz singuliere waarden een variationele betekenis krijgen analoog aan deze van de singuliere waarden. Er wordt aangetoond hoe deze normaalvorm van een toestand alle informatie bevat over het al dan niet aanwezig zijn van entanglement, en er worden op natuurlijke wijze maten voor entanglement bekomen. Daarenboven genereert deze ontbinding een continue parameterizatie van oppervlakken met constant entanglement, waardoor deze LSVD uitermate geschikt is om verschillende maten van entanglement met elkaar te vergelijken. De Lorentz singuliere waardenontbinding is dan ook het centrale gegeven in dit vierde hoofdstuk.

- Paragraaf 4.3 veralgemeent het quantum sturings-theorema van hoofdstuk 2 naar het gemengde geval. Een heel aantrekkelijk geometrisch beeld van alle gemengde toestanden van twee qubits wordt bekomen, waarin lokale transformaties als Lorentz boosts of contracties weerspiegeld worden. De relevantie hiervan ligt in het feit dat dit inzicht geeft in de mogelijke lokale transformaties van quantumtoestanden. Paragraaf 4.1 behandelde ook het quantum sturingstheorema in het geval van hogere dimensies.

- In paragraaf 4.4 worden een heel aantal entanglement maten voor gemengde toestanden van twee qubits besproken en vergeleken. Eerst en vooral wordt een nieuwe afleiding gegeven voor het berekenen van de maat "entanglement of formation" (EoF), wat ook direct de nodige en voldoende voorwaarden oplevert voor een gemengde toestand van twee qubits om entangled te zijn. Een expliciet algoritme voor de berekening van de optimale decompositie wordt bekomen, en er wordt aangetoond hoe dit alles kan veralgemeend worden voor hoger dimensionale systemen. De connectie met de Lorentz singuliere waardenontbinding wordt blootgelegd, waardoor een compleet nieuwe variationele karakterizatie bekomen wordt van de entanglement maat EoF. Dit levert op zijn beurt een nieuw bewijs op voor de nodige en voldoende voorwaarde voor het partieel transpositie criterium van Peres en Horodecki voor entanglement. Verder worden enkele niet-triviale eigenschappen bewezen over de eigenwaarden van de partiële transpose van een densiteitsoperator van twee qubits, en worden de entanglement maten "Negativity", "Relative Entropy of Entanglement", "Fidelity" en "Bell-CHSH ongelijkheden" kort besproken. Vervolgens worden onder- en bovengrenzen voor al deze entanglement maten afgeleid als functie van elkaar. Dit werd mogelijk door expliciet gebruik te maken van de Lorentz singuliere waardenontbinding en van geavanceerde technieken voor het differentiëren van matrices. In het geval van "Bell-CHSH ongelijkheden" was deze afleiding bijzonder interessant aangezien het een antwoord verstrekte op de sinds lang openstaande vraag van het classificeren van alle toestanden die Bell-ongelijkheden schenden. In het laatste stuk van paragraaf 4.4 werden dan uiteindelijk de lokale filtering operaties afgeleid die alle geïntroduceerde entanglement maten maximalizeren. Zoals verwacht zijn dit deze die beantwoorden aan de inverse van de linkse en rechtse Lorentz transformaties die optreden in de Lorentz singuliere waardenontbinding.

- Paragraaf 4.5 behandelt terug een fundamentele vraag: hoe kan optimaal gebruik gemaakt worden van een gemengde toestand van twee qubits om te teleporteren. Teleportatie is een fundamenteel gegeven in quantuminformatietheorie aangezien het toelaat alle globale operaties lokaal uit te voeren indien klassieke communicatie en entanglement beschikbaar zijn. In geval van filtering operaties op gemengde toestanden zijn de optimale operaties direct af te leiden uit de resultaten van paragraaf 4.4, maar het nadeel hierbij is dat geen rekening gehouden wordt met de kans op succes. In het geval wel rekening gehouden wordt met de kans op succes, wordt aangetoond dat het optimale lokale protocol bekomen wordt door 1-wegs klassieke preprocessing toe te laten. Aan de hand van technieken van semidefiniet programmeren wordt het optimale protocol bekomen, waarbij de afleiding op zich interessant is aangezien het aantoont hoe een optimizatie over de PPT-klasse van operaties (een klasse die ook niet-fysische operaties bevat) toch kan leiden tot een optimum dat fysisch implementeerbaar is. De bekomen resultaten worden geïllustreerd aan de hand van het quantum sturingstheorema, en een geometrische interpretatie voor de optimale "fidelity" wordt bekomen. Daarenboven zijn de resultaten van belang bij het construeren van entanglement distillatieprotocols.

- In paragraaf 4.6 worden distillatieprotocols besproken. Hierbij is het de bedoeling toestanden met maximaal entanglement te distilleren uitgaande van een groot aantal gemengde toestanden met entanglement. Eerst tonen we aan hoe de gekende distillatieschema's geoptimalizeerd kunnen worden. Daarna worden deze veralgemeend zodat ze toepasbaar worden voor operaties op meerdere copieën. Een complete karakterizatie van alle lokale permutaties die producten van Bell toestanden omzetten in andere producten van Bell toestanden wordt bekomen, en aan de hand daarvan worden nieuwe distillatieschema's afgeleid die beter presteren dan alle gekende schema's. In een tweede stuk wordt de distillatie van toestanden van lage rang besproken, waarbij ook de unieke toestanden gekarakterizeerd worden waarvoor een eindig aantal copieën volstaat om perfecte Bell toestanden te bekomen. Deze gevallen corresponderen tot de niet-generische klasse die afgeleid werd in het kader van de Lorentz singuliere waardenontbinding.

- De volgende paragraaf handelt over gemengde toestanden met maximaal entanglement. Dit stuk werd geïnspireerd door hetvolgende vraagstuk: gegeven een qubit in een pure toestand (i.e. maximale informatie) en een andere in een maximaal gemengde toestand (i.e. geen informatie), is het mogelijk een globaal unitaire operatie op deze qubits uit te voeren zodanig dat beiden *entangled* worden met elkaar? Meer algemeen, de vraag die we ons stelden was welke eigenvectoren een densiteitsoperator moet hebben zodanig dat zijn entanglement maximaal is voor het gegeven spectrum (i.e. gegeven eigenwaarden). Het antwoord op deze vraag geeft automatisch ook een karakterizatie van toestanden met maximaal entanglement voor gegeven (globale) entropie, of equivalent hiermee de karakterizatie van de toestanden met maximale entropie voor gegeven entanglement.

  In het geval van schending van Bell-CHSH ongelijkheden werd reeds in paragraaf 4.4.5 aangetoond dat de eigenvectoren moeten gekozen worden zodanig dat de toestand Bell-diagonaal is. In het geval van "Entanglement of Formation", "Relative Entropy of Entanglement" en "Negativity" daarentegen geldt een andere unieke oplossing: de eigenvectoren behorende tot de grootste eigenwaarde en de derde grootste eigenwaarde moeten toestanden zijn met maximaal entanglement, terwijl de twee andere eigenvectoren geen entanglement mogen bevatten.

  Een interessant gevolg is het feit dat een volledige karakterizatie van de bol van toestanden zonder entanglement rond de maximaal gemengde toestand bekomen wordt. Daarenboven is het nu een kleine rekenoefening om de toestanden met maximaal entanglement voor een gegeven entropie te bekomen. Indien men de identificatie maakt tussen entanglement en energie enerzijds, en entropie en thermodynamische entropie anderzijds, dan leidt dit tot de conclusie dat de geïntroduceerde toestanden *representatieve* toestanden zijn.

- Paragraaf 4.8 benadert de convexe set van quantumtoestanden vanuit het perspectief van een ruimte met een Hilbert-Schmidt metriek. Dit levert mooie inzichten op over de convexe set van toestanden zonder entanglement (i.e. *separabele* toestanden). De afstand van een toestand tot de convexe ruimte van separabele toestanden wordt gekarakterizeerd. Dit wordt gebruikt om aan te tonen dat het volume van de set van gemengde tripartite W-toestanden niet van maat nul is. Daarnaast worden doorsnedes van de convexe toestandsruimte onderzocht, waarin speciale aandacht besteed wordt aan de grens tussen de toestanden met *entanglement* en de *separabele* toestanden.

- In paragraaf 4.9 gaan we er van uit dat een gemengde toestand van twee qubits eigenlijk een pure toestand is die gedeeld wordt door drie of vier partijen, maar waarbij één of twee partijen onzichtbaar zijn. De vraag die zich dan opdringt is de volgende: welke lokale operaties moeten de *onzichtbare* partijen doen zodanig dat de twee *zichtbare* partijen uiteindelijk zoveel mogelijk entanglement delen? Dit vraagstuk blijkt heel analoog te zijn aan de karakterizering van de entanglement maat "Entanglement of Formation", en daardoor kunnen gelijkaardige technieken toegepast worden als ontwikkeld in paragraaf 4.4. In het geval van vier partijen tonen we aan dat klassieke communicatie tussen de twee onzichtbare partijen niet helpt, en leiden we ook de (unieke) optimale strategie af indien enkel von-Neumann metingen in acht genomen worden. Dit was mogelijk dankzij een lemma dat ook op zich interssant is: elke unitaire matrix kan (uniek) ontbonden worden als het product van drie matrices, waarbij de eerste en de derde (reëel) orthogonaal zijn en de tweede complex diagonaal.

Het tweede deel van de thesis behandelt enkele specifieke onderwerpen in de quantuminformatietheorie. In een eerste hoofdstuk wordt eerst nagegaan hoe klassieke informatie geëncodeerd en geëxtraheerd kan worden in en uit quantumtoestanden. Daarna wordt nagegaan hoe de parameters van een onbekende Hamiltoniaan geïdentificeerd kunnen worden: we zetten de eerste stappen voor het construeren van een algemene quantum-identificatie-theorie. In een volgende hoofdstuk wordt de evolutie van quantumsystemen besproken vanuit het oogpunt van het feit, dat de dynamica beschrijft hoe een systeem op een later tijdstip gecorreleerd is met het systeem op een vroeger tijdstip. Er blijkt een perfecte analogie of dualiteit te bestaan met entanglement, waar niet-lokale correlaties beschreven worden. Er wordt dieper ingegaan op de klassieke- en quantum-capaciteit van een quantum kanaal, en er wordt een expliciete uitdrukking gevonden voor de maximale hoeveelheid entanglement die een niet-lokale Hamiltoniaan kan creëren.

## Hoofdstuk 5

In dit hoofdstuk staat het begrip informatie centraal. In paragraaf 5.1 wordt enige intuïtie gevormd over de hoeveelheid informatie die een qubit kan bevatten door alle mogelijke meetstrategieën te beschouwen. De meest algemene meting levert een 4-dimensionale probabiliteitsverdeling op (i.e. drie vrijheidsgraden wegens normalizatie), maar, in tegenstelling tot het klassieke geval, blijken er een heel aantal beperkingen te bestaan voor de verschillende waarden van de probabiliteiten. Voorgesteld in de probabiliteitssimplex, blijkt dat enkel de toestanden die zich situeren binnenin een ellepsoïde die op haar beurt in de simplex ligt, fysisch zijn. Dit zal impliceren dat een qubit minder dan 1 (klassieke) bit informatie kan bevatten, hoewel zijn parameterruimte continu is.

In een volgende paragraaf worden de intuïtieve resultaten van de eerste paragraaf hard gemaakt aan de hand van een aantal gevierde theorema's. We bespreken het *no-cloning* theorema waarin bewezen wordt dat een onbekende quantumtoestand niet gekloond kan worden. Dit is noodzakelijk aangezien een quantumtoestand anders een oneindig aantal klassiek bits van informatie zou kunnen bevatten. Dit impliceert natuurlijk ook dat een meting de quantumtoestand verstoort. We geven een nieuwe afleiding voor de bovengrens in het aantal elementen dat een POVM-meting moet bevatten om de *wederzijdse* informatie te maximalizeren. We illustreren dit met het centrale theorema van Holevo waarin een bovengrens voor de *toegankelijke* informatie in quantumtoestanden afgeleid wordt. Deze bovengrens is cruciaal in de quantuminformatietheorie aangezien het de strikte bovengrens levert van de klassieke capaciteit van een quantum-communicatie-kanaal. Dit korte overzicht wordt besloten met een korte bespreking van hoe beslissings-problemen in de quantummechanica opgelost kunnen worden aan de hand van semidefiniet programmeren.

De derde paragraaf bespreekt een heel fundamenteel probleem, waarbij de vraag gesteld wordt hoe een meting op een quantumsysteem zodaning kan uitgevoerd worden dat de balans informatie winning - verstoring optimaal is. Meer specifiek beschouwen we een dynamisch systeem met een aantal onbekende parameters in de Hamiltoniaan (i.e. een soort quantum-parameter-identificatie-probleem). De vraag die we ons nu stellen is hoe optimale detectiestrategieën ontwikkeld kunnen worden om de onbekende parameters te schatten. In tegenstelling tot klassieke systemen, waar men in principe oneindig gevoelige metingen kan uitvoeren op een systeem zonder het te verstoren, bestaat er een optimale gevoeligheid van een meting: indien ze te gevoelig is, compromitteren we de volgende metingen door de verstoring geïnduceerd door de meting, en indien te weinig gevoelig, wordt niet voldoende informatie vergaard. De optimale strategie blijkt te bestaan uit het uitvoeren van continue maar oneindig zwakke metingen, waarin men op elk moment het systeem koppelt met een meter (i.e. er ontstaat entanglement tussen het systeem en de meter).

Door gebruik te maken van Ito calculus blijkt er een intrigerende connectie te bestaan tussen dergelijke quantumsystemen en Kalman filtering. Dit is niet geheel verwonderlijk aangezien de quantumtoestand inderdaad een parameterizatie is van een probabiliteitsdistributie, net zoals dit het geval is bij Kalman filtering. Het is echter uitermate interessant het onderliggende equivalente klassieke stochastische systeem te beschouwen. Hieruit blijkt dat de hoeveelheid ruis aanwezig op het meetresultaat volledig geregeld kan worden door de gevoeligheid van de meting, maar dat de hoeveelheid ruis die inwerkt op het systeem omgekeerd evenredig is met deze gevoeligheid. De evenredigheidsconstante blijkt niets anders te zijn dan $\hbar/2$ met $\hbar$ de constante van Planck, wat een heel mooie illustratie oplevert van de onzekerheidsrelatie van Heisenberg.

Door de equivalentie van het quantumprobleem met een klassiek stochastisch systeem kunnen nu welbekende technieken van klassieke identificatie aangewend worden. Deze leiden tot het vinden van nieuwe (i.e. sterkere) *standaard quantum limieten* die een absolute grens (inherent aan de quantummechanica) opleggen aan de nauwkeurigheid waarmee parameters geschat kunnen worden. We besluiten dat tijdscontinue metingen meer informatie kunnen extraheren dan projectieve metingen, wat op zich een heel interessant gegeven is. Expliciete formules worden gevonden in het geval van bv. de identificatie van een onbekende constant inwerkende kracht op een quantumdeeltje, en veralgemeningen naar tijdsafhankelijke parameters worden voorgesteld.

## Hoofdstuk 6

Het laatste hoofdstuk handelt over de tijdsevolutie van quantumsystemen. In een eerste paragraaf wordt nagegaan hoe de meest algemene evolutie van een quantumsysteem beschreven kan worden, inclusief de beschrijving van open quantumsystemen. Mathematisch gezien komt dit neer op de karakterizatie van compleet positieve lineaire afbeeldingen. Een dergelijke afbeelding kan altijd beschreven worden als de koppeling van een input met een output, zodat automatisch een tensorproduct-structuur verkregen wordt. Van zodra de Hilbertruimte een tensorproduct is van lager dimensionale Hilbertruimten komt men natuurlijk terecht bij de notie van entanglement, zodat er een soort dualiteit ontstaat tussen de beschrijving van compleet positieve lineaire afbeeldingen en entanglement. Unitaire evolutie bijvoorbeeld leidt tot maximale correlatie tussen een systeem op verschillende tijdstippen, aangezien deze evolutie reversibel is. Zoals verwacht zijn de corresponderende quantumtoestanden dan deze met maximaal entanglement of maximale correlaties. Deze analogie was ergens wel verwacht vanuit het perspectief van speciale relativiteit, waarin ruimte en tijd een analoge rol spelen.

In de eerste paragraaf leiden we een volledige karakterizatie af van alle mogelijke compleet positieve lineaire afbeeldingen. We illustreren het verschil met positieve maar niet compleet positieve afbeeldingen en leiden enkele interessante eigenschappen af van de Kraus operatoren.

De set van compleet positieve afbeeldingen is convex, en dit blijft zo indien we vereisen dat de afbeelding spoor-behoudend is. Deze spoor-behoudende afbeeldingen worden ook quantum kanalen genaamd. In paragraaf 6.2 worden op een nieuwe en elegante manier alle extreme punten van deze quantumkanalen gekarakterizeerd. Nodige en voldoende voorwaarden voor de extremaliteit van een kanaal worden bekomen, en veralgemeningen besproken. Daarnaast wordt de relevantie besproken van deze extremale kanalen in de context van quantum-informatietheorie.

Paragraaf 6.3 gaat dieper in op de dualiteit tussen quantum kanalen en quantum toestanden, waarin evolutie als een andere vorm van entanglement geïnterpreteerd kan worden. De nodige en voldoende voorwaarden worden afgeleid voor een quantum kanaal om entanglement te kunnen transporteren, en de toestand die door het kanaal moeten gezonden worden om de hoeveelheid entanglement te maximalizeren wordt gekarakterizeerd. Dit is van belang in de studie van de quantum capaciteit van een quantum kanaal. De quantum capaciteit van een kanaal is gegeven door de hoeveelheid quantum informatie dat een gegeven kanaal kan overzenden, en is direct gerelateerd aan de hoeveelheid entanglement dat het kanaal kan overzenden.

Er werd reeds dieper ingegaan op de klassiek capaciteit van een quantum kanaal, en daarin zijn vele niet-triviale resultaten gekend. Nieuw is echter dat men deze resultaten via de dualiteit tussen kanalen en toestanden direct kan gebruiken om de hoeveelheid klassieke correlaties aanwezig in een quantumtoestand (met of zonder entanglement) te karakterizeren. Dit is interessant aangezien men bij toepassingen van de quantuminformatietheorie (e.g. quantumcomputing) uiteindelijk enkel geïnteresseerd is in klassieke correlaties.

Paragraaf 6.4 is in zekere zin een directe vertaling van de resultaten van hoofdstuk 4 in de taal van quantumkanalen. Het is inderdaad een feit dat het duale beeld van qubit-kanalen leidt tot gemengde toestanden van 2 qubits, waardoor interessante decomposities zoals de Lorentz singuliere waardenontbinding ook van direct belang zijn in de studie van quantumkanalen. In analogie met hoofdstuk 4 worden equivalentieklassen van kanalen geformuleerd onder unitaire en onder filtering operaties, en het geometrische beeld van het quantumsturingstheorema is ook hier geldig. Gebruik makend van de normaalvormen afgeleid in het geval van gemengde toestanden van 2 qubits wordt een expliciete en elegante parameterizatie van alle extreme qubit-kanalen bekomen. Dit maakt het mogelijk vooruitgang te boeken op het vlak van de berekening van de klassieke capaciteit van de extreme qubit-kanalen, aangezien dit probleem sterk gerelateerd is aan de berekening van de *entanglement of formation* van gemengde toestanden van 2 qubits. Daarenboven wordt aangetoond hoe dit probleem gerelateerd is aan het probleem om klassieke correlaties tussen twee verschillende partijen te creëren zonder klassieke communicatie. Tenslotte wordt ook dieper ingegaan op de quantum capaciteit van de extreme qubit kanalen, en wordt deze berekend in het geval van bistochastische kanalen van rang 2.

Paragraaf 6.5 veralgemeent de voorgaande resultaten naar afbeeldingen van toestanden die zelf al *entangled* zijn. Er wordt aangetoond hoe een globale unitaire transformatie lokaal kan geïmplementeerd worden door gebruik te maken van een kleine hoeveelheid entanglement (i.e. geen teleportatie), en we bespreken de rol van semidefiniet programmeren in die context.

Tenslotte wordt het volgende fundamentele probleem opgelost: gegeven een niet-lokale Hamiltoniaan, op welke toestanden moet men deze Hamiltoniaan laten inwerken opdat de maximale hoeveelheid entanglement zou gecreëerd worden? Dit is van groot belang voor experimentatoren aangezien het in realistische situaties heel moeilijk is entanglement aan te maken. Tegelijkertijd lost dit het volgende interessante probleem op: gegeven een zekere niet-lokale Hamiltoniaan, wat is de capaciteit van deze Hamiltoniaan om klassieke informatie over te dragen? Het blijkt dat optimale entanglement creatie bekomen wordt door de Hamiltoniaan te laten inwerken op een systeem dat al entanglement bevat. Een volledige oplossing wordt bekomen in het geval van de Ising- en de anisotropische Ising-interactie, waarvoor de entanglement-capaciteit en de klassieke capaciteit bepaald worden.

## CHAPTER 1

# Introduction

The twentieth century gave birth to many scientific revolutions. Two theories that had a major impact on technology and henceforth on our everyday life are quantum mechanics and information theory. Quantum mechanics led to a deep understanding of the basic building blocks of all substances, culminating in many new inventions such as the ubiquitous transistor. Information theory originated from the need of making optimal use of communication devices [**192**], and immediately led to a beautiful abstract mathematical formulation of the concept of information[1].

The fascination for physics originates from our desire to predict and understand the behaviour of all substances. Therefore the language of mathematics was created, and this abstract tool enabled us to transcend the empirical world and to discover beautiful fundamental laws of nature.

As natural as trying to understand the empirical world, however, is our desire to make abstract concepts such as knowledge and reason tangible. This led to the field of logic, the very basics of mathematics, and to the field of information theory, which quantifies information in a mathematical way[2].

It has long been known that there is a deep connection between quantum physics and information theory (see e.g. Jaynes [**131, 132**]). This is especially true in the field of statistical physics, e.g. where the concept of entropy quantifies the amount of disorder of the system; similarly, the Shannon entropy [**192**] is given by the same formula and quantifies our lack of information about a system and therefore the amount of information gathered when a measurement is done. More strongly, since the work of Szilard [**198**], Landauer [**145**] and Bennett [**24**], it has become clear that information is physical an sich [**172**]:

[1]In what follows we will use the term information theory in a very broad context, embracing fields like probability theory and control theory.

[2]It is interesting to note that the role of mathematics in the case of physics and information theory is completely different: in physics, mathematics is the endpoint of abstraction, while in information theory, it is the most concrete form of describing rational knowledge.

erasing information for example is only possible to the cost of an increase of the entropy of the environment.

In quantum mechanics, the connection is even more direct: the wave function or more generally, the density operator, is nothing more than the complete parameterization of the information that an observer has about a quantum system.

Quantum mechanics is essentially a theory that allows us to predict measurement statistics of future experiments, and the laws of quantum mechanics dictate how we have to update the probability distributions associated with all measurements outcomes. Of course, this also holds for classical physics, but the strange thing in quantum mechanics is the fact that it seems impossible to describe measurements on elementary physical systems as closed mechanical systems without invoking the role of the observer: the very act of observation of a quantum system causes a stochastic disturbance of the system that cannot be made arbitrary small by ingenious engineering. This fact lies at the heart of the difference between quantum mechanics and classical mechanics.

We can rephrase this as follows: acquiring information about a quantum system (i.e. a measurement) automatically implies the disturbance of the system in a way that is stochastic but completely conditioned by the measurement outcome. The theory of quantum mechanics is therefore a theory of how we have to update our knowledge (by this we mean the predictive power for future experiments) of a quantum system: if no information is acquired, the evolution will turn out to be unitary and reversible, while a measurement enables us to refine our knowledge (by using Bayes rule). Note that the strange concept of *collapse* in quantum mechanics becomes very natural once the following interpretation of quantum mechanics is accepted: quantum states are states of knowledge, and the laws of quantum mechanics dictate how these probability distributions should be updated. Quantum mechanics is therefore in some sense not a theory that provides a description of reality but rather a theory of how we have to update our knowledge and process our information. In the words of Hartle[3] [**108**]: "... a quantum state is not an objective property of an individual system, but is that information, obtained from a knowledge of how the system was prepared, which can be used for making predictions about future measurements... The *reduction of the wave packet* does take place in the consciousness of the observer, not because of any unique physical process which takes place there, but only because the state is a construct of the observer and not an objective property of the physical system."

[3]See also the reformulation of the words of G. 't Hooft [**199**] by A. Peres[**169**]: "Quantum mechanics is *not* a theory about reality; it is a prescription for making the best possible predictions about the future, if we have certain information about the past."

Another crucial feature of quantum mechanics is the fact that complete knowledge of a system does not imply predictive power about all possible measurement outcomes: the very nature of quantum mechanics is inherently statistical of character[4]. More specifically, consider quantum states that are described in a finite dimensional Hilbert space. It is clear that there exists a continuum of different quantum states. Nevertheless, a quantum measurement can only reveal a limited amount of information: one of the fundamental theorems of quantum information theory dictates that at most $log_2(n)$ bits of information can be revealed by a measurement on a quantum system of an $n$-dimensional Hilbert space [**112**]. This is of course completely compatible with the fact that a quantum measurement disturbs the system: otherwise a refinement of the measurement would ultimately reveal an infinite amount of information. A very appealing theorem in the same spirit says that an unknown quantum system cannot be cloned [**245**]: if this were possible, then one could make an infinite amount of copies and completely determine the state. Note that these considerations lead to a strange asymmetry about quantum systems: in principle one needs a large number of bits to describe a random quantum state, while the state itself can only be used to encode a very limited amount of information.

Let us now briefly recall the fundamental postulates of quantum mechanics[5]. By definition, a quantum state encodes all the information we have about a system. The fundamental postulate is that a (projective) quantum measurement is described in a complex Hilbert space and corresponds to a resolution of the identity in this Hilbert space (i.e. a complete set of orthonormal complex vectors), and that every such possible resolution of the identity corresponds to a feasible measurement. Moreover we require that the probability of getting a specific outcome $\alpha$ is independent on the way the other $n-1$ vectors are chosen; this is the so-called non-contextuality condition. In an amazing paper [**99**], Gleason showed that the above conditions are sufficient to prove that a quantum state must be described by a positive (semi)-definite operator $\rho$ of the dimension of the Hilbert space with trace equal to 1, and that the probability of getting the outcome associated with a specific direction $\Pi_\alpha$ (which is the projector corresponding to this direction) is given by

$$p_\alpha = \mathrm{Tr}(\rho\Pi_\alpha). \tag{1}$$

The crucial point of this equation is that the probabilities obtained are both linear in $\rho$ and in $\Pi$: quantum mechanics is a linear theory. The set of normalized (i.e. Trace 1) positive semidefinite operators is convex, and the extreme points of this set correspond to so-called pure states $\rho = |\psi\rangle\langle\psi|$ (i.e. rank 1

[4]A typical example of the statistical character of quantum mechanics is the following: consider a photon with horizontal polarization and measure its polarization through a polarizer rotated over an angle of $\pi/4$. Then with probability 1/2 the photon will pass, and with probability 1/2 it will be absorbed.

[5]We refer to the article of Chris Fuchs “Quantum Mechanics as Quantum Information”[**90**] for an eloquent exposition on the foundations of quantum mechanics.

operators): these states correspond to the situation where we have maximal possible knowledge about the quantum system under consideration.

A natural question is now whether the evolution of an unobserved quantum system is also described by a linear map. This is what one would expect, as this is the only sensible way of ensuring that probability distributions remain normalized (of course this no longer holds once measurements are done). The linearity of evolution of quantum mechanics is the next postulate of quantum mechanics, and the evolution of quantum states is therefore described by linear maps. A density operator should be mapped onto another one, and this requires that the map should be positive; the map should even be completely positive, as otherwise positivity is not assured if the map acts on a subsystem of a quantum system. The most general map is therefore of the form[**142**]:

$$\Phi(\rho) = \sum_i A_i \rho A_i^\dagger \tag{2}$$

with $\{A_i\}$ a set of Kraus operators obeying the condition $\sum_i A_i^\dagger A_i = I$. If the quantum system is closed however, we expect that a state of maximal knowledge (a pure state) remains pure. This implies that the map must be unitary in that case, and the generator of the unitary group is coined the Hamiltonian of the quantum system. This situation corresponds to evolution as described by Schrödinger's equation.

We still need to specify how a quantum state changes due to the action of a measurement. Suppose a specific von-Neumann measurement has been done and the outcome $\alpha$ has been obtained. If we next immediately repeat exactly the same measurement on the state obtained, then experiments confirm that always the same outcome $\alpha$ arises. This is only possible if the state after a projective measurement is given by the projector associated to the outcome (indeed, this is the only way to assure that all other measurement outcomes have probability zero); this is of course reminiscent of the way Bayes rule [**17**] is used to update a probability distribution.

Let us next describe how to analyse a situation in which two closed systems with respective Hilbert spaces of dimension $n_1$ and $n_2$ are joined together. One larger Hilbert space arises, and this Hilbert space is the tensor product (also called trivial tensor product or Kronecker product) of both original ones [**69, 230, 5**]: this ensures that the global system is still linear in the observables of both systems, that the complete state is still normalized, and that local density operators are unaffected by distant measurements. The Hilbert space of the joint system is therefore of dimension $n_1 \times n_2$, and all possible complete sets of orthogonal projectors in this space correspond to feasible measurements. Moreover, all possible density operators in this larger Hilbert space correspond to physical situations, and not only the tensor products of *local* density operators. This is the origin of the fact that quantum states can be entangled: once two quantum systems have been in contact with each other, interacted (i.e. "rotated" in the larger Hilbert space) and again separated, it is generally not possible anymore to describe the two quantum systems as two different systems, but they are intimately entwined or entangled with each other.

Entanglement is definitely one of the strangest consequences of quantum mechanics. It follows directly from the superposition principle, which is an inherent feature of Hilbert space. The strangeness manifests itself most succinctly once a quantum state of two space-like separated entangled particles is considered [**82**]: a measurement of one particle will immediately (i.e. faster than the speed of light) cause a "collapse" of the state of the other one conditioned on the measurement result. This looks very much like a paradox, and the resolution is intimately connected to the concept of information: although a collapse occurred, no information whatsoever has been transmitted as a quantum measurement is intrinsically stochastic.

The superposition principle and the existence of entanglement are central features of quantum mechanics. Until the beginning of the 1980's, entanglement was mostly considered as something annoying because it caused a system to get entangled with the environment, which leads to decoherence. Following the intuition of Feynman [**85, 86**] however, physicists started to think on how to exploit entanglement as a resource rather than an inconvenience. Following Deutsch [**66**], Deutsch and Jozsa [**68**] and Simon [**195**], it was clear that quantum systems could be used as computational machines (i.e. quantum computers) that could solve some problems much more efficiently than classical computers. The first overwhelming evidence of the power of quantum computers was the invention by Shor of an algorithm that could factor large integers in a polynomial amount of time [**193**]. This was followed by the demonstration by Peter Shor that quantum error correction was possible [**194**], therefore indicating that "it is possible to fight entanglement with entanglement" [J. Preskill] and that it would not be irrealistic in the long term to build quantum computers. Meanwhile, other applications of entanglement and of the superposition principle emerged such as quantum cryptography [**27, 84**] and quantum communication protocols, for which the communication complexity is greatly reduced in comparison with classical models [**46, 178**]. However, in this thesis we will not dwell on these subjects. Instead, we will investigate the phenomenon of entanglement with the aim of characterizing and quantifying it at a fundamental level: despite the fact that quantum entanglement is directly responsible for the appealing power of quantum computation and quantum communication, its nature is still only marginally understood and many of its mysteries remain to be unravelled. In some sense, one could easily justify this fundamental research by the following rule of conduct of J.A. Wheeler[6]: "In any field, find the strangest thing and explore it."

[6]Wheeler had a true gift for timely quotations; one could argue that he created the field of quantum information theory with the following Sybillian prophecy[**238**]: "It from bit."

In a first part of this thesis, we give a unified treatment of how to describe entanglement in pure and mixed distributed quantum states.

- Chapter 2 treats the simplest entangled systems: pure bipartite states. We show that essential results such as teleportation and the necessary and sufficient conditions for single-copy entanglement transformations can be derived almost directly from the quantum steering Theorem of E. Schrödinger [**187**], yielding a new global insight into the structure of pure bipartite entanglement.

- Chapter 3 treats the problem of describing pure multipartite entangled states where the number of parties exceeds two. This problem is much more complicated as it involves multilinear algebra as contrasted to matrix algebra in the bipartite case. We derive a very general normal form for generic multipartite tensors under the action of local $SL(n, \mathcal{C})$-operations; the normal form obtained is interesting as there is supporting evidence for the fact that it is unique up to local unitary operations. Physically, $SL(n, \mathcal{C})$-operations can be implemented probabilistically using local filters. The normal form of a state corresponds to the state with the maximal amount of entanglement (or maximal non-local correlations) to which the given state can be transformed probabilistically. This enables to generalize the concept of Bell-states to the most general multipartite setting. We show how this normal form gives rise to entanglement monotones, and how a similar result holds for the mixed state case. We proceed by giving a complete classification of all possible pure states in the $2 \times 2 \times N$-case under the action of $SL(2, \mathcal{C})$-operations, generalizing the celebrated results on $2 \times 2 \times 2$-states of Dür et al. [**80**]. In a successive section, we derive the analogue of the singular value decomposition under complex orthogonal congruence, which yields the complete classification of four-qubit states under the action of $SL(2, \mathcal{C})$-operations.

- Chapter 4 treats the problem of describing entanglement in mixed bipartite states. We mainly concentrate on the simplest case of two qubits. For an overview of this chapter, we refer to figure 1 at the beginning of chapter 4. The central result of this chapter is the existence of the Lorentz singular value decomposition, being the analogue of the singular value decomposition with proper orthochronous Lorentz transformations instead of with unitary matrices; this enables one to effectively separate the local from the non-local properties of the density operator. As a first application, we generalize the quantum steering Theorem to the case of mixed states of two qubits, yielding an appealing geometrical Bloch-sphere representation of all mixed states of two qubits. Next we give a new derivation of the calculation of the entanglement of formation, briefly discuss the entanglement measures negativity, relative entropy of entanglement, fidelity and Bell CHSH-inequalities, and present a comparison of all these measures. We show how the general results of Chapter 3 apply and give rise to optimal filtering procedures. Next the optimal teleportation protocol with mixed states is derived, and we prove how local operations assisted by classical communication (LOCC) can increase the fidelity of teleportation. In a successive section, it is discussed how pure maximally entangled states can be distilled from a large amount of mixed states, and the best known distillation protocol is derived. Subsequently, the notion of maximally entangled mixed states is introduced, as being the unique class of mixed states whose entanglement cannot be increased anymore by global unitary operations, including the states with maximal entanglement for a given amount of global entropy. Section 4.8 yields a geometric picture of the convex set of separable states, and the final section 4.9 discusses the concept of entanglement of assistance.

The second part of this thesis deals with the description of the evolution of quantum systems when sent through a quantum channel, and with the problem of how to extract (classical) information out of quantum systems. This part naturally follows from the first part, as the description of evolution is equivalent to the description of the correlations between a system at a given time and at a prior time. Mathematically, it turns out that this description is completely equivalent to the problem of characterizing entangled states, and therefore there exists a nice duality between quantum channels (or maps) and entangled quantum states.

- In the first section of chapter 5, we introduce a geometrical picture of the possible measurement outcomes and probabilities associated to a complete POVM-measurement (i.e. a measurement involving ancilla's) of a quantum state. Next we present an (incomplete) overview of classical results about extracting classical information out of quantum systems, emphasizing the physical importance of the fact that finite quantum systems can only reveal a finite amount of information. The last section deals with the following fundamental question: how can one devise a quantum measurement strategy such as to gain as much information as possible while trying to reduce the disturbance introduced by the measurement. More specifically, we devise general quantum parameter estimation schemes based on continuous (indirect) observation of a dynamical system. As an illustrative example, we analyze the canonical scenario of monitoring the position of a free mass or harmonic oscillator to detect weak classical forces, and find out that the use of continuous indirect measurements can reveal more information than direct measurements. Moreover, an intriguing connection between Kalman filtering and the evolution of observed quantum systems is discussed.

- Chapter 6 introduces a unified way of discussing quantum channels through the duality with entangled quantum states. After the characterization of the extreme points of the convex set of completely positive maps (CP-maps), we discuss this duality in the light of the classical and quantum capacity of quantum channels. Next we translate the results derived in the case of mixed states of two qubits to the case of one-qubit channels, yielding appealing representations of the corresponding maps. We end the chapter with some novel results on how to optimally make use of a Hamiltonian to produce entanglement.

In summary, in the first part of this thesis we investigated the fundamental question of describing entanglement, and due to some advanced (multi-)linear algebra we managed to obtain sensible results. The second part is mainly concerned with the dual question of describing quantum evolution, quantum channels and quantum measurement, and due to the duality between maps and states we could shed new light on important aspects of quantum mechanics and quantum information theory. The main interrelations between the different sections are depicted in the flowchart on figure[7] 1.

Parts of this thesis are based on material contained in the papers of Verstraete, Dehaene and De Moor [**211, 212, 213, 214**], Verstraete, Audenaert and De Moor [**210**], Verstraete, Audenaert, Dehaene and De Moor [**209**], Verstraete, Dehaene, De Moor and Verschelde [**215**], Verstraete and Verschelde [**218, 220, 219**], Audenaert, Verstraete and De Moor [**10**], Dehaene, Van der Nest, Verstraete and De Moor [**64**], Verstraete, Doherty and Mabuchi [**216**], Verstraete and Wolf [**221**], Lautsten, Verstraete and van Enk [**146**], Childs, Leung, Verstraete and Vidal [**55**], Wei, Nemoto, Goldbart, Kwiat, Munro and Verstraete [**234**], Verstraete and Rudolph [**217**] and Miyake and Verstraete [**161**].

[7]For an overview of chapter 4, we refer to figure 1 in the introduction of chapter 4.

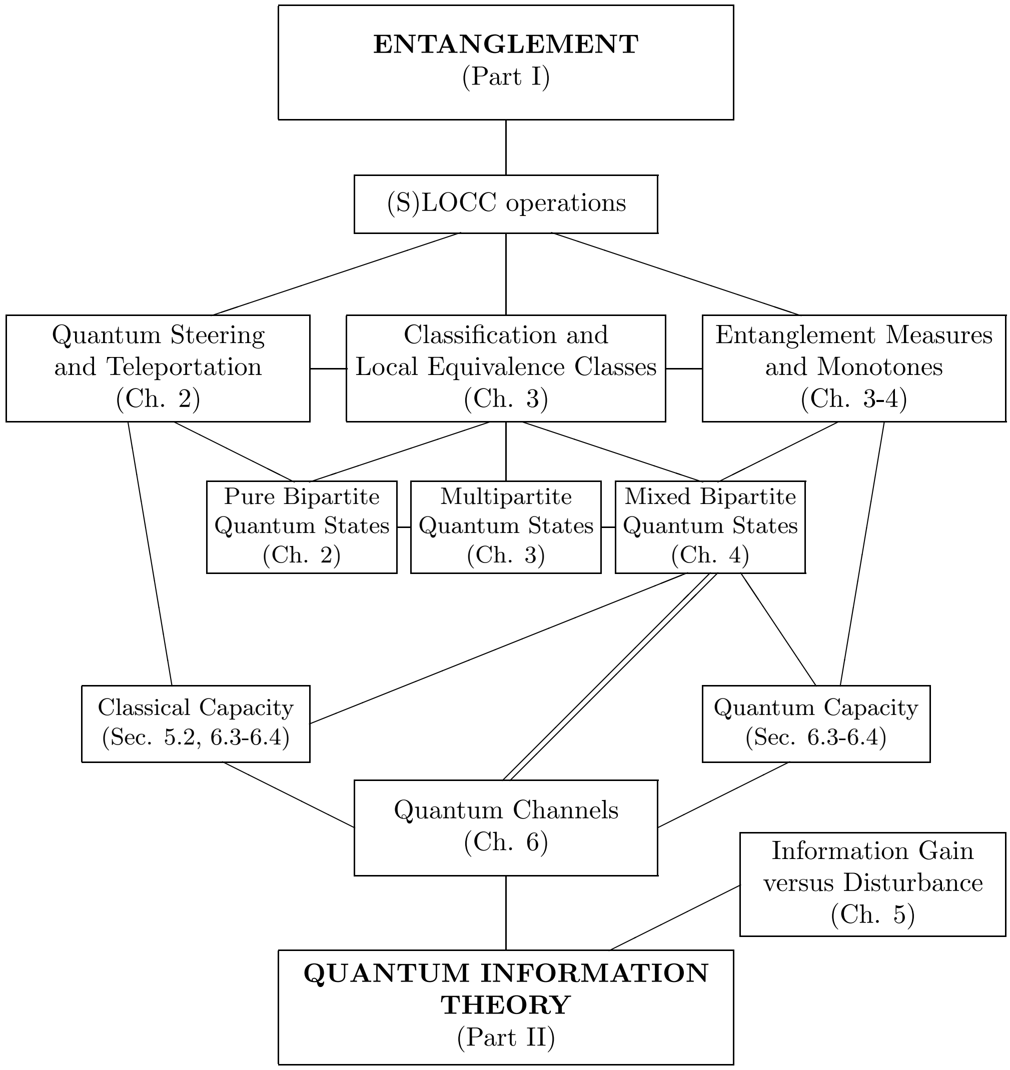



**Figure 1.** Structure of this thesis

# Part 1

# QUANTUM ENTANGLEMENT

## CHAPTER 2

# Entanglement of pure bipartite states

In this chapter, entanglement is discussed in its purest form (i.e. in the pure bipartite case). After a brief historical introduction, the quantum steering Theorem of Schrödinger [**187**] is shown to reveal the basic insight to grasp the strangeness and beauty of entanglement.

### 2.1. The EPR-paradox

Pure quantum states are described in Hilbert space. As a Hilbert space is a complex linear vector space, the superposition of two pure quantum states yields another quantum state. The Hilbert space corresponding to two quantum particles is given by the tensor product of the respective subspaces: $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$: this assures that the superposition principle applies, and that a local measurement does not affect the distant local density operator. A consequence is the fact that the dimension of the Hilbert space scales exponentially with the number of particles [1].

Einstein, Podolsky and Rosen [**82**] pointed out that this superposition principle together with the quantum measurement postulate leads to situations that are in apparent contradiction with the causality principle of relativity. Bohm's [**38**] version of the EPR paradox is as follows: consider two parties $A$ (Alice) and $B$ (Bob) that are very far away from each other. Suppose they both have possession of one qubit, and that their joint state is in the superposition state $|\psi_{AB}\rangle = (|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B)/\sqrt{2}$. If Alice measures her qubit in the $\{|0\rangle, |1\rangle\}$ basis, then the particle of Bob, although very far apart (a time-like interval), will *collapse* into the state $|0\rangle$ or $|1\rangle$ according to the (stochastic) measurement outcome of Alice. Therefore it seems that causality is violated as if there was some kind of "spooky action at a distance". Based on a similar argument [2],

[1] Note that a similar reasoning applies for classical systems: the number of different "states" labelled by $n$ bits scales exponentially with $n$.

[2] A crucial ingredient in the argument was the definition of *elements of physical reality* [**82**]: "If, without in any way disturbing the system, we can predict with certainty the value

Einstein conjectured that quantum theory was incomplete in that the previous paradox would disappear in the presence of extra (hidden) variables not taken into account by quantum theory.

As a reply, Bohr [**39**] introduced the notion of complementarity: "There is no question of a mechanical disturbance of the system under investigation [in the EPR paradox] ... there is essentially the question of an influence on the very conditions which define the possible types of predictions regarding the future behavior of the system." In modern language, Bohr's message was the following: although a *collapse* occurred on Bob's side, there has been no actual transfer of information between Alice and Bob; the only thing that happened is that after the experiment both parties ended up with a perfectly correlated classical bit.

Some thirty years later, John Bell [**20, 21**] made a remarkable discovery that completely changed the EPR-debate; one of the consequences of his discovery implied that either we had to accept the EPR-quantum-nonlocality, or that quantum mechanics predicted measurement outcomes that were in contradiction with experiments. His discovery was the following: the statistical outcomes as predicted by ALL local hidden variable theories (as proposed by Einstein) obey some non-trivial convex constraints (which are now called Bell inequalities) violated by the EPR experiment of Bohm.

Finally, Aspect et al. [**7**] closed the EPR-debate by experimentally confirming the validity of the predictions of quantum theory concerning the EPR-paradox.

The implications of quantum nonlocality are multiple. Shortly after the publication of the EPR-paper, Erwin Schrödinger analyzed the situation very sharply in a series of two papers [**186, 187**]. The paper [**186**] begins as follows:

> When two systems, of which we know the states by their respective representatives, enter into temporary physical interaction due to known forces between them, and when after a time of mutual influence the systems separate again, then they can no longer be described in the same way as before, viz. by endowing each of them with a representative of its own. I would not call that *one* but rather *the* characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought. By the interaction the two representatives (or $\psi$-functions) have become **entangled** ... An other way of expressing the peculiar situation is: the best possible knowledge of a *whole* does not necessarily include the best possible knowledge of all its *parts*, even though they may be entirely separated

of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity."

> and therefore virtually capable of being *best possible known*, i.e. of possessing, each of them, a representative of its own ... It is rather discomforting that the theory should allow a system to be steered or piloted into one or the other type of state at the experimenter's mercy in spite of his having no access to it.

In the previous quotation, Schrödinger coins the term *entanglement*, which has now become ubiquitous in quantum information theory. The mathematical definition of entanglement for pure states is as follows: a pure state $|\psi\rangle_{AB}$ is entangled iff there do not exist local basis in which it can be written as a product state $|\chi\rangle_A|\chi\rangle_B$. If a state is not entangled, than it is separable. An easy way of checking this is to look at the rank of the local density operator $\rho_A = \mathrm{Tr}_B(|\psi\rangle_{AB}\langle\psi|)$ (or $\rho_B$ which has the same eigenvalues): $|\psi\rangle_{AB}$ is entangled iff $\rho_A$ has rank greater than 1. This clarifies the beautiful sentence "the best possible knowledge of a *whole* does not necessarily include the best possible knowledge of all its *parts*": we have the best possible knowledge iff our description of the state is pure, but in the case of entangled states the local density operators are not pure, and nevertheless we have complete knowledge about the state. It is also clear that there should exist a gradation of entanglement: if the local density operator of a pure state is almost pure (i.e. has all small eigenvalues but one) then not much entanglement can be present. This leads to the following natural definition of a state with maximal entanglement: a state is maximally entangled iff its local density operator is proportional to the identity.

## 2.2. Quantum steering and teleportation

In a subsequent paper [**187**], Schrödinger analyzed the following fundamental question: given a bipartite state $|\psi\rangle = \sum_{ij} \tilde{\psi}_{ij}|i\rangle_A|j\rangle_B$, what kind of states can Alice prepare at Bob's side by doing appropriate measurements? Schrödinger called this kind of action quantum steering, and this was exactly the kind of action that Einstein called "spooky". The solution is given by the quantum steering Theorem[3]

**Theorem 1** (Schrödinger [**187**], Hughston, Jozsa and Wootters [**127**])**.** *Given a bipartite state $|\psi\rangle_{AB} = \sum_{ij} \tilde{\psi}_{ij}|i\rangle_A|j\rangle_B$ on a $n\otimes m$ Hilbert space and the local density operator $\rho_B = \mathrm{Tr}_A(|\psi\rangle\langle\psi|)$. Then a POVM measurement on Alice's side can realize the ensemble $\{p_i, \rho_i\}$ at Bob's side if and only if $\rho_B = \sum_i p_i\rho_i$.*

[3] I'm very grateful to the grand grand-son of Schrödinger, Terry Rudolph, for sharing his enthusiasm about quantum steering. Remarkably, the concept of *quantum steering* is almost unknown in the quantum information community.

Proof: Consider the $n \times m$ matrix $\tilde{\psi}_{ij}$; it is readily verified that $\rho_B = \tilde{\psi}^\dagger \tilde{\psi}$. It is sufficient to prove the Theorem for pure state ensembles, as the case of mixed states arises by adding pure elements of the implemented POVM together. Each convex decomposition of $\rho_B$ (and so also $\sum_i p_i \rho_i$) into pure states can be written as $\rho_B = \tilde{\psi}^\dagger X X^\dagger \tilde{\psi}$ with $X$ an isometry ($XX^\dagger = I$). Here the columns of $\tilde{\psi}^\dagger X$ represent the (unnormalized) pure states in the ensemble. Writing the columns of $X$ as $x_i$, it is readily verified that the elements $\{E_i = x_i x_i^\dagger\}$ define a POVM on Alices's side that does the job. The converse of the Theorem is immediate. □

This Theorem has a very wide range of applicability, and in some sense the power of most applications in quantum information theory stems from exploiting this Theorem. As will become apparent in the next section, the quantum steering Theorem is of central importance in the problem of entanglement transformation: if the POVM elements are all chosen to be full rank, then Alice does not destroy the entanglement present in the state, but just transforms it.

One would expect that the power of quantum steering is proportional to the amount of entanglement present in the system; it is immediately obvious that quantum steering is impossible with separable states (the reduced density matrix is pure in that case such that no convex decompositions are possible), that it will be very much biased in the case of barely entangled states (indeed, the weights in the convex decomposition have the effect that with very high probability a state in a particular direction will be created), while the maximally entangled states are the only ones which allow completely symmetric ensembles.

Consider now the following generalization of the quantum steering problem: suppose Alice has another quantum system $C$ in an unknown quantum state (but of the same dimension as $A$) to her disposition and does a measurement on the joint system $AC$ instead of a measurement on only $A$. What happens with the system at Bob's side? This can easily be analyzed using a little notational trick. Consider a pure state

$$|\psi\rangle = \sum_{ij} \tilde{\psi}_{ij} |ij\rangle \tag{3}$$

and define the (unnormalized) maximally entangled state $|I\rangle$ as

$$|I\rangle = \sum_i |ii\rangle, \tag{4}$$

then it is easily verified that the following relations hold:

$$|\psi\rangle = \tilde{\psi} \otimes I |I\rangle = I \otimes \tilde{\psi}^T |I\rangle.$$

Consider now the state

$$|\psi_{ABC}\rangle = |\psi_{AB}\rangle |\chi_C\rangle \tag{5}$$

with $|\chi_C\rangle$ unknown to both Alice and Bob. Suppose Alice applies the POVM with (pure) elements $E_i = A_i \otimes I_C |I_{AC}\rangle\langle I_{AC}| A_i^\dagger \otimes I_C$ and gets outcome $i$. Then the state description at Bob's side becomes:

$$\langle I_{AC}|(A_i \otimes I_B \otimes I_C)(\tilde{\psi} \otimes I_B \otimes I_C)|I_{AB}\rangle|\chi_C\rangle = (A_i\tilde{\psi})^T|\chi_B\rangle. \qquad (6)$$

Up to the matrix $(A_i\tilde{\psi})^T$, Bob ends up with the state $|\chi\rangle$ although Alice nor Bob had any information regarding this state!

A further sophistication arises if classical communication is allowed between Alice and Bob: if Alice communicates her measurement outcome $i$ to Bob (via a classical channel), Bob can apply a further local operation conditioned on this result. Suppose he implements the POVM that has an element that is proportional to $(A_i\tilde{\psi})^{-T}$; then with a certain probability he ends up with the state $|\chi\rangle$ that is still completely unknown to Alice and Bob: the state has been teleported!

As in the quantum steering case, we expect that maximally entangled states will yield a higher probability of success. Indeed, the local POVM by Bob reduces to a local unitary operation if both $\tilde{\psi}$ and $A_i$ are unitary. Then a simple local unitary operation $(A_i\tilde{\psi})^*$ will yield $|\chi\rangle$ with probability 1. The condition that $\tilde{\psi}$ is unitary is equivalent to the condition that the original state $|\psi_{AB}\rangle$ is maximally entangled; the condition that remains to be verified is that there exists a POVM consisting of elements $E_i$ that all correspond to a maximally entangled states. In the case of qubits, such a POVM consists of the four Bell states; in the case of higher dimensional system, the problem reduces to finding an orthonormal set of unitary matrices, which is indeed possible in arbitrary dimensions.

The fact that quantum teleportation is possible was discovered by C. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres and W. Wootters in the seminal paper [**28**], and has fundamentally changed the way we think about entanglement: if entanglement and classical communication are for free, then global unitary operations can be as easily implemented as local unitaries, because we can just teleport whole systems back and forth.

It is remarkable that the idea of quantum teleportation has been developed completely independent of the idea of quantum steering; instead the direct precursor of teleportation was dense coding, which is very analogous to teleportation and by which one is able to communicate $2\log_2(n)$ classical bits of information by sending 1 qudit living in a $n$-dimensional Hilbert space. Indeed, in the teleportation with qudits, one needs to send $2\log_2(n)$ bits of information (corresponding to the number $n^2$ of possible different outcomes of the Bell measurement). Dense coding is just the opposite of this: Alice applies one of $n^2$ orthonormal unitary operations to her qudit, sends her qudit to Bob, and Bob can perform an orthogonal von-Neumann measurement on both qudits together to determine the operation Alice had performed, revealing $2\log_2(n)$ bits of classical information.

It is clear by now that the maximally entangled states will play the central role in distributed quantum information theory tasks. In analogy with the qubit, the maximally entangled state of 2 qubits is therefore called the *ebit*[**34**].

## 2.3. Entanglement monotones and majorization

In this section we will rederive some celebrated results concerning entanglement transformations in a new unified way; it will become clear that all results follow almost trivially from the quantum steering Theorem.

A natural requirement for an entanglement measure is the fact that it should not increase under local operations [**207**]. The definition of an entanglement monotone is given by Vidal [**223**]: "We call a (non-increasing)[4] entanglement monotone (EM), any magnitude $\mu(\rho)$ that does not increase, on average, under local transformations".

In particular, local unitary operations should leave the EM invariant. Consider a pure state $|\psi_{AB}\rangle = \sum_{ij}\tilde{\psi}_{ij}|i_A\rangle|j_B\rangle$. Local unitary operations correspond to left and right multiplication of $\tilde{\psi}$ with unitaries. Every quantity that is left invariant by this action should be a function of the singular values $\{\sigma_i\}$ of $\tilde{\psi}$ (see von Neumann [**231**]). Note that these singular values are the square roots of the eigenvalues (called the Schmidt coefficients) of the local density operators.

The following Theorem is an adapted version of a Theorem of Vidal [**223**]:

**Theorem 2.** *A function of a bipartite pure state is an entanglement monotone iff it can be written as a concave unitarily invariant function of its local density operator.*

Proof: We show that the Theorem is equivalent to the definition of an entanglement monotone. We already showed that an entanglement monotone should be unitarily invariant and therefore a function only of the reduced density operator. The quantum steering Theorem 1 dictates that a state $|\psi_0\rangle$ with reduced density operator $\rho_0$ can be transformed into the ensemble $\{p_i, |\psi_i\rangle\}$ (or locally $\{p_i, \rho_i\}$ iff $\rho_0 = \sum_i p_i\rho_i$. The definition of an entanglement monotone becomes $\mu(\sum_i p_i\rho_i) \geq \sum_i p_i\mu(\rho_i)$, which ends the proof. □

[4]Note that for convenience we use the opposite definition of an entanglement monotone as given in [**223**]

The simplest functions of this kind are given by partial sums of the decreasingly ordered Schmidt coefficients:

$$\mu_m = \sum_{i=n-m+1}^{n} \lambda_i, \qquad 1 \le m \le n, \tag{7}$$

where $n$ is the dimension of the local Hilbert space. Concavity indeed follows from the variational characterization of the eigenvalues[**116**].

An even stronger result can be derived, a result that was obtained by work of Lo and Popescu [**151**], Nielsen [**162**], Vidal [**222**] and Jonathan and Plenio [**134**]:

**Theorem 3.** *Consider a bipartite quantum system. The probabilistic transformation* $|\psi_0\rangle \to \{p_i, |\psi_i\rangle\}$ *can be accomplished using local operations and classical communication*[5] *(LOCC) iff*

$$\lambda(|\psi_0\rangle) \prec \sum_i p_i \lambda(|\psi_i\rangle)$$

*where the symbol* $\prec$ *denotes majorization* [**155**] *and where* $\lambda(|\chi\rangle)$ *denotes the vector of Schmidt coefficients of* $|\chi\rangle$*.*

Proof: The necessity of the majorization condition is immediate from the observation that the $\mu_m$ of equation (7) are entanglement monotones. To prove the converse, we can assume without loss of generality that all local density operators at Bob's side $\rho_i = \mathrm{Tr}_A(|\psi_i\rangle\langle\psi_i|)$ are diagonal and that the eigenvalues are ordered in decreasing order. Let us define $\sigma = \sum_i p_i \rho_i$. The eigenvalues of $\sigma$ majorize those of $\rho_0$. A Theorem of Uhlmann [**202**] (which is a direct consequence of Birkhoff's Theorem) states that the vector of eigenvalues of a matrix $\rho_0$ is majorized by those of a matrix $\sigma$ iff there exist unitary matrices $U_i$ and probabilities $q_j$ such that

$$\rho_0 = \sum_j q_j U_j \sigma U_j^\dagger = \sum_{ij} p_i q_j U_j \rho_i U_j^\dagger. \tag{8}$$

The right hand side is nothing more than a convex decomposition of $\rho_0$, and due to the quantum steering Theorem we know how Alice can to do this by local operations. The only thing that remains to be done is a classical communication of Alice to Bob as an extra local unitary $U_j^\dagger$ has to be applied. $\square$

A weaker version of this Theorem states that a state can be transformed into another one with probability one iff its Schmidt coefficients are majorized by the ones of the other state. The Theorem also gives an answer to the following question: given a pure bipartite state $|\psi\rangle$, what is the strategy to transform

[5] LOCC transformations consist of a sequence of local operations involving for example local POVM's, where both parties can condition the POVM's implemented on the measurement outcomes of the other party.

this state into a maximally entangled one with the largest possible probability? It is easy to show that this probability $p$ is the largest number for which $\lambda(|\psi\rangle \prec p\lambda(|ME\rangle) + (1-p)\lambda(|00\rangle)$ (this amounts to making a maximally entangled state with probability $p$ and a separable one with probability $1-p$). Consider for example the state $|\psi\rangle = a|00\rangle + b|11\rangle$. The largest $p$ fulfilling the previous constraint is given by $\min(2|a|^2, 2|b|^2)$. Note that the maximal probability for transforming a state into another one by LOCC is itself an entanglement monotone; therefore the probability of conversion is bounded above by $\mu(\psi_2)/\mu(\psi_1)$ for an arbitrary entanglement monotone $\mu$. So if an EM $\mu$ and a probability $p$ for conversion of $\psi$ to $\chi$ is found by an explicit protocol satisfying $p = \mu(\psi_2)/\mu(\psi_1)$, then the protocol is assured to be optimal.

Jonathan and Plenio [**133**] found a surprising result in the context of majorization: it is possible that a state $|\psi_1\rangle$ cannot be transformed into $|\psi_2\rangle$ with probability 1, but that there exists a state $|\chi\rangle$ such that $|\psi_1\rangle \otimes |\chi\rangle$ can be transformed with certainty to $|\psi_2\rangle \otimes |\chi\rangle$; for obvious reasons this strange phenomenon has been called entanglement catalysis. In the light of asymptotic entanglement transformations however, it is clear that if arbitrary large catalysts can be used and if not a perfect but asymptotically perfect transformation is required, every state can be transformed into every other one (see also [**204**]).

## 2.4. Asymptotic entanglement transformations

The previous section dealt with entanglement transformations of single systems. If however there is an asymptotic amount of copies of a quantum state available and the transformations are only required to occur with a fidelity (i.e. accuracy) tending to 1, the analysis becomes much more transparent. Because of its importance, we include the following Theorem (without proof):

**Theorem 4** (Bennett et al.[**26**]). *Given an asymptotic number $N$ copies of a bipartite pure state $|\psi_{AB}\rangle$ with local density operator $\rho$, then there exist local transformations that transform this state into $NS(\rho)$ Bell states with fidelity tending to $1$, where $S(\rho)$ denotes the von-Neumann entropy. Conversely, $NS(\rho)$ Bell states can be diluted into $N$ copies of the original state with fidelity tending to $1$.*

It is interesting to note that no communication is needed in the distillation step, but that communication is needed in the dilution step (although the amount of communication is asymptotically vanishing [**150**]).

Therefore the study of pure-state bipartite entanglement is very simple in the asymptotic case: there is only one measure, namely the entropy of the local density operator, that says it all. Note also that the results of the previous section do not lead to the previous Theorem as these were concerned with entanglement transformations occurring with fidelity equal to 1, and not tending to 1. In this exact case, the dilution step does not hold anymore as the number of Schmidt coefficients cannot increase by LOCC transformations.

## 2.5. Conclusion

We identified *quantum steering*, introduced by Schrödinger, as the magic trick of quantum mechanics allowing to give a unified description of the entanglement of pure bipartite systems, and discussed its relation with quantum teleportation, entanglement monotones and entanglement transformations. Many other examples could have been discussed, such as its relation with quantum error correction [**194, 196, 34, 101, 48**], quantum cryptographic protocols [**84**], remote state preparation [**33**], entanglement of assistance [**71**], Bell inequalities [**237**], ...

## CHAPTER 3

# Entanglement of multipartite states

The study of multipartite entanglement, i.e. when more than two parties are involved, is still in its infancy compared to that of the bipartite case. The difficulty stems from the fact that we have to work with multidimensional tensorial objects instead of with matrices, and the fact that no nice analogue of the singular value decomposition exists for tensors: there are too few local degrees of freedom available to bring the tensor into a convenient normal form. Nevertheless, in this chapter we present some first steps into a systematic analysis of higher dimensional tensors and hence of multipartite entanglement.

Multiparticle entanglement exhibits a much richer structure than bipartite entanglement. The first celebrated example of a multiparticle state is the GHZ-state, called after Greenberger, Horne and Zeilinger [**102**]. This state was introduced because it allows to disprove the Einstein locality for quantum systems without invoking statistical arguments such as needed in the arguments of Bell. This follows from the fact that multipartite states permit richer types of correlations between the different subentities than possible with bipartite or classical systems. Since then, much attention has been devoted to the study of multiparticle entanglement. An interesting aspect of multiparticle entanglement was discovered by Wootters et al. [**59**]. They showed that a quantum state has only a limited susceptibility for quantum correlations: the more classical or bipartite correlations that a three-partite state exhibits, the less genuine tripartite entanglement that can be present in the system. DiVincenzo et al. [**70**] (see also [**81**]) found a nice application of all these different kinds of correlations: they showed how these can be used to hide quantum data in such a way that the different parties can only acquire information if they do joint measurements.

A lot of effort has also been devoted to classifying pure states in terms of equivalence classes of local unitary operations (LU) and of local filtering operations (SLOCC). The only complete results however were found in the case of three qubits, where a classification of SLOCC- [**80**] and LU- equivalence classes [**1, 197**] was obtained. Another direction of research consisted of studying equivalence classes in the asymptotic sense. The aim is to find a set of states to which every pure multipartite state can be locally and reversibly transformed in an asymptotic sense (in the bipartite case for example, everything can reversibly be transformed to a Bell state). Identifying these so-called minimal reversible entanglement generating sets (MREGS) [**22**] however has turned out to be very difficult, and only negative results have been obtained until now [**4**].

In this chapter we will mainly concentrate on characterizing all possible local equivalence classes of multipartite systems. In a first section, we develop a new formalism that enables to bring multipartite states into a normal form under the action of SLOCC operations, and show how this enables to quantify multipartite entanglement. We generalize this to mixed states, and in the last sections we give a complete classification of all possible multipartite pure states in $2 \times 2 \times N$ and $2 \times 2 \times 2 \times 2$ dimensions.

## 3.1. The general case

### 3.1.1. Normal forms under local unitary operations

Consider a general multipartite state with $m$ parties defined on an $n_1 \otimes n_2 \cdots n_M$ dimensional Hilbert space:

$$|\psi\rangle = \sum_{i_1 \cdots i_m} \psi_{i_1 \cdots i_m} |i_1\rangle |i_2\rangle \cdots |i_m\rangle. \tag{9}$$

A first natural question is the following: is there a method to verify if two states $|\psi_1\rangle$ and $|\psi_2\rangle$ are equivalent up to local unitary transformations? In the bipartite case, this problem can readily be solved using the singular value decomposition, and we therefore ask for some kind of generalization of this diagonal normal form. Let us state the following Theorem (see also Carteret et al. [**49**]), which is a weak generalization of the SVD:

**Theorem 5.** *Given a general complex tensor $\psi_{i_1 \cdots i_m}$ with dimensions $n_1 = n_2 = \cdots = n_m = n$, then there exist local unitaries $U_i$ such that all the following entries in the tensor $\psi' = U_1 \otimes \cdots \otimes U_m |\psi_{i_1 \cdots i_m}\rangle$ are set equal to zero:*

$$\forall 1 \leq j \leq n, \forall k > j : \qquad \begin{aligned} &\psi'_{j,j,\cdots,j,j,k} = 0 \\ &\psi'_{j,j,\cdots j,k,j} = 0 \\ &\vdots \\ &\psi'_{j,k,j,\cdots,j,j} = 0 \\ &\psi'_{k,j,\cdots,j,j} = 0. \end{aligned}$$

*Moreover all entries $\psi'_{n,n,\cdots,n,i,n,\cdots n}$, $i \leq n$ can be made real and positive. If the number of parties exceeds 2, then the normal form is typically not unique up to permutations, but there exist a discrete number of different normal forms with* *the aforementioned property. The number of zeros however can generically not be increased by further local unitary operations.*

Proof: unlike the proof in [**49**], our proof is constructive and can readily be translated into matlab code to calculate the normal form numerically (see appendix). First consider all entries with at least $m-1$ number of one's in its indices, and define the vectors $x_i^1 = \psi_{i,1,1,\cdots,1}$, $x_i^2 = \psi_{1,i,1,\cdots,1}$, $\cdots$, $x_i^m = \psi_{1,1,\cdots,1,i}$. Define now a recursive algorithm that goes as follows: rotate $x^1$ to $\|x^1\|[1,0,\cdots 0]$ by a unitary transformation, apply the same transformation on the full tensor, and define $x^2 = \psi_{1,i,1,\cdots,1}$ with $\psi$ the transformed tensor. Now do the same thing with $x^2, \ldots x^m$ and then again with $x^1$, until the algorithm converges. This algorithm will certainly converge because at each step the $(1,1,\cdots 1)$ entry of $\psi$ becomes larger and larger, unless all entries $(1,1,\cdots,1,i,1,\cdots 1)$ are equal to zero; moreover its value is bounded above because the unitary group is compact. Next exactly the same algorithm can be applied to the subtensor of $\psi$ defined as the one with all entries larger or equal to 2 (it is easy to check that the zeros obtained in the first step will remain zero by this kind of action). Next we can again do the same thing on another (smaller) subtensor, proving that indeed all zeros quoted in the Theorem can be made.

It is straightforward to prove that the entries $\psi'_{n,n,\cdots,n,i,n,\cdots n}, i \leq n$ can all be made real and positive by further diagonal unitary transformations.

Let us finally prove that no more zeros can be made by whatever unitaries (in the generic case). This follows from the fact that a unitary $n \times n$ matrix has $n^2$ continuous real degrees of freedom, but that only $n^2 - n$ of them can be used to produce zeros as the other $n$ degrees of freedom can be imbedded in a diagonal unitary with just phases. Counting of the number of zeros produced indeed leads to

$$\sum_{j=1}^{m} \sum_{k=1}^{m-1} \max(n-k, 0) = m\frac{n(n-1)}{2} \tag{10}$$

which indeed corresponds to the $m(n^2-n)$ degrees of freedom as the zeros are "complex".

The non-uniqueness of the normal form obtained is surprising but can readily be verified by implementing the algorithm on a generic tensor; typically the algorithm converges to one out of a finite number of possible different normal forms. $\square$

As a first example, consider a system of three qubits. Unfolding the $2 \times 2 \times 2$ tensor in two $2 \times 2$ matrices, the following entries can always be made equal to zero:

$$\left( \begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix} \begin{pmatrix} 0 & x \\ x & x \end{pmatrix} \right) \tag{11}$$

Here $x$ is used to denote a non-zero entry. In this case, it is easy to see that 4 of the remaining 5 entries can be made real by multiplying with appropriate diagonal local unitaries. This is equivalent to the normal form obtained by Acin et al.[**1**].

A more sophisticated example is the $3 \times 3 \times 3$ case, the normal form of which looks like

$$\left( \begin{pmatrix} x & 0 & 0 \\ 0 & x & x \\ 0 & x & x \end{pmatrix} \begin{pmatrix} 0 & x & x \\ x & x & 0 \\ x & 0 & x \end{pmatrix} \begin{pmatrix} 0 & x & x \\ x & 0 & x \\ x & x & x \end{pmatrix} \right) \tag{12}$$

It is also straightforward to generalize the previous Theorem (and constructive proof) to systems with different subdimensions (see Carteret et al.[**49**] for an existence proof); the algorithm of the previous proof can readily be extended to this case. Let us for example consider the normal form of the $N \times 2 \times 2$ case:

$$\left( \begin{pmatrix} x & 0 \\ 0 & x \\ 0 & x \\ 0 & x \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & x \\ x & x \\ x & 0 \\ 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} \right) \tag{13}$$

This case is of particular interest as it is describes a state of two qubits entangled with the rest of the world. We will discuss this case in more detail in a section 3.2.

### 3.1.2. Normal forms under SLOCC operations

Until now we have restricted ourselves to local unitary transformations. These however are not the most general local transformations however: POVM's can be implemented with arbitrary POVM elements $\{A_i\}, E_i = A_i^\dagger A_i$. Therefore it makes sense to study equivalence classes under general local transformations of the kind $|\psi'\rangle = A_1 \otimes \cdots \otimes A_n |\psi\rangle$ with $\{A_i\}$ arbitrary matrices. These kinds of transformations are called SLOCC transformations [**80**] (from stochastic local operations assisted by classical communication), and are also called local filtering operations. It will turn out very useful to restrict ourselves to SLOCC transformations where all $\{A_i\}$ are full rank (note that this is a physically motivated restriction as entanglement is lost whenever an $A_i$ is not full rank; therefore we only consider the reversible operations). We will consider all $\{A_i\}$ to belong to $SL(n, \mathcal{C})$, the group of square complex matrices having determinant equal to 1, and consider unnormalized states.

The following general Theorem appeared in Verstraete et al.[**211**]:

**Theorem 6.** *Consider an $N_1 \times N_2 \times \cdots N_p$ pure multipartite state (or tensor). Then this state (tensor) can constructively be transformed into a normal form by determinant 1 SLOCC operations. The local density operators of the normal* *form are all proportional to the identity. Moreover, the state connected to the original one by determinant 1 SLOCC operations with the minimal possible norm (i.e. trace of the unnormalized density operator) is in normal form.*

*Proof:* We will give a constructive proof of this theorem that can directly be translated into matlab code (see appendix). The idea is that the local determinant 1 operators $A_i$ bringing $\psi$ into its normal form can be iteratively determined by a procedure where at each step the trace of $|\psi\rangle\langle\psi| = \rho$ is minimized by a local filtering operation of one party. Consider therefore the partial trace $\rho_1 = \mathrm{Tr}_{2\cdots p}(\rho)$. If $\rho_1$ is full rank, there exists an operator $X$ with determinant 1 such that $\rho_1' = X\rho_1 X^\dagger \sim I_{N_1}$. Indeed, $X = |\det(\rho_1)|^{1/2N_1}(\sqrt{\rho_1})^{-1}$ does the job[1], and we have $\rho_1' = \det(\rho_1)^{1/N_1} I_{N_1}$. We also have the relation:

$$\mathrm{Tr}\,(\rho') = N_1 \det(\rho_1)^{1/N_1} \leq \mathrm{Tr}\,(\rho_1)\,, \tag{14}$$

where $\rho' = (X\otimes I\cdots\otimes I)|\psi\rangle\langle\psi|(X\otimes I\cdots\otimes I)^\dagger$. This inequality follows from the fact that the geometric mean is always smaller than the arithmetic mean, with equality iff $\rho_1$ is proportional to the identity. Therefore the trace of $\rho$ decreases after this operation. We can now repeat this procedure with the other parties, and then repeat everything iteratively over and over again. After each iteration, the trace of $\rho$ will decrease unless all partial traces are equal to the identity. Because the trace of a positive definite operator is bounded from below, we know that the decrements become arbitrarily small and following equation (14) this implies that all partial traces converge to operators arbitrarily close to the identity.
We still have to consider the case where we encounter a $\rho_i$ that is not full rank. Then there exists a series of $X$ whose norm tends to infinity but has determinant 1 such that $X\rho_i X^\dagger = 0$, leading to a normal form identical to zero, clearly the positive operator with minimal possible trace. This ends the proof of the existence of the normal form.

Consider now the a state that is normal form; then due to the construction of the proof, the trace can always be decreased by determinant 1 SLOCC operations, unless the state is in normal form. This ends the proof. □

There is supporting evidence for the uniqueness of the above normal form. If the normal form were not unique, there would exist a normal $\sigma_q = |\psi_q\rangle\langle\psi_q|$ and $\sigma_r = |\psi_r\rangle\langle\psi_r|$ and diagonal matrices $\{D_i\}$ with determinant 1 such that $\sigma_r = (D_1\otimes\cdots\otimes D_p)\sigma_q(D_1\otimes\cdots\otimes D_p)^\dagger$: if the $D_i$ were not diagonal we could always make them diagonal through local unitary transformations. Writing the diagonal elements of the normal $\sigma_q, \sigma_r$ in the $N_1\times N_2\times\cdots\times N_p$ tensors $T^q, T^r$, it is readily observed that both tensors are, up to a constant, stochastic in all directions: if one arbitrary index $i_j$ of the tensor $T_{i_1\cdots i_p}$ is fixed and the

[1]Note that the numerical algorithm should not calculate $X$ from $\rho_1$ but instead from the singular value decomposition of the $N_1\times(\Pi_{i>1}N_i)$ matrix $\psi_{i_1,(i_2\cdots i_p)} = U\Sigma V^\dagger$: $X$ can be chosen proportional to $\Sigma^{-1}U^\dagger$, and the numerical accuracy will be much higher.

sum is taken over all the other indices, then the result is always equal to the number $\mathrm{Tr}(\sigma_q)/N_j$ (this follows from the fact that all local density operators are proportional to the identity); moreover all elements are greater or equal to zero. Both tensors must be connected by diagonal $\{D_i\}$ working on the respective indices. We will now prove the unicity for the bipartite case, and it is expected that the general case can be derived in an analogous way. We will rewrite the diagonal entries of $D_1, D_2$ as the vectors $x, y$ with respective dimensions $N_x, N_y$, and we define $\alpha = \mathrm{Tr}\,(\sigma_q)\,, \beta = \mathrm{Tr}\,(\sigma_r)$. Moreover we assume without loss of generality that $\forall i : x_1 \geq x_i$. The following relations hold:

$$\sum_i T_{ij} = \frac{\alpha}{N_y} \qquad \sum_i x_i y_j T_{ij} = \frac{\beta}{N_y} \tag{15}$$

$$\sum_j T_{ij} = \frac{\alpha}{N_x} \qquad \sum_j x_i y_j T_{ij} = \frac{\beta}{N_x}. \tag{16}$$

The following (in)equalities are easily derived:

$$\frac{\beta}{N_x} = \frac{\beta}{N_y} \sum_j \frac{x_1 T_{1j}}{\sum_{i'} x_{i'} T_{i'j}} = \frac{\beta}{N_y} \sum_j \frac{T_{1j}}{\sum_{i'} \frac{x_{i'}}{x_1} T_{i'j}} \tag{17}$$

$$\geq \frac{\beta}{N_y} \sum_j \frac{T_{1j}}{\sum_{i'} T_{i'j}} = \frac{\beta}{N_y} \frac{\alpha/N_x}{\alpha/N_y} = \frac{\beta}{N_x}. \tag{18}$$

It is clear that all inequalities have to be equalities, and this is only possible if all elements in the vector $x$ are equal to each other. A similar reasoning applies for $y$, and we conclude that the only possibility is that both $D_1$ and $D_2$ are proportional to the identity. The only exception arises when there are zeros in $T_q$. It turns out that these cases correspond to block doubly stochastic matrices [2] but even then an analogous proof is possible: $D_1 \otimes \cdots \otimes D_p$ will have elements of value 1 on the indices where the (equal) diagonal elements of $\sigma_q$ and $\sigma_r$ do not vanish. Due to the positiveness of $\sigma$ it follows that $\sigma_q = \sigma_r$. Therefore the uniqueness of the normal form is proven in the bipartite case. To generalize this argument to for example tripartite systems, one should prove that the set of equations

$$\sum_{ij} T_{ijk} = 1 \qquad \sum_{ij} x_i y_j z_k T_{ijk} = 1 \tag{19}$$

$$\sum_{ik} T_{ijk} = 1 \qquad \sum_{ik} x_i y_j z_k T_{ijk} = 1 \tag{20}$$

$$\sum_{jk} T_{ijk} = 1 \qquad \sum_{jk} x_i y_j z_k T_{ijk} = 1 \tag{21}$$

$$\forall i, j, k : \qquad T_{ijk} \geq 0 \qquad x_i \geq 0 \qquad y_j \geq 0 \qquad z_k \geq 0 \tag{22}$$

[2]This happens for example in the case of the EPR and GHZ state: there exists local non-trivial operations that map the state onto itself.

in the unknowns $x_i, y_j, z_k$ has only one solution, namely when $\forall i,j,k : x_i y_j z_k = 1$. Extensive numerical investigations confirmed this claim, although we could not complete the proof as in the bipartite case.

We expect that some results of elimination theory [**95**] could complete the proof, although we have not investigated this yet. Note however that the next sections will prove the uniqueness in the pure $2 \times 2 \times N$- and the $2 \times 2 \times 2 \times 2$-case, which is more supporting evidence for the uniqueness. Let us therefore formulate the following conjecture:

**Conjecture 1.** *The normal form defined in Theorem 6 is unique up to local unitary transformations.*

Let us now return to the general Theorem 6. This Theorem is very fundamental in that is states that each pure multipartite state can be transformed into a state with the property that all local density operators are proportional to the identity. States in normal form are clearly expected to be maximally entangled states. As we will argue later, the normal form is the state with the maximal amount of entanglement that can be created locally from the original state with a non-zero probability.

Let us next prove that the normal form is continuous with respect to perturbations of the entries of the original density matrix $\rho$. First of all note that the non-uniqueness due to the local unitaries can be circumvented by imposing all $A_i$ to be Hermitian. The following Lemma shows that the normal form is robust against perturbations or noise:

**Lemma 1.** *If the SLOCC operations bringing the state into the normal form introduced in Theorem 6 are chosen to be Hermitian, and if they turn out to be finite, then the normal form is continuous with respect to the entries of the state.*

Proof: Let us consider $\rho = (A_1 \otimes \cdots \otimes A_p)\sigma(A_1 \otimes \cdots \otimes A_p)^\dagger$ and a perturbation $\dot{\rho}$ resulting in $\{\dot{A}_i\}$ and $\dot{\sigma}$. The following formula is readily verified:

$$\begin{aligned}(A_1 \otimes \cdots \otimes A_p)^{-1}\dot{\rho}(A_1 \otimes \cdots \otimes A_p)^{-\dagger} &= \\ \dot{\sigma} + \sum_{i=1}^{p}\Big((I \otimes \cdots A_i^{-1}\dot{A}_i \cdots \otimes I)\sigma + \text{h.c.}\Big)&. \qquad (23)\end{aligned}$$

As all $\{A_i\}$ are Hermitian and have determinant 1, all $A_i^{-1}\dot{A}_i$ are skew-Hermitian and the second term lives in another subspace $S_2$ than the first term $\dot{\sigma}$ who lives in subspace $S_1$. $\dot{\sigma}$ can therefore be obtained by projecting the left hand side parallel to $S_2$ onto $S_1$. As $\dot{\rho}$ is finite and all $\{A_i\}$ have determinant one and are finite, this projection is of course also finite. This proves that $\dot{\sigma}$ is of the same order of magnitude as $\dot{\rho}$, which ends the proof. $\square$

Note that we have also proven continuity with respect to mixing (i.e. in going over from pure to mixed states).

Let us now discuss some peculiarities. The fact that the algorithm can converge to zero despite the fact that all $A_i$ have determinant equal to 1 is a consequence of the fact that $SL(n,\mathcal{C})$ is not compact: there exist states that can only be brought into their respective normal form by infinite transformations, although the class of states with this property is clearly of measure zero. As an example consider the $W$-state [**80**] $|\psi\rangle = |001\rangle + |010\rangle + |100\rangle$. The following identity is easily checked:

$$\lim_{t\to\infty}\begin{pmatrix} 1/t & 0 \\ 0 & t \end{pmatrix}^{\otimes 3}|W\rangle = 0.$$

The normal form corresponding to the $W$-state is therefore equal to zero, clearly the state with the minimal possible trace. This is interesting, as it will be shown that a normal form is zero iff a whole class of entanglement monotones is equal to zero. Therefore the states with normal form equal to zero are fundamentally different from those with finite normal form, and this leads to the generalization of the $W$-class to arbitrary dimensions.

It thus happens that some states have normal form equal to 0. This also happens if the state does not have full support on the Hilbert space, i.e. if one partial trace $\rho_i$ is rank deficient. Note that states which do not have full support on the Hilbert space, such as pure states from which one party is fully separable, all have normal form equal to zero. It will indeed turn out that the amount of multipartite entanglement present in a state can be quantified by the trace of the obtained normal form, which is clearly zero in the case of separable states. On the other hand, the only normalized states that are already in normal form are precisely the maximally entangled states: in the case of three qubits for example, the only state with the property that all its local density operators are proportional to the identity is the GHZ-state.

As a last remark, we give an example of a state that is brought into a non-zero normal form by infinite SLOCC transformations:

$$|\psi\rangle \simeq a(|0000\rangle + |1111\rangle) + |01\rangle(|10\rangle + |01\rangle) \tag{24}$$

The normal form is just given by the GHZ-state $(|0000\rangle + |1111\rangle)$, but as can be derived from the results to be presented in section 3.3.2, infinite SLOCC transformations are needed to reach this.

### 3.1.3. Entanglement monotones

Until now we contented ourselves to characterize the orbits generated by local unitary or SLOCC operations, but we have not tried to quantify the entanglement present in a state. The SLOCC normal form introduced in the previous section however gives us a strong hint of how to do this. Note that all separable states have a normal form equal to zero, and that the known maximally entangled states such as Bell-states and GHZ-states are the only ones of their SLOCC equivalence class that are in normal form.

This suggests a very general way of constructing entanglement monotones:

**Theorem 7.** *Consider a linearly homogeneous positive function of a pure (unnormalized) state $M(\rho = |\psi\rangle\langle\psi|)$ that remains invariant under determinant 1 SLOCC operations. Then $M(|\psi\rangle\langle\psi|)$ is an entanglement monotone.*

*Proof:* A quantity $M(\rho)$ is an entanglement monotone iff its expected value does not increase under the action of any local operation. It is therefore sufficient to show that for every local $A_1 \leq I_{N_1}$, $\bar{A}_1 = \sqrt{I_{N_1} - A_1^\dagger A_1}$, it holds that

$$\begin{aligned} M(\rho) \geq \quad & \mathrm{Tr}\left((A_1 \otimes I)\rho(A_1 \otimes I)^\dagger\right) M\left(\frac{(A_1 \otimes I)\rho(A_1 \otimes I)^\dagger}{\mathrm{Tr}\left((A_1 \otimes I)\rho(A_1 \otimes I)^\dagger\right)}\right) \\ & + \mathrm{Tr}\left((\bar{A}_1 \otimes I)\rho(\bar{A}_1 \otimes I)^\dagger\right) M\left(\frac{(\bar{A}_1 \otimes I)\rho(\bar{A}_1 \otimes I)^\dagger}{\mathrm{Tr}\left((\bar{A}_1 \otimes I)\rho(\bar{A}_1 \otimes I)^\dagger\right)}\right) \end{aligned}$$

If $A_1$ is full rank, it can be transformed to a determinant 1 matrix by dividing it by $\det(A_1)^{1/N_1}$. Due to the homogeneity of $M(\alpha\rho) = \alpha M(\rho)$ the previous inequality is equivalent to

$$M(\rho) \geq (|\det(A_1)|^{2/N_1} + |\det(\bar{A}_1)|^{2/N_1})M(\rho).$$

As the arithmetic mean always exceeds the geometric mean, this inequality is always satisfied. This argument can be easily completed to the cases where $A_i$ is not full rank due to continuity. The same argument can then be repeated for the other $A_i$, which ends the proof. □

Entanglement monotones of the above class can readily be constructed using the completely antisymmetric tensor $\epsilon_{i_1 \cdots i_N}$.
Indeed, it holds that $\sum A_{i_1 j_1} A_{i_2 j_2} \cdots A_{i_N j_N} \epsilon_{j_1 \cdots j_N} = \det(A)\epsilon_{i_1 \cdots i_N}$, and as $\det(A) = 1$ this leads to invariant quantities under determinant 1 SLOCC operations. These quantities seem to be related to hyperdeterminants [**95, 160**], and those latter seem to be a subclass of the quantities considered here.

Consider for example the case of two qubits. The quantity

$$|\sum_{i_1 j_1 i_2 j_2} \psi_{i_1 j_1} \psi_{i_2 j_2} \epsilon_{i_1 i_2} \epsilon_{j_1 j_2}|$$

is clearly of the considered class, and it happens to be the celebrated concurrence entanglement measure [**244**]. In the case of three qubits, the simplest non-trivial homogeneous quantity invariant under determinant 1 SLOCC operations is given by

$$|\psi_{i_1j_1k_1}\psi_{i_2j_2k_2}\psi_{i_3j_3k_3}\psi_{i_4j_4k_4}\epsilon_{i_1i_2}\epsilon_{i_3i_4}\epsilon_{j_1j_2}\epsilon_{j_3j_4}\epsilon_{k_1k_3}\epsilon_{k_2k_4}|^{1/2}$$

(Note that we use the Einstein summation convention.) This happens to be the square root of the 3-tangle introduced by Wootters et al.[**59**], which quantifies the genuine tripartite entanglement.

More generally, as the considered entanglement monotones are invariant under the determinant 1 SLOCC operations, the number of independent entanglement monotones is equal to the degrees of freedom of the normal form obtained in the case of a pure state, minus the degrees of freedom induced by the local unitary operations. Indeed, this is the number of invariants of the whole class of states connected by SLOCC operations. It is then easily proven that a normal form is equal to zero if and only if all the considered entanglement monotones are equal to zero: the entanglement monotones are homogeneous functions of the normal form, and if the normal form is not equal to zero there always exists an SLOCC invariant quantity that is different from zero.

In the case of 4 qubits for example, parameter counting leads to $(2\cdot2^4-2)-4\cdot6 = 6$ (a state has 32 degrees of freedom $-2$ to an irrelevant phase and the 4 $SL(2,\mathcal{C})$ matrices have each 6 degrees of freedom) independent entanglement monotones. The simplest monotone is given by

$$|\psi_{i_1j_1k_1l_1}\psi_{i_2j_2k_2l_2}\epsilon_{i_1i_2}\epsilon_{j_1j_2}\epsilon_{k_1k_2}\epsilon_{l_1l_2}|, \tag{25}$$

and the other 5 entanglement monotones can be obtained by including more factors, an example being

$$\sqrt{2}|\psi_{i_1j_1k_1l_1}\psi_{i_2j_2k_2l_2}\psi_{i_3j_3k_3l_3}\psi_{i_4j_4k_4l_4}\epsilon_{i_1i_2}\epsilon_{i_3i_4}\epsilon_{l_1l_2}\epsilon_{l_3l_4}\epsilon_{j_1j_3}\epsilon_{j_2j_4}\epsilon_{k_1k_3}\epsilon_{k_2k_4}|^{1/2}. \tag{26}$$

These are clearly generalizations of the concurrence and the 3-tangle to four parties. Note however that the situation here is more complicated due to the existence of multiple independent entanglement monotones. Note also that there exist biseparable states that can be brought into a non-zero normal form by determinant 1 SLOCC operations. Consider for example the tensor product of two Bell states; all local density operators are proportional to the identity, the value of the entanglement monotones (25) and (26) is respectively given by 1 and $1/\sqrt{2}$ (as opposed to 1 and 1 for the GHZ-state $(|0000\rangle + |1111\rangle)/\sqrt{2}$), and nevertheless no true 4-partite entanglement is present.

If the subsystems happen to be of unequal dimension, then the respective subdimensions should be chosen not larger than the maximal allowed dimension such that all local density operators remain full rank. In a $2\times2\times N$ system for example, the state can only have full support on the $2\times2\times4$ subspace, and therefore it makes no sense to calculate the normal form with $N > 4$: one can always first rotate the $N$-dimensional system into a 4-dimensional one by local unitary operations, and proceed by calculating the normal form for the

$2\times2\times4$ system. More generally, if the dimension of the largest subsystem does not exceed the product of all the other ones, then generically the normal form will not be equal to zero, leading to non-trivial entanglement monotones. As an example, consider a $2 \times 2 \times 4$ system; there are more local SLOCC parameters than the number of degrees of freedom, so there will be only one entanglement monotone (as is the case in the $2 \times 2$ and $2 \times 2 \times 2$ case). The $2 \times 2 \times 4$ tangle is given by:

$$\sqrt{4/3}|\psi_{i_1j_1k_1}\psi_{i_2j_2k_2}\psi_{i_3j_3k_3}\psi_{i_4j_4k_4}\epsilon_{i_1i_2}\epsilon_{i_3i_4}\epsilon_{j_1j_3}\epsilon_{j_2j_4}\epsilon_{k_1k_2k_3k_4}|^{1/2} \tag{27}$$

The factor $\sqrt{4/3}$ is included to ensure that the state in normal form

$$(|000\rangle + |011\rangle + |102\rangle + |113\rangle)/2 \tag{28}$$

has a value of the EM given by 1. Indeed, as will be shown in the following section, the maximal value of the tangle is always obtained for states in normal form, and this is the unique state (up to LU) having all its local density operators proportional to the identity. Note that this state is therefore the generalization of the $GHZ$ state to $2 \times 2 \times 4$ systems.

For completeness, let us also give a formula for the $2 \times 2 \times 3$ tangle:

$$\sqrt[3]{\frac{27}{4}}|\sum\psi_{i_1,j_1,k_1}\psi_{i_2,j_2,k_2}\psi_{i_3,j_3,k_3}\psi_{i_4,j_4,k_4}\psi_{i_5,j_5,k_5}\psi_{i_6,j_6,k_6} \\ \epsilon_{i_1i_4}\epsilon_{i_2i_5}\epsilon_{i_3i_6}\epsilon_{j_1j_4}\epsilon_{j_2j_5}\epsilon_{j_3j_6}\epsilon_{k_1k_2k_3}\epsilon_{k_4k_5k_6}|^{1/3} \tag{29}$$

The state maximizing this entanglement monotone (the number is bounded by 1) is the generalization of the $GHZ$ to the $2 \times 2 \times 3$ case:

$$\frac{1}{\sqrt{3}}|000\rangle + \frac{1}{\sqrt{6}}|011\rangle + \frac{1}{\sqrt{6}}|101\rangle + \frac{1}{\sqrt{3}}|112\rangle. \tag{30}$$

Let us finally give a non-trivial example of an entanglement monotone of the considered class in the case of three qutrits:

$$\sqrt{2}|\psi_{i_1j_1k_1}\psi_{i_2j_2k_2}\psi_{i_3j_3k_3}\psi_{i_4j_4k_4}\psi_{i_5j_5k_5}\psi_{i_6j_6k_6} \\ \epsilon_{i_1i_2i_3}\epsilon_{i_4i_5i_6}\epsilon_{j_1j_2j_4}\epsilon_{j_3j_5j_6}\epsilon_{k_1k_5k_6}\epsilon_{k_2k_3k_4}|^{1/3}. \tag{31}$$

The other $(2 \cdot 3^3 - 1) - (3 \cdot 16) - 1 = 4$ independent entanglement monotones can again be constructed by including more factors.

### 3.1.4. Optimal Filtering

A natural question now arises: how do we characterize the optimal SLOCC operations to be performed on one copy of a multipartite system such that, with a non zero chance, a state with maximal possible multipartite entanglement is obtained? This question is of importance for experimentalists as in general they are not able to perform joint operations on multiple copies of the system.

Therefore the procedure outlined here often represents the best entanglement distillation procedure that is practically achievable.

In the previous section a whole class of entanglement monotones that measures the amount of multipartite entanglement was introduced. The following Theorem can easily be proved using the techniques of Theorem 6

**Theorem 8.** *Consider a pure multipartite state, then the local filtering operations that maximize all entanglement monotones introduced in theorem 7 are represented by operators proportional to the determinant 1 SLOCC operations that transform the state into its normal form.*

*Proof:* The proof of this Theorem is surprisingly simple. Indeed, all the quantities introduced in Theorem 7 are invariant under determinant 1 SLOCC operations if the states do not get normalized. The value of an entanglement monotone however only makes sense if defined on normalized states, and due to the linear homogeneity of the entanglement monotones, the following identity holds:

$$M\left(\frac{(\otimes_i A_i)\rho(\otimes_i A_i)^\dagger}{\mathrm{Tr}\left((\otimes_i A_i)\rho(\otimes_i A_i)^\dagger\right)}\right)=\frac{M(\rho)}{\mathrm{Tr}\left((\otimes_i A_i)\rho(\otimes_i A_i)^\dagger\right)}$$

The optimal filtering operators are then obtained by the $\{A_i\}$ which minimize

$$\mathrm{Tr}\left((\otimes_i A_i)\rho(\otimes_i A_i)^\dagger\right). \tag{32}$$

But this problem was solved in theorem 6, where it was proved that the $\{A_i\}$ bringing the state into its normal form minimize this trace. □

It is therefore proved that the (reversible) procedure of washing out the local correlations maximizes the multipartite entanglement as measured by the generalization of the tangle. This is in complete accordance with the results of majorization outlined in Chapter 2, where it was shown that the notion of local disorder is intimately connected to the amount of entanglement present. Therefore we have supporting evidence to call pure states in normal form maximally entangled with relation to their SLOCC orbit.

### 3.1.5. The mixed state case.

The normal form derived in Theorem 6 can readily be generalized to the case where the state is mixed, i.e. the case where the density operator is a convex sum of pure states. Indeed, nowhere in the proof of the Theorem it was used that the state $\rho$ was pure; the same holds for the continuity for the normal form. We have therefore proven:

**Theorem 9.** *Consider an $N_1 \times N_2 \times \cdots N_m$ mixed multipartite state. Then this state can be brought into a normal form by determinant 1 SLOCC operations, where the normal form has all local density operators proportional to the* *identity. Moreover there exists a normal form such that its trace is the minimal one that can be obtained by determinant 1 SLOCC operations. If the SLOCC operations are chosen to be Hermitian, then the normal form is continuous with respect to perturbations of the original state.*

Moreover, it is again conjectured that the normal form obtained is unique up to local unitary operations.

Note that if $\rho$ is full rank, its normal form will never converge to zero: the determinant of the density operator is constant under SLOCC operations.

It is also possible to adopt the results about entanglement monotones. First of all we extend the definition of an entanglement monotone $\mu_p$ that is defined on pure states and that is linearly homogeneous in $\rho$ by the convex roof formalism:

$$\mu_m(\rho) = \min_{\sum_i p_i|\psi_i\rangle\langle\psi_i|=\rho} \sum_i p_i \mu_p(|\psi_i\rangle). \tag{33}$$

Here the optimization has to be done over all pure state decompositions of the state. The fact that the pure state entanglement monotone is linearly homogeneous in $\rho$ ensures that $\mu_m$ is, on average, not increasing under local operations, and therefore assures that $\mu_m$ is an entanglement monotone. Moreover, it is obvious that these entanglement monotones are again invariant under determinant 1 SLOCC operations. The results on optimal filtering for mixed states also readily apply, and therefore we arrive at the following very powerful result:

**Theorem 10.** *The local filtering operations bringing a mixed state into its normal form are exactly the ones which maximize the entanglement monotones that remain invariant under determinant 1 SLOCC operations.*

This result is remarkable, because there does typically not exist a way of actually calculating the value of an entanglement monotone defined on a mixed state: finding the optimal pure state decomposition of a state with relation to the convex roof formalism for a given EM is excessively difficult and has until now only been proven possible for the concurrence (i.e. the case of two qubits). So although we cannot calculate the entanglement monotone, we know how to maximize it! This particularly applies to mixed states of three qubits: we have proven how to maximize the 3-tangle, although we don't know how to calculate it.

Note that this optimal filtering procedure produces highly non-trivial results even in the case of two qubits: it proves that the concurrence and therefore the entanglement of formation of a mixed state of two qubits is maximized by the SLOCC operations bringing the state into its unique (Bell-diagonal) normal form.

## 3.2. The $2 \times 2 \times N$-case

A pure multipartite state with subdimensions $2\times2\times N$ is of particular interest as it describes the system of two qubits (possibly) entangled with the rest of the world. Note that it is sufficient to consider the $2\times2\times4$ case as the third system can only have support on a 4-dimensional subspace (note that this is not longer true in the case of mixed states). As a special case this class also contains the $2\times2\times2$ states, containing the celebrated Greenberger-Horne-Zeilinger [**102**] state $(|000\rangle + |111\rangle)/\sqrt{2}$ introduced by Mermin [**158**].

The following Lemma will be crucial in the next sections:

**Lemma 2.** *Consider* $2\times2$ *matrices* $U_1, U_2$, *then both* $U_1$ *and* $U_2$ *belong to* $SU(2)$ *if and only if* $T(U_1\otimes U_2)T^\dagger \in SO(4)$ *and* $\det(U_1)=\det(U_2)=1$ *where*

$$T = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & i & i & 0 \\ 0 & -1 & 1 & 0 \\ i & 0 & 0 & -i \end{pmatrix}. \tag{34}$$

*Analogously, consider* $2\times2$ *matrices* $A_1, A_2$, *then both* $A_1$ *and* $A_2$ *belong to* $SL(2,\mathcal{C})$ *if and only if* $T(A_1\otimes A_2)T^\dagger \in SO(4,\mathcal{C})$ *and* $\det(A_1)=\det(A_2)=1$.

*Proof:* This Lemma is a consequence of accidents in Lie-group theory: $SU(2)\otimes SU(2)\simeq SO(4)$ and $SL(2,\mathcal{C})\otimes SL(2,\mathcal{C})\simeq SO(4,\mathcal{C})$.
The proof easily follows from the observation that all groups under consideration are simply connected Lie groups, and that all elements are therefore characterized by their logarithms. One can check that the three matrices $iT(\sigma_i\otimes I_2)T^\dagger$ together with the three matrices $iT(I_2\otimes\sigma_i)T^\dagger$ form a complete basis for the antisymmetric real matrices (the $\{\sigma_i\}$ are the three Pauli matrices that are the generators of $SU(2)$, and the antisymmetric real matrices are the generators of $SO(4)$). The same argument holds in the case of $SO(4,\mathcal{C})$. $\square$

Note that the matrix $T$ was chosen such that it is a *false* square root of $T^T T = \epsilon_2\otimes\epsilon_2$ with $\epsilon_2$ the completely antisymmetric tensor of rank 2: the Theorem can easily be seen to hold by observing that $A^T\epsilon_2 A = \det(A)\epsilon_2$. Note also that the equivalence is not one to one but two to one : both $SU(2)$ or $SL(2,\mathcal{C})$ matrices can be multiplied by $-1$ to yield the same orthogonal matrix. As a last remark, we observe that the action of an element of $O(4)$ (or $O(4,\mathcal{C})$) with determinant equal to $-1$ corresponds to the action of local unitaries (or local filtering) preceded by an uneven permutation (physically corresponding to swapping two qubits).

### 3.2.1. LU equivalence classes

Let us first characterize a unique normal form for pure $2 \times 2 \times 4$ states under the action of local unitary operations. Each state is parameterized by a three index tensor $\psi_{i_1 i_2 i_3}$ with $i_1, i_2 \in \{1, 2\}$ and $i_3 \in \{1, 2, 3, 4\}$. This tensor can be rewritten as a $4 \times 4$ matrix $\tilde{\psi}$ by concatenating the indices $(i_1, i_2)$. Next we define the matrix R as

$$R = T\tilde{\psi} \tag{35}$$

where $T$ is defined in lemma 2. It is then straightforward to show that a local unitary transformation $|\psi'\rangle = U_1 \otimes U_2 \otimes U_3|\psi\rangle$ results in a transformation $R' = ORU^T$ with $O = T(U_1 \otimes U_2)T^\dagger \in SO(4)$ (note that without loss of generality $U_1$ and $U_2$ were chosen to have determinant 1). The problem is therefore reduced to define a unique normal form of a $4 \times 4$ matrix under left multiplication with a matrix in $SO(4)$ and right multiplication with one in $U(4)$. This can be accomplished as follows: first multiply $R$ with a phase such that its determinant is real and positive. Then determine the eigenvalue decomposition of the real part of $RR^\dagger$, denoted as $\mathrm{Re}\{RR^\dagger\}$, and call $O$ the matrix containing the eigenvectors (note that this procedure yields a unique $O$ up to a right multiplication with a diagonal orthogonal matrix containing only the elements $\pm 1$ if the eigenvalues have all multiplicity equal to 1). In a similar way, we can calculate the Takagi decomposition of $R^T R = V \Sigma V^T$, determining $V$ uniquely up to right multiplication with a diagonal matrix containing only elements $\pm 1$. If we define $U = V$ and $R' = O^T R U^*$, then it is easy to see that the real part of $R' R'^\dagger$ is diagonal and that $R'^T R'$ is diagonal. This $R'$ is unique up to left and right multiplication with diagonal matrices with entries $\pm 1$, which can be chosen to make all real parts of the elements in the first row and column positive (note that the left diagonal matrix has to be chosen such as to assure that the total left orthogonal transformation has determinant $+1$). Therefore we have proven the existence of a unique normal form:

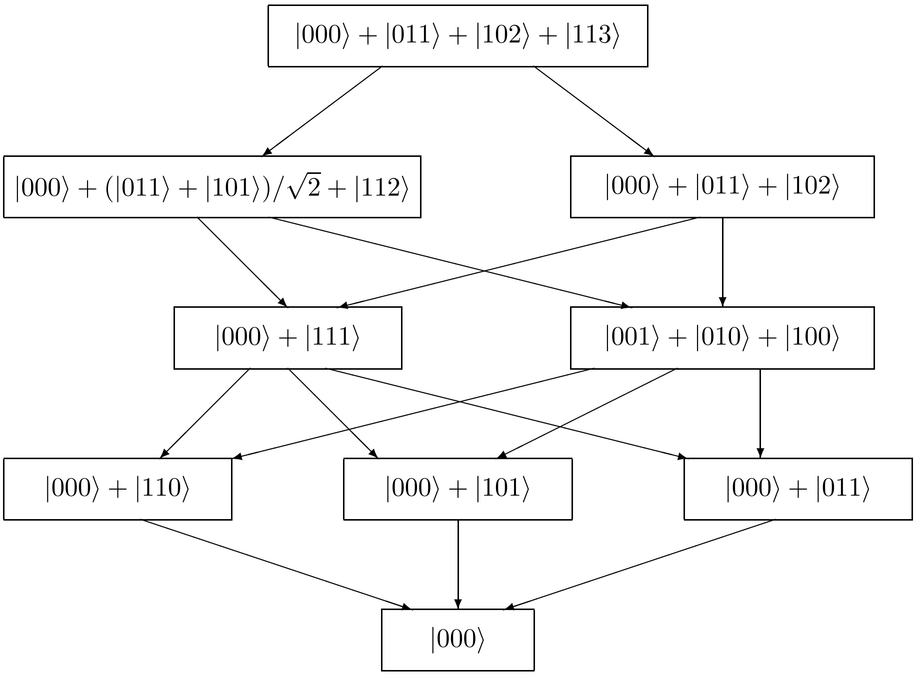
**Theorem 11.** *Each $2 \times 2 \times 4$ state can be brought into an easy to calculate unique normal form. The problem is equivalent to bringing a $4 \times 4$ complex matrix $R$ with real determinant into a unique normal form by multiplying it left with a matrix in $SO(4)$ and right with one in $SU(4)$, which can be done by considering the eigenvalue decomposition of $\mathrm{Re}\{RR^\dagger\}$ and the Takagi decomposition of $R^T R$.*

The obtained normal forms consist of a $32 - 2 - 6 - 15 = 9$ real parameter family of states. Note that the 3-qubit normal form under LU is just a special case of this decomposition.

### 3.2.2. SLOCC equivalence classes

A much simpler classification is possible in the case of SLOCC equivalence classes:

**Theorem 12.** *Consider a pure $2 \times 2 \times 4$ multipartite state, then it can be transformed to exactly one of the following 9 unnormalized states by reversible SLOCC operations:*




*Moreover, every state that is depicted higher in the hierarchy can be transformed by (irreversible) SLOCC operations to every one that is strictly lower in the hierarchy.*

*Proof*[3]*:* Each state is parameterized by a three index tensor $\psi_{i_1 i_2 i_3}$ with $i_1, i_2 \in \{1,2\}$ and $i_3 \in \{1,2,3,4\}$. This tensor can be rewritten as a $4 \times 4$ matrix $\tilde{\psi}$ by concatenating the indices $(i_1, i_2)$. Next we define the matrix R as

$$R = T\tilde{\psi} \tag{36}$$

[3]An alternative proof can be obtained by a direct application of the Lorentz singular value decomposition, which will be introduced in section 4.2

where $T$ is the one defined in lemma 2. Using the results of Lemma 2, the problem is equivalent to finding appropriate normal forms for the complex $4 \times 4$ matrix $R$ under left multiplication with a complex orthogonal matrix $O$ in $SO(4, \mathcal{C})$ and right multiplication with an arbitrary matrix $A$.
If the matrix $R$ is full rank, then $A$ can be chosen to be equal to $T^\dagger R^{-1}$, yielding the state

$$(|000\rangle + |011\rangle + |102\rangle + |113\rangle)/2, \tag{37}$$

the highest state in the hierarchy.
Suppose however that the rank of $R$ is three. As a first step, $R$ can always be multiplied right by $A$ and left by a permutation matrix such as to yield an $R$ of the form

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ a & b & c & 0 \end{pmatrix}. \tag{38}$$

Suppose $a \neq \pm i$, then it can easily be checked that right multiplication with

$$\begin{pmatrix} 1 & -b/(a+1/a) & -c/(a+1/a) & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{39}$$

and left multiplication by the complex orthogonal matrix

$$\begin{pmatrix} 1/\sqrt{1+a^2} & 0 & 0 & a/\sqrt{1+a^2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -a/\sqrt{1+a^2} & 0 & 0 & 1/\sqrt{1+a^2} \end{pmatrix} \tag{40}$$

yields a new $R$ of the form

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & b' & c' & 0 \end{pmatrix}. \tag{41}$$

Exactly the same can be done in the case where $b, c \neq \pm i$, and therefore we only have to consider the case where $a, b, c \in \{0, i, -i\}$. It can however be checked that in the case that when 2 or 3 elements $a, b, c$ are not equal to zero, a new $R$ can be made where all $a, b, c$ become equal to zero: this can be done by first multiplying $R$ with orthogonal matrices of the kind

$$O = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}, \tag{42}$$

and repeating the procedure outlined above. There remains the case where exactly one of the elements is equal to $\pm i$. Without loss of generality, we assume that $(a, b, c) = (i, 0, 0)$ (this is possible because one can do permutations (with

signs) by appropriate $O \in SO(4)$ and $A$). This case is fundamentally different from the one where all $a, b, c$ are equal to zero as the corresponding matrix $R^T R$ has rank 2 as opposed to rank 3 of $R$. There is no way in which this behaviour can be changed by multiplying $R$ left and right with appropriate transformations, and we therefore have identified a second class (which is clearly of measure zero: a generic rank 3 state $R$ will also yield a rank 3 $R^T R$).
It is now straightforward to construct a representative state of each class. As a representative of the generic class, we choose the state

$$|000\rangle + (|011\rangle + |101\rangle)/\sqrt{2} + |112\rangle \tag{43}$$

because it has all local density operators proportional to the identity: as argued before, this state is expected to be most entangled. As a representative of the non-generic class, we choose the state

$$|000\rangle + |011\rangle + |102\rangle \tag{44}$$

as it makes clear that the states in this class have Schmidt number 3 (as opposed to the states in the generic class that have Schmidt number 4).

The case where $R$ has rank 2 can be solved in a completely analogous way. Exactly the same reasoning leads to the following four possible normal forms for $R$:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix} \tag{45}$$

Note that the last two cases cannot be transformed into each other due to the constraint that $O$ has determinant $+1$. The corresponding representative states are easily obtained by choosing symmetric ones:

$$|000\rangle + |111\rangle \tag{46}$$
$$|001\rangle + |010\rangle + |100\rangle \tag{47}$$
$$|000\rangle + |011\rangle \tag{48}$$
$$|000\rangle + |101\rangle \tag{49}$$

The first state is the celebrated $GHZ$-state, the second one the $W$-state introduced by Dür, Vidal and Cirac [**80**], and the remaining ones represent states with only bipartite entanglement.

As a last class, we have to consider the one where $R$ has rank equal to 1. This leads to the following two possible normal forms for $R$:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix} \tag{50}$$

The corresponding states are given by

$$|000\rangle + |110\rangle \tag{51}$$

$$|000\rangle \tag{52}$$

which ends the complete classification.

It remains to be proven that each state that is higher in the depicted hierarchy can be transformed to all the other ones that are strictly lower. The first step downwards is evident from the fact that right multiplication of a rank 4 $R$ with a rank deficient $A$ can yield whatever $R$ of rank 3. In going from a generic rank 3 $R$ to a rank 2, the state $|000\rangle + (|011\rangle + |101\rangle)/\sqrt{2} + |112\rangle$ can be transformed into the GHZ-state by a projection of the third party on the subspace $\{|0\rangle, |2\rangle\}$ and into the $W$-state by the third party implementing the POVM element

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{53}$$

From the non-generic normal form of rank 3, the GHZ state can easily be constructed by a projection of the third party on his $\{|1\rangle, |2\rangle\}$ subspace, while the $W$-state is obtained by projecting on the $\{|0\rangle, |1\rangle + |2\rangle\}$ subspace. Finally, the conversion of GHZ- and W-states to EPR's is straightforward. □

We have therefore obtained a complete classification of all possible $2 \times 2 \times N$ pure states under SLOCC transformations. It is very nice that there is one and only one state presiding[4]. It is also remarkable that there exist incompatible kinds of entanglement at the same level.

Note that in section 3.1 we presented a way of how to quantify the amount of entanglement present in a multipartite state by means of entanglement monotones. In the $2 \times 2 \times 4$ case we obtained the entanglement monotone

$$\sqrt{4/3}\left|\sum \psi_{i_1j_1k_1}\psi_{i_2j_2k_2}\psi_{i_3j_3k_3}\psi_{i_4j_4k_4}\epsilon_{i_1i_2}\epsilon_{i_3i_4}\epsilon_{j_1j_3}\epsilon_{j_2j_4}\epsilon_{k_1k_2k_3k_4}\right|^{1/2} \tag{54}$$

which maximal value 1 is obtained for the "maximally entangled"state on top of the hierarchy. Note that this is the only entanglement monotone for $2 \times 2 \times 4$ systems that is invariant under SLOCC transformations: this follows from the fact that the only parameter remaining after the filtering procedure is exactly the normalization, given by the tangle.

In the $2 \times 2 \times 3$ case, the only entanglement monotone is given by:

$$\sqrt[3]{\frac{27}{4}}\Big|\sum \psi_{i_1,j_1,k_1}\psi_{i_2,j_2,k_2}\psi_{i_3,j_3,k_3}\psi_{i_4,j_4,k_4}\psi_{i_5,j_5,k_5}\psi_{i_6,j_6,k_6} \epsilon_{i_1i_4}\epsilon_{i_2i_5}\epsilon_{i_3i_6}\epsilon_{j_1j_4}\epsilon_{j_2j_5}\epsilon_{j_3j_6}\epsilon_{k_1k_2k_3}\epsilon_{k_4k_5k_6}\Big|^{1/3} \tag{55}$$

[4] It is tempting to conjecture that the whole family of nine depicted states forms a minimal reversible entanglement generating set (MREGS)

Note that again the generalization of the GHZ-state was chosen such as to maximize this entanglement monotone (i.e. value 1), and that the generalization of the W-state has $2-2-3$-tangle equal to zero, as should be.

In the $2\times 2\times 2$ case, the entanglement monotone

$$\sqrt{2}|\psi_{i_1j_1k_1}\psi_{i_2j_2k_2}\psi_{i_3j_3k_3}\psi_{i_4j_4k_4}\epsilon_{i_1i_2}\epsilon_{i_3i_4}\epsilon_{j_1j_2}\epsilon_{j_3j_4}\epsilon_{k_1k_3}\epsilon_{k_2k_4}|^{1/2} \tag{56}$$

is the square root of the 3-tangle introduced by Wootters [**59**], which has a very appealing property: it is equivalent to

$$C^2_{A,(BC)} - C^2_{AB} - C^2_{AC}. \tag{57}$$

Here $C_{A,(BC)}$ is the expression for the concurrence of the pure bipartite state obtained by considering $B$ and $C$ to be one joint party, $C_{AB}$ is the concurrence of the mixed state obtained by tracing out $C$, and similarly for $C_{AC}$. Since the total expression is always positive and since the concurrence is bounded above by 1, this means that the amount of entanglement that $A$ shares with $B$ is bounded by the amount of entanglement that $A$ shares with $C$: a state has only a finite susceptibility for entanglement. The more a system is entangled with another one, the less it can be entangled with all the rest. In the case of a GHZ-state for example, $C_{AB} = C_{AC} = 0$ such that all entanglement is genuinely tripartite; on the contrary, the 3-tangle is zero in the case of $W$-states [5] such that one could argue that all the entanglement present is genuinely bipartite. The previous observation of entanglement susceptibility is also evident from the fact that a maximally entangled bipartite state is pure (a mixed state would be obtained if more entanglement were possible if the particle were entangled with another one).

As is obvious by now, the way of transforming $2\times 2\times N$ states into each other is by no means unique. It would be nice to find the optimal way of converting one state into another one. It turns out that this is much more difficult than in the bipartite case. In the case of the transformation of a $2\times 2\times 2$ state belonging to the GHZ-class, the optimal probability can be calculated explicitly.

### 3.2.3. Optimal distillation of the GHZ-state

The SLOCC operations bringing a generic pure $2\times 2\times 2$ state to the GHZ form are not unique but consist of a four-parameter family. This happens because a pure tripartite state has 14 degrees of freedom and the three Lorentz transformations have 18 independent real parameters. Indeed, if $A\otimes B\otimes C|\psi\rangle = |GHZ\rangle$, then also

$$\begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix} A \otimes \begin{pmatrix} b & 0 \\ 0 & 1/b \end{pmatrix} B \otimes \begin{pmatrix} 1/ab & 0 \\ 0 & ab \end{pmatrix} C|\psi\rangle = |GHZ\rangle \tag{58}$$

[5] This is of course in complete accordance with the fact that its "general" normal form is equal to zero.

with $a$ and $b$ complex numbers. The single-copy distillation of a GHZ state is therefore not unique. The probability by which an SLOCC operation produces the desired result can therefore be optimized such as to yield the optimal single-copy distillation protocol. This optimal procedure was previously found by Acin et al. [**3**], but they did not have a complete proof of the optimality. The main difficulty is to prove that a 1-branch protocol is optimal, meaning that it is sufficient that all parties implement a local filtering operation, and that there is no gain in considering protocols where more than one of the possible outputs is a $GHZ$-state.

Consider a generic pure state of three qubits (automatically belonging to the GHZ-class). Then we will show that the optimal 1-copy distillation protocol, i.e. the protocol making a GHZ-state with the highest possible probability, is a 1-branch protocol.

The problem can now be tackled as follows. The most general local procedure of distilling a GHZ-state out of a single copy of a pure state consists of a multi-branch protocol in which different branches consist of different SLOCC operations connected through equation (58). There is no restriction in taking all $\{A_i\}, \{B_i\}, \{C_i\}$ to have determinant 1, and the SLOCC operations corresponding to each branch are of the form

$$\begin{aligned}
&q_i A_i \otimes B_i \otimes C_i|\psi\rangle = q_i \tau^{1/2}|GHZ\rangle \\
&A_i = D_i^a A_0 \quad D_i^a = \begin{pmatrix} a_i & 0 \\ 0 & 1/a_i \end{pmatrix} \\
&B_i = D_i^b B_0 \quad D_i^b = \begin{pmatrix} b_i & 0 \\ 0 & 1/b_i \end{pmatrix} \\
&C_i = D_i^c C_0 \quad D_i^c = \begin{pmatrix} 1/a_i b_i & 0 \\ 0 & a_i b_i \end{pmatrix}.
\end{aligned}$$

Here $\tau$ is the 3-tangle of $\psi$ and $q_i$ is a real proportionality factor such as to assure that all the branches together are implementable as a part of a POVM. This leads to a necessary (but generally not sufficient) condition:

$$\sum_i q_i^2 A_i^\dagger A_i \otimes B_i^\dagger B_i \otimes C_i^\dagger C_i \leq I_8 \tag{59}$$

Each branch yields the GHZ-state with probability $q_i^2\tau$, and therefore the total probability is given by $\tau \sum_i q_i^2$, which has to be maximized. Due to the condition (59), an upper bound on this probability can be derived. It will turn out that this upper bound is achievable by a 1-branch protocol. Defining $p_i = q_i^2/(\sum_i q_i^2)$, it holds that the total probability is bounded by

$$\max_{\{A_i\},\{B_i\},\{C_i\}} \frac{\tau}{\lambda_{\max}(\sum_i p_i A_i^\dagger A_i \otimes B_i^\dagger B_i \otimes C_i^\dagger C_i)} \tag{60}$$

where $\lambda_{\max}(X)$ denotes the largest eigenvalue of operator $X$. An upper bound is therefore obtained by minimizing this largest eigenvalue.

Therefore the standard techniques for differentiating the eigenvalues of a matrix can be used [**63**]: given a Hermitian matrix $X$, its eigenvalue decomposition $X = UEU^\dagger$ and its variation $\dot{X}$, then the variation on its eigenvalues is given by $\dot{E} = \mathrm{diag}\{U^\dagger \dot{X} U\}$. Here we take

$$\begin{aligned} X &= Z_0^\dagger \underbrace{\sum_i p_i D_i}_{D} Z_0 \\ Z_0 &= A_0 \otimes B_0 \otimes C_0 \\ D_i &= |D_i^a|^2 \otimes |D_i^b|^2 \otimes |D_i^c|^2 \end{aligned}$$

Note that varying the free parameters $\{a_i, b_i, p_i\}$ only affects $D$ and not $Z_0$. In the case of an extremal maximal eigenvalue all variations $\dot{\lambda}_{\max} = \mathrm{Tr}\left(\dot{E}P_{11}\right)$ with $P_{11} = \mathrm{diag}[1;0;0;0;0;0;0;0]$ have to be equal to zero:

$$\mathrm{Tr}\left((\delta D) Z_0 U P_{11} U^\dagger Z_0^\dagger\right) = 0$$

The following identities are easily obtained:

$$\begin{aligned} \frac{\delta D}{\delta a_i} &= \frac{2}{a_i}\mathrm{diag}[0,1,0,1,-1,0,-1,0] D_i \\ \frac{\delta D}{\delta b_i} &= \frac{2}{b_i}\mathrm{diag}[0,1,-1,0,0,1,-1,0] D_i \\ \frac{\delta D}{\delta \sqrt{p_i}} &= 2\sqrt{p_i} D_i \end{aligned}$$

Therefore only the (real and positive) diagonal elements of $Z_0 U P_{11} U^\dagger Z_0^\dagger$ are of importance and let us write them in the vector $z_0$. Similarly, we write the diagonal elements of $D_i$ in the vector $d_i = [1; |a_i b_i|^2; 1/|b_i|^2; |a_i|^2; 1/|a_i|^2; |b_i|^2; 1/|a_i b_i|^2; 1]$, and the extremal relations become:

$$\begin{aligned} \forall i: \quad & 0 = d_i^T \mathrm{diag}[0,1,0,1,-1,0,-1,0] z_0 \\ & 0 = d_i^T \mathrm{diag}[0,1,-1,0,0,1,-1,0] z_0 \\ & \mu = d_i^T z_0 \end{aligned} \tag{61}$$

where $\mu$ is the Lagrange multiplier corresponding to the condition $\sum_i (\sqrt{p_i})^2 = 1$. This forms sets of each time 3 equations for 2 unknowns $a_i, b_i$, which can be shown to have exactly one solution. Indeed, the first and second equation lead to

$$\begin{aligned} |a_i|^4 &= \frac{z_0(5) + z_0(7)/|b_i|^2}{z_0(4) + z_0(2)|b_i|^2} \\ |b_i|^4 &= \frac{z_0(3) + z_0(7)/|a_i|^2}{z_0(6) + z_0(2)|a_i|^2}. \end{aligned} \tag{62}$$

Let us analyze how these equations behave. When $b_i \to 0$ then the solution of the first equation goes like $|a_i| \sim 1/\sqrt{|b_i|}$ and when $a_i \to 0$ then $|b_i| \sim 1/|a_i|^2$. Exactly the opposite happens in the case of the second equation, and due to this different asymptotic behaviour it is assured that both curves cross and therefore at least one solution exists for all (real positive) values of $z_0$. Moreover there is always at most one solution. To prove this, we first note that $|a_i|$ and $|b_i|$ can be scaled such that both curves cross at the value $(1,1)$, and we call these rescaled variables $(x,y)$ and $\bar{z}_0$. The hyperbola $xy = 1$ crosses both rescaled curves (62) at $(1,1)$. Moreover it is trivial to check that the hyperbola does not cross any of the rescaled curves anymore in the first quadrant (this amounts to solving a quadratic equation), and due to the asymptotic behaviour one curve lies below and the other one above the hyperbola (except in $(1,1)$). Therefore both rescaled curves have exactly one crossing. Therefore for all (real positive) values in $z_0$, there is always exactly one real solution for $|a_i|, |b_i|$, and as $z_0$ is independent of the index $i$, all $|a_i|$ are equal to each other and the same applies to the $|b_i|$. Therefore at most the phase of the constants $\{a_i, b_i\}$ varies in different branches, and as this amounts to local unitary operations we conclude that all branches are equivalent and can be implemented by a one-branch protocol.

A serious objection can be raised to the previous arguments: we derived stationarity condition on the eigenvalues of a matrix, but nothing assures that the minimum of the largest eigenvalue actually occurs at a differentiable point; this follows from the fact that an the eigenvalue in function of the free parameters is not necessarily an analytic function anymore at a point where its multiplicity exceeds 1 [**63**]. Let us however for the moment suppose that the extremal value of $\lambda_{\max}$ indeed occurs at a point where the multiplicity of the eigenvalue is just 1, and we will discuss the other case later.

In the case of a one branch protocol, the eigenvectors of $X$ can be calculated analytically as $X$ becomes a tensor product of $2 \times 2$ matrices. Instead of calculating the eigenvalues and eigenvectors of $X$, it is more convenient to calculate them for $YY^\dagger$ with

$$Y = \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix} A \otimes \begin{pmatrix} b & 0 \\ 0 & 1/b \end{pmatrix} B \otimes \begin{pmatrix} 1/ab & 0 \\ 0 & ab \end{pmatrix} C.$$

Similar conditions as the ones described in equations (61) are readily obtained. Given particular determinant 1 transformations $A, B, C$ and taking $a, b$ to be real, let us define $\tilde{A} = AA^\dagger, \tilde{B} = BB^\dagger, \tilde{C} = CC^\dagger$. The eigenvector $v$ corresponding to the *largest* eigenvalue of the matrix $YY^\dagger$ happens to be $v = v_1 \otimes v_2 \otimes v_3$ with

$$\begin{aligned}
v_i &= \begin{pmatrix} -\beta_i + \sqrt{|\alpha_i|^2 + \beta_i^2} \\ \alpha_i \end{pmatrix} \\
\alpha_1 &= -2\tilde{A}_{21} \qquad \beta_1 = \tilde{A}_{11} a^2 - \tilde{A}_{22}/a^2 \\
\alpha_2 &= -2\tilde{B}_{21} \qquad \beta_2 = \tilde{B}_{11} b^2 - \tilde{B}_{22}/b^2 \\
\alpha_3 &= -2\tilde{C}_{21} \qquad \beta_3 = \tilde{C}_{11}/(ab)^2 - \tilde{C}_{22}(ab)^2
\end{aligned}$$

The extremality conditions can be shown to be equivalent to:

$$\frac{\tilde{A}_{11}a^2 - \tilde{A}_{22}/a^2}{|\tilde{A}_{21}|} = \frac{\tilde{B}_{11}b^2 - \tilde{B}_{22}/b^2}{|\tilde{B}_{21}|} = \frac{\tilde{C}_{11}/(ab)^2 - \tilde{C}_{22}(ab)^2}{|\tilde{C}_{21}|}.$$

These equations have to be solved in the unknowns $a$ and $b$. $b$ can readily be written in function of $a$ through one of those, and then a sixth order equation in the remaining unknown $a^2$ results. As shown, only one solution corresponding to a physical solution for $a$ and $b$ exists, and this solution can easily be solved numerically. The optimal local filtering operations and the maximal probability of making a GHZ-state (an entanglement monotone) can then easily be calculated.

We still have to check whether the extremum did not occur at a point where the multiplicity of the largest eigenvalue exceeds 1. We have already shown that only one extremum is obtained along a particular path of an eigenvalue. The proof is therefore complete if we can show that the *maximal* eigenvalue reaches a minimum at the solution obtained; this ensures that no crossings with other eigenvalues occurred. This is indeed true in the generic case: the maximal eigenvalue of the matrix $YY^\dagger$, a pure tensor product, is the product of the largest eigenvalues of its $2\times 2$ components. We therefore have to calculate the eigenvalues of matrices of the kind

$$AA^\dagger \begin{pmatrix} x^2 & 0 \\ 0 & 1/x^2 \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \beta^* & \gamma \end{pmatrix} \begin{pmatrix} x^2 & 0 \\ 0 & 1/x^2 \end{pmatrix},$$

which are given by

$$\frac{1}{2}\left(\alpha x + \frac{\gamma}{x} \pm \sqrt{(\alpha x + \frac{\gamma}{x})^2 - 4(\alpha\gamma - |\beta|^2)}\right).$$

A crossing of these eigenvalues can never occur if $|\beta| > 0$, which is generically indeed always the case. It is therefore proven that the optimal protocol in the generic case consists of a 1-branch protocol, and can be obtained by solving a 6'th order equation in 1 unknown parameter.

A similar non-uniqueness exists in the case of distilling a state in the class of W-states to the W-state. Indeed, if $A\otimes B\otimes C|\psi\rangle = |W\rangle$, then the most general symmetry operations are given by

$$\begin{aligned}
A' \otimes B' \otimes C'|\psi\rangle &= |W\rangle \\
A' &= \begin{pmatrix} x & y \\ 0 & 1/x \end{pmatrix} A \\
B' &= \begin{pmatrix} x & z \\ 0 & 1/x \end{pmatrix} B \\
C' &= \begin{pmatrix} x & -(y+z) \\ 0 & 1/x \end{pmatrix} C
\end{aligned}$$

with $x, y, z$ arbitrary complex numbers. As every matrix can be written as the product of a unitary matrix and an upper triangular matrix (this is the so-called

QR-decomposition), there are enough degrees of freedom to make whatever one out of $A'$,$B'$ or $C'$ equal to a unitary matrix. Numerical investigations reveal that one of these three possibilities is also the optimal choice over all 1-branch protocols in the sense that it will yield a distillation protocol that produces the W-state with the highest possible probability. Therefore the optimal 1-branch distillation protocol of a W-state consists of two parties applying a local filtering operation, while one party performs a local unitary operation. Note however that we don't know if multi-branch protocols can do better, as the methods used in the case of the distillation of GHZ-states cannot readily be generalized.

## 3.3. The $2 \times 2 \times 2 \times 2$ case

The $2 \times 2 \times N$ case of the previous section could essentially be solved because we could transform the problem about tensors into a problem about matrices by using a very useful accident of Lie group theory. The same trick can be applied to the four qubit case, although much more sophistication is needed. Let us first analyze the orbits under the action of local unitary transformations.

### 3.3.1. LU equivalence classes

A pure state of four qubits is parameterized by a four index tensor $\psi_{i_1 i_2 i_3 i_4}$ with $i_j \in \{1, 2\}$. This tensor can be rewritten as a $4 \times 4$ matrix $\tilde{\psi}$ by concatenating the indices $(i_1, i_2)$ and $(i_3, i_4)$. Next we define the matrix R as

$$R = T\tilde{\psi}T^\dagger \tag{63}$$

where $T$ was defined in Lemma 2. Using lemma 2 it is straightforward to show that a local unitary transformation $|\psi'\rangle = U_1 \otimes U_2 \otimes U_3 \otimes U_4|\psi\rangle$ results in a transformation $R' = O_1 R O_2$ with $O_1, O_2 \in SO(4)$ and $O_1 = T(U_1 \otimes U_2)T^\dagger$, $O_2 = T(U_3 \otimes U_4)^T T^\dagger$. A normal form under local unitary operations can now be imposed as follows: make the determinant of $R$ real and positive by multiplying the whole matrix with the appropriate phase[6], and use $O_1$ and $O_2$ to diagonalize the real part of $R$ through the unique real singular value decomposition. This procedure eliminates all 13 degrees of freedom of the local unitary operations, and two states are therefore equivalent up to local unitary operations iff they have the same normal form. In the case of degenerate singular values, a more sophisticated analysis can be done to yield a unique normal form.

[6]Strictly speaking, this operation is only unique up to a factor $\pm 1, \pm i$, but one can take care of this discrete degeneracy by imposing that e.g. the real part of the trace of the final normal form is maximal.

### 3.3.2. SLOCC equivalence classes

Let us now try to characterize the local orbits generated by SLOCC operations of the form

$$|\psi'\rangle = A_1 \otimes A_2 \otimes A_3 \otimes A_4 |\psi\rangle \tag{64}$$

with $\{A_i\}$ full rank and therefore invertible $2\times 2$ matrices. There is no restriction in choosing $\{A_i\} \in SL(2,\mathcal{C})$, and then the accident of Lemma 2 can be exploited:

$$SL(2,\mathcal{C}) \otimes SL(2,\mathcal{C}) \simeq SO(4,\mathcal{C}). \tag{65}$$

Using the same parameterization as in the LU-case, SLOCC operations correspond to left and right multiplication of $R$ with complex orthogonal matrices. The challenge is now to exploit the two times 12 degrees of freedom of these complex orthogonal matrices to bring $R$ into an unique normal form with maximal 8 real degrees of freedom left. This will be possible using some advanced techniques of linear algebra. Let us first recall the structure of a symmetric Jordan block (see Horn and Johnson [**116**] 4.4.9). If e.g. the dimension of the Jordan block is 5, the symmetric Jordan block is given by

$$S_5 = \frac{1}{2}\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} + \frac{i}{2}\begin{pmatrix} 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ 0 & -1 & 0 & 1 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}. \tag{66}$$

The following Theorem is a generalization of the singular value decomposition to complex orthogonal equivalence:

**Theorem 13.** *Given a complex $n\times n$ matrix $R$, then there always exist complex square orthogonal matrices $O_1$ and $O_2$ such that $R' = O_1 R O_2$ is a unique direct sum of blocks of the form:*
*1. $m \times m$ blocks of the form $(\lambda_j I_m + S_m)$ with $S_m$ symmetric Jordan blocks of the kind considered in equation (66). Here $\lambda_j$ is a complex parameter (note that the case $m = 1$ corresponds to the scalar case, and that nothing prevents that $\lambda_j$ is equal to zero).*
*2. $m\times m$ blocks consisting of an upper left $m_1 \times (m_1+1)$ part being the matrix obtained by taking the even rows and odd columns of an $(2m_1+1)\times(2m_1+1)$ symmetric Jordan block of the kind (66), and a lower right $(m-m_1)\times(m-m_1-1)$ part being the transpose of the matrix obtained by taking the odd rows and even columns of a $(2(m-m_1)-1)\times(2(m-m_1)-1)$ symmetric Jordan block of the kind (66).*

Proof: Consider the $2n \times 2n$ complex symmetric matrix

$$P = \begin{pmatrix} 0 & R \\ R^T & 0 \end{pmatrix}. \tag{67}$$

Due to Theorem 5 in ch.XI of Gantmacher [**94**], there exists a complex orthogonal $W$ such that $P = WP'W^T$ with $P'$ a direct sum of symmetric $m \times m$ Jordan blocks $J_i$ with eigenvalue $\lambda_i$. Note that the eigenspaces corresponding to different Jordan blocks are orthogonal to each other. Next we observe that whenever $[v_1; v_2]$ ($v_1$ and $v_2$ both have $n$ rows such that $[v_1; v_2]$ has $2n$ rows) is the eigenspace of $P$ corresponding to a symmetric Jordan block $J_i$, then $[v_1; -v_2]$ is the eigenspace of $P$ corresponding to a Jordan block $-J_i$. Due to the uniqueness of the Jordan canonical decomposition, these eigenspaces will be either orthogonal (this holds for example for sure if the corresponding eigenvalue is different from zero), or equal to each other (which implies that the corresponding eigenvalue is equal to zero). If the first case applies, both $v_1$ and $v_2$ are orthogonal isometries: orthogonality of $W$ implies $v_1^T v_1 + v_2^T v_2 = I$ and $v_1^T v_1 - v_2^T v_2 = 0$, and therefore $v_1^T v_1 = v_2^T v_2 \simeq I$.

The second degenerated case however is more difficult. In this case, it holds that $[v_1; v_2] = [v_1; -v_2]Q$ for some orthogonal $Q$ (indeed, if $[v_1; v_2]$ is an orthogonal isometry, then also $[v_1; -v_2]$). The structure of $P$ dictates that $Q$ is an orthogonal matrix for which $Q^T J Q = -J$ with $J$ a symmetric Jordan block of the appropriate dimension. We now proceed to prove that these conditions completely determine $Q$.

Let us first calculate the standard non-symmetric Jordan canonical form $\tilde{J}$ of the symmetric Jordan block with eigenvalue 0: $\tilde{J} = U^\dagger J U$ with $U$ unitary and symmetric; this $U$ has the special property that $U^2 = \text{Sip}$, where the Sip-matrix is the permutation matrix permuting all vectors $[x_1, x_2 \cdots x_n]$ to $[x_n, x_{n-1} \cdots x_1]$. In the $5 \times 5$ case for example, it holds that

$$\text{Sip} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad U = \frac{1}{2} \begin{pmatrix} 1+i & 0 & 0 & 0 & 1-i \\ 0 & 1+i & 0 & 1-i & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 1-i & 0 & 1+i & 0 \\ 1-i & 0 & 0 & 0 & 1+i \end{pmatrix}.$$

If we define $\tilde{Q} = U^\dagger Q U$, the following identities hold: $\tilde{Q}^T \text{Sip} \tilde{Q} = \text{Sip}$, $\tilde{Q}\tilde{J} = -\tilde{J}\tilde{Q}$ and $[x_1; x_2]^T [x_1; x_2] = \text{Sip}$. The condition $\tilde{Q}\tilde{J} = -\tilde{J}\tilde{Q}$ implies that $Q$ is of the form

$$\text{diag}[1, -1, 1, -1, \cdots] \begin{pmatrix} a & b & c & \cdots & & \\ 0 & a & b & c & \cdots & \\ 0 & 0 & a & b & c & \cdots \\ \cdots & & & & & \end{pmatrix}$$

and it is easy to show that the condition $\tilde{Q}^T \text{Sip} \tilde{Q} = \text{Sip}$ on its turn implies that $b = c = \cdots = 0$. Therefore $\tilde{Q}$ is equal to the matrix $\tilde{Q}_{\alpha\beta} = \pm(-1)^\alpha \delta_{\alpha\beta}$.

Retransforming to the picture with symmetric Jordan forms, it turns out that the situation is different for blocks with even and odd dimension. Let us first treat the case of even dimension. Then $Q_{\alpha\beta} = i(-1)^\alpha \text{Sip}_{\alpha\beta}$. This $Q$ is such that

$[v_1; v_2] = [v_1; -v_2]Q$. Writing out these equations explicitly, it is immediately clear that this system of equations can only hold if $v_1 = v_2 = 0$; this proves that it is not possible that $P$ has Jordan blocks of the degenerated kind of even dimension. In the case of odd dimension, $Q$ is given by $Q_{\alpha\beta} = \pm(-1)^\alpha \delta_{\alpha\beta}$, and this conditions $v_1$ and $v_2$ to be either of the form

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_1 & 0 & b_1 & 0 & c_1 & \cdots \\ 0 & a_2 & 0 & b_2 & 0 & \cdots \end{pmatrix} \tag{68}$$

or

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 & a_1 & 0 & b_1 & 0 & \cdots \\ a_2 & 0 & b_2 & 0 & c_2 & \cdots \end{pmatrix}. \tag{69}$$

Just as in the non-degenerate case, the vectors $[a_1, b_1, \cdots]$ form an orthogonal set that is furthermore orthogonal to all other isometries $v_i$ (the same holds of course for $[a_2, b_2, \cdots]$).

As the dimension of a $J_i$ giving rise to the degenerated case has to be odd, it is compulsory that there is an even number of degenerated cases (indeed, the non-degenerate cases give rise to two times a similar block and the total dimension of $P$ is even). More precisely, for each $[v_1; v_2]_j$ of the form (68), there has to exist a $[v_1; v_2]_k$ of the form (69) (eventually of different dimension), as this is the only way one can assure that $W$ will be orthogonal. The eigenstructure of such pairs of degenerate cases can then be brought into the form

$$\begin{pmatrix} a_1^i & b_1^i \cdots & a_1^k & b_1^k & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots \\ \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots a_2^i & b_2^i \cdots & a_2^k & b_2^k & \cdots \end{pmatrix}$$

by right multiplication with a permutation matrix $S$. The effect on $J_i$ and $J_k$ is to transform them as

$$S^T \begin{pmatrix} J_i & 0 \\ 0 & J_k \end{pmatrix} S = \begin{pmatrix} 0 & 0 & K_i & 0 \\ 0 & 0 & 0 & K_k^T \\ K_i^T & 0 & 0 & 0 \\ 0 & K_k & 0 & 0 \end{pmatrix} \tag{70}$$

where $K_\nu$ represents the matrix obtained by taking the even rows and odd columns of the symmetric Jordan block $J_\nu$ (if $J_\nu$ is a $1 \times 1$ matrix, then $K_\nu$ is just the empty matrix). This can be seen as follows: the entries of a symmetric Jordan block $J_{\alpha\beta}$ of odd dimension are equal to zero wherever $\alpha, \beta$ are both even or odd. As an example, consider the case of a $5 \times 5$ Jordan Block; then $K$ is given by

$$K = \frac{1}{2} \begin{pmatrix} 1 & -i \\ 1-i & 1+i \\ i & 1 \end{pmatrix}.$$

Collecting all the pieces, it is now easily verified that we have obtained a decomposition in the form

$$P = \begin{pmatrix} O_1 & O_3 & 0 & O_1 \\ O_2 & 0 & O_4 & -O_2 \end{pmatrix} \begin{pmatrix} J & 0 & 0 & 0 \\ 0 & 0 & \begin{pmatrix} K_1 & 0 \\ 0 & K_2^T \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} K_1^T & 0 \\ 0 & K_2 \end{pmatrix} & 0 & 0 \\ 0 & 0 & 0 & -J \end{pmatrix} \begin{pmatrix} O_1^T & O_2^T \\ O_3^T & 0 \\ 0 & O_4^T \\ O_1^T & -O_2^T \end{pmatrix}$$

where $O_1^T O_1 = O_2^T O_2 = I; O_3^T O_3 = O_4^T O_4 = I; O_1^T O_3 = 0 = O_2^T O_4$ and where we considered the situation with 1 non-degenarate Jordan block and two complementary degenarate Jordan blocks. This leads exactly to a normal form as stated in the theorem:

$$R = O_1 J O_2^T + O_3 \begin{pmatrix} K_1 & 0 \\ 0 & K_2^T \end{pmatrix} O_4^T.$$

Remark that we can always choose the real part of the $\{\lambda_i\}$ to be positive, yielding a unique normal form. $\square$

Note that the proof was constructive, and that the exact structure of the obtained normal form of a matrix $R$ can readily be derived from calculating the eigenstructure of the matrix

$$P = \begin{pmatrix} 0 & R \\ R^T & 0 \end{pmatrix}.$$

Let us next introduce the following notation: non-degenerate Jordan blocks of dimension $n$ (including the scalar case) associated to a number $\lambda_i$ are written as $J_n(\lambda_i)$, while the non-generic blocks are denoted as $K_{n\oplus\bar{m}}$ where $n, m$ denote the dimension of the Jordan blocks appearing in the eigenvalue decomposition of $P$. Here $n$ is the dimension of the degenerated Jordan block of $P$ with eigenvectors of the kind given in equation (68) and $\bar{m}$ the dimension of the degenerated Jordan block of $P$ with eigenvectors as in (69). Note that the dimension of the matrix $K_{n\oplus\bar{m}}$ is given by $(n+m)/2 \times (n+m)/2$. The normal form of a generic matrix will always be something of the kind $J_1(\lambda_1) \oplus J_1(\lambda_2) \oplus J_1(\lambda_3) \oplus J_1(\lambda_4)$, while special cases are for example of the form $J_2(\lambda_1) \oplus K_{3\oplus\bar{1}}$.

If we consider for example of a $4 \times 4$ matrix $R$, then its normal form under complex orthogonal equivalence is a direct sum of blocks of the following kind:

$$
\begin{aligned}
J_1(\lambda) &= \lambda \\
J_2(\lambda) &= \lambda I_2 + \frac{1}{2}\begin{pmatrix} -i & 1 \\ 1 & i \end{pmatrix} \qquad J_3(\lambda) = \lambda I_3 + \frac{1}{2}\begin{pmatrix} 0 & 1-i & 0 \\ 1-i & 0 & 1+i \\ 0 & 1+i & 0 \end{pmatrix} \\
J_4(\lambda) &= \lambda I_4 + \frac{1}{2}\begin{pmatrix} 0 & 1 & -i & 0 \\ 1 & -i & 1 & i \\ -i & 1 & i & 1 \\ 0 & i & 1 & 0 \end{pmatrix} \qquad K_{3\oplus\bar{1}} = \frac{1}{2}\begin{pmatrix} 1-i & 0 \\ 1+i & 0 \end{pmatrix} \\
K_{5\oplus\bar{1}} &= = \frac{1}{2}\begin{pmatrix} 1 & -i & 0 \\ 1-i & 1+i & 0 \\ i & 1 & 0 \end{pmatrix} \qquad K_{7\oplus\bar{1}} = \frac{1}{2}\begin{pmatrix} 1 & 0 & -i & 0 \\ 1 & 1-i & i & 0 \\ -i & 1+i & 1 & 0 \\ i & 0 & 1 & 0 \end{pmatrix} \\
K_{3\oplus\bar{3}} &= \frac{1}{2}\begin{pmatrix} 1-i & 0 & 0 \\ 1+i & 0 & 0 \\ 0 & 1-i & 1+i \end{pmatrix} \qquad K_{5\oplus\bar{3}} = \frac{1}{2}\begin{pmatrix} 1 & -i & 0 & 0 \\ 1-i & 1+i & 0 & 0 \\ i & 1 & 0 & 0 \\ 0 & 0 & 1-i & 1+i \end{pmatrix} \\
K_{1\oplus\bar{7}} &= \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & -i & i \\ 0 & 1-i & 1+i & 0 \\ -i & i & 1 & 1 \end{pmatrix} \qquad K_{1\oplus\bar{3}} = \frac{1}{2}\begin{pmatrix} 0 & 0 \\ 1-i & 1+i \end{pmatrix} \\
K_{1\oplus\bar{5}} &= \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 1 & 1-i & i \\ -i & 1+i & 1 \end{pmatrix} \qquad K_{3\oplus\bar{5}} = \begin{pmatrix} 1-i & 0 & 0 & 0 \\ 1+i & 0 & 0 & 0 \\ 0 & 1 & 1-i & i \\ 0 & -i & 1+i & 1 \end{pmatrix}
\end{aligned}
$$

This yields a complete classification of all possible normal forms for $4 \times 4$ matrices. Note that it is easily seen how this generalized to arbitrary dimensions.

Let us now look how the present results apply to the study of SLOCC equivalence classes of 4 qubits. Due to the equivalence of $SL(2,\mathcal{C}) \otimes SL(2,\mathcal{C})$ and $SO(4,\mathcal{C})$, the normal forms arising in the above classification will immediately yield a natural representative state for each class of 4-qubit states connected by SLOCC operations. The normal form encodes the genuine non-local properties of the state, while the SLOCC operators needed to bring the state into normal form characterize the local information. The only problem still left is the fact that the orthogonal matrices considered in the previous theorem were not restricted to have determinant $+1$. As already remarked however, a complex orthogonal matrix with determinant equal to $-1$ corresponds to a SLOCC operation together with a permutation of two qubits, and so we can just use the theorem and keep in mind that maybe a permutation has to be done.

The following classification is obtained, where we had to choose a representative state out of each class:

**Theorem 14.** *A pure state of 4 qubits can, up to permutations of the qubits, be transformed into one of the following 9 families of states by determinant 1 SLOCC operations (64):*

$$\begin{aligned}
G_{abcd} &= \frac{a+d}{2}(|0000\rangle+|1111\rangle)+\frac{a-d}{2}(|0011\rangle+|1100\rangle)\\
&\quad+\frac{b+c}{2}(|0101\rangle+|1010\rangle)+\frac{b-c}{2}(|0110\rangle+|1001\rangle)\\
L_{abc_2} &= \frac{a+b}{2}(|0000\rangle+|1111\rangle)+\frac{a-b}{2}(|0011\rangle+|1100\rangle)\\
&\quad+c(|0101\rangle+|1010\rangle)+|0110\rangle\\
L_{a_2b_2} &= a(|0000\rangle+|1111\rangle)+b(|0101\rangle+|1010\rangle)\\
&\quad+|0110\rangle+|0011\rangle\\
L_{ab_3} &= a(|0000\rangle+|1111\rangle)+\frac{a+b}{2}(|0101\rangle+|1010\rangle)\\
&\quad+\frac{a-b}{2}(|0110\rangle+|1001\rangle)\\
&\quad+\frac{i}{\sqrt{2}}(|0001\rangle+|0010\rangle+|0111\rangle+|1011\rangle)\\
L_{a_4} &= a(|0000\rangle+|0101\rangle+|1010\rangle+|1111\rangle)\\
&\quad+(i|0001\rangle+|0110\rangle-i|1011\rangle)\\
L_{a_2 0_{3\oplus\bar{1}}} &= a(|0000\rangle+|1111\rangle)+(|0011\rangle+|0101\rangle+|0110\rangle)\\
L_{0_{5\oplus\bar{3}}} &= |0000\rangle+|0101\rangle+|1000\rangle+|1110\rangle\\
L_{0_{7\oplus\bar{1}}} &= |0000\rangle+|1011\rangle+|1101\rangle+|1110\rangle\\
L_{0_{3\oplus\bar{1}}0_{3\oplus\bar{1}}} &= |0000\rangle+|0111\rangle
\end{aligned}$$

*The complex parameters $a,b,c,d$ are the unique eigenvalues of $P$ (67) with non-negative real part, and the indices $L_{\alpha\beta\ldots}$ are representative for the Jordan block structure of $P$.*

*Proof:* If Theorem 13 is applied to a $4\times4$ $R$, it is easily checked that 12 different families arise where a family is defined as having Jordan and degenerated Jordan blocks of specific dimension. Note however that the orthogonal matrices obtained by application of the Theorem can have determinant equal to $-1$, while the SLOCC operations correspond to an orthogonal matrix with determinant $+1$; this is however not a problem as these operations correspond to SLOCC operations followed by a permutation of the qubits $(1\leftrightarrow2)$ or $(3\leftrightarrow4)$. One can proceed by checking that permutations of qubits $(2\leftrightarrow3)$ or $(1\leftrightarrow4)$ transform different families into each other. It is indeed true that $R=J_1(a)\oplus J_1(b)\oplus K_{3\oplus\bar{1}}$ transforms into $R'=J_2(a)\oplus J_2(b)$ if qubit 2 and 3 are permuted. This also happens in the case $J_1(a)\oplus K_{5\oplus\bar{1}}\ \rightarrow J_4(a)$. Moreover it can be shown that $J_1(a)\oplus K_{3\oplus\bar{3}}$ is equivalent to $J_1(a)\oplus J_3(0)$. Therefore only 9 essentially different normal forms are retained. □

A generic pure state of 4 qubits can always be transformed to the $G_{abcd}$ state. This state is peculiar in the sense that all local density operators, obtained by tracing out all parties but one, are proportional to the identity. As shown in section 3.1 treating the general case, this is the unique state (up to local unitary operations) with this property of all states connected by SLOCC operations. In the light of the results of Gisin [**98**] and of the results of Nielsen about majorization [**162, 164**], we claim that this is the state with maximal 4-partite entanglement on the complete orbit generated by SLOCC operations: the more entanglement, the more local entropy.

It is interesting to note that the 3-tangle (56) of the mixed states obtained by tracing out one party of this $G_{abcd}$ state is always equal to zero. Indeed, if the right-unitary matrix U

$$\begin{aligned} U &= \frac{1}{\sqrt{2(1+|\beta|^2)}} \begin{pmatrix} 1 & \beta & 1 & -\beta \\ \beta & 1 & -\beta & 1 \end{pmatrix} \\ \beta &= \sqrt{\frac{-q+\sqrt{q^2-r^2}}{r}} \\ q &= 2\,a^2 d^2 + 2\,b^2 c^2 - a^2 b^2 - a^2 c^2 - d^2 b^2 - d^2 c^2 \\ r &= (a^2-d^2)(b^2-c^2) \end{aligned}$$

is applied to the $8 \times 2$ matrix

$$\begin{pmatrix} a+d & 0 & 0 & a-d & 0 & b+c & b-c & 0 \\ 0 & b-c & b+c & 0 & a-d & 0 & 0 & a+d \end{pmatrix}^T$$

being the square root of the density operator obtained by tracing out the first qubit, 4 3-qubit W-states are obtained. If we define the mixed 3-tangle as the convex roof of the square root of the 3-tangle, this quantity is clearly equal to zero. Therefore the SLOCC operations maximizing the 4-partite entanglement result in a loss of all true 3-partite entanglement. This is reminiscent to the case of 3 qubits where the 2-qubit state obtained by tracing out one particle of a GHZ-state is separable.

Let us next discuss some specific examples. A completely separable state belongs to the family $L_{abc_2}$ with $a=b=c=0$. If only two qubits are entangled, an EPR state arises belonging to the family $L_{a_2b_2}$ with $a=b=0$. A state consisting of two EPR-pairs belongs to $G_{abcd}$ with $(a=1; b=c=d=0)$ or $a=b=c=d$, depending on the permutation. The class $L_{0_{3\oplus\bar{1}}0_{3\oplus\bar{1}}}$ consists of all 3-qubit GHZ states accompanied with a separable qubit, while the 3-qubit W-state belongs to the family $L_{a_2 0_{3\oplus\bar{1}}}$ with $a=0$.

The 4-qubit $|\Phi_4\rangle$-state exhibiting *persistent entanglement* [**45**] belongs to the generic family, while the 4-qubit W-state $(|0001\rangle+|0010\rangle+|0100\rangle+|1000\rangle)/2$ belongs to the family $L_{ab_3}$ with $a=b=0$. This W-state can be shown to have a mixed 3-tangle equal to zero, but has a concurrence of 1/2 when whatever two qubits are traced out. On the contrary the state $L_{O_{7\oplus\bar{1}}}$ has all concurrences equal to zero if two qubits are traced out. This state is completely symmetric in the permutation of the qubits 2,3 and 4. It has the property of having a mixed 3-tangle equal to 1/2 if particle 2,3 or 4 is traced out. This can be proven by considering the $8 \times 2$ "square root"

$$\frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}^T .$$

Some straightforward calculations show that the average square root of the 3-tangle of the vectors obtained by multiplying this matrix with whatever $2 \times n$ right-unitary matrix is equal to 1/2. Similar arguments show that only three-qubit W-type entanglement ($\tau = 0$) is retained if the first qubit is traced out.

The state $L_{0_{5\oplus\bar{3}}}$ is somehow a hybrid of both the 4-qubit W-state and $L_{O_{7\oplus\bar{1}}}$. Again a mixed 3-tangle of 1/2 is obtained if qubit 2,3 or 4 is traced out, a mixed 3-tangle equal to zero if qubit 1 is traced out, but now the mixed state obtained by tracing out qubit 1 and (3 or 4) has a concurrence equal to 1/2, while the other concurrences vanish.

Another interesting state belongs to the family $L_{a_4}$ with $a = 0$: $|\psi\rangle = (|0001\rangle + |0110\rangle + |1000\rangle)/\sqrt{3}$. Its mixed 3-tangle equals 2/3 in the case of tracing out qubit 1 or 4 and vanishes otherwise. Moreover the concurrence vanishes everywhere if 2 qubits are traced out except in the case of tracing out qubit 2 and 3, resulting in a concurrence of 2/3.

After this zoological survey, let us next move on to the topic of entanglement monotones. The complex eigenvalues of P (67), given by $\pm(a, b, c, d)$, are the only invariants under all determinant 1 SLOCC operations (note that an eigenvalue 0 is associated to the degenerated Jordan blocks). We have already proven that all real positive functions of the parameters of a pure state that are linearly homogeneous in $\rho$ and remain invariant under determinant 1 SLOCC operations, are entanglement monotones (in the case of mixed states they are defined by the convex roof formalism). Therefore all real positive homogeneous functions of $(a^2, b^2, c^2, d^2)$ are entanglement monotones, such as

$$M_\alpha(\psi) = |a^\alpha + b^\alpha + c^\alpha + d^\alpha|^{2/\alpha}.$$

Taking into account one degree of freedom due to the phase, this gives rise to a seven-parameter family of entanglement monotones. Note that the entanglement monotones can be described in terms of the original pure states by making use of completely antisymmetric tensors, as shown in equations (25,26). All these entanglement monotones are maximized by the operations making the density matrix locally stochastic (meaning that the identity is obtained when all qubits but one are traced out). The optimal single-copy distillation procedure for a generic pure state is therefore to implement the SLOCC operations bringing it into its normal form $G_{abcd}$. Note that all the other normal forms can only be brought into the local stochastic normal form (i.e. the $G_{abcd}$ class) by a filtering procedure whose probability of success tends to zero, as the $SL(2, \mathcal{C})$ matrices will tend to infinity.

Note that this study of SLOCC operations on states of four qubits is particularly interesting as the current experimental state of the art allows to entangle four photons [**168, 235, 143**] or ions [**182**]. Furthermore SLOCC operations can relatively easily be implemented on photons, and it is therefore of interest to implement the optimal SLOCC operations such as to yield a state with maximal 4-partite entanglement.

In summary, we have identified all different families of pure states of 4 qubits generated by SLOCC operations. Only one family is generic, and all states in it can be made locally stochastic by SLOCC operations. The same SLOCC operations represent the optimal single-copy distillation protocol. The eight other families correspond to states having some kind of degenerated 4-partite entanglement and are the 4-partite generalizations of the 3-partite W-state. In contrast to the three qubit case however, almost all normal forms still have some continuous parameters, implying that effectively an infinite number of SLOCC orbits exists.

## 3.4. Higher dimensional cases

We have seen that the complexity of characterizing SLOCC equivalence classes grows considerably with the dimensions of the subsystems. In general, there will be an infinite amount of SLOCC equivalence classes parameterized by some continuous parameters. Consider the case of a $n_1 \times \cdots \times n_p$ dimensional system. Then the generic number of parameters remaining in the SLOCC normal form (also taking into account the LU transformations and an additional multiplication with an arbitrary complex number) is given by

$$\max\left(0, 2(-\sum_i n_i^2 + \prod_i n_i + p - 1)\right). \tag{71}$$

In the case of four qubits for example, this number is indeed 6. Note however that this number becomes equal to zero for a $2 \times 2 \times 2 \times N$ dimensional system with $N \geq 7$: this means that, similarly to the results obtained in the case of $2 \times 2 \times N$ systems, the $2 \times 2 \times 2 \times N$ family will be presided by one and only one state given by

$$|0000\rangle + |0011\rangle + |0102\rangle + |0113\rangle + |1004\rangle + |1015\rangle + |1106\rangle + |1117\rangle, \tag{72}$$

and that the states lower in the hierarchy need to be parameterized by continuous parameters. A similar result holds in arbitrary dimensions: by making one subdimension arbitrary large (note that the dimension need not be larger than the product of the other ones), exactly one generic normal form is obtained, presiding all states that are lower in the hierarchy. Surprisingly, these *lower* states cannot be described by a finite number of equivalence classes.

We have already proven that the $2 \times 2 \times N$ system exhibits a finite number of SLOCC equivalence classes. More generally, one expects this to be true for all $2 \times N_1 \times N_2$ systems: the formula (71) is indeed zero for all possible values of $N_1$ and $N_2$. It is a nice but highly non-trivial open problem to determine these finite number of equivalence classes for states in a $2 \times N_1 \times N_2$ dimensional Hilbert space. Note that for all other dimensions, an infinite number of equivalence classes will exist.

As a last example, we treat pure states consisting of three qutrits, having a normal form with 4 real and continuous parameters plus 2 for the normalization and the phase. Of course the general Theorem 6 applies. We generated a lot of random $3\times3\times3$ states, applied the general procedure for bringing it into normal form by SLOCC operations, and finally applied the algorithm on the obtained state for bringing it into normal form under LU operations. A remarkable but unproven result arose, which we formulate as a conjecture:

**Conjecture 2.** *Given a generic complex* $3 \times 3 \times 3$ *tensor* $\psi$*, then there exist local matrices* $A, B, C \in SL(3, \mathcal{C})$ *such that the tensor* $\chi = A \otimes B \otimes C\psi$ *is of the form*

$$\chi_{ijk} = \alpha\delta_{ijk} + \beta\epsilon_{ijk} + \gamma|\epsilon_{ijk}| \tag{73}$$

*with* $\alpha, \beta, \gamma$ *complex numbers, with* $\delta_{ijk}$ *equal to* 1 *iff* $i = j = k$ *and zero elsewhere, and* $\epsilon_{ijk}$ *the completely antisymmetric tensor.*

If this conjecture is true, then a nice normal form for all generic $3 \times 3 \times 3$ states exists under SLOCC operations. Note however that we have observed that one and the same tensor $\psi$ can lead to different coefficients $\alpha, \beta, \gamma$: the decomposition is not unique and we observed that this is due to the non-uniqueness of the LU normal form (recall that we conjectured the uniqueness of the SLOCC normal form).

## 3.5. Conclusion

We have investigated pure and mixed multipartite entangled states in a new unified way by characterizing local LU and SLOCC equivalence classes. This has enabled us to identify different kinds of entanglement. A general formalism was developed to bring a pure or mixed multipartite state into a normal form by SLOCC transformations in a constructive way. We argued that this normal form is unique and corresponds to the maximally entangled state of all possible local orbits and is therefore the generalization of the singlet state to higher dimensions and multiple parties. The introduced formalism lead to a natural way of defining entanglement monotones, for which the ubiquitous entanglement measures concurrence and 3-tangle are special cases. Moreover, we proved that the filtering operations maximizing these entanglement monotones were exactly the operations bringing a state into normal form. This could be of great practical value in realistic entanglement distillation protocols. Next we made a complete analysis of all possible pure states in a $2 \times 2 \times N$ dimensional Hilbert space. This led to the introduction of 9 different states to which every possible $2\times2\times N$ state can be transformed by SLOCC operations. These states exhibit some kind of hierarchy, and they can all be made from one maximally entangled state (the $GHZ$- and $W$-states are lower in the hierarchy). Finally, we gave a complete classification of all $2\times2\times2\times2$ states. We identified unique normal forms for all $2 \times 2 \times 2 \times 2$ states under SLOCC operations, and came to the conclusion that there exist 9 families (although only 1 generic family) of normal forms each parameterized by at most 8 real continuous parameters. Much of the properties of these states still await to be discovered.

Meanwhile, some highly non-trivial results in linear and multilinear algebra were derived. We have shown how to generalize the singular value decomposition to tensors in a constructive way, and argued why it failed to be unique. The central result however is the fact that a unique normal form of a generic tensor can be obtained if we enlarge the class of possible operations from $SU(N)$ to $SL(N,\mathcal{C})$. Of central importance was the fact that a solution was found to the variational characterization of the minimal possible trace of a tensor under the action of all $SL(N,\mathcal{C})$-operations. Along similar lines, we were able to give a complete characterization of all $2 \times 2 \times N$ and $2 \times 2 \times 2 \times 2$-tensors under the action $SU(N)$ and of $SL(N,\mathcal{C})$ operations. Moreover, the analogue of the singular value decomposition for complex orthogonal equivalence has been derived: unlike the SVD, diagonalization is not always possible, and we identified all possible normal forms.

CHAPTER 4

# Entanglement of mixed States of two Qubits

Mixed quantum states arise in nature due to the (mostly unwanted) coupling of quantum systems with the environment. This implies that quantum systems become entangled with the environment, and therefore that imperfectly isolated quantum systems are described by mixed states. In this chapter we will study the structure of mixed density operators and try to get insight into the way entanglement manifests itself. We will almost exclusively concentrate on the mixed states of two qubits, the simplest of all entangled quantum systems[1].

In a first section, we investigate how the quantum steering theorem can be generalized to mixed states. Next we move on to describe normal forms for mixed states of two qubits, and introduce the Lorentz Singular Value Decomposition (LSVD). This normal form separates in some sense the classical correlations from the quantum correlations, and therefore gives a lot of insight into the entanglement characteristics of mixed states of two qubits. We then apply this formalism to get an interesting geometrical picture of density operators motivated by the quantum steering theorem, and indicate how to characterize LOCC operations. Next we discuss some successful entanglement measures, specifically: entanglement of formation, negativity, relative entropy of entanglement, fidelity and Bell-CHSH-violation. In the case of entanglement of formation, we present a new approach, and make the connection with the results of the previous chapter. Then we move on to compare all those popular entanglement measures, which gives insight into the different ways entanglement can manifest itself. Finally we prove that all the discussed entanglement measures are maximized by exactly the same filtering operations bringing the state into the normal form of the chapter 3.

In section 4.5 we show how to do optimal teleportation with mixed states of two qubits: it turns out that some pre-processing can typically enhance the fidelity

[1]Note that this study is also relevant for higher dimensional systems as these always have $2 \times 2$ dimensional subspaces.

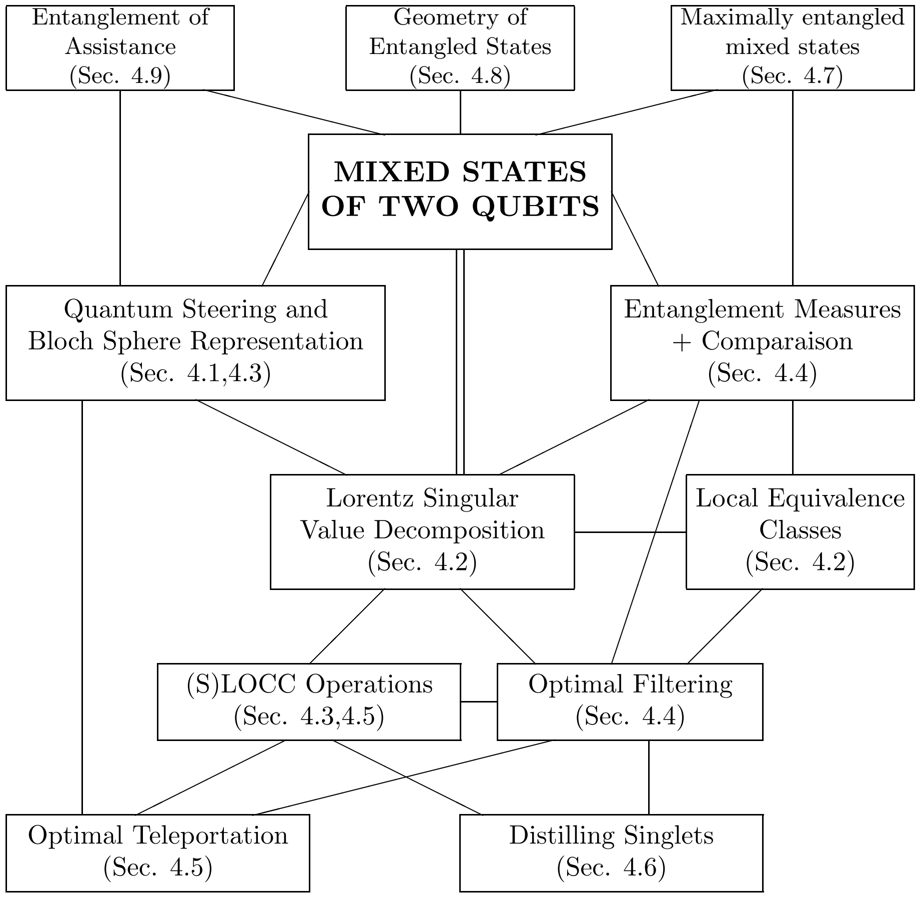



**Figure 1.** Structure of Chapter 4

of teleportation. The proof of the optimality is very interesting in its own right: although an optimization over all (unphysical) PPT-protocols was done, it turned out that the optimal solution was physically implementable. Section 4.6 discusses the important issue of entanglement distillation: as the basic resource for quantum communication is the ebit (i.e. a maximally entangled state of two qubits), it is natural to find optimal local protocols to transform entangled mixed states into ebits. We will present the best available protocol for distilling bipartite entanglement. Section 4.7 introduces an interesting class of mixed states: the maximally entangled mixed states of two qubits. These are the states whose entanglement cannot be increased by any global unitary operation, and turn out to have a whole lot of interesting properties, such as being the states with maximal entanglement for given entropy. Section 4.8 discusses entangled states in the light of the Hilbert-Schmidt metric, yielding a geometrical picture of the convex set of separable states, and the final section presents some novel results in the context of entanglement of assistance.

As this is quite a long chapter, the main interrelations are depicted in figure 1.

## 4.1. Quantum steering with mixed states

The quantum steering Theorem 1 turned out to be of central importance in the study of pure bipartite quantum systems. It is therefore natural to investigate the same problem in the context of mixed states. In the case of pure states of two qubits, the local density operators can be represented in the Bloch sphere. The quantum steering Theorem says that a local POVM measurement can create an ensemble of local density operators $\{p_\alpha, \rho_\alpha\}$ at Bob's side iff the convex sum of the elements of the ensemble is equal to the original local density operator: every decomposition inside the Bloch sphere and with fixed ensemble average can be realized.

In the case of mixed $n \times n$ states (i.e. local dimensions of the Hilbert spaces are $n$), we expect a similar picture, but with one major difference: there will be a restriction on the states $\{\rho_\alpha\}$ allowed in the convex decomposition. Let us first of all see what happens when Alice performs a POVM measurement $\{E_\alpha = A_\alpha^\dagger A_\alpha\}$ on a state $\rho$. Recall that we can always parameterize a state by operators $X_i$:

$$\rho = \sum_i (I \otimes X_i)|I\rangle\langle I|(I \otimes X_i)^\dagger = \sum_i (X_i^T \otimes I)|I\rangle\langle I|(X_i^* \otimes I) \qquad (74)$$

with $|I\rangle = \sum_i |ii\rangle$. A POVM measurement of Alice $\{E_\alpha\}$ results in a convex decomposition of the state at Bob's side given by[2]

$$\sum_\alpha \mathrm{Tr}_A\left(\rho(E_\alpha \otimes I)\right) = \sum_\alpha \left(\sum_i X_i E_\alpha^T X_i^\dagger\right). \qquad (75)$$

The question of which local density operators can be obtained at Bob's side is equivalent to the following: what is the convex hull of all states of the form

$$\frac{\sum_i X_i E_\alpha^T X_i^\dagger}{\mathrm{Tr}(\sum_i X_i E_\alpha^T X_i^\dagger)} \qquad (76)$$

with $0 \leq E_\alpha \leq I$? More generally, we would like to formulate answers to the following questions:

- Given a decomposition $\{p_\alpha, \rho_\alpha^B\}$ of $\mathrm{Tr}_A(\rho_{AB})$, does there exist a simple procedure for determining if it can be generated by Alice?

[2] Note that this looks very much like the description of the action of a quantum channel acting on states $E_\alpha^T$. As will be shown later, there is indeed a duality between maps and entangled states.

- Given a particular state $\rho_\alpha^B$, what is the largest probability that Alice can generate it?
- Can we easily characterize the set of all possible convex decompositions which Alice can generate?

The answer to these questions will turn out to be remarkably simple if the "Bloch sphere parameterization"of the density operator is used, defined as:

$$\begin{aligned} \rho_{AB} &= \frac{1}{n^2}\sum_{i,j=0}^{n^2-1} R_{ij}(\sigma_j \otimes \sigma_i) \\ R_{ij} &= \mathrm{Tr}\left(\rho_{AB}(\sigma_j \otimes \sigma_i)\right) \end{aligned}$$

Here $\{\sigma_i\}$ denote a complete orthogonal basis of the $n \times n$ Hermitian matrices and $\sigma_0$ is the identity . Note that the $n^2 \times n^2$ matrix $R_{ij}$ is real and that there is a one to one correspondence between $R$ and $\rho$. In an analogous way we define

$$\begin{aligned} E_\alpha &= \frac{1}{n}\sum_{j=0}^{n^2-1} x_j^\alpha \sigma_j; \qquad & x_j^\alpha = \mathrm{Tr}(E_\alpha \sigma_j) \\ \rho_B^\alpha &= \frac{1}{n}\sum_{i=0}^{n^2-1} y_\alpha \sigma_i; \qquad & y_i^\alpha = \mathrm{Tr}(\rho_A^\alpha \sigma_i) \end{aligned}$$

It is easy to check that the convex decomposition(75) becomes:

$$\sum_\alpha y^\alpha \tag{77}$$

with the components of the vector $y^\alpha$ defined as

$$y_i^\alpha = \sum_j R_{ij} x_j^\alpha .$$

In this new parameterization the action of the measurement by Alice is very appealing: it corresponds to right multiplication of the matrix $R$ with the vector $x_\alpha$ corresponding to the POVM element.

The answer to the above questions is now straightforward if $R$ is a full rank matrix. Given an ensemble $\{p_\alpha, \rho_B^\alpha\}$ or equivalently $\{p_\alpha, y^\alpha\}$, let us define $\{x^\alpha = p_\alpha R^{-1} y^\alpha\}$ and the corresponding $E_\alpha$. Then the ensemble can be realized iff for every element $E_\alpha$ it holds that $0 \leq E_\alpha \leq I$ and $\sum_\alpha E_\alpha = I$. Similarly, the maximal probability by which one can steer Bob's system into a particular state $y$ is given by the following: it is the maximal probability $p$ such that the $E_\alpha$ corresponding to $pR^{-1}y$ fulfills the constraint $0 \leq E_\alpha \leq I$ (note that $p$ will be equal to zero if $y$ is not feasible). It the state $\rho_{AB}$ is pure, then it can easily be verified that the maximal value of $p$ is given by

$$\frac{1}{\lambda_{\max}(\rho_B^{-1}\sigma_B)} \tag{78}$$

where $\rho_B$ is the original local density operator, $\sigma_B$ is the local state we are trying to generate, and $\lambda_{\max}$ is the maximal eigenvalue.

We still have to consider the case where $R$ is not full rank. In that case, it is easy to see that all $y^\alpha$ have to belong to the column space of $R$. Now many different $x^\alpha$ will correspond to the same $y^\alpha = Rx^\alpha$, but in this specific case a simple semidefinite program will answer the questions about feasibility of a certain decomposition and about maximizing the probability of a certain state in the convex decomposition. Indeed, the constraints upon the matrices $\{E_\alpha\}$ are convex linear matrix inequalities, and the cost function linear.

We have therefore proven:

**Theorem 15.** *Consider a mixed $n \times n$ state parameterized by $R$. Then the ensemble at Bob's side $\{p_\alpha, \rho_\alpha\} := \{p_\alpha, y_\alpha\}$ can be realized by a POVM measurement of Alice $\{E_\alpha\} := \{x^\alpha\}$ iff the following set of equations is feasible:*

$$\forall\alpha : p_\alpha y^\alpha = Rx^\alpha \qquad 0 \leq E_\alpha \leq I \qquad \sum_\alpha E_\alpha = I \tag{79}$$

*This problem is a semidefinite program in the variables $\{E_\alpha\}$. However, it reduces to a problem that involves only testing whether matrices are positive definite in the generic case where $R$ is full rank.*

Note that this Theorem combined with the quantum steering theorem for pure states 1 implies that a matrix $R$ corresponding to a pure entangled state with maximal Schmidt rank (i.e. its local density operator is full rank), is full rank. This is the only way one can ensure that all possible local density operators can be created by appropriate measurements of Alice. Using the Lorentz singular value decomposition which will be introduced later, it is moreover easy to show that every $R$ corresponding to an entangled mixed state of two qubits is full rank. It is also easy to show that a separable state with the property that it is a Kronecker product of two local ones has a corresponding $R$ of rank 1.

Another observation reveals an interesting fact: it is possible to specify a state represented by a $n^2 \times n^2$ matrix $R$ completely if we know the action of $n^2$ linearly independent POVM elements $x^\alpha$; in other words, $n^2$ pairs $\{x^\alpha, y^\alpha\}$ form a complete parameterization of a state $R$. Thus a state is completely specified if one knows the "image" $y$ at Bob's side under the action of a measurement $x$ at Alice's side.

In the case of mixed states of two qubits, many more results can be obtained, but we will postpone that discussion until the Lorentz singular value decomposition has been introduced.

## 4.2. Equivalence classes under local operations

In analogy with the study of multipartite pure state entanglement, the natural starting point for a study of mixed states of two qubits is to look for equivalence classes under local unitary and under SLOCC operations. The main tool will again consist of exploiting some accidents in Lie group theory: $SU(2) \simeq SO(3)$, $SL(2,\mathcal{C}) \simeq SO(3,1)$ and $SL(2,\mathcal{C}) \otimes SL(2,\mathcal{C}) \simeq SO(4,\mathcal{C})$. It will be very useful to work in the R-picture:

$$\begin{aligned} \rho &= \frac{1}{4}\sum_{ij=0}^{3} R_{ij}\sigma_i \otimes \sigma_j && (80) \\ R_{ij} &= \mathrm{Tr}(\rho\sigma_i \otimes \sigma_j) && (81) \end{aligned}$$

Here $\sigma_0 = I$ and the $\sigma_i$ are the Pauli matrices. This R-picture is very appealing because local operations by Alice and Bob correspond to left respectively right multiplication of $R$ with appropriate matrices: operations of the form $(A \otimes B)\rho(A \otimes B)^\dagger$ in the $\rho$-picture correspond to operations of the form $L_A R L_B^T$ in the $R$-picture.

### 4.2.1. LU equivalence classes

It is easy to check that the local unitary operations

$$\rho' = (U_1 \otimes U_2)\rho(U_1 \otimes U_2)^\dagger \tag{82}$$

correspond to left and right multiplication of $R$ with orthogonal matrices:

$$R' = \begin{pmatrix} 1 & 0 \\ 0 & O_1 \end{pmatrix} R \begin{pmatrix} 1 & 0 \\ 0 & O_2^T \end{pmatrix}. \tag{83}$$

Here $O_1$ and $O_2$ are real $3\times 3$ orthogonal matrices belonging to $SO(3)$. Indeed, the $4 \times 4$ orthogonal matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & O_\alpha \end{pmatrix} \tag{84}$$

has coefficients $\mathrm{Tr}(U_\alpha\sigma_i U_\alpha^\dagger\sigma_j)$.

It is now straightforward to bring $R$ into normal form: calculate the singular value decomposition of the lower diagonal $3 \times 3$ block $R_{1:3,1:3} = O_1\Sigma O_2^T$ (note that we label the elements of $R$ from 0 to 3), divide $O_1$ and $O_2$ by their respective determinant to make sure that they have determinant $+1$, and calculate $R^N$ as

$$R^N = \begin{pmatrix} 1 & 0 \\ 0 & O_1^T \end{pmatrix} R \begin{pmatrix} 1 & 0 \\ 0 & O_2 \end{pmatrix}. \tag{85}$$

The lower $3 \times 3$ block of $R^N$ is diagonal, and the final normal form is obtained by imposing that the elements $R_{11}^N$ and $R_{22}^N$ are negative. We therefore obtain a unique normal form of the kind

$$R^N = \begin{pmatrix} 1 & x_1 & x_2 & x_3 \\ y_1 & -\sigma_1 & 0 & 0 \\ y_2 & 0 & -\sigma_2 & 0 \\ y_3 & 0 & 0 & -\sigma_3 \end{pmatrix} \tag{86}$$

with $\{\sigma_i\}$ the singular values in $\Sigma$, $\sigma_1, \sigma_2 \geq 0$ and the sign $\sigma_3$ equal to the sign of the determinant of $R_{1:3,1:3}$. Note that if a singular value in $\Sigma$ had multiplicity larger than 1, an additional zero can be made in the first row of $R^N$.

A very useful quantity in quantum information theory is the fidelity of a state. The fidelity is defined as the maximal overlap of a state with all maximally entangled states:

$$F(\rho) = \max_{|\psi\rangle = \mathrm{ME}} \langle\psi|\rho|\psi\rangle. \tag{87}$$

It measures how close the state $\rho$ is to a maximally entangled state, and plays a crucial role in distillation protocols. This fidelity can explicitly be calculated for a state in normal form (86):

**Theorem 16.** *The fidelity of a state $R$ in normal form (86) is given by* $(1 + \sigma_1 + \sigma_2 + \sigma_3)/4$.

*Proof:* Given two states $\rho_1$ and $\rho_2$ with corresponding $R_1$ and $R_2$, then it can easily be checked that $\mathrm{Tr}(\rho_1\rho_2) = \mathrm{Tr}(R_1 R_2^T)/4$. All maximally entangled states have an associated $R$-picture of the form:

$$\begin{pmatrix} 1 & 0 \\ 0 & -O_\psi \end{pmatrix} \tag{88}$$

with $O_\psi \in SO(3)$. The optimization involved in calculating the fidelity is therefore equivalent to:

$$F(R) = \max_{O_\psi \in SO(3)} \frac{1}{4}\mathrm{Tr}\left(R^T \begin{pmatrix} 1 & 0 \\ 0 & -O_\psi \end{pmatrix}\right). \tag{89}$$

All elements in the group $SO(3)$ can be written as $\exp(G_\psi)$ with $G_\psi$ antisymmetric. The derivatives of something of the form $\mathrm{Tr}(O_\psi Q)$ over $O_\psi$ are zero iff $O_\psi Q$ is symmetric. If $R$ is in normal form and the diagonal part is called $Q$, then $O_\psi Q$ can only be symmetric if $O_\psi$ is diagonal (this follows for example from the uniqueness of the singular value decomposition). We have therefore obtained the result that $O_\psi$ has to be chosen diagonal, and the Theorem follows easily. □

### 4.2.2. SLOCC equivalence classes: the Lorentz Singular Value Decomposition

We will again consider the R-picture of states of two qubits. Let us begin by formulating the following useful lemma:

**Lemma 3.** *The* 4x4 *matrix $R$ with elements $R_{ij} = \mathrm{Tr}(\rho(\sigma_i \otimes \sigma_j))$ transforms, up to normalization, under SLOCC operations $(A \otimes B)\rho(A \otimes B)^\dagger$ as*

$$R' = L_A R L_B^T \tag{90}$$

*where $L_A$ and $L_B$ are proper orthochronous Lorentz transformations*[3] *given by*

$$L_A = T(A \otimes A^*)T^\dagger/|\det(A)| \tag{91}$$

$$L_B = T(B \otimes B^*)T^\dagger/|\det(B)| \tag{92}$$

$$T = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & i & -i & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix}. \tag{93}$$

*Proof:* This Theorem can be proven by introducing the matrix $\tilde{\rho}_{kl,k'l'} = \rho_{kk',ll'}$ and noting that $R = 4T\tilde{\rho}T^T$. It is easy to check that under SLOCC operations $\tilde{\rho}$ transforms as $\tilde{\rho}' = (A \otimes A^*)\tilde{\rho}(B \otimes B^*)^T$. Therefore $R$ transforms as $R' = L_A R L_B^T |\det(A)||\det(B)|$ with $L_A = T(A \otimes A^*)T^\dagger/|\det(A)|$, $L_B = T(B \otimes B^*)T^\dagger/|\det(B)|$. Using the identities $A\sigma_y A^T = \det(A)\sigma_y$ and $T^\dagger M T^* = -\sigma_y \otimes \sigma_y$ with $M$ the matrix associated with the Lorentz metric

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \tag{94}$$

it is easily checked that $L_A M L_A^T = M = L_B M L_B^T$. Furthermore the determinant of $L_A$ and of $L_B$ is equal to $+1$, and the $(0,0)$ element of $L$ is positive, which completes the proof. □

As the complex 2×2 matrices with determinant one indeed form the spinor representation of the Lorentz group, there is a 1 to 2 correspondence between each $L_A$ and $A/\sqrt{\det(A)}$. It is interesting to note that when both $A$ and $B$ are unitary, the Theorem reduces to the LU-case treated in the previous section; $SO(3)$ is indeed a subgroup of the Lorentz group.

As an example, let us consider a pure state. Every pure state can be written as $|\psi\rangle = A \otimes I|S\rangle$ with $|S\rangle$ the singlet state $|S\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$. The $R$-picture of the singlet state is exactly given by $M$ (equation (94)), and therefore a pure state of two qubits is proportional to a proper orthochronous Lorentz transformation times $M$ (it is easy to check that the proportionality factor is the concurrence encountered in section 3.1.3), and therefore itself an improper Lorentz transformation. Note however that a problem is encountered when the state $|\psi\rangle$ is separable; then $A$ is not full rank, and the corresponding

[3] A proper Lorentz transformation is one for which the determinant is $+1$, and an orthochronous Lorentz transformation one for which the $(0,0)$ element is positive.

Lorentz transformation would tend to infinity. We will indeed see that the non-compactness of the Lorentz group is responsible for different types of normal form of states of two qubits.

For later reference, let us recall the $R$-picture of the four Bell states:

$$\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

It is clear that these four matrices form a complete basis for all diagonal matrices. This proves that a diagonal matrix in the $R$-picture corresponds to a mixture of Bell states and therefore to a Bell diagonal state. It is easy to parameterize all possible Bell diagonal states in the $R$-picture: one transforms the state back to the $\rho$-picture and verifies whether this state is positive (semi)definite. Given a diagonal $R$ of the form

$$\Sigma = \begin{pmatrix} s_0 & 0 & 0 & 0 \\ 0 & -s_1 & 0 & 0 \\ 0 & 0 & -s_2 & 0 \\ 0 & 0 & 0 & -s_3 \end{pmatrix}$$

The eigenvalues of the corresponding Bell diagonal state are given by

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} s_0 \\ s_1 \\ s_2 \\ s_3 \end{pmatrix}. \tag{95}$$

This immediately leads to the following necessary and sufficient conditions for a diagonal $R$ to correspond to a positive (semi)definite Bell diagonal state:

$$s_0 - |\tau_1| - |\tau_2| + \tau_3 \geq 0 \tag{96}$$

where $\tau_1, \tau_2, \tau_3$ is the unique permuted version of $s_1, s_2, s_3$ such that $|\tau_1| \geq |\tau_2| \geq |\tau_3|$.

A natural question is now to find a decomposition of $R$ as $R = L_1 \Sigma L_2^T$ with $\Sigma$ diagonal and $L_1, L_2$ proper orthochronous Lorentz transformations. This would be the analogue of a singular value decomposition but now in the Lorentz instead of the Euclidean metric[4].

[4]Note that this decomposition is different from the hyperbolic SVD as defined by Bojanczyk et al. [**40**], where they considered a normal form by multiplying left with a Lorentz matrix and right with an orthogonal one. During the writing of this thesis, H. Woerdeman also mentioned a similar theorem as the one stated here in the context of transformation of Stokes parameters as encountered in the study of the polarization of light [**205, 177**]; the

**Theorem 17.** *The* $4 \times 4$ *real matrix* $R$ *with elements* $R_{ij} = \mathrm{Tr}\,(\rho\sigma_i \otimes \sigma_j)$ *can be decomposed as*

$$R = L_1 \Sigma L_2^T$$

*with* $L_1, L_2$ *finite proper orthochronous Lorentz transformations, and* $\Sigma$ *either of unique real diagonal form with Lorentz singular values* $\{s_i\}$

$$\Sigma = \begin{pmatrix} s_0 & 0 & 0 & 0 \\ 0 & -s_1 & 0 & 0 \\ 0 & 0 & -s_2 & 0 \\ 0 & 0 & 0 & -s_3 \end{pmatrix}$$

*with* $s_0 \geq s_1 \geq s_2 \geq |s_3|$ *and* $s_3$ *positive or negative, or of one of the following four degenerate normal forms*

$$\begin{pmatrix} a & . & . & a/2 \\ . & b & . & . \\ . & . & b & . \\ a/2 & . & . & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \qquad (97)$$

*with unique coefficients* $a, b$ *with* $0 \leq b \leq a/2$.

*Proof:* The original proof that we published in [**212**] was very technical and is given in the appendix, but since then we have been able to produce a much more elegant proof that we will give here. The proof is similar to the one where we derived the $2 \times 2 \times 2 \times 2$ normal form. We will make use of the accident $SL(2,\mathcal{C}) \otimes SL(2,\mathcal{C}) \simeq SO(4,\mathcal{C})$ encountered in Lemma 2. Let us define $\tilde{\rho} = T\rho T^\dagger$ with $T$ defined in equation (34). Then the SLOCC operations correspond to operations of the form

$$\tilde{\rho}' = O\tilde{\rho}O^\dagger \qquad (98)$$

with $O$ complex orthogonal. But as $\tilde{\rho}$ is positive semidefinite and Hermitian, we can use corollary 2.19 in Hong [**115**], which states that there exists a complex orthogonal matrix $O$ bringing $\tilde{\rho}$ into one of three possible normal forms:

$$\begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{pmatrix} \quad \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & 1 & -i & 0 \\ 0 & i & 1 & 0 \\ 0 & 0 & 0 & \lambda_2 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 & -i \\ 0 & 1 & -i & 0 \\ 0 & i & 1 & 0 \\ i & 0 & 0 & 1 \end{pmatrix} \qquad (99)$$

Moreover the first normal form is obtained iff rank($\tilde{\rho}$)=rank($\tilde{\rho}\tilde{\rho}^T$), the second iff rank($\tilde{\rho}$)>rank($\tilde{\rho}\tilde{\rho}^T$)> 0, and the third iff rank($\tilde{\rho}\tilde{\rho}^T$)= 0. However, we still

---

present theorem could yield a very useful parameterization of all so-called Mueller matrices, but we did not check the exact relation yet. After completion of this work, an extensive literature study also revealed a related paper by Y. Bolshakov and B. Reichstein [**41**], although we did not see how to translate their unfathomable results to the current setting of proper orthochronous Lorentz transformations. On the other hand, it seems more likely that the results of Horn and Merino [**118**] on contragredient equivalence relations could yield a different derivation of the current theorem.

have to assure that the complex orthogonal matrices have determinant $+1$, and so eventually $\pm 1$ signs have to be added at the appropriate places. Note however that if the orthogonal matrix has determinant $-1$, this corresponds to a permutation of the qubits, and we can just keep on working with the given normal forms and eventually permute the qubits at the end. A straightforward conversion of the given normal forms to the $R$-picture then yields the given normal forms of the theorem; more specifically, the diagonal one of (99) corresponds to the diagonal generic case of the theorem, the second one in (99) corresponds to the first case in (97) except when $\lambda_1 = \lambda_2 = 0$ when it corresponds to the last one in (97), and the last one in (99) corresponds either to the second or the third one in (97) depending on the determinant of $O$. □

In essence, the Lorentz singular value decomposition separated the local degrees of freedom (i.e. the Lorentz transformations) from the global ones (encoded into the normal forms). The normal form fully encodes the information whether a state is separable or entangled: as determinant 1 SLOCC operations are probabilistically invertible, they cannot change the property whether a state is entangled or separable. Note also that the diagonal normal form is exactly of the kind discussed in the last chapter: all local density operators become proportional to the identity. We will therefore be able to repeat the arguments about entanglement monotones and optimal filtering operations, which is very convenient.

The normal forms presented in the Theorem can be computed numerically by calculating the Jordan canonical decomposition of $C = MRMR^T$ and of $C' = MR^TMR$. It is easy indeed to show that for example in the case of diagonalizable $R$ the eigenvectors of $C$ form a Lorentz matrix, and $|s_i| = \sqrt{\lambda_i(C)}$. Note that we always order the diagonal elements such that $s_0 \geq s_1 \geq s_2 \geq |s_3|$. Note also that the Lorentz transformations in the Lorentz singular value decomposition are unique up to signs (and up to local orthogonal transformations iff there are Lorentz singular values with multiplicity larger than 1). Of course the diagonal normal form can also be calculated using the numerical algorithms developed in the previous chapter.

Let us now analyze more closely the different kinds of normal forms. The diagonalizable case is generic, and a diagonal $R$ corresponds to a Bell-diagonal state. The existence of the non-diagonal normal forms is a consequence of the fact that the Lorentz group is not compact: these non-diagonal normal forms can only be brought into diagonal form by infinite Lorentz transformations of the form A Lorentz transformation tending to infinity is of the form

$$L = \lim_{t\to\infty} \begin{pmatrix} 1 & 0 \\ 0 & O_1 \end{pmatrix} \begin{pmatrix} \sqrt{1+t^2} & 0 & 0 & t \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ t & 0 & 0 & \sqrt{1+t^2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & O_2 \end{pmatrix}. \tag{100}$$

This indeed allows to bring $R$ asymptotically into diagonal form with diagonal elements given by $[a/2, -b, -b, -a/2]$ or by $[0, 0, 0, 0]$. So even in those cases the Lorentz singular values are well defined and given by:

$$[s_0, s_1, s_2, s_3] = [a/2, a/2, b, b]$$

for the first case and by

$$[s_0, s_1, s_2, s_3] = [0, 0, 0, 0]$$

in all the other ones.

The four distinct non-diagonal normal forms correspond to the following states:

- first case: these are rank 3 states (rank 2 iff $(b = a/2)$) with the strange property that their entanglement cannot be increased by any global unitary operation (see section about maximally entangled mixed states).
- second and third case: $\rho$ is separable and a tensor product of the projector diag$[1; 0]$ and the identity.
- fourth case: $\rho$ is the separable pure state diag$[1; 0; 0; 0]$.

These states with a non-diagonal normal form can only be made Bell diagonal by Lorentz transformations tending to infinity. Note that only the case where the parameters $a, b$ are still present should be considered, as the other ones correspond to states that are tensor products. Applying diagonalizing Lorentz transformations of the kind (100), the magnitude of the off-diagonal elements becomes of the order $1/t^2$. Therefore a state is obtained that is infinitesimally close to a Bell diagonal state of rank 2 (if $b \neq a/2$; note that the original state is rank 3) or infinitesimally close to a pure Bell state (if $b = a/2$; note that the original state is rank 2). This last observation is very surprising: given one copy of a state of rank 2 with non-diagonal normal form and $a/2 = b$, there exist SLOCC operations that bring this state infinitesimally close to a maximally entangled state, irrespective of the values of $a$. This is clearly the only state with this property and will therefore be called a quasi-distillable state (see also Horodecki [**124**]). Note however that the SLOCC operation corresponding to this Lorentz transformation has singular values $\lim_{t\to\infty}[t, 1/t]$. This element has to be implemented as part of a POVM, and should therefore be divided by $t$ such as to yield an element $A \leq I$. This however implies that the probability of actually achieving the filtering operation scales as $1/t^2$: the probability of distilling a perfect EPR-state out of a quasi-distillable one decreases as $1 - F$ with $F$ the fidelity of the "distilled"state. In later sections we will see that these quasi-distillable states also have many other strange properties. It turns out that two copies of them can for example be distilled to an EPR-state with a finite probability.

Let us now return to the mathematics of the Lorentz singular value decomposition. The success of the ordinary singular value decomposition is to a large extent the consequence of the nice variational properties of the singular values: the sum of the $n$ largest singular values is equal to the maximal inner product of the matrix with whatever n orthonormal vectors. Interestingly, a similar property holds for the Lorentz singular values:

**Theorem 18.** *The Lorentz singular values $s_0 \geq s_1 \geq s_2 \geq |s_3|$ of a state $R$ are variationally defined as:*

$$s_0 = \min_{L_1,L_2} \mathrm{Tr}\left(L_1 R L_2^T \begin{pmatrix} 1 & . & . & . \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \end{pmatrix}\right)$$

$$s_0 - s_1 = \min_{L_1,L_2} \mathrm{Tr}\left(L_1 R L_2^T \begin{pmatrix} 1 & . & . & . \\ . & 1 & . & . \\ . & . & . & . \\ . & . & . & . \end{pmatrix}\right)$$

$$s_0 - s_1 - s_2 = \min_{L_1,L_2} \mathrm{Tr}\left(L_1 R L_2^T \begin{pmatrix} 1 & . & . & . \\ . & 1 & . & . \\ . & . & 1 & . \\ . & . & . & . \end{pmatrix}\right)$$

$$s_0 - s_1 - s_2 - s_3 = \min_{L_1,L_2} \mathrm{Tr}\left(L_1 R L_2^T\right)$$

*where $L_1, L_2$ are proper orthochronous Lorentz transformations.*

*Proof:* We will give a proof of the fourth identity and the other proofs follow in a completely analogous way. An arbitrary Lorentz transformation can be written as

$$L = \begin{pmatrix} 1 & . \\ . & V \end{pmatrix} \begin{pmatrix} \cosh(\alpha) & \sinh(\alpha) & . & . \\ \sinh(\alpha) & \cosh(\alpha) & . & . \\ . & . & 1 & . \\ . & . & . & 1 \end{pmatrix} \begin{pmatrix} 1 & . \\ . & W \end{pmatrix},$$

where $V$ and $W$ are orthogonal 3x3 matrices with determinant 1. There is no restriction in letting $R$ be in normal diagonal form, and therefore we have to find the minimum of

$$\mathrm{Tr}\left(\begin{pmatrix} \cosh(\alpha) & \sinh(\alpha) & . & . \\ \sinh(\alpha) & \cosh(\alpha) & . & . \\ . & . & 1 & . \\ . & . & . & 1 \end{pmatrix} \begin{pmatrix} 1 & . \\ . & W \end{pmatrix} \Sigma \begin{pmatrix} 1 & . \\ . & V \end{pmatrix}\right)$$

over all $V, W, \alpha$. Using the variational properties of the ordinary singular value decomposition and the fact that the Lorentz singular values are ordered, it is immediately clear that an optimal solution will consist in choosing $W = I_3 = V$ and $\alpha = 0$ as $\cosh(\alpha) > \sinh(\alpha)$ and $s_0 \geq s_1$. This ends the proof. $\square$

This property fully justifies the name Lorentz singular value decomposition. Note that the 4 Lorentz singular values are the only SLOCC invariants, just as the ordinary singular values are the only invariants under unitary operations.

In analogy with the entanglement monotones defined on pure states as the SLOCC invariants(see section (3.1.3)), this immediately raises the question: do these Lorentz singular values give rise to entanglement monotones for mixed states? With the help of the previous theorem, the answer is readily obtained. Let us define the quantities $M_1(\rho) = \max(0, -(s_0 - s_1 - s_2)/2)$ and $M_2(\rho) = \max(0, -(s_0 - s_1 - s_2 - s_3)/2)$.

**Theorem 19.** *$M_1(\rho) = \max(0, -(s_0 - s_1 - s_2)/2)$ and $M_2(\rho) = \max(0, -(s_0 - s_1 - s_2 - s_3)/2)$ are entanglement monotones.*

*Proof:* A quantity $M(\rho)$ is an entanglement monotone iff its expected value does not increase under the action of every possible local operation. Due to the variational characterization of the quantities $(s_0 - s_1 - s_2)$ and $(s_0 - s_1 - s_2 + s_3)$, it is immediately clear that both $M_1$ and $M_2$ are decreasing under the action of mixing. But now the proof of Theorem 7 can be repeated word for word, ending the proof. □

Note that these entanglement monotones were directly defined upon a mixed state without invoking the convex roof formalism: they can be calculated analytically. $M_2(\rho)$ will turn out to be the celebrated concurrence as introduced by Wootters [**111, 244**].

The existence of entanglement monotones is very interesting as it gives necessary conditions for one state to be convertible into another one by LOCC operations with probability 1, and upper bounds on the probability of conversion of one state to another one. Moreover, the general formalism developed in section 3.1 about optimal filtering can be applied to determine the SLOCC operations that will maximize the introduced entanglement monotones: these are the SLOCC operations making the local density operators proportional to the identity, which is precisely the action of bringing the state into Bell-diagonal normal form. A similar result was obtained by Kent, Linden, Massar and Popescu [**135, 148, 136**], where they showed that single-particle distillation of full rank mixed states is not possible and that the local filtering operations maximizing the entanglement of formation are given by the ones that make a state Bell-diagonal.

We conclude this section with the following remark: the Lorentz singular value decomposition induces a continuous parameterization of the manifold of density operators with constant values of the entanglement monotones $M_1$ and $M_2$(note that $M_2$ is the concurrence): given a state $\rho$, then any SLOCC operation acting on it that leaves the trace of $\rho$ invariant leaves also $M_1$ and $M_2$ invariant. This follows from the fact that the following holds for SLOCC invariant entanglement monotones:

$$M\left(\frac{(A\otimes B)\rho(A\otimes B)^\dagger}{\mathrm{Tr}\left((A\otimes B)\rho(A\otimes B)^\dagger\right)}\right) = M(\rho)\frac{|\det A||\det B|}{\mathrm{Tr}\left((A\otimes B)\rho(A\otimes B)^\dagger\right)} \tag{101}$$

Even stronger, this enables to write down a continuous parameterization of the boundary between separable and entangled states: this follows from the fact that we will show in section 4.4.1 that a state is on the boundary between the separable and entangled states iff $-s_0 + s_1 + s_2 + s_3 = 0$ (note that the trace-condition is not necessary in this case). A continuous parameterization of this boundary is therefore obtained by acting with Lorentz transformations on all Bell diagonal states whose maximal eigenvalue is exactly $1/2$.

The Lorentz singular value decomposition will turn out to be an extremely useful tool in the study of mixed states of two qubits. This stems from the fact that it effectively separates the local parameters (i.e. the Lorentz transformations) from the global ones (i.e. the Lorentz singular values). Note for example that the complete classification of all pure $2\times 2\times N$ states under SLOCC operations could readily have been obtained from it. As a first new application, we will study the quantum steering problem of mixed states of two qubits.

## 4.3. Quantum steering with mixed states of 2 qubits

The general framework of quantum steering developed in section 4.1 can be refined considerably in the case of qubits. Of course we will do the analysis in the $R$-picture. The problem is the following: given a mixed state of two qubits, what kind of local density operators can be created at Bob's side by POVM measurements done by Alice?

Parameterizing the local operators as

$$x_i^\alpha = \mathrm{Tr}(E^\alpha\sigma_i), \qquad y_i = \mathrm{Tr}(\rho_B\sigma_i),$$

with $0\le i\le 3$, we obtained the result that the POVM element $E^\alpha$ implemented by Alice gave rise to the local density operator $\rho_B$ of Bob:

$$y^\alpha = Rx^\alpha.$$

Here $R$ is the representation in the $R$-picture of the mixed state of two qubits, $y^\alpha = p_\alpha[1;\vec{y}^\alpha]$ with $p_\alpha$ the probability of getting outcome $\alpha$ and $\vec{y}^\alpha$ the Bloch vector of the obtained state.

The only constraint on $x_i^\alpha = \mathrm{Tr}(\sigma_i E_\alpha)$ is the fact that $0\le E_\alpha$. This condition becomes much more appealing in terms of $x^\alpha$, where it is equivalent to the conditions

$$(x^\alpha)^T M(x^\alpha)\ge 0 \qquad \text{and} \qquad x_0^\alpha \ge 0.$$

POVM elements can therefore be associated to four-vectors (or particles) with a certain positive mass; this reduces to light-like particles in the case of a POVM element that is pure [5].

Let us investigate the effect of this condition more closely in the case of quantum steering with a pure state $R$. Assume that Alice applied a projective POVM element for which $(x^\alpha)^T M(x^\alpha) = 0$. Bob's local density operator becomes proportional to $y^\alpha = Rx^\alpha$. But as the original state $R$ was pure, we know that a projective measurement of Alice results in a pure local density operator of Bob. This has to hold for all possible projective measurements, and this is only possible iff $R^T MR = M$, implying that $R$ is a Lorentz transformation. This is of course in complete correspondence with the results of the Lorentz singular value decomposition. In this pure state case, all possible local density operators can be generated inside the Bloch sphere, and the only condition is that the ensemble average of all obtained density operators of Bob is given by the original one.

In the case of mixed states, the possible density operators in the ensemble generated by the POVM $\{E_\alpha\}$ will be constrained to lie inside an ellipsoid in the Bloch sphere:

**Theorem 20.** *Given a mixed state of two qubits parameterized by $R$. Then the ensemble $\{p_\alpha, \vec{y}^\alpha\}$ at Bob's side can be created by a POVM measurement of Alice $\{E^\alpha\}$ or $\{x^\alpha\}$ if and only if all the $\vec{y}^\alpha$ lie inside the "quantum steering ellipsoid"specified by*

$$[1; \vec{y}^T]R^{-T}MR^{-1}[1; \vec{y}] \geq 0$$

*and if*

$$\sum_\alpha p_\alpha y^\alpha = R \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

*where $y^\alpha = [1; \vec{y}^\alpha]$.*

*Proof:* The proof is immediate with the above observation about the condition on POVM elements: $x^T Mx \geq 0$. It is easily verified that the obtained equation indeed defines an ellipsoid in the Bloch sphere when we normalize the obtained states $y^\alpha$. □

Note that the quantum steering ellipsoid is completely contained within the Bloch sphere. This Theorem is a direct generalization of the quantum steering Theorem in the case of pure states: there the ellipsoid corresponds to the complete Bloch sphere. It is clear that the shape of the ellipsoid will reveal

[5] It is tempting to conclude that faster than light transmission of information would be possible if one would allow $POVM$ elements that are indefinite, but unfortunately this seems not to be the case as faster than light communication is not possible in any linear theory.

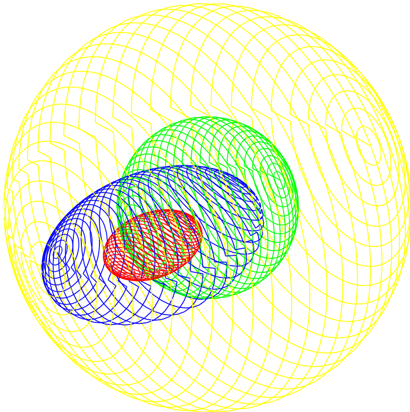


**Figure 2.** The Bloch sphere, the steering ellipsoid of a random state of two qubits, the ellipsoid of local density operators of Bob compatible with the steering ellipsoid for a random choice of the proportionality factor, and the scaled Bloch sphere of local density operators of Alice compatible with the proportionality factor.

many of the properties of the considered mixed state of two qubits. Indeed, the ellipsoid encodes almost all the information about the original mixed state:

**Lemma 4.** *Up to a local unitary operation that leaves the local density operator at Alice's side invariant, a density operator of two qubits is completely specified by the coordinates of the "quantum steering ellipsoid" at Bob's side, supplemented by the knowledge of the two local density operators. Moreover, suppose that the local density operator (i.e. the Bloch vector) of Alice is situated on a sphere (centered around the origin) with radius $\beta$ inside the Bloch sphere. Then the local density operator of Bob is situated on an ellipsoid inside the quantum steering ellipsoid that is a scaled version of it with a scaling factor $\beta$ (see Figure (2)).*

*Proof:* The proof is constructive. The knowledge of the ellipsoid means that we know the matrix

$$E = R^{-T} M R^{-1}$$

up to a proportionality factor. This symmetric matrix has a unique Lorentz singular value decomposition of the form $E = LDMDL^T$ where $DMD$ are the Lorentz singular values (the existence of this decomposition follows from the fact that we know that $R$ itself also has a LSVD). Note that we explicitly introduced the matrix $M$ such as to ensure that the signature of $E$ is conserved. If $R = L_B \Sigma L_A^T$, then it is obvious from the previous expression that we can uniquely[6] determine $L_B$ and that we can define $\Sigma$ up to a proportionality factor.

[6]Note that if $\Sigma$ has Lorentz singular values with multiplicity larger than 1, $L_B$ is not uniquely determined; this is however not a problem as we still have to determine $L_A$. Note

It remains to determine $L_A$ which has 6 degrees of freedom, exactly the number of degrees of freedom of the two local density operators.

Without loss of generality, we assume that the local density operator of Alice is diagonal (this can always be achieved by LU transformations). Let us next define a $L_1$ with the property that the first row of $L_B\Sigma L_1$ is given by $(\alpha, 0, 0, 0)$ (note that $\alpha$ is the proportionality factor and therefore a parameter, and that $L_1$ is only unique up to LU). It is now easy to find a $L_2$ such that the first column of $L_B\Sigma L_1 L_2$ coincides with the Bloch vector of Alice's local density operator, which we parameterize as $(1, 0, 0, \beta)$, and it is then easy to see that this local density operator completely determines the proportionality factor. The first row of $L_2$ is therefore uniquely determined, and to assure that $L_2$ is a proper orthochronous Lorentz transformation, $L_2$ must be of the form

$$L_2 = \begin{pmatrix} 1/\alpha & 0 & 0 & \beta/\alpha \\ \beta x & \sqrt{1-x^2(1-\beta^2)} & 0 & x \\ \beta y & -\frac{xy(1-\beta^2)}{\sqrt{1-x^2(1-\beta^2)}} & \sqrt{\frac{1-(x^2+y^2)(1-\beta^2)}{1-x^2(1-\beta^2)}} & y \\ \beta z & -\frac{xz(1-\beta^2)}{\sqrt{1-x^2(1-\beta^2)}} & \pm\sqrt{\frac{1-(x^2+z^2)(1-\beta^2)}{1-x^2(1-\beta^2)}} & z \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & O_2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{102}$$

with $x, y, z$ parameters left to be determined and constrained under $1/\alpha^2 - x^2 - y^2 - z^2 = 1$, $\pm$ a sign that is a function of $(x, y, z)$ such as to assure that $L_2$ is a Lorentz transformation and $O_2$ an arbitrary orthogonal $2 \times 2$ matrix with the only constraint that its determinant must be chosen such that $L_2$ has determinant $+1$. The goal is now to identify $L_A$ with the transpose of $L_1 L_2$. Therefore the following condition must be fulfilled: $L_B\Sigma L_1$ times the first column of $L_2$ must be equal to the local density operator of Bob. Of course, this uniquely specifies $(x, y, z)$, specifying $L_2$ and therefore $L_A$ up to the local unitary $O_2$ with one degree of freedom. Note that this is precisely the subgroup of the unitaries leaving the local density operator of Bob invariant.

The first column of $L_2$ is of the form

$$\begin{pmatrix} 1/\alpha \\ \frac{\sqrt{1-\alpha^2}}{\alpha}\vec{n} \end{pmatrix} \tag{103}$$

with $\alpha$ determined by $\rho_A$ and $\vec{n}$ a unit vector (having two degrees of freedom). This implies that $\rho_B$ is determined to lie in a two-dimensional manifold inside the ellipsoid once the ellipsoid is specified. Note that $\rho_B$ can be chosen to lie anywhere in this $2D$ submanifold independent of the choice of $\rho_A$. Geometrically, this submanifold is the surface of an new ellipsoid. This follows from the fact that all possible local density operators are generated by multiplying the matrix $L_B\Sigma L_1$ with the vector in equation (103); the condition on $\vec{n}$ translates into an equation of a new ellipsoid in terms of the Bloch vector of $\rho_B$. Writing

---

also that if the corresponding state $\rho$ is separable, there is no way of distinguishing $\rho$ and $\rho^{T_B}$ with $T_B$ the notation for partial transpose: $\Sigma$ is then uniquely defined up to one sign.

down the explicit expression for this ellipsoid results in:

$$y^T(R^{-T}MR^{-1})y = 1/\alpha^2(\rho_A). \tag{104}$$

The left hand side is exactly the same expression as in the case of the quantum steering ellipsoid, while the right hand side is $1/\alpha^2$ instead of 0: the new ellipsoid has exactly the same center as the old one, and is just a scaled (i.e. smaller) version of it. Moreover, the scaling factor is completely determined by $\rho_A$: if $\rho_A$ is almost pure, then $\alpha$ becomes very large, meaning that the local density operator of Bob will be located very close to the steering ellipsoid (note that this implies that only the states on the ellipsoid close to the density operator will have a large weight in a convex decomposition). More specifically, let us draw a scaled version of the Bloch sphere containing $\rho_A$. Then this scaling factor is exactly the same scaling factor as the one of the two ellipsoids. Of course this implies that if we first fix $\rho_B$ (wherever in the steering ellipsoid), then $\rho_A$ will be constrained to be situated on a scaled version of the Bloch sphere, with the scaling factor determined by $\rho_B$. □

It is amazing that the ellipsoid encodes all the information about $L_B$ and $\Sigma$ while the local density operators encode the information about $L_A$: in some sense a highly nonlocal density operator of two qubits can completely be specified by local characteristics. Note however that the mixed state case is much more involved than the pure state case, where the quantum steering ellipsoid reduces to the Bloch sphere itself, and where the inner ellipsoid coincides with the sphere on which Alice's local density operator is situated. In complete analogy with this pure state case, a mixed state of two qubits will contain more entanglement the deeper the local density operators lie inside the ellipsoid or Bloch sphere; the deeper they are situated, the more local disorder, and we have already argued that this maximizes a whole family of entanglement monotones. Using the techniques of the previous chapter, it is indeed easy to prove that the local filtering operation to be performed by Alice such as to yield a new state with maximal entanglement is the one that transforms the local density operator of Alice to one that is proportional to the identity (implying that the local density operator of Bob moves to the center of the steering ellipsoid). Note that a local filtering operation by Alice that is full rank does not change the steering ellipsoid and just changes the position of the local density operators: this follows from the fact that this local filtering operation is probabilistically reversible.

To complete the geometrical picture of all possible states, we would like to have a characterization of all possible steering ellipsoids. The solution to this depends on equation (96) which characterizes the allowed combinations of Lorentz singular values. This immediately implies that not all ellipsoids inside the Bloch sphere correspond to physical (i.e. positive semidefinite) density operators: take just an example for which all $\{s_i\}$ have absolute value smaller than $s_0$ but for which $s_3 \leq s_1 + s_2 - s_0$. Geometrically, equation (96) corresponds

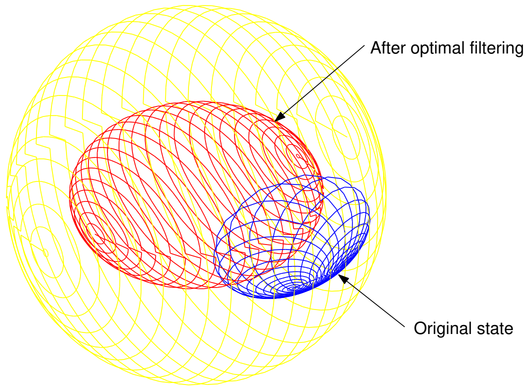



**Figure 3.** The Bloch sphere, the steering ellipsoid of a random state $L_A\Sigma L_B^T$ of two qubits, and the steering ellipsoid after the local filtering operation (i.e. a boost) of Bob corresponding to $L_B^{-T}$.

to the fact that the ellipsoid can typically not be squeezed too much in one direction while being broad in the two others: the length of the smallest axis $s_3$ has to remain larger than $s_1 + s_2 - s_0$.

We have now obtained a complete geometrical picture of all mixed states of two qubits. We have also identified all states that can be created locally by the action of local filtering operations of Alice: of course the steering ellipsoid does not change if the applied filter was full rank, and only the local density operators of Alice and Bob change. More specifically, the only thing that changes is the proportionality factor of the inner ellipsoid (and Bloch sphere). On the other hand, if Bob is applying a filtering operation, this amounts to a Lorentz transformation of the steering ellipsoid: depending on the direction of the boost, the ellipsoid will undergo some squeezing (recall the Fitzgerald contraction of special relativity). Meanwhile, the inner ellipsoid transforms in a similar way, but the proportionality factor will change (unless no actual boost was applied but only a LU). We illustrate this behaviour in Figure 3.

Let us now look for more general local operations. The most general kind of trace-preserving 1-local operations (without classical communication) that Alice can perform are described by a POVM $\{E_i\}$ where one does not keep track of the measurement outcome. In the R-picture, this amounts to a transformation of the kind:

$$R' = R(\sum_i c_i L_i) = RQ \tag{105}$$

where $c_i$ are positive constants, $L_i$ proper orthochronous Lorentz transformations corresponding to the POVM-elements and $Q$ is defined as $Q = \sum_i c_i L_i$. This operation is trace preserving iff the first column of $Q$ is given by $[1;0;0;0]$. As will be shown in the later chapter on quantum channels, a given matrix $Q$

corresponds to a convex sum of proper orthochronous Lorentz transformations iff it is corresponds to the R-picture of the partial transpose of a state $\rho$. As a partial transpose corresponds to multiplying the third column (or row) with $-1$, we can easily use the formalism used for describing quantum states to determine the class of $1-$local operations. In complete analogy with the Lorentz singular value decomposition, every $Q$ can be brought into a normal form by left and right multiplication with proper orthochronous Lorentz transformations (the only difference it that the sign convention of all the Lorentz singular values has to be reversed).

To determine the steering ellipsoid of $R' = RQ$, one has to determine the image of the vectors $Qx$ with $x^T Mx \geq 0$ under the map $R$. Of course this defines a new steering ellipsoid inside the original one. It seems very plausible that every physical steering ellipsoid inside the original one can be constructed this way. This is however not the case. To see this, we first note that it is sufficient to consider the Bell diagonal case (diagonal $R$) which defines an ellipsoid with the axis corresponding to the Lorentz singular values $s_1, s_2, |s_3|$. We consider the case where a Bell diagonal state is mapped onto a Bell diagonal state, and therefore we have to choose $Q$ also diagonal: $Q = \mathrm{diag}[1, \tau_1, \tau_2, \tau_3]$ with $1 \geq \tau_1 \geq \tau_2 \geq |\tau_3|$ and $1 - \tau_1 - \tau_2 + \tau_3 \geq 0$. To check whether there exists such a $Q$ that transforms one Bell diagonal into another one, it is necessary and sufficient that the vector of Lorentz singular values of the second one $x_2$ is a permutation of the vector of Lorentz singular values of the first $x_1$ one multiplied by a feasible $Q$:

$$\exists P : x_2 = PQx_1. \tag{106}$$

It is however easy to find situation in which all ordered singular values $x_2$ are smaller than the respective ones of $x_1$, and nevertheless no $P$ and $Q$ exist that are in correspondence with the previous equation. This indicates how difficult it is to characterize local operations.

The previous situation is very much related to a conjecture we made in [**213**]. There we considered the problem of determining if two density operators $\rho_1$ and $\rho_2$ could be converted into each other by the class of SLOCC operations when also mixing is allowed. It is of course necessary and sufficient to show that the normal forms (i.e. Bell diagonal states) can be transformed into each other by SLOCC. Numerical investigations indicated that a given Bell diagonal state can only be converted into another one iff this last one is a mixture of the original Bell-diagonal state with a separable state, although a general proof has not been found. The conjecture was as follows:

**Conjecture 3.** *A two-qubit state $\rho_1$ can probabilistically be converted into the state $\rho_2$ iff the Bell-diagonal normal form of $\rho_2$ is a convex sum of a separable state and the Bell-diagonal normal form of $\rho_1$.*

It is clear that a trivial procedure exists to implement this conversion with unit efficiency: mix the state with one that can be locally made. Let us for example investigate whether the Bell-diagonal $\rho_1$ with ordered eigenvalues $\{\lambda_i\}$ can be transformed into the Bell-diagonal $\rho_2$ with ordered eigenvalues $\{\mu_i\}$. We can restrict ourselves to mixing with separable Bell diagonal states lying on the boundary of the entangled and separable states, and these have their largest eigenvalue equal to 1/2. Under the assumption of our conjecture, conversion is possible iff the following constrained system of equations in $x, y, z, t, P$ has a solution:

$$\begin{pmatrix} 1 & 0 \\ 0 & P_3 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix} = (1-x) \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{pmatrix} + x \begin{pmatrix} 1/2 \\ y \\ z \\ t \end{pmatrix}$$
$$(0 \leq x \leq 1) \qquad (y, z, t \geq 0) \quad (y + z + t = 1/2)$$

where $P_3$ is a $3 \times 3$ permutation matrix. This system can readily be solved. Note that there is a close relation between the above set of equations and majorization. This condition turns out the be the only known necessary and sufficient non-trivial condition for probabilistically transforming a state to another one. This is a first step in characterizing all possible 2-LOCC protocols (allowing local operations by both Alice and Bob and also allowing back and forth classical communication), which is a major open problem.

## 4.4. Entanglement measures

Entanglement measures quantify the degree of entanglement present in a quantum state. There are essentially two types of entanglement measures that one can define in an operational way: the one is the entanglement of formation, related to the minimal asymptotic amount of EPR-pairs needed to construct many copies of the state. The other one is the entanglement of distillation, defined as the maximal fraction of EPR-pairs (or a state asymptotically close to it) one can distill out of a large amount of copies of the state. We have already seen that in the case of pure states, the two quantities coincide. In the case of mixed states however, there is a definite gap between the two measures: intuitively this follows from the observation that we have to "forget" information to create a mixed state out of EPR-pairs, and this information cannot be recovered during a distillation protocol (see e.g. [**226**]).

Much more other useful entanglement measures do exist: we will also describe the entanglement measures negativity, relative entropy of entanglement, fidelity (maximal singlet fraction) and amount of violation of the Bell inequalities. These measures all highlight a different aspect of entanglement for mixed states, and are very interesting for e.g. obtaining lower and upper bounds of the classical or quantum capacity of a quantum channel.

### 4.4.1. Entanglement of Formation and Concurrence

The entanglement of formation is related to the minimal amount of ebits needed to prepare a given state. This leads to the following mathematical expression of the entanglement of formation:

$$E_f(\rho) = \min_{\{p_i,|\psi_i\rangle\}} \sum_i p_i E(|\psi_i\rangle) \tag{107}$$

where $E(|\psi_i\rangle)$ is given by the von-Neumann entropy of the local density operator of $|\psi_i\rangle$. The entanglement of formation is therefore variationally defined over all ensembles $\{p_i, |\psi_i\rangle\}$ for which $\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|$. It is not known whether $E_f(\rho^{\otimes n}) = nE_f(\rho)$, but despite huge efforts by many people not a single counterexample has been found that disproves additivity. Note also that the proof of lemma 11 in the appendix can readily be applied to this case to yield a bound on the number of pure states necessary in the optimal decomposition: $N \leq d^4$ with $d$ the dimension of the Hilbert space of each particle.

The entanglement of formation can be calculated exactly in the case of mixed states of two qubits. The following nice theorem[7] is due to Wootters [**244**]:

**Theorem 21.** *Given a mixed state of two qubits $\rho = XX^\dagger$ with $X$ a* false *square root of $\rho$. Then the entanglement of formation is given by a convex monotonously increasing function $f(C)$ of the concurrence $C$ given by*

$$f(C) = H((1+\sqrt{1-C^2})/2) \tag{108}$$

$$H(p) = -p\log p - (1-p)\log(1-p). \tag{109}$$

*The concurrence is defined as*

$$C = \max(\sigma_1 - \sigma_2 - \sigma_3 - \sigma_4, 0)$$

*where $\{\sigma_i\}$ are the (ordered) singular values of the matrix $X^T\epsilon\otimes\epsilon X$ with $\epsilon$ the completely antisymmetric $2\times 2$ tensor. It follows that a mixed state of two qubits is entangled if and only if its concurrence is larger than zero. The number of pure states in the optimal decomposition is given by the rank of the density operator, except if $\rho$ is separable and has rank 3 with $\sigma_1 - \sigma_2 - \sigma_3 < 0$ where a decomposition with 4 pure separable pure states is needed.*

*Proof:* Given a pure state of two qubits $|\psi\rangle$, it is easy to check that $E(|\psi\rangle) = f(C(|\psi\rangle))$ where $C(|\psi\rangle)$ is defined as

$$C(\psi) = |\psi^T\epsilon\otimes\epsilon\psi|.$$

As $f(C)$ is a convex monotonously increasing function, we will certainly have succeeded the minimization if we succeed to minimize the average concurrence

[7] Note that the statement about the number of pure states in the decomposition was not given by Wootters, and that we present a different derivation.

and moreover can show that all pure states in the decomposition have the same concurrence:

$$E_f(\rho) = \min_{p_i,\psi_i} \sum_i p_i f(C(\psi_i)) \geq \min_{p_i,\psi_i} f\left(\sum_i p_i C(\psi_i)\right)$$

Let us therefore try to minimize the average concurrence over all possible pure state decompositions. Given a square root $X$ of $\rho$, all pure state decomposition can be written as the columns of $XU$ with $U$ an arbitrary isometry (note that the norm of the columns corresponds to the square root of the weight of this particular state in the decomposition). It is now straightforward to see that the minimization of the average concurrence is equal to the following optimization problem over the manifold of unitary matrices:

$$\min_U \sum_i |U^T(X^T\epsilon \otimes \epsilon X)U|_{ii}.$$

The problem is therefore to find the isometry $U$ such that the sum of the absolute values of the diagonal elements of the above matrix is minimal. Observe that $X^T\epsilon \otimes \epsilon X$ is complex symmetric, so the Takagi decomposition[8] dictates that $X^T\epsilon\otimes\epsilon X = V^T\Sigma V$ with $\Sigma$ the diagonal matrix containing the ordered singular values of $X^T\epsilon\otimes\epsilon X$ and $V$ unitary. Absorbing $V$ into $U$, the optimization problem becomes equivalent to:

$$\min_U \sum_i |U^T\Sigma U|_{ii}.$$

A lower bound is obtained as follows, where we used the notation $U_{ij}^2 = p_{ij}\exp(i\phi_{ij})$, $\sum_i p_{ij} = \sum_j p_{ij} = 1$:

$$\begin{aligned}
\sum_i |U^T\Sigma U|_{ii} &= \sum_i |p_{1i}\sigma_1 + \sum_{j=2}^n p_{ji}\exp(i(\phi_{ji}-\phi_{j1}))\sigma_j| \\
&\geq \sum_i \left(p_{1i}\sigma_1 - \sum_{j=2}^n p_{ji}\sigma_j\right) \\
&= \sigma_1 - \sum_{j=2}^n \sigma_j
\end{aligned}$$

This lower bound is sharp if $\sigma_1 - \sum_{j=2}^n \sigma_j \geq 0$, and we then call it the concurrence. Indeed, $U$ can be chosen such as to yield the diagonal elements of $U^T\Sigma U$ zero everywhere except on the $(1,1)$ entry which is equal to $\sigma_1 - \sum_{i=2}^n \sigma_i$. This can be achieved in three steps: consider the $2\times 2$ submatrix

$$\begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}, \tag{110}$$

[8] The Takagi singular value decomposition states that each complex symmetric matrix $X$ can be decomposed as $X = U\Sigma U^T$ with $\Sigma$ diagonal (containing the singular values) and $U$ unitary.

define $\phi = \arctan(\sqrt{\sigma_2/\sigma_1})$ and

$$U = \begin{pmatrix} \cos(\phi) & i\sin(\phi) \\ -\sin(\phi) & i\cos(\phi) \end{pmatrix}. \tag{111}$$

It follows that

$$U^T \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} U = \begin{pmatrix} \sigma_1 - \sigma_2 & \sqrt{\sigma_1\sigma_2} \\ \sqrt{\sigma_1\sigma_2} & 0 \end{pmatrix}. \tag{112}$$

Repeating this 2 times more indeed yields a real symmetric matrix with all diagonal elements equal to zero except the $(1,1)$ element, which becomes equal to the lower bound. Let us define $U$ as the unitary matrix fulfilling this task. Note that $C = \mathrm{Tr}(U^T\Sigma U)$ in this case.

If $\sigma_1 - \sum_{j=2}^n \sigma_j < 0$, then it is trivial to find another lower bound: $\sum_i |U^T\Sigma U|_{ii} \geq 0$. This bound is actually achievable by infinitely many different $U$: it suffices to observe that if $\sigma_1 - \sum_{j=2}^n \sigma_j < 0$, then it is always possible to find phases $\phi_i$ such that

$$\sum_i \sigma_i \begin{pmatrix} \cos(\phi_i) \\ \sin(\phi_i) \end{pmatrix} = 0. \tag{113}$$

This follows because four $2D$-vectors can always form a closed path iff the length of the largest is smaller than the sum of the length of all the other ones. Suppose the set of phases $\{\phi_i\}$ does the job. Then $U$ can be chosen to be

$$U = \frac{1}{2}\begin{pmatrix} e^{i\phi_1} & e^{i\phi_2} & e^{i\phi_3} & e^{i\phi_4} \\ e^{i\phi_1} & e^{i\phi_2} & -e^{i\phi_3} & -e^{i\phi_4} \\ e^{i\phi_1} & -e^{i\phi_2} & e^{i\phi_3} & -e^{i\phi_4} \\ e^{i\phi_1} & -e^{i\phi_2} & -e^{i\phi_3} & e^{i\phi_4} \end{pmatrix} \tag{114}$$

and then the lower bound equal to 0 will be saturated. We have therefore proven that the entanglement of formation is zero if $\sigma_1 - \sum_{i=2}^n \sigma_i \leq 0$, as we found an explicit decomposition where all pure states have concurrence equal to zero and are therefore separable.

It remains to be proven that in the case $\sigma_1 - \sum_{i=2}^n \sigma_i > 0$, all the concurrences of the pure states in the decomposition can be chosen equal to each other. First of all we note that $U^T$ as chosen is not the only $U$ minimizing the average concurrence, but that any $O^TU^T$ would do as well with $O$ orthogonal and leaving all the diagonal elements of $O^TU^T\Sigma UO$ positive: this follows from the fact that the trace $\mathrm{Tr}(O^TU^T\Sigma UO)$ is left invariant and therefore the sum of the absolute values. We would like to have all concurrences of the pure states in the decomposition equal to their average; this means that we would like to get the diagonal elements of $O^TU^T\Sigma UO$ to be proportional to their weight in the ensemble times the concurrence: we want all the diagonal elements in the matrix

$$O^T(\underbrace{U^T\Sigma U - U^\dagger VX^\dagger XV^\dagger U}_{Q})O \tag{115}$$

to be zero. Note that $V$ is the unitary matrix appearing in the Takagi decomposition, and that the pure state decomposition under consideration is given by $XV^\dagger UO$. The remaining task is to tune the orthogonal matrix $O$ such that all the diagonal values vanish. Note that $Q$ is Hermitian and has trace equal to zero. An orthogonal matrix doing the job can be found in three steps: as the trace is zero, there must certainly be a pair of diagonal entries for which one is positive and the other one is negative. It is easy to see that a $2 \times 2$ rotation on the block under consideration can always make one of these entries equal to zero. Next we can repeat this step on another pair, and the third time this has been done will yield a matrix with complete zero diagonal. Note that we are assured that the entries of $O^T U^T \Sigma U O$ remain positive as $O^T U^\dagger V X^\dagger X V^\dagger U O$ is a positive matrix.

Finally, it remains to be proven that the number of pure states in the decomposition is given by the rank, with the exception of one case. If $C > 0$, this follows from the fact that nowhere in the proof we used the fact that $X$ was square. If $C = 0$ however, we explicitly made use of this fact, as we defined the matrix (114) as a $4 \times 4$ matrix. If the rank of $\rho$ is 4, then there is obviously no problem. If the rank of $\rho$ is 2 there is no problem neither, as in this separable case $\sigma_1 = \sigma_2$ and we could therefore have used the techniques of $C > 0$. There remains the rank 3 case with $\sigma_1 - \sigma_2 - \sigma_3 < 0$; then no unitary $3 \times 3$ matrix can be found such as to yield all zero values on the diagonal. This follows from the following nice Theorem of Thompson [**201**]:

> Let $d_1, \cdots, d_n$ be complex numbers and $s_1, \cdots s_n$ nonnegative real numbers, enumerated so that $|d_1| \geq \cdots \geq |d_n|$ and $s_1 \geq \cdots \geq s_n$. A complex symmetric matrix exists with $d_1, \cdots, d_n$ as its diagonal elements and $s_1, \cdots, s_n$ as its singular values, if and only if

$$\begin{aligned}
\sum_{i=1}^{k} |d_i| &\leq \sum_{i=1}^{k} s_i, & 1 \leq k \leq n \\
\sum_{i=1}^{k-1} |d_i| - \sum_{i=k}^{n} |d_i| &\leq \left( \sum_{i=1, i\neq k}^{n} \right) - s_k & 1 \leq k \leq n \\
\sum_{i=1}^{n-3} |d_i| - |d_{n-2}| - |d_{n-1}| - |d_n| &\leq \sum_{i=1}^{n-2} s_i - s_{n-1} - s_n
\end{aligned}$$

> The last inequality does not apply when $n < 3$.

The inequality of interest is the last one with $n = 3$: it becomes $|d_1| + |d_2| + |d_3| \geq -\sigma_1 + \sigma_2 + \sigma_3$ and the right hand side is larger than zero by assumption. If we let $n = 4$ however, corresponding to considering 4 pure states in the decomposition, we already know we could make all diagonal entries equal to zero by the matrix (114). This ends the proof. □

The following is an immediate corollary:

**Corollary 1.** *A mixed state of two qubits is entangled if and only if its concurrence is larger than* 0.

Note that the above proof was constructive and yields a robust way of calculating the optimal decomposition numerically. The central tool in the Theorem was the introduction of the concurrence. This concurrence was defined during the proof as the convex roof of the concurrence defined on pure states. The definition of the concurrence on pure states coincides exactly with the way we defined entanglement monotones in section 3.1.3, which are invariant under determinant 1 SLOCC operations (note that the mixed state version defined as a convex roof of the linearly homogeneous EM is of course also invariant under these operations):

$$C(\rho) = C((A \otimes B)\rho(A \otimes B)^\dagger) \qquad \forall A, B \in SL(2, \mathcal{C}). \tag{116}$$

This also follows immediately from the following identity:

$$X^T(A \otimes B)^T \epsilon_2 \otimes \epsilon_2 (A \otimes B)X = \det(A)\det(B)X^T \epsilon_2 \otimes \epsilon_2 X$$

As the concurrence in invariant under determinant 1 SLOCC operations, it should solely depend on the Lorentz singular values. Note that we have seen that every mixed state of two qubits can be brought into Bell diagonal form by SLOCC operations (even if the states are quasi-distillable and the transformations involved tend to infinity, the Lorentz singular values are still well defined). But in the case of (unnormalized) Bell diagonal states, the concurrence can readily be expressed in function of the eigenvalues: $C = \max(0, \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4)$. Using equation (95), this implies that the concurrence of a state is given by

$$C(\rho) = \max(0, (-s_0 + s_1 + s_2 + s_3)/2) \tag{117}$$

with $\{s_i\}$ the Lorentz singular values of the state. We note that this expression is completely equivalent to the entanglement monotone $M_2$ introduced in Theorem 19. This means that we get a new variational characterization of the concurrence of a state $R$:

$$C(R) = \max\left(0, -\min_L \mathrm{Tr}(LR)/2\right) \tag{118}$$

with $L$ a proper orthochronous Lorentz transformation. This variational characterization will turn out to be very useful in the sequel.

Before discussing other entanglement measures for mixed states, let us briefly have a look of how the concept of concurrence generalizes to higher dimensions (see also [**180, 11**]). First of all we observe that the entanglement of formation is a function of the eigenvalues of the local density operator, but unlike in the case of qubits where one parameter suffices (i.e. the concurrence), there are more parameters, and the complexity of the optimization problem grows drastically (note that typically the number of states needed in an optimal decomposition grows like $n^4$ with $n$ the dimension of the local Hilbert space; the qubit case is an exception). The numerical optimization using an efficient gradient search algorithm suffers from local minima, but extended numerical search indicated that the global minimum is typically reached after a few trials. For details we refer to the paper of Audenaert et al.[**10**].

But let's be less ambitious and try to define a concurrence for higher dimensional systems that enables to distinguish separable from entangled states. Given a pure state $\psi$ in a $n \times m$ dimensional Hilbert space, checking whether this state is separable can be done by checking if all the $2 \times 2$ minors of the reshaped $n \times m$ matrix containing the entries of $\psi$ are equal to zero. It is easy to see that it is enough to check $(n-1)(m-1)$ minors. Each of these $(n-1)(m-1)$ minors gives rise to a concurrence $C_\alpha$ expressible in a form like

$$\psi^T \underbrace{((\epsilon_2 \oplus 0) \otimes (\epsilon_2 \oplus 0))}_{\hat{C}_1} \psi. \tag{119}$$

In the case of qubits, only one concurrence had to be defined, and the separability problem was equivalent to checking if isometries $U$ existed such that the diagonal elements of $U^T X^T \epsilon_2 \otimes \epsilon_2 X U$ are all equal to zero. In the higher dimensional case, we have to look for isometries $U$ such that the diagonal elements of all $(n-1)(m-1)$ matrices $U^T X^T \hat{C}_\alpha X U$ are equal to zero. This is the necessary and sufficient condition for a mixed state $\rho = XX^\dagger$ to be separable. Unlike in the case of qubits however, no constructive way of finding the optimal $U$ has been found, as the problem has become a tensor problem (i.e. extra index $\alpha$) instead of a matrix problem. Nevertheless, a numerical optimization problem over the manifold of isometries $U$ can easily be constructed [**63**], although there is again no guarantee that the algorithm will converge to the global minimum.

### 4.4.2. Negativity

Let us first define the partial transpose map of a state naturally endowed with a tensor product structure:

$$\rho^{T_2}_{ij,kl} = \rho_{il,kj} \qquad\qquad \rho^{T_1}_{ij,kl} = \rho_{kj,il}. \tag{120}$$

We have shown that a state is entangled iff its concurrence exceeds 0. We have also given a variational characterization of the concurrence in terms of Lorentz transformations. We can readily translate the variational characterization of the concurrence to the $\rho$-picture by the following lemma:

**Lemma 5.** *Given a mixed state of two qubits $\rho$. Taking the partial transpose of this matrix with respect to A or B amounts in the R-picture to multiplying the third row or third column with minus one.*

*Proof:* The proof is immediate by noting that the only Pauli spin operator that is changing sign due to the partial transposition is $\sigma_y$. □

More specifically, a proper orthochronous Lorentz transformation in the $R$-picture corresponds to the partial transpose of a rank 1 operator of the form

$$X = (A \otimes I)|\phi\rangle\langle\phi|(A \otimes I)^\dagger \tag{121}$$

with $|\phi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ and $A \in SL(2,\mathcal{C})$: this follows from the fact that the representation of a pure state in the $R$-picture is given by a proper orthochronous Lorentz transformation times the representation of a maximally entangled state diag$[1,1,-1,1]$, which corresponds to the action of taking a partial transpose.

Using the fact that $\mathrm{Tr}(R_1 R_2^T) = 4\mathrm{Tr}(\rho_1\rho_2)$, the variational characterization of the concurrence in the $R$-picture (118) becomes:

$$C(\rho) = \max\left\{0, -\min_{A\in SL(2,\mathcal{C})} \mathrm{Tr}\left((|A\rangle\langle A|)^{T_B}\rho\right)\right\} = \max\left\{0, -\min_{A\in SL(2,\mathcal{C})}\langle A|\rho^{T_B}|A\rangle\right\} \tag{122}$$

where we slightly abused notation: $|A\rangle$ denotes the unnormalized state $(A \otimes I)|I\rangle$ with $|I\rangle = \sum_i |ii\rangle$, $\det(A) = 1$.

This characterization is very nice and immediately points out the following non-trivial fact:

**Theorem 22.** *A mixed state of two qubits $\rho$ is entangled if and only if its partial transpose has a negative eigenvalue.*

*Proof:* Because we have already proven that a state is entangled iff $C > 0$, it suffices to see that the vector $|A\rangle$ in equation (122) spans the whole space. $C > 0$ therefore implies that $\rho^{T_B}$ cannot be positive semidefinite, while $C = 0$ implies that $\rho^{T_B}$ is positive semidefinite. □

The previous celebrated Theorem was first conjectured by Peres [**170**] and proven by the Horodecki's [**120**] by showing that the Theorem follows from a result of Woronovic [**246**] concerning positive but not completely positive maps (we will investigate the relation between entanglement and positive maps more closely in the last chapter). Here we presented a completely different approach that emphasized the connection between concurrence and the partial transpose criterion.

The partial transpose condition gives rise to an entanglement monotone called the negativity [**250, 83, 229**]:

**Theorem 23.** *Given a bipartite state $\rho$, then we define the negativity to be twice the sum of the absolute values of all negative eigenvalues of $\rho^{T_B}$. The negativity is an entanglement monotone.*

*Proof:*[Vidal and Werner [**229**]] Consider a convex set $\mathcal{S}$ that spans the whole Hilbert space. The set of PPT states, i.e. the set of states that remain positive after partial transposition, is clearly of this kind as all separable pure states span the whole space. It is then always possible to write down a state as

$$\rho = a_+\rho_+ - a_-\rho_-$$

with $a_+, a_- \geq 0$, $\rho_+, \rho_- \in PPT$. It is easily checked that the minimal possible value of $a_-$ is a norm and hence a convex function of $\rho$. Moreover, a state in the $PPT$-class remains in the class after arbitrary LOCC operations, and henceforth $a_-$ cannot increase under LOCC operations. It remains to be shown that $a_-$ is the sum of the negative eigenvalues of $\rho^{T_2}$, which follows trivially from the fact both $\rho_-$ and $\rho_+$ are PPT-states. □

This negativity is very appealing as it is easy to calculate. The negativity can however be zero while the state is entangled if the system is not defined over a $2 \times 2$ or $2 \times 3$ Hilbert space. This has to do with the existence of positive nondecomposable maps[9] in higher dimensional systems. This led to the discovery of bound entangled states by the Horodecki's [**123**], which are states that cannot be distilled. This follows from the fact that the negativity is an entanglement monotone: these states can never be distilled to singlet pairs because these last ones have negativity absolutely larger than 0 while the original ones have negativity equal to zero. The states that remain positive after transposition are called $PPT$-states (from positive partial transposition), and include the separable states as a subclass.

Let us now return to the case of two qubits. The partial transpose criterion is equivalent to the following: given the Lorentz singular values $s_0 \geq s_1 \geq s_2 \geq |s_3|$, then a state is entangled iff changing the sign of $s_3$ still yields a valid state $R$. It is straightforward to show that this implies that $s_3$ is always positive in the case of entangled states. Note that this has a rather strange consequence for the description of the measurement ellipsoids introduced in the quantum steering section: the steering ellipsoid and the corresponding local density operators are left completely invariant under partial transposition (note that we choose the local density operators to be diagonal). This is no problem if $\rho$ is entangled, as in this case there is only one valid choice for the signs of the Lorentz singular values. In the case of a separable state however, two equally valid choices exist. In other words: there is no way of locally distinguishing a separable state from its partial transpose! This is in complete accordance with the intuition of Peres [**170**], where he introduced the partial transposition operator because he remembered from his quantum field theory classes that there is no way why two space-like observers should use the same definition of $\sqrt{-1}$: Alice could use $i$ and Bob $-i$, and still everything should be consistent.

[9]A positive nondecomposable map is a map that cannot be written as the convex sum of a completely positive map and a completely positive map preceded by a partial transposition.

This picture has to break down however in the presence of entanglement, as $+i$ corresponds to evolution forward in time and $-i$ to backwards evolution [**183**].

As a next remark concerning the negativity in the case of two qubits, we observe the following:

**Lemma 6.** *If $\rho$ describes a mixed state of two qubits, then at most one eigenvalue of $\rho^{T_B}$ can be negative. Moreover, if $\rho$ is entangled, then $\rho^{T_B}$ has no zero eigenvalues (i.e. is full rank).*

*Proof:* The signature (i.e. number of negative eigenvalues) of the matrix

$$\left((A\otimes B)\rho(A\otimes B)^\dagger\right)^{T_B} = (A\otimes B^*)\rho^{T_B}(A\otimes B^*)^\dagger \tag{123}$$

is equal to the signature of $\rho^{T_B}$ due to the law of inertia of Sylvester. So it is sufficient to consider the Bell diagonal case or to look at the Lorentz singular values. It is readily verified that changing the sign of $s_3$ can make at most one of the eigenvalues $\{\lambda_i\}$ of formula (95) negative, but then all the other eigenvalues are assured to be strictly positive. Indeed, in the case that entanglement is present $s_0 \geq s_1 \geq s_2 \geq s_3$ are all positive, and the eigenvalues of the partially transposed Bell diagonal operator read:

$$\begin{aligned}
\lambda_1 &= s_0 + s_1 + s_2 - s_3 && (124)\\
\lambda_2 &= s_0 + s_1 - s_2 + s_3 && (125)\\
\lambda_3 &= s_0 - s_1 + s_2 + s_3 && (126)\\
\lambda_4 &= s_0 - s_1 - s_2 - s_3 && (127)
\end{aligned}$$

$\lambda_3$ is strictly larger than zero, except in the case where $s_0 = s_1$ and $s_2 = s_3 = 0$. But then no entanglement is present. □

Observe also that the vector $|A\rangle$ appearing in equation (122) always obeys the relation $\langle A|A\rangle = TrA^\dagger A \geq 2$. Therefore the concurrence is always larger than the negativity, and is equal to it iff the previous inequality is fulfilled. This occurs iff the optimal $A$ is proportional to a unitary matrix, and therefore this occurs iff the eigenvector of $\rho^{T_B}$ corresponding to the negative eigenvalue is maximally entangled. This is for example always the case when $\rho$ itself is pure or maximally entangled. More generally, consider the pure state $|\psi\rangle$ that has the coordinates of the eigenvector of $\rho^{T_B}$ corresponding to its negative eigenvalue. Then one can prove the following:

$$C(|\psi\rangle) \geq N(\rho)/C(\rho). \tag{128}$$

This equation holds because one can verify that the $|A\rangle$ defined as $|A\rangle\langle A| = 2|\psi\rangle\langle\psi|/C(|\psi\rangle)$ is of the form $|A\rangle$ used in the variational characterization of the concurrence; indeed, the largest eigenvalue of $A^\dagger A$ is exactly given by $2/C(|\psi\rangle)$. As this $|A\rangle$ is not necessarily optimal, the stated inequality holds (and is saturated iff the optimal $|A\rangle$ lies along the direction of the "negative"eigenvector).

As a last remark, we would like to mention that in the case of mixed states of two qubits, there exists a different way of calculating the negativity. This can be seen as follows:

$$\rho = \sum_i (A_i \otimes I)|I\rangle\langle I|(A_i \otimes I)^\dagger \tag{129}$$

$$\rho^{T_B} = \sum_i (A_i \otimes I) \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{=I_4 - I_2\otimes\sigma_y|I\rangle\langle I|I_2\otimes\sigma_y} (A_i \otimes I)^\dagger \tag{130}$$

$$= \rho_A \otimes I - (I_2 \otimes \sigma_y)\rho(I_2 \otimes \sigma_y) \tag{131}$$

The reduction criterion first formulated by Cerf et al.[**52**] follows easily from the Peres-Horodecki criterion: a mixed entangled state of two qubits is separable iff $\rho_A \otimes I - \rho \geq 0$. It can also be shown that in higher dimensions the reduction criterion is only necessary but not sufficient [**119, 52**].

### 4.4.3. Relative Entropy of Entanglement

The entanglement measure with the not very elegant name "Relative Entropy of Entanglement"(RelEnt) was introduced by Vedral and Plenio [**208, 207**]. They were looking for an entanglement measure that reduces to the unique asymptotic one for pure states but decreases under LOCC operations. From a geometric point a view, it is easy to see that the set of separable states remains invariant under all LOCC operations, while the set of entangled states should shrink. A natural entanglement measure to define in this context is the relative entropy of entanglement:

$$E_R(\rho) = \min_{\sigma\in\mathrm{Sep}} S(\rho||\sigma) = \min_{\sigma\in\mathrm{Sep}} \mathrm{Tr}(\rho\log\rho) - \mathrm{Tr}(\rho\log\sigma). \tag{132}$$

The notation $S(\rho||\sigma)$ denotes the Umegaki relative entropy which can be shown to be positive and jointly convex in its arguments. These properties ensure that the relative entropy of entanglement is indeed an entanglement monotone [**207**]. Moreover, the RelEnt is one of the best known upper bounds of the entanglement of distillation (there is only one mixed state for which the entanglement of distillation is known, and this is the mixture of two Bell states; in that case $E_D = E_R$, but in general it is not expected that the bound is sharp).

For classical probability distributions, the relative entropy has a nice operational characterization as the error exponent in the optimal strategy for distinguishing two probability distributions. The natural generalization from the classical relative entropy to the quantum case would be the quantity

$$\max_{E_\alpha} \sum_\alpha \mathrm{Tr}\,(\rho E_\alpha) \log\left(\frac{\mathrm{Tr}\,(\rho E_\alpha)}{\mathrm{Tr}\,(\sigma E_\alpha)}\right) \tag{133}$$

where the optimization is done over all possible POVM's. It turns out that the quantum Umegaki relative entropy plays exactly the same role as the classical relative entropy in that case: the error exponent corresponding to the optimal discriminating measurement strategy is given by the Umegaki relative entropy[**165**].

The relative entropy of entanglement is therefore a measure for the extent that we would confuse a separable state with the given entangled one.

In the case of pure states, the relative entropy is equal to the von-Neumann entropy of the reduced density operator. Let us therefore rewrite the variational definition of the RelEnt as:

$$E_R(\rho) = -S(\rho) + \min_{\{p_i,\psi_i\},\sigma\in\mathrm{Sep}} -\sum_i p_i \mathrm{Tr}(|\psi_i\rangle\langle\psi_i|\log(\sigma)), \tag{134}$$

where $\sum_i p_i|\psi_i\rangle\langle\psi_i| = \rho$. This definition has to be contrasted with the definition of the entanglement of formation, which can be written as

$$E_f(\rho) = \min_{\{p_i,\psi_i\},\sigma_i\in\mathrm{Sep}} -\sum_i p_i \mathrm{Tr}(|\psi_i\rangle\langle\psi_i|\log(\sigma_i)), \tag{135}$$

due to the stated property about the RelEnt of a pure state. The two variational expressions look very similar: the only distinction is the fact that in the case of entanglement of formation, we are free to vary over different states $\sigma_i$, while in the case of Relent we are not; this indicates of course that the problem of finding an analytical expression of the Relent will be even much harder than in the case of the entanglement of formation: in the case of Relent $\sigma$ is coupled to the whole decomposition, unlike in the case of EoF.

Nevertheless, the above expressions reveal an interesting inequality (see also [**173**]):

$$E_f(\rho) \leq E_R(\rho) + S(\rho). \tag{136}$$

Let us now focus on the case of two qubits. Due to the resemblance of both variational definitions and the fact that we know how to find the optimal decomposition for the EoF, it appears that one should be able to formulate a good lower bound for the Relent. A good choice would be the following: as shown in the section on entanglement of formation, we know how to calculate the optimal $\{\sigma_i\}$, and moreover we know the optimal weights $p_i$ of the pure states $\psi_i$. A good candidate for the optimal $\sigma$ in the case of Relent is therefore given by $\sum_i p_i\sigma_i$. Numerical investigations indeed indicate that this choice of $\sigma$ leads to a value of the Relent that is very close to the optimal one, but not equal to it. This $\sigma$ is however very useful as a starting point of a numerical optimization program to find the exact value of the Relent.

### 4.4.4. Fidelity

The fidelity was already defined in section 4.2.1:

$$F(\rho) = \max_{|\psi\rangle\in\mathrm{ME}} \langle\psi|\rho|\psi\rangle \tag{137}$$

where the optimization is done over all maximally entangled states. In the case of mixed states of two qubits, it has been shown how to calculate its value in theorem 16. Instead of the term fidelity, the term maximum singlet fraction is also used.

The square root of the number $\langle\chi|\rho|\chi\rangle$ with $|\chi\rangle$ an arbitrary state has a nice property[**87**]:

$$\sqrt{\langle\chi|\rho|\chi\rangle} = \min_{\{E_\alpha\}} \sum_\alpha \sqrt{\mathrm{Tr}\,(|\chi\rangle\langle\chi|E_\alpha)}\sqrt{\mathrm{Tr}\,(\rho E_\alpha)}, \tag{138}$$

where $\{E_\alpha\}$ denotes a POVM. Keeping in mind that the quantity $\sum_\alpha \sqrt{p_\alpha}\sqrt{q_\alpha}$ measures the distinguishability between two probability distributions, this implies that $\langle\chi|\rho|\chi\rangle$ is a measure of how well we can distinguish both quantum states with an optimally chosen POVM: the larger the fidelity, the closer the associated measurement outcomes.

The fidelity is the central entanglement measure in the context of entanglement distillation: the fidelity is by far the easiest quantity to calculate, and if the fidelity tends to 1, then all other entanglement measures will also reach their maximal value. In a seminal paper of Bennett, DiVincenzo, Smolin and Wootters [**34**], it was shown that every mixed state of two qubits can be distilled with a finite yield if its fidelity exceeds 1/2. This value of 1/2 is indeed the maximal achievable fidelity of a separable state. Note however that there exist entangled states with fidelity smaller than 1/2. In a later section, it will be shown that the fidelity of these states can also made larger than 1/2 by LOCC operations. The fidelity is therefore certainly not an entanglement monotone.

The fidelity is also important in the context of teleportation with mixed states. The fidelity of teleportation is defined as

$$f(\rho) = \int d\chi \langle\chi|\Phi_\rho(|\chi\rangle\langle\chi|)|\chi\rangle. \tag{139}$$

Here the integration is done over all possible states using the natural Haar measure, and $\Phi_\rho$ denotes the teleportation channel[10] using the standard teleportation protocol $\Phi$ but using the mixed state $\rho$ instead of the singlet state [**124**]. Some algebra leads to the following nice relation between the fidelity $F$ and the teleportation fidelity[**124**]:

$$f(\rho) = \frac{F(\rho)d+1}{d+1}, \tag{140}$$

[10] It is indeed easy to see that teleportation of a state is equal to sending the state through a unital channel (see e.g. last chapter).

with $d$ the dimension of the Hilbert space (strictly speaking, this relation only holds if standard teleportation is done using $\rho$ after is has been brought into its LU normal form). Therefore the quality of teleportation is linear in the fidelity of the state used to teleport.

Let us now prove an upper bound for the fidelity:

**Lemma 7.** *Given a mixed state of two qubits $\rho$ with negativity equal to $N$ and concurrence equal to $C$, then its fidelity $F$ is bounded above by*

$$F \leq \frac{1+N}{2} \leq \frac{1+C}{2}.$$

*Moreover, the first inequality becomes an equality iff $N = C$, and this condition is equivalent to the condition that the eigenvector corresponding to the negative eigenvalue of the partial transpose of $\rho$ is maximally entangled.*

Proof: The fidelity of a state $\rho$ is given by

$$\max_{U_A, U_B \in SU(2)} \operatorname{Tr}\left((U_A \otimes U_B)|\psi\rangle\langle\psi|(U_A \otimes U_B)^\dagger \rho\right) =$$

$$\frac{1}{2} \max_{U_A, U_B} \operatorname{Tr}\left(\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} (U_A \otimes U_B^*)^\dagger \rho^\Gamma (U_A \otimes U_B^*)\right)$$

with $|\psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$. An upper bound is readily obtained by extending the maximization over all unitaries instead of all local unitaries, and it follows that $F \leq \operatorname{Tr}\left(|\rho^\Gamma|\right) = (1+N)/2$. Equality is achieved iff the eigenvector of $\rho^{T_\Gamma}$ corresponding to the negative eigenvalue is maximally entangled. But this condition is exactly equivalent to the condition for $N$ to reach its upper bound $C$, which ends the proof. $\square$

Note that the upper bound is always achieved for pure states and Bell diagonal states.

### 4.4.5. Bell-CHSH inequalities

For a long time, discussions about entanglement were purely meta-theoretical. However, this appeal was changed dramatically in 1964 by John Bell's [**20**] observation that the EPR dilemma could be formulated in the form of assumptions naturally leading to a falsifiable prediction. The experimental fact that these *Bell inequalities* can indeed be violated [**7**] has not only ruled out a single theory, but the very way theories had been formulated for quite a long time. For a long time, entanglement was widely believed to be equivalent to the violation of a Bell inequality.

In the case of two qubits, the definitive version of the Bell inequalities has been formulated by Clauser, Horne, Shimony and Holt [**58**]. Their argument runs as follows: consider two space-like separated spin 1/2 systems, and two measurement apparatus (e.g. polarization filters followed by a photon counter) each of which can measure the spin randomly into one of two assigned directions (we call these directions $\vec{a}, \vec{b}$ at Alice's side, and $\vec{c}, \vec{d}$ at Bob's side). To invoke locality, it is essential in the argument that the random choice of Alice is independent of the random choice of Bob. Note that each measurement produces a binary output (yes or no), which we label $\pm 1$ (this type of measurement is called dichotomic). Let us label the binary measurement outcome of $\vec{a}$ as $a$, and similar for the other ones. Suppose we would like to describe this experiment with a deterministic local hidden variable theory (note that it can easily be shown that a probabilistic hidden variable theory is no stronger). Then for each value $\lambda$ of the local hidden variable, $a^\lambda, b^\lambda, c^\lambda, d^\lambda$ are all well defined and $+1$ or $-1$; the locality requirement resides in the fact that for example the value of $a^\lambda$ is independent of the choice of the measurement direction of Bob, and therefore should be attributed reality (in a non-local hidden variable theory, the products $(ac)^\lambda, (ad)^\lambda, (bc)^\lambda, (bd)^\lambda$ could be defined independently of each other). It is clear that the following relation holds:

$$|a^\lambda(c^\lambda + d^\lambda) + b^\lambda(c^\lambda - d^\lambda)|/2 = 1. \tag{141}$$

As we don't know the actual value of the hidden variable, we can only predict the average, and it follows that

$$| < a(c+d) + b(c-d) > |/2 \leq 1, \tag{142}$$

where we made use of the fact that the average of the absolute value exceeds the absolute value of the average. This is the CHSH inequality, and as shown this relation has to be fulfilled for all measurement statistics predicted by all local hidden variable theories.

Let us now consider the predictions of quantum mechanics concerning the experiment considered. The dichotomic observables corresponding to the (unit) directions $\vec{a}, \cdots$ are given by $a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3, \cdots$ with $\sigma_i$ the Pauli matrices. Given the density operator $\rho$, the expected value $< a(c+d) + b(c-d) > /2$ is of the form $\mathrm{Tr}(\mathcal{B}\rho)$ with

$$\mathcal{B} = \frac{1}{2} \sum_{ij=1}^{3} \left[a_i(c_j + d_j) + b_i(c_j - d_j)\right] \sigma_i \otimes \sigma_j, \tag{143}$$

where $(\vec{a}, \vec{b}, \vec{c}, \vec{d})$ are real unit vectors. In the case of a singlet for example, an appropriate choice of these directions will lead to a violation of the CHSH-bound: as will be shown later, the expected value can grow to $\sqrt{2}$ in this case. This means that not one local hidden variable theory is able to predict the same measurement statistics as quantum mechanics!

We would like to have a general way of determining whether a given (pure or mixed) state violates the CHSH bound. In [**125**]the Horodecki family showed that the maximal violation of the CHSH inequality can be calculated by considering the $3 \times 3$ matrix

$$\tilde{R}_{kl} = \mathrm{Tr}\left(\rho\sigma_k \otimes \sigma_l\right) \qquad 1 \leq k, l \leq 3$$

just as was the case in calculating the fidelity. We will give an alternative derivation of this result in a way that will be very useful in a later section:

**Lemma 8.** *(Horodecki [**125**]) Given the decreasingly ordered singular values $\{\sigma_i\}$ of $\tilde{R}$, then the maximal CHSH violation $\beta(\rho) = \max_{\mathcal{B}} \mathrm{Tr}\left(\rho\mathcal{B}\right)$ is given by $\sqrt{\sigma_1^2 + \sigma_2^2}$.*

*Proof:* Translated into the $R$-picture, calculating the maximal expected value of $\mathcal{B}$ under the constraint that $(\vec{a}, \vec{b}, \vec{c}, \vec{d})$ are real unit vectors, amounts to maximizing $\mathrm{Tr}\left(\tilde{R}X\right)$ with

$$X = \begin{pmatrix} \vec{c} & \vec{d} \end{pmatrix} \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \vec{a}^T \\ \vec{b}^T \end{pmatrix}. \tag{144}$$

It is an easy exercise to show that $X$ is a real $3 \times 3$ matrix, subjected to the only constraints that it be of rank 2 and that $\mathrm{Tr}\left(X^T X\right) = 1$. Standard linear algebra then dictates that $\mathrm{Tr}\left(\tilde{R}X\right)$ is maximized iff $X$ is chosen to be proportional to the best rank 2 least-squares approximation of the matrix $\tilde{R}$. In the basis where $\tilde{R}$ is diagonal ($\tilde{R} = \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)$), $X$ is therefore given by $X = \mathrm{diag}(\sigma_1, \sigma_2, 0)/\sqrt{\sigma_1^2 + \sigma_2^2}$, which immediately leads to $\beta = \sqrt{\sigma_1^2 + \sigma_2^2}$. $\square$

This Theorem immediately implies the following result: each pure entangled state of two qubits violates the CHSH inequality by an amount $\sqrt{1 + C^2} - 1$ where $C$ is the concurrence of the state (N. Gisin [**96**] was the first to observe that all pure entangled states violate a Bell inequality). The proof is immediate as the singular values of $\tilde{R}$ corresponding to a pure state with concurrence $C$ are given by $1, C, C$.

We would therefore expect that the amount of violation of the Bell inequality is a good measure of the amount of entanglement present in a state. In a seminal paper of Werner [**236**] however, where the concept of separability was introduced, it was shown that there exist states that do not violate any Bell inequality although they are entangled. We will give a complete classification of these two-qubit states in section 4.4.6.4. Following Popescu [**174**], Gisin [**97**] showed that there exist states that do not violate any CHSH inequality, but do violate it after an appropriate filtering equation. This implies that the amount of violation of the CHSH inequality cannot be an entanglement monotone. In section 4.4.7, we will give a complete classification of all two-qubit states that do violate a Bell inequality after an appropriate filtering equation.

It is clear that Bell diagonal states will play a prominent role in the case of CHSH inequalities: then $R$ and henceforth $\tilde{R}$ is diagonal, and it is thus straight forward to show that the maximal Bell violation is

$$\beta = \sqrt{2}\sqrt{(\lambda_2-\lambda_3)^2+(\lambda_1-\lambda_4)^2} \tag{145}$$

with $\{\lambda_i\}$ the eigenvalues of the Bell diagonal state. Since the concurrence is given by $C = \max[0, 2\lambda_1 - 1]$, the region of possible violations is in this case

$$\sqrt{2}(2C+1)/3 \leq \beta \leq \sqrt{1+C^2}, \tag{146}$$

where the lower bound is sharp for *Werner states* and the upper bound is attained for rank 2 Bell diagonal states and is equal to the relation for pure states; this is remarkable, as it says that the maximal amount of Bell violation for a given amount of entanglement can be obtained for mixed states.

Note that the Bell operator $\mathcal{B}$ in Eq.(143) is itself Bell diagonal due to the relation

$$\mathrm{Tr}\,(\mathcal{B}\sigma_i\otimes\sigma_0) = \mathrm{Tr}\,(\mathcal{B}\sigma_0\otimes\sigma_i) = 0. \tag{147}$$

In fact, Bell diagonal states exhibit a special property:

**Theorem 24.** *For any given spectrum of the density matrix, the respective Bell diagonal state $\rho$ maximizes the Bell violation, i.e. $\forall U \in U(4) : \beta(\rho) \geq \beta(U\rho U^*)$.*

*Proof:* First note that as we have to calculate a supremum over all unitary rotations of the state $\rho$, we can without loss of generality assume that the initial state commutes with the Bell operator $\mathcal{B}$. The proof of the Theorem is then based on the fact that if $u_{ik}$ are the matrix elements of a unitary matrix, then $|u_{ik}|^2$ is a doubly stochastic matrix, i.e., a convex combination of permutations $\tau$. If $\{\lambda_i\}, \{b_i\}$ are the decreasingly ordered eigenvalues of $\rho$ resp. $\mathcal{B}$, then

$$\begin{aligned} \mathrm{Tr}\,(U\rho U^*\mathcal{B}) &= \sum_{ik}\lambda_i b_k|u_{ik}|^2 = \sum_\tau p_\tau \sum_i \lambda_i b_{\tau(i)} \\ &\leq \sum_i \lambda_i b_i = \mathrm{Tr}(\rho\mathcal{B}). \end{aligned} \tag{148}$$

This immediately implies that if we fix any spectral property of the state, such as the purity $\mathrm{Tr}\left(\rho^2\right)$ or the entropy $-\mathrm{Tr}\,(\rho\log\rho)$, the maximal violation of the CHSH inequality will always be attained for Bell diagonal states. We refer to the section of maximally entangled mixed states for a more elaborate discussion.

### 4.4.6. A Comparison of Entanglement Measures on mixed states of two qubits.

It is natural to compare the introduced entanglement measures. This is of importance as different measures lead to different upper and lower bounds for e.g. the capacity of a quantum channel. It has already been proven that

$F(\rho) \leq N(\rho) \leq C(\rho)$ with equality iff the eigenvector corresponding to the negative eigenvalue of $\rho^{T_B}$ is maximally entangled. Lower bounds turn out to be much harder to prove.

4.4.6.1. *Concurrence versus negativity.*

Let us first compare the concurrence with the negativity:

**Theorem 25.** *The negativity $N$ of a mixed state with given concurrence $C$ is always smaller than $C$ with equality iff the eigenvector of $\rho^\Gamma$ corresponding to its negative eigenvalue is a Bell state (up to local unitary transformations). Moreover the negativity is always larger than $\sqrt{(1-C)^2+C^2}-(1-C)$, with equality iff the state is a rank 2 quasi-distillable state.*

*Proof:* For the fun of mathematics, we will also give an alternative proof for the upper bound. To prove the upper bound, we make use of the fact that a state with a given concurrence can always be decomposed as a convex sum of four pure states all having the same concurrence. It is readily checked that the negativity of a pure state is exactly equal to its concurrence. Due to linearity of the partial trace operation, the negativity of a mixed state is now obtained by calculating the smallest eigenvalue of the matrix obtained by making the convex sum of the partial transposes of the four pure states which have all one equal negative eigenvalue. It is a well-known result due to Weyl that the minimal eigenvalue of the sum of matrices always exceeds the sum of the minimal eigenvalues, which concludes the proof of the upper bound.

The lower bound is much harder to prove. To this end we heavily make use of the parameterization of the manifold of states with constant concurrence. It was shown before how the concurrence changes under the application of a SLOCC operation of the type

$$\rho' = \frac{(A\otimes B)\rho(A\otimes B)^\dagger}{\mathrm{Tr}\left((A\otimes B)\rho(A\otimes B)^\dagger\right)} \tag{149}$$

The transformation rule is:

$$C(\rho') = C(\rho)\frac{|\det A||\det B|}{\mathrm{Tr}\left((A\otimes B)\rho(A\otimes B)^\dagger\right)} \tag{150}$$

It is then straightforward to obtain the parameterization of the surface of constant concurrence (and hence constant entanglement of formation): it consists of applying all complex full rank $2\times 2$ matrices $A$ and $B$ on states with a given concurrence, under the constraint that

$$\mathrm{Tr}\left(\left(\frac{A^\dagger A}{|\det(A)|}\otimes\frac{B^\dagger B}{|\det B|}\right)\rho\right) = 1.$$

It is clear that we can restrict ourselves to matrices $A$ and $B$ having determinant 1 ($A, B \in SL(2, C)$), as will be done in the sequel.

The extremal values of the negativity can now be obtained in two steps: first find the state with extremal negativity for given Lorentz singular values by varying A and B , and then do an optimization over all these optimal states with equal concurrence.

The first step can be done by differentiating[11] the following cost function over the manifold of $A, B \in SL(2, C)$:

$$\begin{aligned}\Phi(A,B) &= \lambda_{min}\left(\left((A\otimes B)\rho(A\otimes B)^\dagger\right)^\Gamma\right) &(151)\\ &= \lambda_{min}\left((A\otimes B^*)\rho^\Gamma(A\otimes B^*)^\dagger\right) &(152)\end{aligned}$$

under the constraint

$$\mathrm{Tr}\left((A\otimes B^*)\rho^\Gamma(A\otimes B^*)^\dagger\right) = 1,$$

where the notation $\Gamma$ is used to denote partial transposition.

There exists a very elegant formalism for differentiating the eigenvalues of a matrix: given the eigenvalue decomposition of a Hermitian matrix $X = U\Lambda U^\dagger$, it is easy to proof that $\dot{\Lambda} = \mathrm{diag}(U^\dagger \dot{X} U)$, where 'diag' means the diagonal elements of a matrix. We can readily apply this to our Lagrange constrained problem. Indeed, the complete manifold of interest is generated by varying $A$ and $B$ as $\dot{A} = KA$ and $\dot{B} = LB$ with $K, L$ arbitrary complex 2x2 traceless matrices (the trace condition is necessary to keep the determinants constant). Moreover the minimal eigenvalue is given by $\mathrm{Tr}(\mathrm{diag}[0;0;0;1]D)$ where $D$ is the diagonal matrix containing the ordered eigenvalues of $C = PDP^\dagger = (A\otimes B^*)\rho^\Gamma_{BD}(A\otimes B^*)^\dagger$ and $P$ the eigenvectors of $C$. We proceed as

$$\begin{aligned}\dot{\Phi} &= \mathrm{Tr}\left(P^\dagger \dot{C} P\left(\underbrace{\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&1\end{pmatrix} - \mu I_4}_{=J(\mu)}\right)\right)\\ \dot{C} &= \left((K\otimes I_2)+(I_2\otimes L)\right)C + C\left((K^\dagger\otimes I_2)+(I_2\otimes L^\dagger)\right)\end{aligned}$$

where $\mu$ is the Lagrange multiplier. An extremum is obtained if $\dot{\Phi}$ vanishes for all possible traceless $K$ and $L$. Some straightforward algebra shows that this condition is fulfilled iff $CPJ(\mu)P^\dagger = P(DJ(\mu))P^\dagger$ is Bell diagonal (up to local unitary transformations).

Next we have to distinguish two cases, namely when the Lagrange multiplier $\mu = 0$ and $\mu \neq 0$. The first case leads to the condition that the eigenvector of

[11] Note that eigenvalues are differentiable whenever the multiplicity of the eigenvalue is equal to 1. As we have proven that in the case of the partial transpose of states of two qubits there is only one negative eigenvalue (Lemma 6) and due to the fact that filtering properties do not change the signature, no *crossings* between eigenvalues can occur and the negative eigenvalue is differentiable everywhere.

$\rho^\Gamma$ corresponding to the negative eigenvalue is a Bell state. It is indeed easily checked that all density matrices with this property have negativity equal to the concurrence, and this is clearly an extremal case. We have therefore again identified the class of states for which the negativity is equal to the concurrence.

The problem becomes much more subtle when the Lagrange multiplier does not vanish. We have already proven that the partial transpose of an entangled state is always full rank and has at most one negative eigenvalue. $P(DJ(\mu))P^\dagger$ will therefore be Bell diagonal either if the eigenvectors of $C$ are Bell states, or possibly if $DJ(\mu)$ contains eigenvalues with a multiplicity of 2: in this last case the two eigenvectors corresponding to the multiple eigenvalue are not uniquely defined and can be rotated to Bell states if the two other eigenvectors were already Bell states. As the first case was already treated in the previous paragraph, we concentrate on the second case. Denoting the eigenvalues of $C$ as $\lambda_1, \lambda_2, \lambda_3 > 0 > \lambda_4$, the eigenvector corresponding to $\lambda_4$ can be different from a Bell state iff we choose the Lagrange multiplier such that $-\mu\lambda_3 = (1-\mu)\lambda_4$. The eigenvectors corresponding to $\lambda_1$ and $\lambda_2$ can always be chosen to be Bell states. Therefore all states for which the eigenvectors of the partial transposes are, up to local unitary transformations, of the form

$$P = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \end{pmatrix} \begin{pmatrix} I_2 & 0 \\ 0 & U_2 \end{pmatrix}, \tag{153}$$

with $U_2$ an arbitrary 2x2 unitary matrix, will give extremal values of the negativity. The next step is therefore to find the state belonging to this class with minimal negativity for fixed concurrence, or equivalently the one with the largest concurrence for fixed negativity. Parameterizing the unitary $U$ as $\begin{pmatrix} a & -b \\ b^* & a^* \end{pmatrix}$, the class of states we are considering is parameterized as:

$$\begin{pmatrix} \frac{\lambda_1+\lambda_2}{2} & 0 & 0 & ab(\lambda_3-\lambda_4) \\ 0 & \lambda_3|a|^2+\lambda_4|b|^2 & \frac{\lambda_1-\lambda_2}{2} & 0 \\ 0 & \frac{\lambda_1-\lambda_2}{2} & \lambda_3|b|^2+\lambda_4|a|^2 & 0 \\ a^*b^*(\lambda_3-\lambda_4) & 0 & 0 & \frac{\lambda_1+\lambda_2}{2} \end{pmatrix}$$

The concurrence of this state can be calculated by finding the Cholesky decomposition of $\rho = XX^\dagger$ and calculating the singular values of $X^T(\sigma_y \otimes \sigma_y)X$. As

$\rho$ is a direct sum of two 2x2 matrices, this can be done exactly:

$$\begin{aligned}
\sigma_1 &= \frac{\lambda_1+\lambda_2}{2} + |ab|(\lambda_3-\lambda_4) && (154)\\
\sigma_3 &= \frac{\lambda_1+\lambda_2}{2} - |ab|(\lambda_3-\lambda_4) && (155)\\
\sigma_2 &= \sqrt{(\lambda_3|a|^2+\lambda_4|b|^2)(\lambda_3|b|^2+\lambda_4|a|^2)} + \frac{\lambda_1-\lambda_2}{2} && (156)\\
\sigma_4 &= \sqrt{(\lambda_3|a|^2+\lambda_4|b|^2)(\lambda_3|b|^2+\lambda_4|a|^2)} - \frac{\lambda_1-\lambda_2}{2} && (157)
\end{aligned}$$

The concurrence is therefore given by:

$$C = 2(\lambda_3-\lambda_4)|ab| - 2\sqrt{(\lambda_3|a|^2+\lambda_4|b|^2)(\lambda_3|b|^2+\lambda_4|a|^2)} \tag{158}$$

The task is now reduced to finding $a, b, \lambda_1, \lambda_2, \lambda_3$ such that $C$ is maximized for fixed $\lambda_4$. Some long but straightforward calculations lead to the optimal solution:

$$\begin{aligned}
|a|^2 &= 1-|b|^2 = \frac{\lambda_3}{|\lambda_4|} && (159)\\
\lambda_1 &= \lambda_2 = \sqrt{\lambda_3|\lambda_4|} && (160)\\
1 &= \lambda_1+\lambda_2+\lambda_3+\lambda_4 && (161)
\end{aligned}$$

This solution corresponds to a state with two vanishing eigenvalues, while the remaining two eigenvectors are a Bell state and a separable state orthogonal to it:

$$\rho = \begin{pmatrix} C/2 & 0 & 0 & C/2 \\ 0 & 1-C & 0 & 0 \\ 0 & 0 & 0 & 0 \\ C/2 & 0 & 0 & C/2 \end{pmatrix} \tag{162}$$

The concurrence $C$ is then related to the negativity $N = 2|\lambda_4|$ by the equation

$$N^2 + 2N(1-C) - C^2 = 0. \tag{163}$$

This equation defines the lower bound we were looking for, as it relates the minimal possible value of the negativity for given concurrence. □

The family of states minimizing the negativity for given concurrence turn out to be exactly the quasi-distillable ones of rank 2. A scatter plot of the negativity versus the concurrence for all entangled mixed states of two qubits is shown in Figure 4.

4.4.6.2. *Entanglement of formation versus Relative Entropy of Entanglement.*

A similar analysis can be performed to compare the entropic entanglement measures entanglement of formation and the relative entropy of entanglement. It is well-known that they coincide for pure states, and that the relative entropy of entanglement can never exceed the entanglement of formation. Due

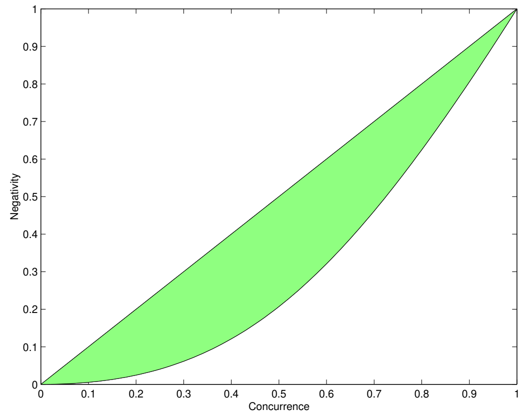



**Figure 4.** Range of values of the negativity for given concurrence.

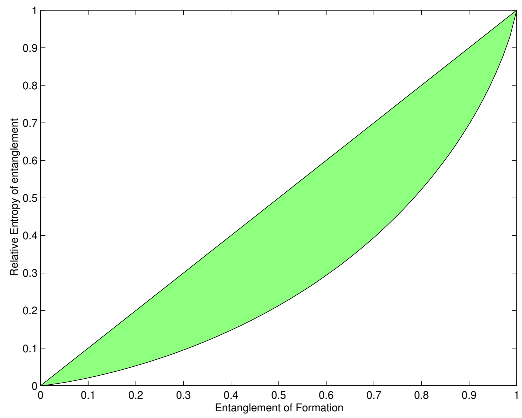



**Figure 5.** Range of values of the Relative Entropy of Entanglement for given Entanglement of formation.

to the logarithmic nature of these quantities however, finding the states with minimal relative entropy of entanglement for given entanglement of formation is very hard to do analytically. Numerical investigations however showed that again the same quasi-distillable rank 2 states minimize the relative entropy of entanglement. It is indeed possible to show that these states are local minima to the optimization problem. Using the results of section 4.7 on maximally entangled mixed states of two qubits, this minimal value is then given by:

$$E_R(\rho) = (C-2)\log(1-C/2) + (1-C)\log(1-C). \qquad (164)$$

A scatter plot of the range of values of the relative entropy of entanglement is given in Figure 5.

4.4.6.3. *Concurrence versus Fidelity.*

Let us now compare the fidelity with the concurrence, for which the upper bound was already proven in Lemma 7. Proving a lower bound turned out to be very hard. A central lemma needed is the following:

**Lemma 9.** *Consider the density operator* $\rho$ *and the real* $3 \times 3$ *matrix* $\tilde{R}$ *with coefficients* $\tilde{R}_{ij} = \mathrm{Tr}\,(\rho\sigma_i \otimes \sigma_j)$ *with* $1 \le i, j \le 3$. *Then* $\rho$ *is as a convex sum (i.e. mixture) of rank* 2 *density operators all having exactly the same coefficients* $\tilde{R}_{ij}$.

Proof: Consider the real $4 \times 4$ matrix $R$ with coefficients $R_{\alpha\beta} = \mathrm{Tr}\,(\rho\sigma_\alpha \otimes \sigma_\beta)$, parameterized as

$$R = \left( \begin{array}{cccc} 1 & x_1 & x_2 & x_3 \\ y_1 & & & \\ y_2 & & \tilde{R} & \\ y_3 & & & \end{array} \right).$$

If $\rho$ is full rank, then a small perturbation on the values $\{x_i\}, \{y_i\}$ will still yield a full rank density operator. Consider a perturbation on $x_1' = x_1 + \epsilon$ and the corresponding $\rho'$. As the set of density operators is compact, there will exist a lower bound $lb < 0$ and an upper bound $ub > 0$ such that $\rho'$ is positive iff $lb < \epsilon < ub$. Call $\rho_{lb}$, $\rho_{ub}$ the rank three density operator obtained when $\epsilon = lb$ and $\epsilon = ub$ respectively. It is easy to see that $\rho = (ub\rho_{lb} + lb\rho_{ub})/(lb + ub)$, such that it is proven that a rank four density operator can always be written as a convex sum of two rank three density operators with the same corresponding $\tilde{R}$.

Consider now $\rho$ rank three and its associated "square root"$\rho = XX^\dagger$ with $X$ a $4 \times 3$ matrix. A small perturbation of the form $\rho' = \rho + \epsilon XQX^\dagger$, with $Q$ an arbitrary Hermitian $3 \times 3$ matrix $Q = \sum_{i=1}^{9} q_i G_i$ and $G_i$ generators of U(3), will still yield a state of rank three. Moreover, there always exists a non-trivial $Q$ such that $\tilde{R}$ is left unchanged by this perturbation. This is indeed the case if the following set of equations is fulfilled:

$$\sum_i q_i \mathrm{Tr}\left(G_i X^\dagger (\sigma_\alpha \otimes \sigma_\beta) X\right) = 0$$

for $(\alpha, \beta) = (0, 0)$ and $\alpha, \beta \ge 1$. It can easily be verified that this set of 10 equations only contains at most 8 independent ones irrespective of the $4 \times 3$ matrix $X$ , and as $Q$ has nine independent parameters there always exists at least one non-trivial solution to this set of homogeneous equations. A similar reasoning as in the full rank case then implies that one can always tune $\epsilon$ such that $\rho$ can be written as a convex sum of two rank two density operators with the same $\tilde{R}$, which concludes the proof. □

This Lemma is interesting if one wants to maximize a convex measure of a density operator (such as the entropy or an entanglement monotone) under the constraint that the fidelity is fixed: indeed, the fidelity is only a function of $\tilde{R}$, and by the previous Lemma we immediately know that states with maximal entropy for given fidelity will have rank two. Note that exactly the same reasoning applies when one wants to maximize a convex measure under the constraint that the CHSH Bell-violation is fixed, as this CHSH Bell-violation is also solely a function of $\tilde{R}$.

We are now ready to prove a tight lower bound on the fidelity:

**Theorem 26.** *Given a mixed state of two qubits $\rho$ with concurrence equal to $C$, then a tight bound for its fidelity $F$ is given by:*

$$\max\left(\frac{1+C}{4}, C\right) \leq F \leq N \leq C.$$

*Proof:* A direct consequence of Lemma (9) is that to find states with minimal fidelity for given concurrence (i.e. maximal concurrence for given fidelity), it is sufficient to look at states of rank two. Consider therefore a rank 2 state $\rho$ and associated to it the real $4\times4$ matrix $R$ with coefficients $R_{\alpha\beta} = \mathrm{Tr}\,(\sigma_\alpha \otimes \sigma_\beta \rho)$. If $R$ is multiplied right and left by proper orthochronous Lorentz transformations leaving the $(0,0)$-element equal to 1, then a new state is obtained with the same concurrence. Moreover the fidelity of a state $\rho$ is variationally defined as

$$F(\rho) = \max_{O_A, O_B \in SO(3)} \mathrm{Tr}\left(M \begin{pmatrix} 1 & 0 \\ 0 & O_A \end{pmatrix} R \begin{pmatrix} 1 & 0 \\ 0 & O_B^T \end{pmatrix}\right)$$

with $M = \mathrm{diag}(1,-1,-1,-1)$ ($M$ is the representation of the singlet in the R-picture). The minimal fidelity for given concurrence can therefore be obtained by minimizing the following constrained cost-function over all proper orthochronous Lorentz transformations $L_1, L_2$:

$$K = \mathrm{Tr}\left(ML_1RL_2^T\right) - \lambda \mathrm{Tr}\left(L_1RL_2^T \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}\right).$$

Note that $\lambda$ is a Lagrange constraint. Without loss of generality we can assume that the lower $3\times3$ block $\tilde{R}$ of $R$ is diagonal and of the form $\tilde{R} = \mathrm{diag}(-|s_1|, -|s_2|, -s_3)$ with $|s_1| \geq |s_2| \geq |s_3|$, as this is precisely the form needed for maximizing the fidelity over all local unitary operations. The cost-function $K$ can be differentiated[12] over $L_1, L_2$ by introducing the generators of the Lorentz group, and this immediately yields the optimality conditions $(\lambda = 0, MRM = R^T)$ or $(\lambda = 2, R = R^T)$. Note however that the above argument breaks down in the case that $|s_2| = -s_3$. Indeed, the fidelity cannot

[12] Note that we have to take into account that it is not guaranteed that the minimum of the fidelity occurs at a point where the derivative is equal to zero: this is a problem inherent in min-max problems. A sufficient condition for optimality is however obtained if we can show that the extremum derived is indeed an extremum of the maximal value, as will be the case in the following derivation.

be differentiated in this case as for example a perturbation of $s_3$ of the form $s_3' = s_3 + \epsilon$ always leads to a perturbation of the fidelity $F' = F + |\epsilon|$. In this case the conditions $x_2 = y_2, x_3 = y_3$ or $x_2 = -y_2, x_3 = -y_3$ vanish, and if also $|s_1| = |s_2| = -s_3$ there are no optimality conditions on $\{x_i, y_i\}$ left.
Let us first treat the case with $R$ symmetric and $s_1 \geq s_2 \geq |s_3|$:

$$R = \begin{pmatrix} 1 & x_1 & x_2 & x_3 \\ x_1 & -s_1 & 0 & 0 \\ x_2 & 0 & -s_2 & 0 \\ x_3 & 0 & 0 & -s_3 \end{pmatrix}.$$

The condition that $\rho$ corresponding to this state is rank 2 implies that all $3 \times 3$ minors of $\rho$ are equal to zero. Due to the conditions $s_1 \geq s_2 \geq |s_3|$, it can easily be shown that a state of rank 2 (and not of rank 1!) is obtained iff $x_1 = 0 = x_2$ and $x_3 = \pm\sqrt{(1-s_1)(1-s_2)}$ and $1 - s_1 - s_2 + s_3 = 0$. In this case the concurrence is equal to $C = s_2$ and the fidelity is given by $F = (s_1 + s_2)/2$, and the constraints become $1 \geq s_1 \geq s_2 \geq (1 - s_1)/2$ what implies that $C \geq 1/3$. The minimal fidelity for given concurrence occurs when $s_1 = s_2$ and then $C = F$ which gives the lower bound of the Theorem in the case of $C \geq 1/3$.
Let us now consider the case where $R = MR^TM$:

$$R = \begin{pmatrix} 1 & x_1 & x_2 & x_3 \\ -x_1 & -s_1 & 0 & 0 \\ -x_2 & 0 & -s_2 & 0 \\ -x_3 & 0 & 0 & -s_3 \end{pmatrix}$$

with again $s_1 \geq s_2 \geq |s_3|$. Let us first note that, due to the symmetry, $R$ has a Lorentz singular value decomposition [**211**] of the form $R = L_1 \Sigma \tilde{M} M L_1^T M$ with $\Sigma$ of the form $\mathrm{diag}(|\sigma_0|, -|\sigma_1|, -|\sigma_2|, -|\sigma_3|)$ and $\tilde{M}$ of the form $\mathrm{diag}(1,1,1,1)$ or $\mathrm{diag}(1,-1,-1,1)$ or $\mathrm{diag}(1,-1,1,-1)$ or $\mathrm{diag}(1,1,-1,-1)$. It follows that $\mathrm{Tr}\,(R) = \mathrm{Tr}\left(\Sigma\tilde{M}\right)$, and due to the ordering of the Lorentz singular values, $\tilde{M}$ has to be equal to the identity if $\mathrm{Tr}\,(R) \leq 0$. But $\mathrm{Tr}\,(\Sigma)$ is just $-2C$ with $C$ the concurrence of the state, and $\mathrm{Tr}\,(R) = 2 - 4F$ with $F$ the fidelity of the state. Therefore it holds that $F = (1 + C)/2$ if $\mathrm{Tr}\,(R) \leq 0$ which corresponds to the upper bound of the fidelity. Therefore only the case where $\mathrm{Tr}\,(R) > 0$ has to be considered for finding lower bounds of the fidelity. The condition that the state be rank 2 (and not rank 1) immediately yields: $x_3 = 0$, $s_1 + s_2 - s_3 = 1$ and $s_1 + s_2 = x_1^2/(1 - s_2) + x_2^2/(1 - s_1)$. If we only consider the case with $\mathrm{Tr}\,(R) > 0$, it holds that $s_3 < 0$ and the inequality constraints become $(1 - s_1)/2 \leq s_2 \leq (1 - s_1) \leq 2/3$. The concurrence can again be calculated analytically and is given by $C = (1 - s_1 - s_2 - s_3)/2$, and it follows that $F = (1 - C)/2$. Note that the inequality constraints limit $C$ to be in the interval $C \in \{0, 1/3\}$, and so this bound is less stringent than the one stated in the theorem.

Let us now move to the degenerate case where $s_1 > s_2 = -s_3$:

$$R = \begin{pmatrix} 1 & x_1 & x_2 & x_3 \\ y_1 & -s_1 & 0 & 0 \\ y_2 & 0 & -s_2 & 0 \\ y_3 & 0 & 0 & s_2 \end{pmatrix}.$$

As $s_1 > s_2$, optimality requires $x_1 = \pm y_1$. We first treat the case $x_1 = y_1$. Defining $\alpha = x_3/y_3$, a set of necessary and sufficient conditions for being rank 2 is given by:

$$\begin{aligned} 0 &= x_1 = y_1 \\ 0 &= x_2 + \alpha y_2 \\ 0 &= \alpha^2 - \alpha\frac{1-s_1}{s_2} + 1 \\ 0 &= (x_2^2 + x_3^2) - \alpha s_2(1+s_1). \end{aligned}$$

Under these conditions the concurrence can again be calculated exactly and is given by $C = s_2$, while the fidelity is given by $F = (1+s_1)/4$. Note that the above set of equations only has a solution if $(1-s_1)/2 \geq s_2$, implying that $C \leq 1/3$. The fidelity will now be minimal when $s_2 = s_1$, and then $F = (1+C)/4$ which is the second bound stated in the theorem.
Let us now consider the degenerate case with $s_1 > s_2 = -s_3$ but $x_1 = -y_1$. The rank 2 condition implies that $s_1 + 2s_2 = 1$ and $x_2 = -y_2$ and $x_3 = y_3$. Some straightforward algebra leads to the condition

$$4\frac{1-s_1}{1+s_1}x_1^2 + 1 - s_1^2 - 2x_2^2 - 2x_3^2 = 0.$$

Taking into account the constraints, the concurrence is again given by $C = s_2 = (1-s_1)/2$ and bounded above by $1/3$, while the fidelity if equal to $F = (1+s_1)/4 = (1-C)/2$. This bound always exceeds the previously derived bound $F \geq (1+C)/4$ for $C \leq 1/3$, and is therefore useless.
It only remains to consider the case where $s_1 = s_2 = -s_3$:

$$R = \begin{pmatrix} 1 & x_1 & x_2 & x_3 \\ y_1 & -s_1 & 0 & 0 \\ y_2 & 0 & -s_1 & 0 \\ y_3 & 0 & 0 & s_1 \end{pmatrix}.$$

Defining $\alpha = x_1/y_1$, the rank 2 constraint leads to the following set of necessary and sufficient conditions:

$$\begin{aligned} 0 &= x_2 - \alpha y_2 \\ 0 &= x_3 + \alpha y_3 \\ 0 &= s_1\alpha^2 + \alpha(1-s_1) + s_1 \\ 0 &= \alpha(x_1^2 + x_2^2 + x_3^2) + s_1(1+s_1). \end{aligned}$$

The inequality constraint reads $s_1 \leq 1/3$, and the concurrence can again be calculated exactly and is given by $C = s_1$. Therefore the fidelity of these

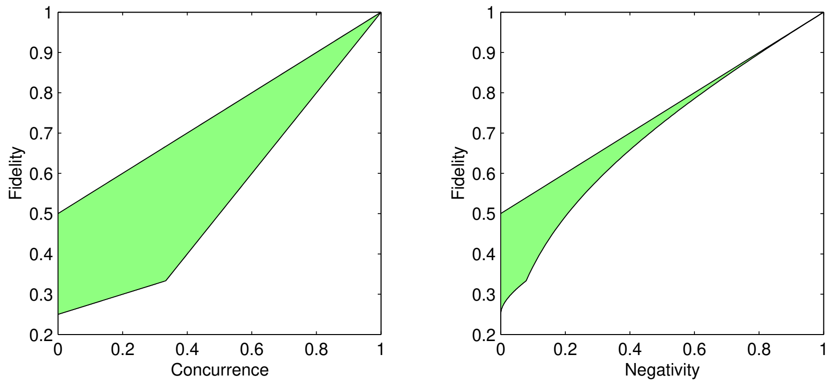



**Figure 6.** Range of values of the fidelity for given concurrence and negativity.

states obeys the relation $F = (1+C)/4$ for $C \le 1/3$, which is the sharp lower bound. $\square$

It might be interesting to note that all rank 2 states minimizing the fidelity for given concurrence are quasi-distillable [**124, 211**] and have one separable and one entangled eigenvector. More specifically, the states minimizing the fidelity for $C \le 1/3$ are, up to local unitaries, of the form

$$\rho = \begin{pmatrix} \frac{1+C}{2} & 0 & 0 & 0 \\ 0 & \frac{1-C+\sqrt{1-2C-3C^2}}{4} & -\frac{C}{2} & 0 \\ 0 & -\frac{C}{2} & \frac{1-C-\sqrt{1-2C-3C^2}}{4} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

and those for $C \ge 1/3$ of the form

$$\rho = \begin{pmatrix} 1-C & 0 & 0 & 0 \\ 0 & C/2 & -C/2 & 0 \\ 0 & -C/2 & C/2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Extensive numerical investigations revealed that exactly the same states also minimize the fidelity for given negativity. This leads to the following sharp bounds for the fidelity versus negativity:

$$\begin{aligned} F &\ge \frac{1}{4} + \frac{1}{8}\left(N + \sqrt{5N^2+4N}\right) \\ F &\ge \sqrt{2N(N+1)} - N \\ F &\le \frac{1+N}{2}. \end{aligned}$$

The first condition applies when $N \le (\sqrt{5}-2)/3$ and the second when $N \ge (\sqrt{5}-2)/3$. A plot of these bounds is given in Figure (6). One observes that the difference between the lower bound and the upper bound in terms of the negativity becomes very small ($\simeq \epsilon^2/16$) for large negativity $N = 1 - \epsilon$. Moreover the fidelity is always larger than 1/2 if the negativity exceeds $(\sqrt{2} - 1)/2$. The discontinuity in the lower bound can be understood from the fact that at this point the contribution to the fidelity of the separable part becomes larger then the contribution of the entangled fraction. It is indeed the case that, unlike the concurrence for example, the fidelity takes also into account the amount of classical correlations.

4.4.6.4. *Concurrence versus CHSH violation.*

Let us at last compare the amount of violation of the CHSH inequality with the concurrence.

In the following we will derive the extremal violations for a given amount of entanglement plotted in Fig.7.

**Theorem 27.** *The maximal violation of the CHSH inequality for given concurrence $C$ is $\beta(\rho) = \sqrt{1 + C^2(\rho)}$.*

*Proof:* As shown by Wootters [**244**], it is possible to decompose a mixed state of two qubits $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ into a convex sum of pure states, all with concurrence equal to the concurrence of the mixed state. Since the extremal violation is moreover a convex function, i.e., $\max_{\mathcal{B}} \mathrm{Tr}\,(\rho\mathcal{B}) \leq \sum_i p_i \max_{\mathcal{B}} \langle\psi_i|\mathcal{B}|\psi_i\rangle$, it is sufficient to have a look at pure states, which can always be written in their Schmidt form as $|\psi\rangle = \lambda_+|00\rangle + \lambda_-|11\rangle$ with $\lambda_\pm = (\sqrt{1+C} \pm \sqrt{1-C})/2$. The corresponding $\tilde{R}$-matrix is diagonal with singular values $(1, C, C)$ leading to $\beta = \sqrt{1 + C^2}$. □

It is interesting to note that there also exist mixed states of rank 2 for which the violation is as strong as for pure states. These are, up to local unitary operations, all of the form

$$\rho = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1-a & C & 0 \\ 0 & C & 1+a & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{165}$$

with $C$ being the concurrence and $a$ a free real parameter constrained by $|a| \leq \sqrt{1 - C^2}$, where equality leads to pure states and *Bell diagonal states* (see section below) are obtained for $a = 0$.

**Theorem 28.** *The minimal violation of the CHSH inequality for given concurrence $C$ is given by $\beta(\rho) = \max[1, \sqrt{2}C(\rho)]$.*

*Proof:* We will use similar techniques as used in the previous proofs, where it was shown that surfaces of constant concurrence can be generated by transforming $R \mapsto R' = L_1 R L_2^T$ by left and right multiplication with proper orthochronous Lorentz transformations, taken into account the constraint that the $(0,0)$ element of $R$ (representing the trace of $\rho$) does not change under these transformations. They leave the Lorentz singular values invariant, and the concurrence is a function of these four parameters only.

Using the variational characterization used in lemma 8, the first step consists of varying the Lorentz transformations $L_1, L_2$ and the $3 \times 3$ rank 2 matrix $X$ (with constraint $\mathrm{Tr}\left(X^T X\right) = 1$), and imposing that these variations be zero (i.e. we have an extremum). The object function is given by

$$\mathrm{Tr}\left(L_1 R L_2^T \left(\begin{array}{cc} 0 & 0 \\ 0 & X \end{array}\right)\right) \tag{166}$$

under the constraints $\mathrm{Tr}\left(X^T X\right) = 1$ and

$$\mathrm{Tr}\left(L_1 R L_2^T \left(\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)\right) = 1. \tag{167}$$

The orthogonal degrees of freedom of $X$ can be absorbed into $L_1, L_2$, such as to yield a diagonal $X$ of rank 2: $X = \mathrm{diag}(q, r, 0)$ with $q^2 + r^2 = 1$. Variation of the Lorentz transformations yields the extremal conditions

$$\mathrm{Tr}\left(G_1 R' \left(\begin{array}{cc} \lambda & 0 \\ 0 & X \end{array}\right)\right) = \mathrm{Tr}\left(R' G_2 \left(\begin{array}{cc} \lambda & 0 \\ 0 & X \end{array}\right)\right) = 0. \tag{168}$$

for all possible generators $G_1, G_2$ of the Lorentz group and $\lambda$ being a Lagrange parameter. The generators are all of the form

$$G = \left(\begin{array}{cc} 0 & \vec{v} \\ \vec{v}^T & A \end{array}\right) \tag{169}$$

with $\vec{v} \in \mathcal{R}^3$ and $A$ a real and antisymmetric $3 \times 3$ block. A detailed discussion of the case $\lambda \neq 0$ along the lines of the proof of Theorem 26 shows, that this leads to suboptimal solutions. The minimal value of the Bell violation turns out to correspond to the case where $\lambda = 0$ and yields the condition that $R'$ is of form

$$R' = \left(\begin{array}{cccc} 1 & 0 & 0 & a \\ 0 & x & 0 & 0 \\ 0 & 0 & y & 0 \\ b & 0 & 0 & z \end{array}\right). \tag{170}$$

The extremal violation of the Bell inequality is then directly found by varying the remaining diagonal elements of $X$, leading to a violation given by $\sqrt{x^2 + y^2}$. The concurrence of the extremal state can be calculated explicitly, and is given

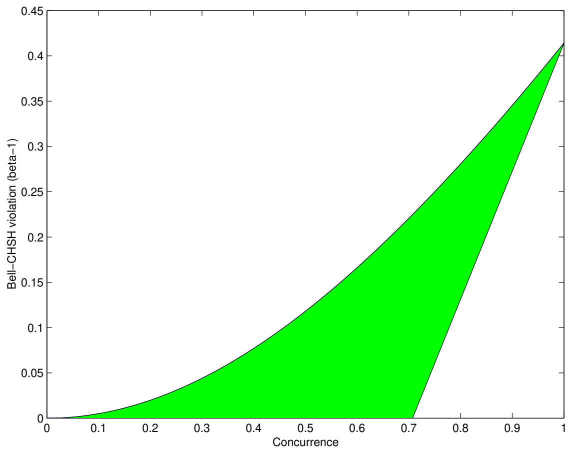



**Figure 7.** The region of possible maximal Bell violation for given concurrence.

by:

$$C = \frac{1}{2}\max\left[0, |x-y| - \sqrt{(1-z)^2-(a-b)^2},\right.$$
$$\left. |x+y| - \sqrt{(1+z)^2-(a+b)^2}\right]. \quad (171)$$

The constraints that $R$ corresponds to a (positive) state are expressed by the inequalities

$$-1 \leq \quad z \quad \leq 1, \quad (172)$$
$$(1+z)^2-(a+b)^2 \geq (x-y)^2, \quad (173)$$
$$(1-z)^2-(a-b)^2 \geq (x+y)^2. \quad (174)$$

Applying these to the expression of the concurrence, this immediately leads to the sharp inequality $C \leq \min(|x|,|y|)$. The Bell violation, given by $\beta = \sqrt{x^2+y^2}$, will then be minimal for given concurrence if $|x| = |y|$, leading to final result: $\beta(\rho) \geq \sqrt{2}C(\rho)$. To complete the proof, we still have to check if there indeed exists a state with the properties that $x = y$, $(1+z)^2 = (a+b)^2$, $(1-z)^2-(a-b)^2 \geq (x+y)^2$,$-1 \leq z \leq 1$ and $|z| \leq |x|$. Choosing for example $a = b = (1+z)/2$ and $z = -|x|$ indeed leads to a possible result, which is a convex combination of a maximally entangled and an orthogonal separable pure state. Note that all parameters fulfilling the above constraints lead to states with the same minimal possible amount of $\beta$ for given concurrence. □

The states minimizing the Bell violation for given entanglement of formation are all rank 2 and are exactly the states minimizing the negativity, the RelEnt and the fidelity for given amount of concurrence (or entanglement of formation). Figure 7 depicts the derived bounds. This is the first time that quantitative relations between the amount of entanglement present in the a mixed system and the amount of Bell violation have been derived: as expected, the larger the amount of entanglement, the larger the amount of Bell violation typically is. This shows that the folklore of identifying a large Bell violation with a large amount of entanglement is indeed reasonable.

### 4.4.7. Optimal filtering

An interesting question that arises is the following: what local filtering operations do I have to implement such as to get, with a certain probability, a state with the maximal possible amount of entanglement. This question is important in an experimental setup involving e.g. photons, as it is in general very hard to implement joint operations on multiple qubits such as required in general distillation schemes, but very simple to implement a local filter (using e.g. a polarization filter).

The following Theorem holds:

**Theorem 29.** *The entanglement of formation, concurrence, negativity, fidelity and the CHSH-violation of a mixed state $\rho$ are all maximized under the same local filtering (SLOCC) operations bringing $\rho$ into its unique Bell diagonal normal form.*

*Proof:* In the case of concurrence and the entanglement of formation, the proof is a direct consequence of theorem 10 which said that every entanglement monotone defined as the convex roof of an entanglement monotone that is invariant under determinant 1 SLOCC operations, is maximized by bringing it into local stochastic form. As the negativity and the fidelity are bounded above by the concurrence, and become equal to it in the case of a Bell diagonal state, we also have completed that part of the proof. The CHSH-case remains to be proven. This can be proven using techniques completely similar to the ones used in the proof of Theorem 28, so we will only repeat the major steps. In the $R$-picture, filtering operations correspond to left and right multiplication with Lorentz transformations, followed by renormalization. The function, which we have to maximize with respect to $L_1, L_2$ and $X = \mathrm{diag}(q, r, 0)$ in order to obtain the maximal Bell violation, therefore becomes

$$\mathrm{Tr}\left(\frac{L_1 R L_2^T}{\left(L_1 R L_2^T\right)_{00}} \begin{pmatrix} 0 & 0 \\ 0 & X \end{pmatrix}\right), \tag{175}$$

with the constraint $q^2 + r^2 = 1$ and the normalization factor $(L_1 R L_2^T)_{00}$. Variation leads to the condition

$$\mathrm{Tr}\left(G_1 R' \begin{pmatrix} -\beta & 0 \\ 0 & X \end{pmatrix}\right) = \mathrm{Tr}\left(R' G_2 \begin{pmatrix} -\beta & 0 \\ 0 & X \end{pmatrix}\right) = 0,$$

where again this has to hold for arbitrary $G_1, G_2$, and where $\beta$ is equal to Eq. (175), i.e., the Bell expectation value for given $q, r, L_1, L_2$. If $\beta > 1$ (i.e. Bell violation), it holds that $\beta$ cannot be equal to $|q|$ or $|r|$, and the form of the generators in Eq.(169) implies that the above equations can only be satisfied iff $R'$ is Bell diagonal. □

The above Theorem stresses once more the particular physical significance of the diagonal normal form obtained by the Lorentz singular value decomposition: reversibly washing out the local information by making the local density operators proportional to the identity is indeed the optimal way of maximizing the entanglement present in the system. The higher the entanglement, the larger the local disorder for two states out of the same SLOCC class.

The previous Theorem is also of historical importance as it finally characterizes all entangled states of two qubits that do not violate any CHSH inequality even after a filtering operation: you just have to look at the Bell-diagonal normal form of the state.

## 4.5. Optimal teleportation with a mixed state of two qubits

The concept of teleportation plays a crucial role in the field of quantum information theory (QIT). Indeed, the success of QIT stems from the fact that we have the extra resource of entanglement to our disposition. And using entanglement and teleportation, it is possible to implement all possible non-local quantum operations using only classical communication and local measurements. This is clearly extremely powerful, as it is well known that e.g. global quantum operations on a quantum computer can process information exponentially faster than is possible classically. There also exist quantum communication protocols known with an exponential speed-up compared to their classical counterparts.

In a realistic setting however, maximally entangled states as needed for perfect teleportation do not exist. If collective operations on large blocks of states can be implemented easily, then one could use a distillation protocol to produce states with a very high fidelity. However, collective measurements are typically very difficult to implement, and it is therefore very reasonable to ask whether the quality of teleportation can be increased by doing a LOCC preprocessing on the state used to teleport.

The quality of teleportation is measured by its teleportation fidelity introduced in section 4.4.4. This teleportation fidelity was a simple linear function of the fidelity of a mixed state, and as this last one is not an entanglement monotone, the teleportation fidelity isn't neither. This is good news, as it tells us that typically the quality of teleportation can be enhanced by some LOCC preprocessing.

In section 4.4.6.3, we have proven that the fidelity of an entangled mixed state of two qubits can be smaller than 1/2. This means that this state would yield a teleportation fidelity that is worse than the one that can be achieved with a separable pure state, and this suggests of course that one should be able to do better. We will indeed prove that a mixed state of two qubits is entangled iff it can yield a teleportation fidelity larger than 1/2 by allowing LOCC preprocessing. This finally answers the following open question by Badziag and Horodecki [**12**] in the case of two qubits: "Can any entangled state provide better than classical fidelity of teleportation?"

In the previous section, the optimal local filtering operations such as to maximize the fidelity of the filtered state were derived. The optimal filter is the one that transforms the state into Bell diagonal form, and the fidelity of this state always exceeds 1/2 if the original one was entangled. This was expected in the light of the work of Horodecki [**121**] where it was shown that there always exist a filter such that the filtered state has fidelity larger than 1/2. The optimality of the derived filter is of great interest in devising optimal distillation protocols.

The drawback of filtering operations is that these operations can only be implemented with a certain probability. It is therefore an interesting question whether trace preserving local operations can also enhance the fidelity. In a surprising paper of Badziag et al. [**12**], it was shown that there exist mixed states with fidelity smaller than 1/2, for which local protocols exist that do not require any communication between Alice and Bob (LO), and that transform this state into a state with fidelity larger than 1/2. Motivated by this example, we looked for the optimal LOCC protocols such as to transform an entangled state into one with fidelity as large as possible allowing classical communication. We will prove that the optimal trace-preserving protocol for maximizing the fidelity of a given state always belongs to a very simple class of 1-LOCC operations, and provide a constructive way of obtaining this optimal (state-dependent) LOCC operation. We conclude by giving a geometrical interpretation of the maximum achievable fidelity by LOCC, and show how the derived result yields the optimal protocol for teleportation with mixed states of two qubits.

The central question is whether there always exist trace-preserving local operations such that the fidelity of the obtained state exceeds 1/2 if the original state is entangled. The crucial point is to incorporate the previously described filtering operation as part of a trace preserving LOCC operation. The idea is that it is always possible to make a trace-preserving LOCC operation out of a SLOCC filtering operation by making a pure separable state if the state did not pass the filter. Then with a certain probability a Bell diagonal state $\rho_f$ arises, and with the complementary probability a pure separable state $|\chi\rangle$ (note that $|\chi\rangle$ must be chosen such that $|\langle\chi|\psi\rangle|^2 = 1/2$ with $|\psi\rangle$ the maximally entangled state obeying $F(\rho_f) = \langle\psi|\rho_f|\psi\rangle$). Using the filter proposed by Horodecki [**121**], it is even sufficient that only one party implements the filter. This proves that for each entangled mixed state of two qubits there exists a trace-preserving

1-LOCC protocol that transforms it into a state with fidelity larger than 1/2.

Let us now try to optimize the trace-preserving operation used in the protocol just described such as to maximize the fidelity of a given state. Note that in general the optimal filter bringing the state into Bell diagonal form will not be optimal in the trace-preserving setting as in that case the probability of obtaining the state was not taken into account. The setting is now as follows: we want to find the filter, such that the probability of success $p_{AB}$ of the filter multiplied by the fidelity $F$ of the state coming out of this filter, plus $(1-p_{AB})$ times the fidelity of the pure separable state given by 1/2, is maximal. For a given filter $-I \leq A, B \leq I$, the cost-function $K_{AB}$ is therefore given by

$$K_{AB} = p_{AB}F(\rho_f) + \frac{1-p_{AB}}{2}$$

where

$$\begin{aligned} p_{AB} &= \mathrm{Tr}\left((A\otimes B)\rho(A\otimes B)^\dagger\right) \\ \rho_f &= \frac{(A\otimes B)\rho(A\otimes B)^\dagger}{p_{AB}} \end{aligned}$$

Now some tricks will be applied. Due to the presence of $A, B$, we can replace $F(\rho_f)$ by $\langle\psi|\rho_f|\psi\rangle$ with $|\psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$, and we use the fact that the trace of the product of two matrices is equal to the trace of the product of the partial transpose of two matrices. This leads to the following expression:

$$K_{AB} = \frac{1}{2} - \langle\psi|(B^\dagger A\otimes I)\rho^\Gamma(A^\dagger B\otimes I)|\psi\rangle. \tag{176}$$

This cost-function has to be maximized over all complex $2\times 2$ matrices $-I \leq A, B \leq I$, and this leads to a lower bound on the maximum achievable fidelity by LOCC operations. Note that the above expression implies that the optimal LOCC operations will be 1-LOCC: if $-I \leq A, B \leq I$ then certainly $-I \leq B^\dagger A \leq I$ and it is therefore sufficient for one party to implement a filter.

Let us next try to obtain an upper bound. In the light of the inequalities derived in the previous section, it is very easy to obtain an upper bound in terms of the negativity. Indeed, unlike the fidelity, the negativity is an entanglement monotone and cannot increase under LOCC operations. If we define $F^*(\rho)$ the maximal LOCC achievable fidelity of $\rho$, this leads to the following upper bound that was also obtained by Vidal and Werner [**229**]:

$$F^*(\rho) \leq \frac{1+N(\rho)}{2} \leq \frac{1+C(\rho)}{2}.$$

Note that this implies that LOCC operations cannot be used to enhance the fidelity for all states for which $F = (1+N)/2$. These are the states for which the eigenvector corresponding to the negative eigenvalue of the partial transpose of $\rho$ is maximally entangled (see previous section). All pure and Bell diagonal states belong to this class.

A much better upper bound can be obtained by using the techniques developed by Rains [**176**]. Indeed, if we enlarge the class of allowed operations from trace-preserving LOCC operations to trace-preserving PPT-operations [**176**], a simple optimization problem arises. A quantum operation $\Lambda$ is PPT iff the dual state $\rho_\Lambda$ associated to this operation [**129, 124, 57**] is PPT (see chapter on quantum channels). The dual state $\rho_\Lambda$ corresponding to a map $\Lambda$ on two qubits is defined in a $2\otimes 2\otimes 2\otimes 2$ Hilbert space and the following relation holds:

$$(\Lambda(\rho))^T_{A'B'} = \mathrm{Tr}_{AB}\left(\rho_\Lambda^{AA'BB'}(\rho_{AB}\otimes I_{A'B'})\right).$$

An upper bound on $F^*$ can now be obtained by considering the following optimization problem: maximize

$$\mathrm{Tr}\left(\rho_\Lambda(\rho\otimes|\psi\rangle\langle\psi|)\right)$$

under the constraints

$$\begin{aligned} \rho_\Lambda &\geq 0 \\ \rho_\Lambda^{T_{BB'}} &\geq 0 \\ \mathrm{Tr}_{A'B'}(\rho_\Lambda) &= I_{AB} \end{aligned}$$

and with $|\psi\rangle$ a maximally entangled state. This is a convex semidefinite program and can be solved numerically. Exploiting symmetries however, it is possible to reduce the complexity drastically. Indeed, $|\psi\rangle$ remains invariant under a twirl operation[13] and this twirl can be applied on $\rho_\Lambda$, leading to a state of the form with $X$ such that the same constraints apply:

$$\rho_\Lambda = \frac{1}{4}\left(I_4\otimes I_4 + (4X^\Gamma - I_4)\otimes(4|\psi\rangle\langle\psi| - I_4)\right).$$

The optimization problem now reduces to: maximize

$$\mathrm{Tr}\,(X\rho)$$

under the constraints

$$\begin{aligned} 0 &\leq X \leq I_4 \\ -\frac{I_4}{2} &\leq X^\Gamma \leq \frac{I_4}{2}. \end{aligned}$$

This is again a convex semidefinite program but now of low dimension and can therefore be solved very efficiently. The following equivalent optimization problem is obtained by a simple change of variables $X' = -X^\Gamma + I_4/2$: maximize

$$1/2 - \mathrm{Tr}\left(X\rho^\Gamma\right) \tag{177}$$

[13]A twirl operation [**124**] consists of applying correlated random local unitary operations such as to obtain a Werner state [**236**] with the same fidelity as the original one.

under the constraints

$$\begin{aligned} 0 &\leq X \leq I_4 \\ -\frac{I_4}{2} &\leq X^\Gamma \leq \frac{I_4}{2}. \end{aligned}$$

Note that the maximal value of the cost-function will certainly yield a value larger than 1/2 if $\rho$ is entangled as $X$ can be chosen to be parallel with the eigenvector corresponding to the negative eigenvalue of $\rho^\Gamma$. Note also that the constraint $-\frac{I_4}{2} \leq X^\Gamma$ will automatically be satisfied if the other three constraints are satisfied: this follows from the fact that $X^\Gamma$ has at most one negative eigenvalue $\lambda_-$ and that $|\lambda_-| \leq \max(\lambda(X^\Gamma))$ (see also previous section). Suppose now that $X$ fulfills the constraints and has rank larger than one. Then $X$ has a separable state $S$ in its support, as each two-dimensional subspace contains at least one separable state. Consider now $y^2$ the largest real positive scalar such that $X - y^2 S \geq 0$. It is easy to verify that the matrix $Y = X - y^2 S$ also fulfils the four constraints, as $S^\Gamma$ is positive due to its separability. Moreover the value $\mathrm{Tr}\left(S\rho^\Gamma\right) = \mathrm{Tr}\left(S^\Gamma \rho\right)$ with $S$ separable and $\rho$ entangled is assured to be positive. Therefore the matrix $Y$ will yield a larger value of the cost-function. This argument implies that the maximal value of the cost-function will be obtained for $X$ rank one. $X$ can therefore be written in the form:

$$X = (A \otimes I_2)|\psi\rangle\langle\psi|(A^\dagger \otimes I_2),$$

and the constraints become $-I_2 \leq A \leq I_2$.

But then the variational characterization of the upper bound (177) becomes exactly equal to the variational characterization of the lower bound (176)! This is very surprising as it implies that the proposed 1-LOCC protocol used in deriving the lower bound was actually optimal over all possible LOCC protocols!

We have therefore proven:

**Theorem 30.** *The optimal trace-preserving LOCC protocol for maximizing the fidelity of a given state $\rho$ consists of a 1-LOCC protocol where one party applies a state-dependent filter. In case of success, the other party does nothing, and in case of failure, both parties make a pure separable state. The optimal filter and fidelity $F^*$ can be found by solving the following convex semidefinite program: maximize*

$$F^* = \frac{1}{2} - \mathrm{Tr}\left(X\rho^\Gamma\right)$$

*under the constraints:*

$$\begin{aligned} 0 &\leq X \leq I_4 \\ -\frac{I_4}{2} &\leq X^\Gamma \leq \frac{I_4}{2}. \end{aligned}$$

*$F^* > 1/2$ if $\rho$ is entangled and the optimal $X_{opt}$ will be of rank 1, and the filter $A$ can be obtained by making the identification*

$$X_{opt} = (A \otimes I_2)|\psi\rangle\langle\psi|(A^\dagger \otimes I_2)$$

*with* $|\psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$.

The given semidefinite program can be solved exactly if $\rho$ has some symmetry. Indeed, if $\rho^\Gamma$ remains invariant under certain symmetry operations, the optimal $X$ can always be chosen such that it has the same symmetry, by exactly the same argument previously used for the twirling step in the proof. As an example, we will calculate $F^*$ for the family of states

$$\rho(F) = F|\psi\rangle\langle\psi| + (1-F)|01\rangle \tag{178}$$

with $F \geq 1/3$ the fidelity of the state (these are precisely the states with minimal fidelity for given concurrence). The partial transpose $\rho^\Gamma(F)$ is given by

$$\rho^\Gamma(F) = \frac{1}{2}\begin{pmatrix} F & 0 & 0 & 0 \\ 0 & 2(1-F) & F & 0 \\ 0 & F & 0 & 0 \\ 0 & 0 & 0 & F \end{pmatrix}.$$

The symmetry under transposition and under the local operations $\sigma_z \otimes \sigma_z$ and $\mathrm{diag}[1,i] \otimes \mathrm{diag}[1,i]$ implies that $X$ will be real and of the form

$$X = \begin{pmatrix} x_1 & 0 & 0 & 0 \\ 0 & x_2 & x_3 & 0 \\ 0 & x_3 & x_4 & 0 \\ 0 & 0 & 0 & x_5 \end{pmatrix}.$$

Moreover $x_1$ and $x_5$ will be equal to zero in the case of an optimal $X$ as otherwise $X$ cannot be rank 1, and a simple optimization problem remains. The optimal filter is readily obtained as $A = \mathrm{diag}[F/(2(1-F)); 1]$, and the maximal achievable fidelity $F^*$ becomes equal to:

$$\begin{array}{rcl} F^*(\rho(F)) = & \frac{1}{2}\left(1 + \frac{F^2}{4(1-F)}\right) & (\text{if} 1/3 \leq F \leq 2/3) \\ F^*(\rho(F)) = & F & (\text{if} F \geq 2/3) \end{array}$$

So if $F \geq 2/3$, no LOCC protocol exists that can increase the fidelity for this class of states. Note that this is something we observed in general: the gain of the optimal LOCC operations is not very large anymore for states with high fidelity.

For the state of the class just described with $F = 1/3$, figure 8 presents a Bloch sphere picture of the states that Alice can steer Bob's system into before ("Original state") and after LOCC. One can indeed immediately see that the LOCC processed state will have a higher teleportation fidelity.

Figure 9 represents a similar pictures, but for a randomly generated state.

A quantum state used for teleportation is a special kind of unital or bistochastic quantum channel (see e.g. [**218, 42**]). A unital quantum channel is completely characterized by looking at the image of the Bloch sphere under the action of the channel[**140**]. In Figure 10, we depict the images of the Bloch sphere

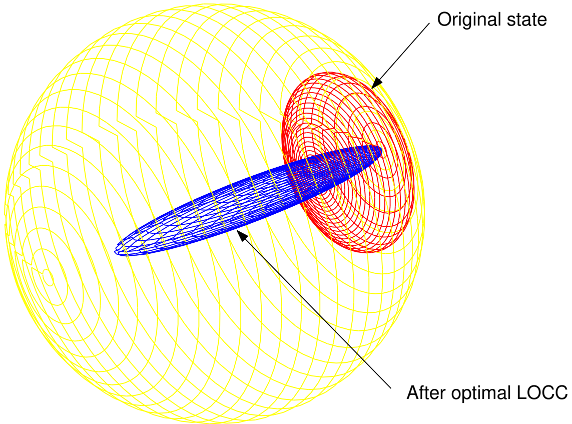



**Figure 8.** The Bloch sphere picture of the states that Alice can steer Bob into, before and after optimal LOCC. The original bipartite state is the one in equation (178) with $F = 1/3$.

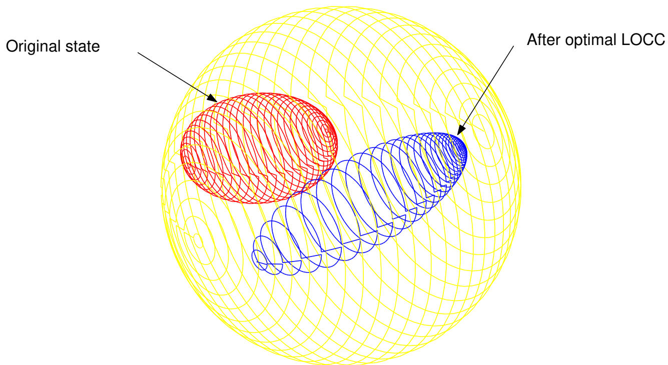



**Figure 9.** The Bloch sphere picture of the states that Alice can steer Bob into, before and after optimal LOCC. The original bipartite state was randomly generated.

under the action of the teleportation channel obtained by the states $\rho(F)$ of eq. 178 with $F = .4$ when the following preprocessing was done: 1. optimal LU-preprocessing; 2. optimal trace-preserving LOCC transformations; 3. optimal filtering operations (probabilistic). This gives a nice illustration of the results derived.

For general two-qubit states, no analytical method for obtaining an expression of $F^*$ is known, and as shown in the previous theorem, a (simple) semidefinite program has to be solved. It is however easy to obtain an explicit lower bound on the optimal $F^*$ in terms of the negativity and the concurrence of the original state. This lower bound is obtained by choosing $X$ to be a constant times the

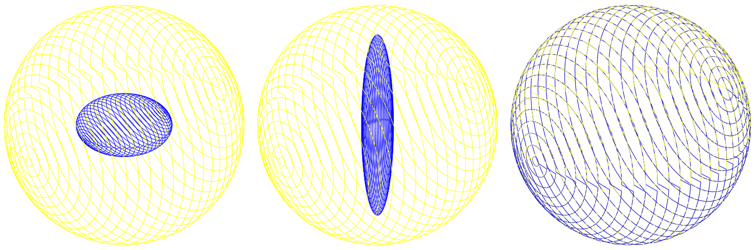


**Figure 10.** The image of the Bloch sphere induced by the teleportation channel with the state $\rho(.4)$ (eq. 178) under optimal LU (left), LOCC (middle) and SLOCC (right) local preprocessing.

subspace spanned by the negative eigenvector $v_-$ of $\rho^\Gamma$. This constant has to be chosen such that the largest eigenvalue of $X^\Gamma$ does not exceed 1/2, and it can be shown that this implies that this constant is equal to $1/(1+\sqrt{1-C_{v_-}^2})$ with $C_{v_-}$ the concurrence of $v_-v_-^\dagger$. The following identity has been proven in equation (128):$C_{v_-} \geq N(\rho)/C(\rho)$. Putting all the pieces together, we arrive at the following lower bound for the maximum achievable fidelity $F^*$ for an arbitrary state $\rho$:

$$\frac{1}{2}\left(1+\frac{N(\rho)}{1+\sqrt{1-\left(\frac{N(\rho)}{C(\rho)}\right)^2}}\right) \leq F^*(\rho) \leq \frac{1}{2}(1+N(\rho)).$$

It is possible to present a simple geometrical picture of the maximum achievable fidelity $F^*$. To that purpose, we use the fact that to each formulation of a semidefinite program, there exists a dual formulation that yields exactly the same value for the extremum. The dual of (177) can be shown to reduce to: minimize

$$G = \frac{1}{2} + \frac{1}{2}\mathrm{Tr}\,(Z)$$

subject to the constraints

$$\begin{aligned} Z &\geq 0 \\ (\rho+Z)^\Gamma &\geq 0. \end{aligned}$$

Note that we explicitly made use of the fact that the set of conditions in (177) for the particular cost function is equivalent to the set of conditions: $0 \leq X$, $X^\Gamma \leq I/2$. This is indeed the case as the condition $X^\Gamma \geq -I/2$ was already shown to be redundant; the condition $X \leq I$ is redundant in a similar way. Defining the state $\rho_Z = Z/\mathrm{Tr}\,(Z)$, this dual problem is equivalent to: minimize

$$G = \frac{1}{2(1-p)}$$

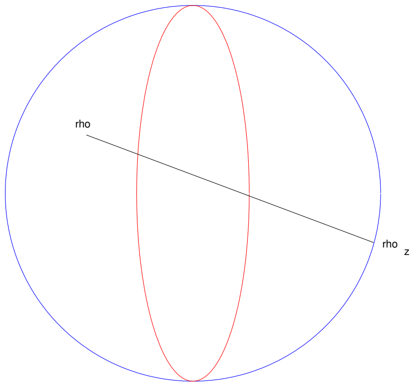



**Figure 11.** The convex set of states (depicted as a circle) and separable states (depicted as the inner ellipse), and the optimal $\rho_Z$ for a given $\rho$.

over all $0 \leq p < 1$ and over all states $\rho_Z$, subject to the constraint that the state $\rho'$

$$\rho' = (1-p)\rho + p\rho_Z$$

is separable. The minimum value obtained is the maximum achievable fidelity $F^*$. As $1/(1-p)$ is monotonously increasing over $0 \leq p < 1$, this problem amounts to finding the state $\rho_Z$ such that the weight in the mixture of this state with the original state $\rho$ is minimal, under the constraint that this mixture is separable. The maximal achievable fidelity $F^*(\rho)$ is therefore a measure of the minimal amount of mixing required of $\rho$ with another state such that a separable state is obtained (see figure 11).

It is trivial to prove that the obtained $F^*$ is an entanglement monotone and has all nice properties expected from a good entanglement measure. Furthermore, $F^*(\rho)$ leads to the exact expression for the optimal teleportation fidelity that can be achieved with a mixed state $\rho$: this follows from the fact that the teleportation fidelity of a state $\rho$ depends linearly on the fidelity of the state [**124**].

In conclusion, we have proven that a state is entangled iff its fidelity can be made larger than 1/2 by trace-preserving LOCC operations. Optimal teleportation can thus be achieved by first doing some pre-processing on the state consumed during the teleportation process by the procedure derived above. This is due to the fact that the fidelity is both dependent on the quantum correlations and on the classical correlations, and enhancing the classical correlations by mixing (and hence losing quantum correlations) can lead to a higher fidelity. One of the most surprising results is that the optimization was done over all PPT-operations, and nevertheless a physically realizable protocol was optimal: this is supporting evidence for the fact that good bounds on entanglement distillation can be found allowing the class of PPT-protocols [**176**]. It would be very interesting to apply the same techniques to the multiple copy case, as this would ultimately lead to an explicit expression for the entanglement of distillation. Remark also that the previous derivation is of direct interest in the construction of distillation protocols: instead of implementing the optimal filter (that does not take into account the probability of success), it could be much more advantageous to implement the filter associated to the given semidefinite program (of course without implementing the mixing with the separable state).

A similar reasoning can be applied to maximize the fidelity for entangled states in higher dimensional systems. Repeating exactly the same arguments, an upper bound of the fidelity is given by the solution of the following semidefinite program: maximize

$$\mathrm{Tr}\,(\rho X) \tag{179}$$

under the conditions

$$0 \leq X \qquad\qquad X^{\Gamma} \leq \frac{I}{n}. \tag{180}$$

$n$ is the dimension of the Hilbert space (e.g. 2 in the case of qubits), and we again made explicitly use of the fact that the conditions $-I/n \leq X^{\Gamma}$ and $X \leq I$ are redundant. This semidefinite program can readily be seen to be equivalent to:

$$\max_X \left(\frac{1}{n} - \mathrm{Tr}\left(X\rho^{\Gamma}\right)\right) \qquad 0 \leq X \qquad X^{\Gamma} \leq \frac{I}{n} \tag{181}$$

Following the lines of the arguments in the qubit case, this immediately points out the following facts:

- A PPT-state can never yield a fidelity exceeding $1/n$, and is therefore not useful for teleportation.
- The optimal $X$ will be such that it has no PPT-states and therefore certainly no separable states in its range (cfr. the two-qubit case, where PPT-states are equivalent to separable states). It is an open question whether this fact can be used to show that the PPT-operation is actually separable (and maybe even LOCC as in the case of two qubits).
- The same geometrical picture depicted in Figure (11) applies as in the case of qubits.

## 4.6. Distilling singlets

In the case of pure states, the entanglement of distillation is asymptotically equal to the entanglement of formation. In the case of mixed states however, some information is lost during the preparation procedure (one has to artificially “forget”information) and therefore is would be a real miracle if all the entanglement could be recovered. It has indeed been proven that the procedure of preparing a mixed state from singlets is irreversible in some specific examples [**226**].

In some sense, the entanglement of distillation is the central property of a mixed bipartite state if it is to be used to process quantum information: the only way of faithfully teleporting a quantum state is by states of high fidelity. But the entanglement of distillation is also by far the entanglement measure from which we know the least as it is very difficult to create physical distillation protocols. This problem has to do with the fact that not much is known of how to characterize LOCC operations for mixed states (note that this was the main motivation of studying quantum steering in the mixed state case). One of the only sensible results on entanglement of distillation are the ones obtained by Rains, where he shows how to bound the entanglement of distillation from above by allowing distillation protocols allowing unphysical PPT-operations[**175, 176**].

The actual proof that it is possible to distill singlets out of some entangled mixed states was given by Bennett et al.[**30**]. The procedure outlined in that paper was considerably extended and streamlined in the seminal paper by Bennett, DiVincenzo, Smolin and Wootters [**34**], where they showed that the concept of hashing [**29**] could be used to distill all mixed states of two qubits with fidelity larger than 1/2 with a non-zero yield. More specifically, they showed that a local measurement on two pairs of entangled Bell diagonal states could result in one pair with a larger amount of entanglement. By repeating this iteratively, the entanglement can be increased until the global entropy of the Bell diagonal state is smaller than 1. At that moment, they showed that hashing methods could be used to transform a large amount $N \gg 1$ of equal states with entropy $S$ into $N(1-S)$ singlets. The hashing step is pretty efficient, but the recurrence method however consumes an exponential number of states. The bottleneck is clearly the recurrence step, and we will show shortly how to obtain more efficient versions.

4.6.0.1. *Recurrence schemes.* The simplest recurrence scheme involves two states, and the aim is to do joint local measurements on both pairs such as to yield one new pair with a larger amount of entanglement. Consider two identical Bell-diagonal states $\rho$

$$\rho = \lambda_1|\Phi^+\rangle\langle\Phi^-| + \lambda_2|\Psi^+\rangle\langle\Psi^+| + \lambda_3|\Psi^-\rangle\langle\Psi^-| + \lambda_4|\Phi^-\rangle\langle\Phi^-| \tag{182}$$

with $\lambda_1 \geq 1/2 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$. If both parties apply a CNOT to their respective pairs, do a projective measurement in the computational basis on the second qubit, then with probability $p = (\lambda_1+\lambda_4)^2 + (\lambda_2+\lambda_3)^2$ a new Bell diagonal state is obtained with eigenvalues [**67**]:

$$\begin{aligned}
\lambda_1' &= \frac{\lambda_1^2+\lambda_4^2}{(\lambda_1+\lambda_4)^2+(\lambda_2+\lambda_3)^2}\\
\lambda_2' &= \frac{\lambda_2\lambda_3}{(\lambda_1+\lambda_4)^2+(\lambda_2+\lambda_3)^2}\\
\lambda_3' &= \frac{\lambda_1\lambda_4}{(\lambda_1+\lambda_4)^2+(\lambda_2+\lambda_3)^2}\\
\lambda_4' &= \frac{\lambda_2^2+\lambda_3^2}{(\lambda_1+\lambda_4)^2+(\lambda_2+\lambda_3)^2}
\end{aligned}$$

Deutsch et al. [**67**] showed that a recursive use of this protocol ultimately leads to a state very close to the singlet state if and only if the original $\lambda_1 > 1/2$. Macchiavello [**154**] reported that the convergence was not guaranteed to be monotonic however. This defect can easily be cured by doing a local permutation between each recurrence step such as to reorder the eigenvalues in decreasing order. Indeed, if $\lambda_1 > 1/2 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$, then the following holds:

$$\begin{aligned}
\frac{\lambda_1^2+\lambda_4^2}{(\lambda_1+\lambda_4)^2+(\lambda_2+\lambda_3)^2}-\lambda_1 &= \frac{(2\lambda_1-1)(\lambda_1-(\lambda_1+\lambda_4)^2)}{1+2(\lambda_1+\lambda_4)^2-2(\lambda_1+\lambda_4)}\\
&\geq \frac{(2\lambda_1-1)(1-\lambda_1)(4\lambda_1-1)}{9(1+2(\lambda_1+\lambda_4)^2-2(\lambda_1+\lambda_4))}\\
&\geq 2\frac{(2\lambda_1-1)(1-\lambda_1)(4\lambda_1-1)}{9+(4\lambda_1-1)^2}\\
&> 0 \qquad (183)
\end{aligned}$$

In the second and the third step we made explicit use of the fact that $\lambda_4 \leq (1-\lambda_1)/3$. As the previous expression in always strictly positive unless $\lambda_1 = 1$, this proves that the convergence is monotonic if between each recurrence step a reordering of the eigenvalues is done.

After a certain number of recurrence steps, the entropy of the global state will be lower than 1, and then the hashing scheme can be implemented.

It is however possible to devise better recurrence schemes. The main idea of the recurrence scheme sketched above was the following: the local CNOT is responsible for a local permutation of all 16 states $|\psi_i\rangle|\psi_j\rangle$ (with $\{|\psi_\alpha\rangle\}$ all Bell states) into each other (see also [**34**]). The projection that follows is such that a new Bell diagonal state arises with a larger $\lambda_1$. A generalization of this would give the following protocol:

(1) Start from $n$ identical Bell diagonal state that are entangled. This yields a mixture of $4^n$ tensor product of Bell states.

(2) Apply a local permutation of these $4^n$ products of Bell states. As a result the $n$ qubit pairs get statistically dependent.

(3) Check whether the last $n-m$ qubit pairs are $|\Phi^{\pm}\rangle$ states. This can be accomplished locally by measuring both qubits of each pair in the $|0\rangle, |1\rangle$ basis, and checking whether both measurements yield the same result. If all measured pairs were $|\Phi\rangle$-states, keep the first $m$ pairs. This is a new mixture of $4^m$ products of Bell states.

Jeroen Dehaene [**64**] found a very interesting way of characterizing all local permutations that map products of Bell states to products of Bell states:

**Theorem 31.** *A tensor product of n Bell states can be represented by a binary number $x$ of $2n$ digits, in which the digits $2i-1, 2i$ represent the $i$'th Bell state. Then every local unitary operation that results in a permutation of the possible $4^n$ tensor products of $n$ Bell states can be represented as an affine transformation*

$$\phi : \mathbb{Z}_2^{2n} \to \mathbb{Z}_2^{2n} : x \to Ax + b \tag{184}$$

*where*

$$b \in \mathbb{Z}_2^{2n} \qquad A \in \mathbb{Z}_2^{2n\times 2n} \qquad A^T P A = P \qquad P = \oplus_{i=1}^n \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{185}$$

*Conversely, any such permutation can be realized by local unitary operations.*

*Proof:* see Dehaene et al. [**64**].

The previous Theorem is very interesting as it gives an exhaustive way of generating all possible permutations of Bell states that can be performed locally and therefore of all permutations that can be used in the generalized recurrence protocols.

As an example, we considered the question of making a recurrence protocol that maps 4 qubit pairs to 1. We did an exhaustive search over all local permutations that could be realized by four elementary two-qubit operations, and found out that the local permutations $U_A, U_B$ corresponding to the unitary matrices

$$U_A = U_B^* = e^{i\pi/4\sigma_z^1\otimes\sigma_x^2} e^{i\pi/4\sigma_x^1\otimes\sigma_x^4} e^{i\pi/4\sigma_z^1\otimes\sigma_y^3} e^{i\pi/4\sigma_x^2\otimes\sigma_z^3}, \tag{186}$$

followed by a measurement of the last three pairs, generally performed very well if the eigenvalues corresponding to the Bell states $|\Phi^+\rangle, |\Psi^+\rangle, |\Phi^-\rangle, |\Psi^-\rangle$ were ordered in decreasing order. Note however that one protocol that is optimal for all possible input states is not possible: there are typically a few possible optimal solutions for the local permutations, depending on the eigenvalues of the considered Bell diagonal states.

Figure 12 shows the performance of this method for initial Werner states. Note that the local permutations of equation (186) were chosen, that a reordering of the eigenvalues was done between the different recurrence steps, that we switched to the 2 to 1 recurrence protocol once the fidelity exceeded $\simeq .82$ (the 4 to 1 protocol is too expensive in that region), and that we switched to

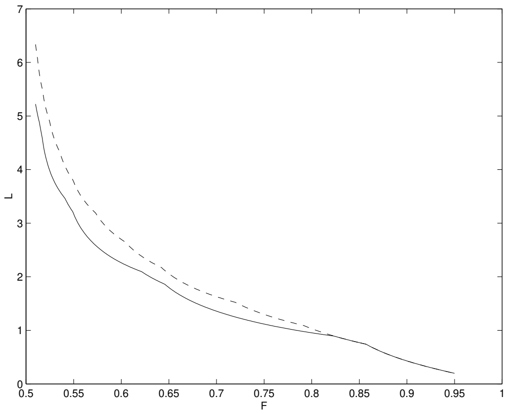



**Figure 12.** Comparison of 10-logarithm of inverse asymptotic yield $L$ for input Werner states with fidelity $F$ for proposed protocol (full line) and existing recurrence protocol (dashed line).

the hashing protocol once the fidelity was around $\simeq .86$. Also depicted is the previously best known protocol involving only 2 to 1 schemes: there is a clear gain using the new method for low fidelities.

An even higher gain is expected with a 8 to 1 protocol for initial states with low fidelity, and so further. Ultimately, such methods should lead to generalizations of the hashing method that work from fidelity 1/2 on. Note however that the computational cost of finding good protocols in higher dimensions scales exponentially; it is however plausible that huge simplifications can be done in the asymptotic region and when only Werner states are considered.

4.6.0.2. *Distillation of low rank states.* The only mixed entangled states for which the entanglement of distillation is known are the Bell diagonal states of rank 2[**175**]: it is easy to see that their entropy is always smaller than 1, so that the hashing method can readily be applied. This leads to an entanglement of distillation given by

$$E_d(\rho) = 1 - \lambda_1 \log \lambda_1 - \lambda_2 \log \lambda_2. \tag{187}$$

The results of Rains however imply that the relative entropy of entanglement is an upper bound for the entanglement of distillation, and it turns out that the Relent of a Bell diagonal state is given by exactly the same expression[14].

As a second example, we note that distilling pure states out of full rank mixed states is inherently only possible in the asymptotic regime and moreover in the limit of fidelity going to one: if the ensemble were finite, the only local

[14]note that this implies that the regularized (asymptotic)version of the Relent [**8**] is equal to the Relent itself in that case, as the regularized version can only be smaller.

operations that can make a pure state of a mixed state are projective and rank 1 and these destroy all entanglement [**135**]. In the case of low rank states however, it is in principle possible to distill singlets even with a finite number of states. Indeed, we have already introduced the quasidistillable states, which could be brought arbitrary close to the singlet state with SLOCC operations. In the case of two copies of quasi-distillable states, one can even distill a perfect singlet, as shown by Dür (unpublished) and Jané [**130**]. The quasi-distillable states are the unique states with this property.

This can be proven as follows: if a perfect singlet has to be distilled out of a mixed state $\rho^{\otimes 2}$ by a filtering operation $A \otimes B$, then all the states in the range of $\sigma$ have to be mapped or to the singlet state or to zero. Suppose that the basis vectors of the vector space associated to the range of $\sigma$ can be chosen separable. Obviously none of them can be mapped to the singlet state, and this implies that such a state can never be mapped onto the singlet state. Therefore we have to look for states whose range cannot be spanned with separable states. Obviously, it is sufficient to consider the normal forms of the density operators, as local filtering operations are not able to change the fact that the range is spanned by separable states. The states that can be brought into Bell diagonal form can immediately ruled out: by changing the eigenvalues, it is always possible to make the state separable. There remains the non-diagonalizable cases of rank 3 and 2, but the rank 3 case can also ruled as the range of a rank 3 state is spanned by three product vectors. Therefore the only states that can possibly do the job are the quasi-distillable ones. Let us write the quasi-distillable states as

$$x/2(|00\rangle + |11\rangle)(\langle 00| + \langle 11|) + (1-x)|01\rangle\langle 01|. \tag{188}$$

If we take two copies, the local filters $A, B$ (parameterized as $4 \times 4$ matrices) have to fulfill the following conditions:

$$\begin{aligned} 0 &= (A \otimes B)|01\rangle|01\rangle & (189) \\ 0 &= (A \otimes B)|01\rangle(|00\rangle + |11\rangle) & (190) \\ 0 &= (A \otimes B)(|00\rangle + |11\rangle)|01\rangle & (191) \end{aligned}$$

The first condition is obvious: $A \otimes B$ times a separable state can never result in a singlet. It follows that or the first column of $A$ is zero, or the last of $B$. By the same reasoning but taking into account these zeros, the two other equalities follow because separable states cannot become entangled. Moreover, we do not want that all the entanglement in

$$(A \otimes B)(|00\rangle + |11\rangle)(|00\rangle + |11\rangle) \tag{192}$$

is destroyed. This together with the 3 other conditions implies that the only possible choice for $A$ and $B$ is the one where the first column of $A$ is zero and the last column of $B$ is zero. Indeed, if we choose

$$A=\begin{pmatrix}0&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&0\end{pmatrix}\qquad B=\begin{pmatrix}0&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&0\end{pmatrix}\tag{193}$$

then the singlet $|00\rangle|11\rangle+|11\rangle|00\rangle$ is created. Note that the total probability of this filtering procedure to work is given by $x^2/2$, which is indeed finite[15]: it is thus possible to distill perfect singlets out of (very special) mixed entangled states.

## 4.7. Maximally entangled mixed states of two qubits

This section is devoted to the study of the following basic question: given a mixed state of two qubits, what global unitary operations[16] do I have to apply to the system such as to maximize the amount of entanglement present in the system? As an example, consider a tensor product state of one qubit in a pure state (i.e. maximal knowledge) and another one in the maximally mixed state (i.e. no knowledge); is it possible to create entanglement in this system through unitary evolution despite the fact that we do not know anything about the second qubit? Surprisingly, the answer is yes[17]

The class of states for which no more entanglement can be created by global unitary operations is clearly a generalization of the class of Bell states to mixed states, and gives strict bounds on how the mixedness of a state limits its entanglement. It will indeed be shown that the states with maximal entropy for given amount of entanglement belong to this class of maximally entangled mixed states. Also closely related to the issue of maximally entangled mixed states is the question of characterizing the largest ball of separable density

---

[15] One can easily generalize the previous reasoning to the case where more copies (i.e. $N$) are available. However, this turns out to be not advantageous as a yield in the order of

$$\sqrt{\frac{2N}{\pi}}x^N\tag{194}$$

is obtained for large $N$, which is exponentially smaller than the yield obtained if done on blocks of two copies.

[16] If not only unitary operations but also measurements were allowed, it is clear that a Von Neumann measurement in the Bell basis would immediately yield a singlet. Here however we restrict ourselves to unitary operations. Obviously, these unitary operations must be global ones, that is, acting on the system as a whole, since any reasonable measure of entanglement must be invariant under local unitary operations, acting only on single qubits.

[17] Note however that the corresponding entangled state does not violate CHSH inequalities; this follows from the fact that we have proven in theorem (24) that the maximal violation for states with a given spectrum occurs for Bell diagonal states, and that a Bell diagonal state with eigenvalues $[1/2, 1/2, 0, 0]$ is not entangled.

matrices surrounding the maximally entangled state [**44, 228**]. Indeed, the entangled states closest to the maximally mixed state necessarily have to belong to the proposed class of maximally entangled mixed states. We will thus give a complete characterization of all nearly entangled states lying on the boundary of the sphere of separable states surrounding the maximally mixed state, and therefore derive the best known lower bound on the volume of separable states.

Let us now formulate the central Theorem of this section:

**Theorem 32.** *Let the eigenvalue decomposition of* $\rho$ *be*

$$\rho = \Phi\Lambda\Phi^\dagger$$

*where the eigenvalues* $\{\lambda_i\}$ *are sorted in non-ascending order. The entanglement of formation is maximized if and only if a global unitary transformation of the form*

$$U = (U_1 \otimes U_2)\begin{pmatrix} 0 & 0 & 0 & 1 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} D_\phi \Phi^\dagger$$

*is applied to the system, where* $U_1$ *and* $U_2$ *are local unitary operations and* $D_\phi$ *is a unitary diagonal matrix. This same global unitary transformation is the unique transformation maximizing the negativity and the relative entropy of entanglement. The entanglement of formation and negativity of the new state* $\rho' = U\rho U^\dagger$ *are then given by*

$$\begin{aligned} E_f(\rho') &= f\left(\max\left(0, \lambda_1 - \lambda_3 - 2\sqrt{\lambda_2\lambda_4}\right)\right) \\ E_N(\rho') &= \max\left(0, \sqrt{(\lambda_1-\lambda_3)^2 + (\lambda_2-\lambda_4)^2} - \lambda_2 - \lambda_4\right) \end{aligned}$$

*respectively, while the expression for the relative entropy of entanglement is given by*

$$\begin{aligned} &E_R(\rho') = \mathrm{Tr}(\rho\log\rho) - \lambda_1\log((1-a)/2) - \\ &\quad \lambda_2\log((a+b+2(\lambda_2-\lambda_4))/4) - \\ &\quad \lambda_3\log((1-b)/2) - \lambda_4\log((a+b-2(\lambda_2-\lambda_4))/4) \\ a &= (d - \sqrt{d^2 - 4(1-\lambda_1)(1-\lambda_3)(\lambda_2-\lambda_4)^2})/(2(1-\lambda_3)) \\ b &= (d + \sqrt{d^2 - 4(1-\lambda_1)(1-\lambda_3)(\lambda_2-\lambda_4)^2})/(2(1-\lambda_1)) \\ d &= \lambda_2 + \lambda_4 + (\lambda_2-\lambda_4)^2 \end{aligned}$$

The class of states obtained are defined as the maximally entangled mixed states (MEMS-states).

*Proof:* The proof of this Theorem is very long and uses some advanced linear algebra. The cases of entanglement of formation, negativity and relative entropy of entanglement will be treated independently. We start with the entanglement of formation.

As the function $f(x)$ is monotonously increasing, maximizing the EoF is equivalent to maximizing the concurrence. The problem is now reduced to finding:

$$C_{\max} = \max_{U \in U(4)} (0, \sigma_1 - \sigma_2 - \sigma_3 - \sigma_4) \tag{195}$$

with $\{\sigma_i\}$ the singular values of

$$Q = \Lambda^{1/2} \Phi^T U^T S U \Phi \Lambda^{1/2}. \tag{196}$$

Now, $\Phi$, $U$ and $S$ are unitary, and so is any product of them. It then follows that

$$C_{max} \le \max_{V \in U(4)} (0, \sigma_1 - \sigma_2 - \sigma_3 - \sigma_4) \tag{197}$$

with $\{\sigma_i\}$ the singular values of $\Lambda^{1/2} V \Lambda^{1/2}$. The inequality becomes an equality if there is a unitary matrix $U$ such that the optimal $V$ can be written as $\Phi^T U^T S U \Phi$. A necessary and sufficient condition for this is that the optimal $V$ be symmetric ($V = V^T$): as $S$ is symmetric and unitary, it can be written as a product $S_1^T S_1$, with $S_1$ again unitary. This is known as the Takagi factorization of $S$ [**116**]. This factorization is not unique: left-multiplying $S_1$ with a complex orthogonal matrix $O$ ($O^T O = I$) also yields a valid Takagi factor. An explicit form of $S_1$ is given by:

$$S_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -i & i & 0 \\ i & 0 & 0 & i \end{pmatrix}. \tag{198}$$

If $V$ is symmetric it can also be factorized like this: $V = V_1^T V_1$. It is now easy to see that any $U$ of the form

$$U = S_1^\dagger O V_1 \Phi^\dagger, \tag{199}$$

with $O$ real orthogonal, indeed yields $V = V_1^T V_1$.

To proceed, we need two inequalities concerning singular values of matrix products. Henceforth, singular values, as well as eigenvalues will be sorted in non-ascending order. The following inequality for singular values is well-known [**117**]:

> Let $A \in M_{n,r}(\mathcal{C})$, $B \in M_{r,m}(\mathcal{C})$. Then,
>
> $$\sum_{i=1}^{k} \sigma_i(AB) \le \sum_{i=1}^{k} \sigma_i(A) \sigma_i(B), \tag{200}$$
>
> for $k = 1, \ldots, q = \min\{n, r, m\}$.

Less known is the following result by Wang and Xi [**233**]:

Let $A \in M_n(\mathcal{C})$, $B \in M_{n,m}(\mathcal{C})$, and $1 \le i_1 < \cdots < i_k \le n$. Then

$$\sum_{t=1}^{k} \sigma_{i_t}(AB) \ge \sum_{t=1}^{k} \sigma_{i_t}(A)\sigma_{n-t+1}(B). \tag{201}$$

Set $n = 4$ in both inequalities. Then put $k = 1$ in the first, and $k = 3, i_1 = 2, i_2 = 3, i_3 = 4$ in the second. Subtracting the inequalities then gives:

$$\begin{aligned}&\sigma_1(AB) - (\sigma_2(AB) + \sigma_3(AB) + \sigma_4(AB)) \le \\ &\sigma_1(A)\sigma_1(B) - \sigma_2(A)\sigma_4(B) - \sigma_3(A)\sigma_3(B) - \sigma_4(A)\sigma_2(B)\end{aligned}$$

Furthermore, let $A = \Lambda^{1/2}$ and $B = V\Lambda^{1/2}$, with $\Lambda$ positive diagonal and with the diagonal elements sorted in non-ascending order. Thus, $\sigma_i(A) = \sigma_i(B) = \sqrt{\lambda_i}$. This gives:

$$(\sigma_1 - (\sigma_2 + \sigma_3 + \sigma_4))(\Lambda^{1/2}V\Lambda^{1/2}) \le \lambda_1 - (2\sqrt{\lambda_2\lambda_4} + \lambda_3).$$

It is easy to see that this inequality becomes an equality iff $V$ is equal to the permutation matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \tag{202}$$

multiplied by an arbitrary unitary diagonal matrix $D_\phi$. Therefore, we have proven:

$$\begin{aligned}\max_{V \in U(4)} \quad (\sigma_1 - (\sigma_2 + \sigma_3 + \sigma_4))(\Lambda^{1/2}V\Lambda^{1/2}) = \\ \lambda_1 - (2\sqrt{\lambda_2\lambda_4} + \lambda_3).\end{aligned} \tag{203}$$

We can directly apply this to the problem at hand. The optimal $V$ is indeed symmetric, so that it can be decomposed as $V = V_1^T V_1$. A possible Takagi factor is:

$$V_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 0 & 1 & 0 \\ 0 & i/\sqrt{2} & 0 & -i/\sqrt{2} \end{pmatrix} \tag{204}$$

The optimal unitary operations $U$ are thus all of the form: $U = S_1^\dagger O V_1 D_\phi^{1/2} \Phi^\dagger$ with $O$ an arbitrary orthogonal matrix. It has to be emphasized that the diagonal matrix $D_\phi$ will not have any effect on the state $\rho' = U\Phi\Lambda\Phi^\dagger U^\dagger$.

To proceed we exploit a well-known accident in Lie group theory:

$$SU(2) \otimes SU(2) \cong SO(4). \tag{205}$$

It now happens that the unitary matrix $S_1$ is exactly of the form for making $S_1(U_1 \otimes U_2)S_1^\dagger$ real for arbitrary $\{U_1, U_2\} \in SU(2)$. It follows that $S_1(U_1 \otimes U_2)S_1^\dagger$ is orthogonal and thus is an element of $SO(4)$. Conversely, each element $Q \in SO(4)$ can be written as $Q = S_1(U_1 \otimes U_2)S_1^\dagger$. On the other hand the orthogonal matrices with determinant equal to $-1$ can all be written as an orthogonal matrix with determinant 1 multiplied by a fixed matrix of determinant $-1$. Some calculations reveal that

$$S_1^\dagger \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} V_1 = (\sigma_y \otimes \sigma_y) S_1^\dagger V_1 \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

We conclude that for each $O \in O(4)$ and $D_\phi$ unitary diagonal, there exist $U_1, U_2 \in SU(2)$ and $D_{\phi'}$ unitary diagonal, such that $U = S_1^\dagger O V_1 D_\phi \Phi^\dagger = (U_1 \otimes U_2) S_1^\dagger V_1 D_{\phi'} \Phi^\dagger$.

It is now easy to check that a unitary transformation produces maximal entanglement of formation if and only if it is of the form

$$(U_1 \otimes U_2) \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} D_\phi \Phi^\dagger. \tag{206}$$

This completes the proof of the first part of the Theorem.

We now proceed to prove the second part of the Theorem concerning the negativity. Here we will look for the eigenvectors $\Psi$ such that the negativity of the operator $\Psi \Lambda \Psi^\dagger$ is maximal. This proof is based on the Rayleigh-Ritz variational characterization of the minimal eigenvalue of a Hermitian matrix:

$$\begin{aligned} \lambda_{min}(\rho^{T_A}) &= \min_{x:||x||=1} \operatorname{Tr} \rho^{T_A} |x\rangle\langle x| \\ &= \min_{x:||x||=1} \operatorname{Tr} \rho (|x\rangle\langle x|)^{T_A} \end{aligned} \tag{207}$$

It is easy to see that the partial transpose of a pure state is of the form $VDV^\dagger$ where

$$D = \begin{pmatrix} \cos(\alpha)^2 & 0 & 0 & 0 \\ 0 & \cos(\alpha)\sin(\alpha) & 0 & 0 \\ 0 & 0 & -\cos(\alpha)\sin(\alpha) & 0 \\ 0 & 0 & 0 & \sin(\alpha)^2 \end{pmatrix} \tag{208}$$

$$V = (U_1 \otimes U_2^*) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{209}$$

where we choose $0 \leq \alpha \leq \pi/4$. The maximal negativity for given spectrum $\Lambda$ can therefore be found by solving the following optimization problem:

$$\min_W \operatorname{Tr}(W \Lambda W^\dagger D) = \min_W \sum_{ij} D_{ii} \Lambda_{jj} |W_{ij}|^2 \tag{210}$$

where the optimization is done over all unitary $W = V^\dagger\Psi$. This problem can easily be solved using Birkhoff's Theorem which says that the set of doubly-stochastic matrices is the convex closure of the set of permutation matrices. The optimal solution clearly corresponds to choosing

$$|W_{ij}|^2 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \tag{211}$$

leading to the minimal value

$$-(\lambda_1 - \lambda_3)\cos(\alpha)\sin(\alpha) + \lambda_2 \sin(\alpha)^2 + \lambda_4 \cos(\alpha)^2$$

Minimizing this expression over the remaining free parameter $\alpha$ gives, after a few basic calculations:

$$\begin{aligned} \cos(2\alpha) &= \frac{\lambda_2 - \lambda_4}{\sqrt{(\lambda_1 - \lambda_3)^2 + (\lambda_2 - \lambda_4)^2}} \\ \min &= \left(\lambda_2 + \lambda_4 - \sqrt{(\lambda_1 - \lambda_3)^2 + (\lambda_2 - \lambda_4)^2}\right)/2. \end{aligned}$$

This corresponds to the conjectured formula for the optimal negativity.

We now have to find the $\Psi$ for which this optimum is reached. As $\Psi = VW$ and $V$ and $W$ are determined, it is trivial to show that the optimal $\Psi = U\Phi$ indeed corresponds to the one claimed in the theorem.

Next we move to the third part of the Theorem concerning the relative entropy of entanglement, variationally defined as

$$\min_{\sigma \in \mathrm{Sep}} S(\rho||\sigma) \tag{212}$$

We first observe the following fact: the density operator corresponding to the maximally mixed state is invariant under transposition, under permutation of the qubits and under the local unitary operations $\sigma_z \otimes \sigma_z$ and $\mathrm{diag}[1; i] \otimes \mathrm{diag}[1; i]$. Moreover, it is easily checked that if a density operator obeys these symmetry conditions, then its eigenvectors must correspond to the one of the MEMS-states. Moreover, all these operations map separable states to separable states. Following Rains [**175**], we conclude that the optimal separable state in the calculation of the relative entropy has the same eigenvectors as the MEMS states. The calculation of the relative entropy of entanglement of a MEMS state then amounts to a simple (but very long) optimization problem over the eigenvalues of $\sigma$, and the result is given in the theorem. It follows that $\sigma$ itself is a maximally entangled mixed state (i.e. its eigenvalues have the right ordering). It remains to be proven that the MEMS-states have the highest possible relent; this can be done as follows. For $\rho$ a maximally entangled mixed state, and for an arbitrary unitary $U$, it holds:

$$\begin{aligned} E_R(\rho) &= \min_{\sigma\in\mathrm{Sep}} S(\rho||\sigma) \\ &= \min_{\sigma\in\mathrm{Sep}} S(U\rho U^\dagger||U\sigma U^\dagger) \\ &\geq \min_{\sigma'\in\mathrm{Sep}} S(U\rho U^\dagger||\sigma') \\ &= E_R(U\rho U^\dagger), \end{aligned}$$

The important step arose in the third step, where we used the crucial fact that for a maximally entangled mixed state that is separable (namely the optimal $\sigma$), it holds that it remains separable after an arbitrary unitary operation. The inequality arose because the minimization domain has been enlarged. Therefore the MEMS states have larger relative entropy of entanglement than all states that can be obtained from it by doing global unitary operations. This finally completes the proof of the theorem. □

Note that a different result was obtained in the case of violation of CHSH inequalities: in Theorem 24 we have proven that the maximal violation for all states with an equal spectrum occurs for Bell diagonal states.

Let us now analyze more closely the newly defined class of maximally entangled mixed states (MEMS). Note first that the only Bell diagonal state belonging to this class is the Werner state.

We already know that $U$ is unique up to local unitary transformations. It is easy to check that the ordered eigenvalues of the MEMS-states for given entanglement of formation $f(C)$ are parameterized by two independent variables $\alpha$ and $\beta$:

$$\begin{aligned} 0 &\leq \alpha \leq 1 \\ \beta &\geq \sqrt{1-\frac{\alpha^2}{9}} - \sqrt{\frac{8}{9}}\alpha \\ \beta &\leq \min(\sqrt{\frac{1+C}{1-C}-\frac{\alpha^2}{9}} - \sqrt{\frac{2}{9}}\alpha, \sqrt{3-\alpha^2}-\sqrt{2}\alpha) \\ \lambda_1 &= 1-\frac{1-C}{6}(3+\beta^2) \\ \lambda_2 &= \frac{1-C}{6}(\alpha+\sqrt{2}\beta)^2 \\ \lambda_3 &= \frac{1-C}{6}(3-(\sqrt{2}\alpha+\beta)^2) \\ \lambda_4 &= \frac{1-C}{6}\alpha^2 \end{aligned} \tag{213}$$

For given EoF there is thus, up to local unitary transformations, a two dimensional manifold of maximally entangled states. In the case of concurrence $C = 1$ the upper and lower bounds on $\beta$ become equal and the unique pure

Bell states arise. Another observation is the fact that $\lambda_4$ of all MEMS-states is smaller than 1/6. This implies that if the smallest eigenvalue of whatever two-qubit state exceeds 1/6, this state is separable.

A natural question is now how to characterize the entangled states closest to the maximally mixed state. A sensible metric is given by the Frobenius norm $\|\rho - I/4\|_2 = \sqrt{\sum_i \lambda_i^2 - 1/4}$. This norm is only dependent on the eigenvalues of $\rho$ and it is thus sufficient to consider the MEMS states at the boundary of entangled states where both the concurrence and the negativity become zero. This can be solved using the method of Lagrange multipliers. A straightforward calculation leads to a one-parameter family of solutions:

$$\begin{aligned} & 0 \leq \quad x \leq \quad \frac{1}{6} \\ \lambda_1 &= \frac{1}{3} + \sqrt{x\left(\frac{1}{3} - x\right)} \qquad \lambda_2 = \frac{1}{3} - x \\ \lambda_3 &= \frac{1}{3} - \sqrt{x\left(\frac{1}{3} - x\right)} \qquad \lambda_4 = x \end{aligned} \tag{214}$$

The Frobenius norm $\|\rho - I/4\|_2$ for all these states on the boundary of the sphere of separable states is given by the number $\sqrt{1/12}$. This criterion is exactly equivalent to the well-known criterion of Życzkowski et al. [**250**]: $\mathrm{Tr}\,\rho^2 = 1/3$. Here, however, we have the additional benefit of knowing exactly all the entangled states on this boundary as these are the MEMS states with eigenvalues given by the previous formula. Furthermore, Życzkowski et al. [**250**] proposed a lower bound on the volume of separable states by considering the ball of states that remain separable under all global unitary transformations. Clearly the criterion $\sum_i \lambda_i^2 \leq 1/3$ can be strengthened to $\lambda_1 - \lambda_3 - 2\sqrt{\lambda_2\lambda_4} \leq 0$. Some tedious integration then leads to a better lower bound for the volume of separable states relative to the volume of all states: $0.3270\ldots$ (as opposed to $0.3023\ldots$ of [**250**]).

Further interesting properties of the maximally entangled mixed states include the fact that the states with maximal entropy for given entanglement all belong to this class. This can be seen as follows: the global entropy of a state is a function of the eigenvalues of the density matrix only. Therefore the states with maximal entanglement for given entropy can be found by first looking for the states with maximal entanglement for fixed eigenvalues, followed by maximizing the entropy of the obtained class of (maximally entangled) mixed states. Observe that this implies that the states with maximal entropy for given entanglement also belong to the same class. One would expect that Werner states always maximize the entropy for given amount of entanglement. Surprisingly, this turns out to be not the case.

Let us consider the case where the concurrence is fixed and we are looking for the state with the largest amount of entropy. This leads to the following Lagrange constrained problem: maximize

$$-(\lambda_1\log(\lambda_1)+\lambda_2\log(\lambda_2)+\lambda_3\log(\lambda_3)+\lambda_4\log(\lambda_4))$$

under the constraints

$$\begin{aligned} C &= \lambda_1-\lambda_3-2\sqrt{\lambda_2\lambda_4} && (215)\\ 1 &= \lambda_1+\lambda_2+\lambda_3+\lambda_4 && (216)\\ \lambda_1 &\geq \lambda_2\geq\lambda_3\geq\lambda_4\geq 0. && (217)\end{aligned}$$

If we assume that $\lambda_4>0$, then we introduce $x=\sqrt{\lambda_2/\lambda_1}, y=\sqrt{\lambda_3/\lambda_1}, z=\sqrt{\lambda_4/\lambda_1}$, and the optimization problem can shown to lead to the extremal conditions:

$$\begin{aligned} x\log(x) &= z\log(z) && (218)\\ x\log(x)+z\log(z) &= (x+z)\log(y). && (219)\end{aligned}$$

This set of transcendent equations has two different solutions, as the function $x\log(x)$ is one-to-two in the interval $[0:1]$. The obvious solution is when $x=y=z$, giving rise to a Werner state. The other solution cannot be written down analytically, but can easily be calculated numerically, and will turn out to lead to a suboptimal extremum. We still have to consider the case where $\lambda_4=0$. The optimal solution can easily be derived in this case, and gives the condition that $\lambda_2=\sqrt{\lambda_1\lambda_3}$; this implies that

$$\lambda_3=\left(\frac{-\sqrt{\lambda_1}+\sqrt{4-3\lambda_1}}{2}\right)^2 \qquad (220)$$

and recall that the concurrence is given by $\lambda_1-\lambda_3$ in this case. The fourth possible solution corresponds to the case where $\lambda_3=\lambda_4=0$, for which $C=\lambda_1=1-\lambda_2$.

The entropy of these four extremal solutions in function of their concurrence has been plotted in Figure 13. We see that the Werner states maximize the entropy if the concurrence is smaller than $\simeq .3$, but for higher values of the concurrence, the extremal rank 3 states that were explicitly derived maximize the entropy; this is rather surprising and arises because of the strange functional behaviour of the entanglement of formation for given eigenvalues of the maximally entangled mixed states. Moreover, it is therefore proven that a state with entropy exceeding

$$1+\frac{1}{2}\log_2(3)\simeq 1.7925 \qquad (221)$$

cannot be entangled: this is the maximal possible entropy for an entangled state of two qubits.

It can be shown that a similar behaviour occurs in the case of the maximization of the entropy for given amount of relative entropy of entanglement: the Werner states maximize the entropy for states with few entanglement, and rank 3

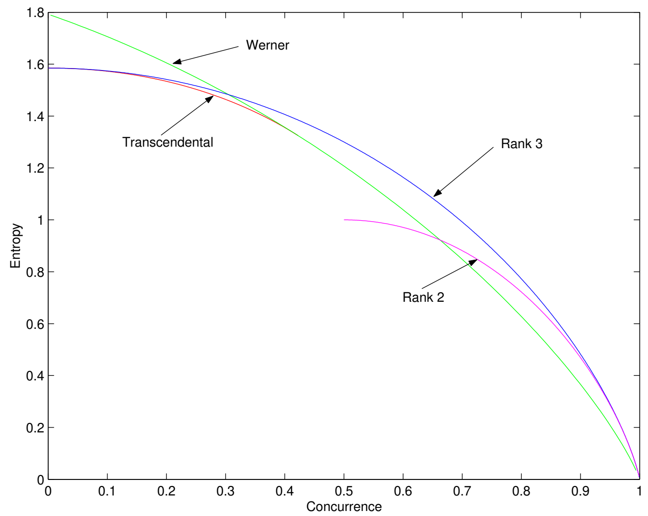



**Figure 13.** The entropy of all two qubit states with a fixed concurrence or entanglement of formation is bounded above by the highest curves. Conversely, the entanglement of formation for states with fixed entropy is bounded above by the same curves. For small concurrence, the Werner states are extremal, while for larger concurrence, specific rank 3 states are.

maximally entangled mixed states maximize the entropy for higher values. In the case of negativity versus entropy however, the Werner states maximize the entropy over the whole range of values. We refer to the paper [**234**] for more details.

The previous properties are fundamental in the light of the fact that we have derived the ultimate amount of entanglement that can be present in a 2-qubit state given a fixed entropy. In analogy with thermodynamics[**122**], it is tempting to call the obtained states representative states: if one associates entanglement to energy, and entropy to entropy, and one is willing to apply Jaynes principle to an ensemble in equilibrium with given amount of entanglement, then the most probable state is the one just derived. This physical interpretation should however not be taken too far.

As a last remark, it is interesting to note that the entanglement measures Negativity, Relative Entropy of Entanglement, Fidelity and CHSH-violation are all minimized for given entanglement of formation by the rank 2 maximally entangled mixed states. We have also just proven that their entanglement cannot be increased by any global unitary transformation. One could therefore argue that these states are the “worst”ones to prepare: one needs a lot of entanglement of formation to prepare them, but all the entanglement monotones related to distillation are minimal for this given amount. This picture corresponds with the fact that their entropy is maximal for given EoF. However, these states are also exactly the quasi-distillable ones with a non-diagonal normal form, and are therefore very useful as they can easily be distilled.

## 4.8. The geometry of separable and entangled states

In this section we try to get some insight into the geometrical structure of separable and entangled states. The main goal will be to characterize the distance of an entangled state to the set of separable states as measured with the Hilbert-Schmidt distance. This will yield a nice illustration of the convex set of separable states.

The study of the convex set of separable states turns out to be remarkably difficult[18]. Only for the $2 \times 2$ and the $2 \times 3$ case, necessary and sufficient conditions for separability of a quantum state have been found based on the partial transpose criterion [**120**]. In higher dimensional systems, no easy way of determining the separability of a state exists due to the existence of bound entangled states [**123**]. The concept of negativity, being a quantitative measure to what extent a quantum state violates the partial transpose criterion, will turn out to be very much related to the Hilbert-Schmidt distance of a state to the set of PPT-states. As the PPT-criterion is the only sensible separability criterion known, we will content ourselves to calculate the Hilbert Schmidt distance of an entangled state to the set of PPT-states (Note that in the case of two qubits no bound entangled states exist). Related questions were addressed in the papers of Zyczkowski et al.[**250, 249**] and Witte and Trucks [**241**] (see also Ozawa [**167**]), although these papers mainly focused on different issues and therefore only solved very special instances of the general results presented here.

The problem we want to tackle is highly related to calculating the distance of an entangled state to the set of partially transposed states, as the intersection of the set of all states with the set of all partially transposed states is equal to the set of all PPT-states. This is visualized in Figure (14), where the boundary of the convex set of states $H$ consists of rank deficient states. The set of partially transposed states is completely isomorphic with the set of states, and can be seen as some kind of reflection of the set of states. The intersection of both sets is the convex set of PPT-states.

From Figure (14) it is immediately clear that the distance of an entangled state to the PPT ones is equal to the distance of an entangled state to the set of partially transposed states iff the closest partially transposed state is positive (semi)-definite; this condition will turn out to be almost always true.

[18] We refer to Doherty et al. [**76**] and Gurvits [**103**] for some recent interesting efforts to attack the problem from a linear algebraic perspective.

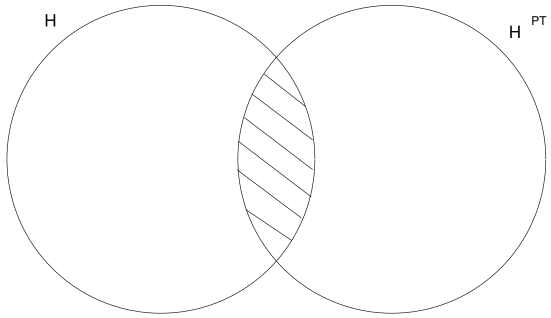



**Figure 14.** The set of all states is depicted by $H$ and the set of all partial transposed states by $H^{PT}$. The intersection of both is the set of all PPT-states.

Let us now calculate the closest partially transposed state to an entangled state. The key observation is the fact that the Hilbert Schmidt norm is preserved under the partial transpose map. Therefore the proposed measure can be defined in the space of partial transposed density matrices as the minimal Hilbert Schmidt distance of $\rho^{PT}$ to the surface of positive (semi)-definite matrices with trace 1, this surface being the partial transpose of the boundary of PPT-states.

We are therefore looking for the best positive semidefinite approximation of the indefinite matrix $\rho^{PT}$ in the Hilbert-Schmidt norm:

$$\min_{\sigma\in H} \|\rho^{PT} - \sigma\|_F = \sqrt{\mathrm{Tr}\left((\rho^{PT} - \sigma)^2\right)} \tag{222}$$

Writing the eigenvalue decomposition $\rho^{PT} = UDU^\dagger$, and absorbing $U$ in $\sigma$, this problem is equivalent to finding $\sigma$ such that $\|D-\sigma\|$ is minimal. Using the eigenvalue decomposition $\sigma = VE^2V^\dagger$ with $\mathrm{Tr}\left(E^2\right) = 1$, this can be written as a Lagrange constrained problem with cost function:

$$K = \|D - VE^2V^\dagger\|_F - \lambda\left(\mathrm{Tr}\left(E^2\right) - 1\right) \tag{223}$$

It is immediately clear that the optimal unitary $V$ is given by the identity: a positive definite matrix remains positive definite if off-diagonal elements are made zero. Differentiation leads to the result that the $e_i^2$ are either equal to 0, either equal to $d_i + \lambda$. Normalization fixes the value of $\lambda$. Straightforward calculations show that the $e_j^2$ corresponding to the negative eigenvalues $d_j$ have to be chosen equal to zero and the other ones either equal to $d_i+\lambda$ either equal to 0, depending on the sign of $d_i + \lambda$. The algorithm for finding the closest partially transposed state therefore becomes:

(1) Calculate the eigenvalue decomposition of $\rho^{PT} = UDU^\dagger$

(2) Define $E^2$ as the unique diagonal positive (semi)-definite normalized matrix such that its elements are $e_i^2 = d_i + \lambda$ or $e_i^2 = 0$.

(3) The closest partially transposed state $\rho_s$ is given by $\rho_s = \left(UE^2U^\dagger\right)^{PT}$. The Hilbert Schmidt distance between both states is given by

$$\|\rho - \rho_s\|_2 = \sqrt{\frac{(\sum_{i\in I_p} d_i + \sum_{i\in I_n} d_i)^2}{n_p} + \sum_{i\in I_n} d_i^2}, \tag{224}$$

where $I_n$ is the set of all indices corresponding to the negative eigenvalues of $\rho^{PT}$, $I_p$ is the set of indices corresponding to positive eigenvalues of $\rho^{PT}$ but for which $e_i^2 = 0$, and $n_p$ denotes the rank of $E^2$.

If $\rho_s$ is a state, it is guaranteed to be the closest PPT-state. Numerical investigations show that for example in the two qubit case the positiveness of $\rho_s$ happens in approximately 97% of the cases. If $\rho_s$ is not positive, then the distance to the set of partially transposed states calculated is a (fairly good) lower bound on the distance of the entangled state to the set of PPT-states.

Let us illustrate the above procedure with an example. Say we want to find the closest biseparable $2 \times 4$ state to the three qubit $W$-state [**80**] $|W\rangle = (|001\rangle + |010\rangle + |100\rangle)/\sqrt{3}$. The eigenvalue decomposition of $(|W\rangle\langle W|)^{PT} = UDU^\dagger$, with the partial transpose operation taken over the 4-dimensional Hilbert space, is given by:

$$\begin{aligned} D &= \mathrm{diag}\begin{pmatrix} 2/3 & \sqrt{2}/3 & 1/3 & 0 & 0 & 0 & 0 & -\sqrt{2}/3 \end{pmatrix}, \quad (225)\\ U &= \begin{pmatrix} 0 & 1/\sqrt{2} & 0 & 0 & 0 & 0 & 0 & -1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 0 & 1/\sqrt{2} & 0 & 0 & 0 & 0 \\ 1/\sqrt{2} & 0 & 0 & -1/\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 0 & 1/\sqrt{2} & 1/2 \\ 0 & 1/2 & 0 & 0 & 0 & 0 & -1/\sqrt{2} & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}. \quad (226) \end{aligned}$$

The eigenvalues $E^2$ are readily obtained:

$$E^2 = \mathrm{diag}\begin{pmatrix} 2/3 - \sqrt{2}/9 & 2\sqrt{2}/9 & 1/3 - \sqrt{2}/9 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{227}$$

Taking the partial transpose leads to the state $\rho_s$, where we used the notation $c = \sqrt{2}/18$:

$$\begin{pmatrix} 2c & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/3-c & 1/3-c & 0 & 1/9 & 0 & 0 & 0 \\ 0 & 1/3-c & 1/3-c & 0 & 1/9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/9 & 1/9 & 0 & 1/3-2c & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & c & c & 0 \\ 0 & 0 & 0 & 0 & 0 & c & c & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{228}$$

The eigenvalues of $\rho_s$ are non-negative and it is possible to show that $\rho_s$ is separable. We have therefore found the closest biseparable $2 \times 4$ state to the $|W\rangle$-state, and the Hilbert Schmidt distance to it is equal to $(2/3)^{3/2}$. Recently, the question arose whether the set of W-type states is of measure zero. Using the language of the Hilbert-Schmidt distance, this problem is readily solved. Indeed, the question is solved if we can prove that the state obtained by mixing the W-state with a small random completely separable mixed state remains outside the set of all convex combinations of biseparable states (with relation to whatever partition). As there is no biseparable pure state infinitesimally close to the W-state, and a mixed state not infinitesimally close to a pure state is always at a finite distance from whatever pure state, it is proved that the set of W-type states is indeed not of measure zero. A different proof was given by Acin et al. [**2**]. Note that the above proof is very general and can be used in systems of arbitrary dimensions: whenever there exists a pure state $\psi_1$ that can probabilistically be converted into another one $\psi_2$ but not vice-versa, the set of $\psi_1$-like states minus the set of the $\psi_2$-like states is of finite measure if there does not exist a $\psi_2$-like state infinitesimally close to $\psi_1$!

The concept of negativity is also connected to the concept of robustness of entanglement [**228**]. Indeed, let us calculate how much an entangled bipartite state of whatever dimension has to be mixed with the identity before it gets PPT. In analogy with the previous derivation of the Hilbert-Schmidt distance, this amounts to the equivalent problem of how much one has to mix the partial transpose of $\rho$ with the identity before it gets positive semi-definite:

$$\min_t (1-t)\rho^{PT} + \frac{t}{4} I_4 \geq 0 \tag{229}$$

This problem is readily solved, and the solution is

$$t = \frac{|d_{min}|}{|d_{\min}| + \frac{1}{4}} \tag{230}$$

where $d_{\min}$ is the minimal negative eigenvalue of $\rho^{PT}$. The minimal $t$ is therefore only a function of the negative eigenvalues. A geometrical implication of this fact is that all surfaces of constant $d_{\min}$ are similar to the boundary of separable and entangled states: the set of all states with constant $d_{\min}$ can be generated by extrapolating all lines from the identity to the boundary of separable states such that the distance of the extrapolated state to the identity is a constant factor ($> 1$) of the distance of the separable state to the identity.

Let us now move to the case of two qubits. In this case $\rho^{PT}$ has at most one negative eigenvalue, as was proven earlier. Numerical investigations indicate that in a vast majority of the states the optimal rank of $E^2$ is equal to three, and if the rank is equal to two it implies that $\rho_s$ has a negative eigenvalue. For the states for which $E^2$ is rank 3, it follows that their distance to the set of partially transposed states is given by

$$\|\rho - \rho_s\| = \frac{2}{\sqrt{3}}|d_{\min}| \tag{231}$$

where $d_{\min}$ is the negative eigenvalue of $\rho^{PT}$. Surfaces of two-qubit states with constant negativity, defined as $N = 2|d_{\min}|$, have therefore two distinct properties: they are all similar to each other and the Hilbert-Schmidt distance between them is constant almost everywhere.

Let us illustrate the above findings by explicitly calculating some two-dimensional intersections of the set of all bipartite qubit states including the maximally mixed state. In the following figures we use the metric based on the Hilbert-Schmidt distance $\|\rho_1 - \rho_2\|^2 = \mathrm{Tr}\left((\rho_1 - \rho_2)^\dagger(\rho_1 - \rho_2)\right)$, and directions represented orthogonal to each other are orthogonal in the sense that $\mathrm{Tr}\,(A_1 A_2) = 0$. Rank deficient density operators always lie on the boundary of the intersection.

Note that an explicit parameterization of the boundary between the entangled and separable states can easily be obtained: it is at most a quartic function of the mixing parameters of the states, as their analytic expression can be obtained by setting the determinant of the partial transpose equal to zero.

As a first example we consider the plane containing the maximally mixed state and the pure states

$$\rho_1 = \frac{1}{2}\begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & 1 & 0 \end{pmatrix} \qquad \rho_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} \tag{232}$$

The plane is plotted in Figure (15) and the boundary of all (rank-deficient) states is given by the solid envelope. The starred line is the boundary between the convex set of separable states and the convex set of all states. The surfaces of constant negativity are indeed all similar to this boundary. The fact that the distance between these surfaces is not constant throughout the picture indicates that the closest separable states lie in other planes. Note that the Werner states lie along the line between the maximally mixed state and the maximally entangled state $\rho_1$. The third extremal point in the undermost left

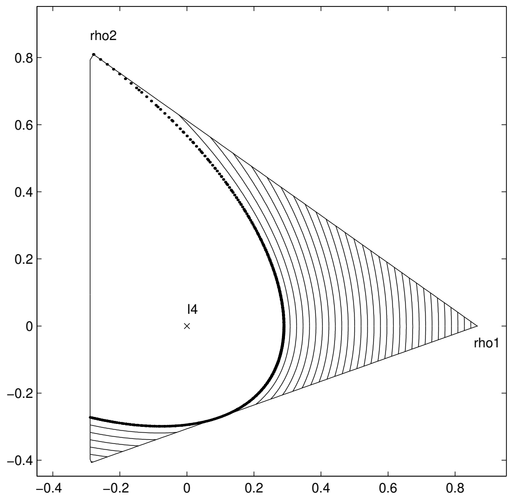



**Figure 15.** Intersection of the convex set of all states including states (232) and the maximally mixed state. The contours represent surfaces of constant negativity, the starred line is the boundary between separable and entangled states. The blank inner space depicts the shape of the convex set of separable states.

corner is given by the rank 2 state

$$\rho = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & -\frac{1}{4} & 0 \\ 0 & -\frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{pmatrix} \tag{233}$$

This state is the quasi-distillable state with a whole lot of remarkable properties, such as the fact that no global unitary operation can increase its entanglement.

Let us now consider a different plane including the maximally mixed state and the pure states

$$\rho_1 = \frac{1}{2}\begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & 1 & 0 \end{pmatrix} \qquad \rho_2 = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}\begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix} \tag{234}$$

This plane is obtained by rotating the previous plane around the axis $\rho_1 - I_4$. In this case $(\rho_1 - I_4)$ is orthogonal to $(\rho_2 - I_4)$, and a completely different picture is obtained as shown in Figure (16, left). Further rotation of the plane leads to the following states:

$$\rho_1 = \frac{1}{2}\begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & 1 & 0 \end{pmatrix} \qquad \rho_2 = \frac{1}{2}\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix} \tag{235}$$

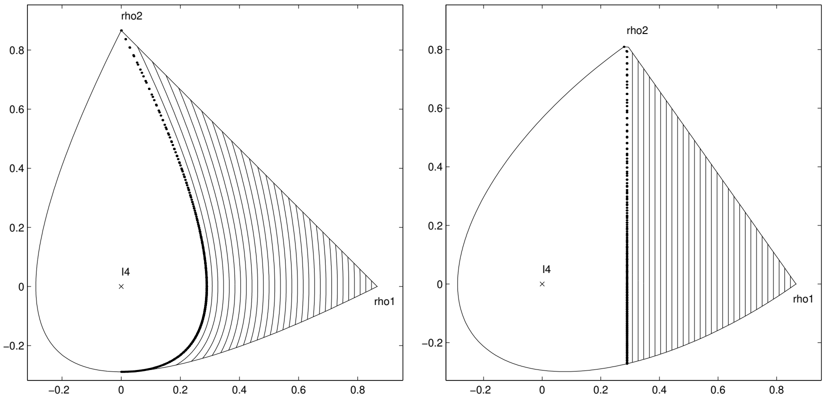



**Figure 16.** Intersection of the convex set of all states including the maximally mixed state and states (234) for the left picture or states (235) for the right one. The blank inner space depicts the shape of the convex set of separable states.

The intersection of the state space by this plane is shown in Figure (16, right).

The surfaces of constant negativity become straight lines, implying that the closest separable states lie in the same plane: the Hilbert-Schmidt distance between the surfaces of constant negativity has to be constant if they consist of parallel planes. Using the procedure previously outlined, it is indeed trivial to check that the separable state closest to the maximally entangled state $\rho_1$ lies in the defined plane and is given by

$$\rho_s = \begin{pmatrix} \frac{1}{6} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{6} & 0 \\ 0 & \frac{1}{6} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{6} \end{pmatrix} \tag{236}$$

Let us rotate the plane further over the $(\rho_1 - I_4)$-axis:

$$\rho_1 = \frac{1}{2}\begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & 1 & 0 \end{pmatrix} \qquad \rho_2 = \frac{1}{101}\begin{pmatrix} 10 \\ 0 \\ 0 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 10 \end{pmatrix} \tag{237}$$

The resulting Figure (17, left) combines the features of the previous figures. Three entangled disconnected regions arise, and once more we observe the strange shape of the boundary between entangled and separable states.

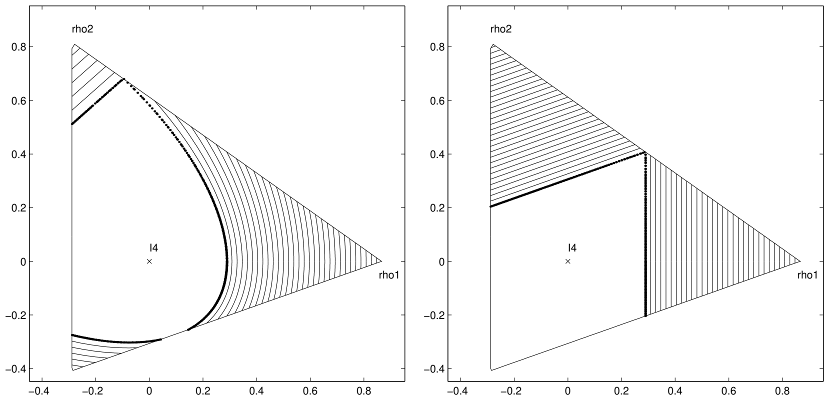



**Figure 17.** Intersection of the convex set of all states including the maximally mixed state and states (237) for the left picture or (238) for the right one. The blank inner space depicts the shape of the convex set of separable states.

A plane with a highly symmetric contour lines is obtained if $\rho_1$ and $\rho_2$ are both taken to be maximally entangled states:

$$\rho_1 = \frac{1}{2}\begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & 1 & 0 \end{pmatrix} \qquad \rho_2 = \frac{1}{2}\begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & -1 & 0 \end{pmatrix} \tag{238}$$

Indeed, only straight lines are obtained in Figure (17, right). The third extremal state is in this case given by $\rho = \mathrm{diag}[1/2; 0; 0; 1/2]$.

At last, we choose two random planes through the maximally mixed state and plotted them in Figure (18). The similarity of all planes with constant negativity is clearly illustrated.

In conclusion, we have analyzed the space of density operators from the perspective of the Hilbert-Schmidt metric. Because of the nice property that the Hilbert-Schmidt between two states is equal to the Hilbert-Schmidt metric between the partial transposes of the states, a very simple technique arose for calculating the closest PPT state to an entangled state. Moreover, we obtained some intuition about the shape of the space of separable states by plotting $2D$-surfaces through the convex set of states.

## 4.9. Entanglement of Assistance

The concept of entanglement of assistance was introduced by DiVincenzo et al. [**71**] and is strongly related to a problem raised by Cohen [**60**]. They raised

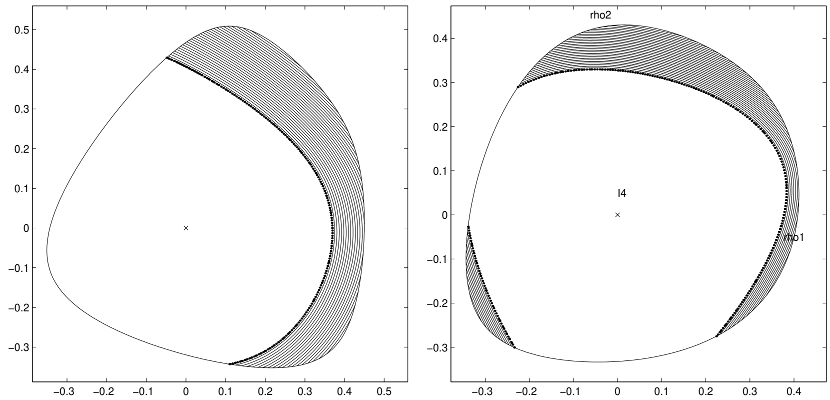



**Figure 18.** Contour plots of the negativity on random planes including the maximally mixed state. The blank inner space depicts the shape of the convex set of separable states.

the following interesting question: if a mixed bipartite system $\rho$ is viewed as a pure state in a tripartite system where one party is traced out, how much entanglement can this third party (C) create between Alice and Bob by doing local measurements? By the quantum steering theorem, C can induce whatever convex decomposition of the density operator of Alice and Bob, and therefore this leads to the following optimization problem:

$$E_A(\rho) = \max_{\{p_i,|\psi_i\rangle\}} \sum_i p_i E(|\psi_i\rangle) \tag{239}$$

where $\{p_i, \rho_i\}$ is a convex decomposition of $\rho$. This problem looks very much like the problem of determining the entanglement of formation, but with a maximization instead of a minimization, and therefore very similar techniques can be used.

Let us try to calculate the entanglement of assistance explicitly. Using exactly the same lines of thought as in the proof of the entanglement of formation (theorem 21), we first optimize the average concurrence over all isometries $U$:

$$\max_U \sum_i |U^T X^T \epsilon_2 \otimes \epsilon_2 X U|_{ii} \tag{240}$$

where $X$ is a square root of $\rho$. The average concurrence is immediately seen to be optimized by the unitary $U = V^*$ with $V\Sigma V^T$ the Takagi decomposition of $Q = X^T\epsilon_2 \otimes \epsilon_2 X$. Therefore the concurrence of assistance $C_A$ is defined as the sum of all singular values or as the trace-norm of $Q$ (as opposed to the difference of the largest one and the sum of the smallest ones in the case of EoF). In the case of EoF, we proceeded to prove that all concurrences could be chosen to be equal to each other, such that the entanglement of formation (being a convex monotonously increasing function of the concurrence for a pure state) would also be minimized. Here however we want to obtain the opposite: we want to maximize the expected entanglement, which is much more difficult.

The unitary $U$ bringing $Q$ into diagonal form is not unique in achieving the largest possible average concurrence; indeed, we can multiply $U$ at the right with whatever real orthogonal $O$ and still the same average concurrence will be obtained. A good choice of $O$ would be such that the pure states in the decomposition have or a large amount of entanglement, or a small amount: this way the convex sum will indeed be maximized. Using again similar (but opposite) tricks as in the proof of entanglement of formation, we have to choose $O$ such that the sum of the absolute value of the diagonal elements of the matrix

$$O^T(\Sigma - \mathrm{Tr}\,(\Sigma)\,U^\dagger X^\dagger XU)O \tag{241}$$

is as large as possible (indeed, this corresponds to the situation where the individual concurrences are as far away possible from the average concurrences). This can easily be done by calculating the eigenvalue decomposition of the real part of $R = \Sigma - \mathrm{Tr}\,(\Sigma)\,U^\dagger X^\dagger XU$, and associating $O$ to the orthogonal matrix of eigenvectors.

It turns out that this procedure does only lead to a lower bound on the entanglement of assistance, but a bound that is a very good one (maximal only $\pm 2\%$ away of the true value obtained by numerical investigation). It is not clear however how to find the value of the $E_A$ analytically, as we will have to take into account the exact shape of the convex monotonously increasing function $f$, and one also has to take into account that POVM measurements corresponding to isometries can do better (note that in the case of maximizing $C_A$ von-Neumann measurements were sufficient). Nevertheless, $f(C_A)$ generally already produces a very good lower bound (maximal only $\pm 2\%$ lower than the actual value). Moreover the concurrence of assistance is an entanglement monotone, and therefore the question of maximizing the concurrence of assistance is of interest on its own.

Note that the above derivation was constructive in that it explicitly gave an expression of the measurement to be done by $C$ to maximize $C_A$: $C$ has to implement the von-Neumann measurement corresponding to $UO$ (on his quartit) and classically send his measurement outcome to both Alice and Bob.

As an example, consider the maximally mixed state $\rho = I/4$. This is the density operator obtained by tracing out $C$, and as this is the locally stochastic matrix, we indeed obtain the result that the concurrence of assistance and henceforth the entanglement of assistance is 1, the maximal achievable value. It is therefore perfectly possible that the original density operator $\rho$ is not entangled while the entanglement of assistance is maximal.

Let us now consider the following variant of the entanglement of assistance problem, a problem raised by Steven van Enk [**146**]: a mixed state of two qubits can also be seen as a pure fourpartite state of qubits where two parties (C and D) are traced out. The problem we want to address is the following: given a pure fourpartite state of qubits $|\psi\rangle$, what measurements do $C$ and $D$ have to implement such that $A$ and $B$ end up with the maximal amount of entanglement. In some sense, one could call this a multipartite version of quantum steering: we want to locally steer the states of Alice and Bob into the direction with most entanglement. This work has an interesting connection with the work of Walgate et al.[**232**], where it was shown that every two orthogonal pure multipartite states can be distinguished locally if classical communication between the parties is allowed. The mathematics involved in the current problem is very similar.

The following questions arise:

- Can local measurement of C and D do as well as global measurement strategies (as in the original EoA problem)?
- Will communication between C and D help them?
- Can POVM measurements do better than von-Neumann measurements?

We will only consider the case of maximizing the concurrence of assistance. Note that in this setting $X$ is not just a square root of $\rho$ anymore but is a reshaped version of the fourpartite state $|\psi\rangle$ (note that in the tripartite case we did not need this information as all bipartite purifications are unitarily equivalent). We will have to make heavily use of the fact that the von-Neumann measurement by $C$ in the previous setting was not unique: given one such an optimal $U$, then any measurement $V$ of the form

$$V = UO \underbrace{\begin{pmatrix} e^{i\phi_1} & 0 & 0 & 0 \\ 0 & e^{i\phi_2} & 0 & 0 \\ 0 & 0 & e^{i\phi_3} & 0 \\ 0 & 0 & 0 & e^{i\phi_4} \end{pmatrix}}_{D} \tag{242}$$

with $O$ real orthogonal will do as good (note that this characterization exhausts all optimal von-Neumann measurements). The question is whether $O$ and $D$ can be chosen such that $V$ becomes of the form

$$V = (W \otimes I) \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix}. \tag{243}$$

Indeed, this type of $V$ corresponds to the situation where C measures in the $W$-basis, communicates his result to D, who measures in the $V_1$ or $V_2$ basis according to the result of C. We will first show that given a $W, V_1, V_2$, there exists a $W', V_1', V_2'$ that does as good but where $V_1' = V_2'$. Indeed, $V_1$ and $V_2$ can be multiplied at the right side by a real $2 \times 2$ orthogonal and a diagonal $2 \times 2$ unitary $D$. We now make use of the following nice lemma:

**Lemma 10.** *Given a unitary $n \times n$ matrix $U$, then there always exist real orthogonal $O_1$ and $O_2$ such that $O_1^T U O_2 = D$ is diagonal (and of course unitary). Therefore each unitary can be written as the product of an orthogonal matrix multiplied by a unitary diagonal matrix only containing phases and another orthogonal one.*

*Proof:* The proof is simple: take the real part of $U_r = \mathcal{R}(U)$ and calculate the singular value decomposition $U_r = O_1 \Sigma O_2$. Now define $Q_i = \mathcal{I}(O_1^T U O_2)$, then the conditions that $U$ is unitary translate into

$$\begin{aligned} I &= \Sigma^2 + Q_i^T Q_i = \Sigma^2 + Q_i^T Q_i &(244)\\ 0 &= \Sigma Q_i^T - Q_i \Sigma &(245) \end{aligned}$$

The first equation implies that $Q_i$ is normal and henceforth has real orthogonal eigenvectors, and moreover $\Sigma^2 + Q_i^T Q_i$ has to be diagonal; this is only possible if $Q_i$ itself is diagonal. Moreover the decomposition is unique if all phases are different. □

We can readily apply this Lemma to our problem: it can easily be checked that given two unitary matrices $V_1, V_2$, there always exist a $O_1, O_2, D$ such that $V_1 O_1 = V_2 O_2 D$. We have therefore proven that in the setup where both C and D do von-Neumann measurements, communication will not enhance the concurrence of assistance (for more information concerning degenerate cases, we refer to [**146**]).

We have not yet considered the problem whether joint measurements can do better than local measurements. This would not be the case if each unitary matrix could be written in the form

$$U = (V \otimes W) D O_1. \tag{246}$$

Using the accident $SU(2) \otimes SU(2) \simeq SO(4)$, this is equivalent to

$$TU = O_2 T D O_1 \tag{247}$$

with $T$ defined in (34). However, it can easily be checked that the decomposition of $TD$ along Lemma 10 yields a diagonal $D'$ with phases $(\phi_1, -\phi_1, -\phi + \pi/2, \phi - \pi/2)$. This is clearly not the generic case, and therefore we have proven that global unitary operations will typically do (a little bit) better than local. Note however that we only came two degrees of freedom short to implement the global unitaries with local unitaries. Further numerical studies by T. Laustsen [**146**] indicated that POVM measurements of C and D can close the gap a little bit, but no nice analytical results have been found in this direction.

## 4.10. Conclusion

Using some advanced linear algebra, we have been able to unlock a whole lot of secrets about mixed states of two qubits. The central result was the introduction of the Lorentz singular value decomposition, which enabled to separate the local degrees of freedom from the global ones responsible for entanglement. Based on this decomposition, we were able to give a unified description of entanglement measures. It also enabled to generalize the quantum steering Theorem to mixed states of two qubits, revealing an appealing geometrical representation of all possible states. We discussed a whole range of different entanglement measures, and were able to derive lower and upper bounds for them. Next the optimal teleportation scheme with mixed states was derived, the best available distillation protocols were devised, and the concept of maximally entangled mixed states was introduced. Finally, we approached the problem of separability based on the geometry induced by the Hilbert-Schmidt norm, and discussed some results concerning the entanglement of assistance.

# Part 2

# QUANTUM INFORMATION

In the second part of this thesis, we focus on quantum information and on quantum channels.

In chapter 5, we give a brief introduction of how to use quantum systems to encode classical information, and how measurements have to be done on quantum systems to extract the classical information. Next we treat a problem that lies at the heart of quantum information theory: how can one extract information from a quantum system without disturbing it too much? Optimal strategies will be devised to do parameter estimation of dynamical systems, and it turns out that some gain is obtained by adopting continuous (but infinitesimally weak) measurement strategies. Moreover, an intriguing connection between Kalman filtering and the evolution of observed quantum systems will be discussed.

In the last chapter, we use the techniques developed in the first part of this thesis to present a unified description of quantum channels. Quantum channels describe the evolution of quantum systems during their transportation of one party to another one, and a good understanding of their action is therefore compulsory if one wants to distribute quantum information. The problem turns out to be equivalent (or at least dual) to the one of describing entangled systems, and therefore the results of the first part of this thesis are translated to this dual picture. The extreme points of the convex set of trace-preserving maps are derived, we discuss new issues about the classical and quantum capacity of a quantum channel, and especially in the case of qubit channels nice normal forms are given. At the end of the chapter, the asymptotic entanglement capability of a certain class of Hamiltonians is calculated explicitely.

CHAPTER 5

# Classical Information by Quantum Measurements

## 5.1. Measurement of Qubits

In this section we will consider the simplest of all quantum systems: a quantum state described in a Hilbert space of dimension 2, corresponding to a spin 1/2 object. In quantum information processing, this system plays a role comparable to a bit in classical information processing; therefore B. Schumacher coined the term "qubit"[**188**].

A density operator $\rho$ corresponding to a qubit is represented by a $2 \times 2$ Hermitian positive-semidefinite matrix with trace equal to 1. The Pauli matrices

$$\sigma_0 = I_2 \qquad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{248}$$

form a complete orthogonal basis for all Hermitian $2 \times 2$ matrices (i.e. $\mathrm{Tr}\sigma_i\sigma_j = 2\delta_{ij}$), and it follows that a density operator $\rho$ has a unique decomposition of the form

$$\rho = \frac{1}{2}\left(I_2 + \sum_{i=1}^{3} x_i\sigma_i\right) \tag{249}$$

$$x_i = \mathrm{Tr}(\sigma_i\rho) \tag{250}$$

with the coefficients $x_i$ real. This defines the Bloch vector $x = (x_1, x_2, x_3)$. The condition of positivity corresponds to the fact that the Bloch vector obeys the relation

$$\|x\|_2 \leq 1,$$

and therefore a state is uniquely parameterized by a point inside a sphere of radius 1 (note that pure states lie on the boundary, and orthogonal pure states have the opposite coordinates).

An operational way of parameterizing a density operator is by specifying the probabilities of obtaining results $1 \cdots n$ of a certain measurement on the system. The most general measurement corresponds to a POVM with elements $\{E_\alpha\}, E_\alpha > 0, \sum_\alpha E_\alpha = I$. As all $E_\alpha$ are Hermitian and positive definite, specifying the probabilities $\{p_\alpha = \mathrm{Tr} E_\alpha \rho\}$ of a POVM with 4 linearly independent elements $E_\alpha$ uniquely characterizes the state $\rho$. Given whatever POVM with 4 linearly independent elements $E_\alpha$, the density operator of a qubit can therefore be represented by a 4-dimensional probability distribution $\{p_\alpha\}$. Note that in this setting it makes no sense to consider POVM's with more elements as the probabilities associated to them would be linearly dependent.

Quantum mechanics however places severe constraints on the possible probabilities $p_\alpha$ corresponding to a fixed measurement $\{E_\alpha\}$. We would like to identify all feasible probability distributions $p_\alpha$ (see also [**171**]). Let us first of all parameterize the 4D vector $p_\alpha$ (with three degrees of freedom) by the 3D-vector $q_\alpha$:

$$\begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & -1/3 & -1/3 & -1/3 \\ 0 & 2\sqrt{2}/3 & -\sqrt{2}/3 & -\sqrt{2}/3 \\ 0 & 0 & \sqrt{2/3} & -\sqrt{2/3} \end{pmatrix}}_{S} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix}. \tag{251}$$

Note that the matrix $S$ contains the coordinates of the vertices of a tetrahedron. In this new parameterization, all possible probability distributions (so even the ones that are not allowed by quantum mechanics) live inside a simplex centered around the origin. As all $E_\alpha$ are Hermitian and positive, we can associate a real 4D-vector $e_\alpha$ with components $e_i^\alpha = \mathrm{Tr}(E_\alpha \sigma_i), i = 0 \cdots 3$ to them. If the state under consideration has Bloch vector $x$, then the probabilities $p_\alpha(x)$ to obtain outcome $p_\alpha(x)$ can easily be shown to be given by the expressions $(e^\alpha)^T[1; x]$. The corresponding $q_\alpha$ are therefore given by

$$\begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} = S \begin{pmatrix} e_1^T \\ e_2^T \\ e_3^T \\ e_4^T \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} = q_0 + Rx. \tag{252}$$

As all elements of the POVM are linearly independent, $R$ is full rank and we can invert this expression to obtain $x = R^{-1}(q - q_0)$. The condition that $x$ corresponds to a physical state is just given by the fact that its Bloch vector lies inside the Bloch sphere. Therefore the probabilities $p_\alpha$ parameterized by $q_\alpha$ can be realized iff

$$(q - q_0)(RR^T)^{-1}(q - q_0) \leq 1. \tag{253}$$

This is the expression of an ellipsoid. We have therefore obtained the result that all possible probability outcomes when a particular POVM is implemented are contained within an ellipsoid inside the probability simplex (see figure (1)). Not surprisingly, the equation of the ellipsoid encodes all the information of the complete POVM-measurement performed, up to a local unitary operation (this last fact follows from the fact that $RR^T$ and $R'R'^T$ are equal to each other iff $R' = RO$ with $O$ orthogonal, and that a unitary matrix acting on a density operator corresponds to a rotation in the Bloch sphere). Conversely, each point inside the ellipsoid corresponds to exactly one state if the POVM was complete. Note that the center of the ellipsoid is solely determined by the weights attributed to each POVM-element. Note also that the number of points where the ellipsoid touches the probability simplex is given by the number of pure elements of the POVM: this follows from the observation that for each pure POVM there exists a state orthogonal to it for which the probability of obtaining this measurement result is equal to zero, and that the only points in the probability simplex corresponding to a zero probability lie on the faces of the simplex.

As an example, consider the following extreme cases. An orthogonal measurement corresponds to an ellipsoid that is completely squeezed in two directions: the ellipsoid is reduced to a line connecting two vertices of the tetrahedron. The two vertices are the images of two orthogonal states. On the other hand, the measurement corresponding to the ellipsoid with the largest possible volume is the one corresponding to a sphere with radius $1/3$ touching the tetrahedron at each face. This measurement corresponds to the completely symmetric POVM with 4 elements each having an image in the Bloch sphere at a vertex of a tetrahedron. On the other hand, a projective measurement corresponds to a completely squeezed ellipsoid that coincides with one of the vertices.

This geometrical picture is a nice illustration of a central feature of an observed quantum mechanical system: quantum measurements can only reveal a limited amount of information, and the more precise they are on one pair of orthogonal directions (e.g. a projective measurement), the less it is in all other ones (this can easily be seen by the fact that the ellipsoid has to fit into the tetrahedron). As will be shown in the following section, a qubit can at most encode one bit of information[1].

## 5.2. Classical Information encoded in Quantum Systems

The transmission of classical information always occurs in the following way: one encodes the bits one wants to send in a physical system (e.g. a light pulse), the physical system is transported (e.g. through a fibre), and finally a measurement is performed on the system. The ultimate physical limit by which we can transmit information occurs when a single quantum particle (e.g.

[1]It should be possible to find an explicit proof of this fundamental issue solely based on the fact that all classical probabilities are confined to lie in an ellipsoid.

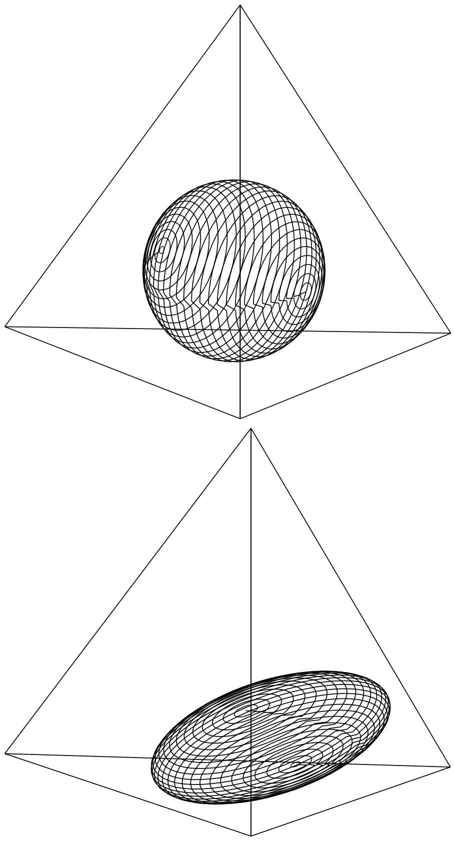


**Figure 1.** Classical probability simplex and the allowed quantum probabilities of measurements of qubits by the completely symmetric POVM (left) or by a randomly chosen complete POVM with four pure elements (right).

a photon) is the information carrier. It is clear that in that case we have to take into account typical quantum effects, and that future telecommunication engineers will have to master the tools to describe those effects.

The aim of this section is to highlight some important ideas discovered in the context of extraction of classical information from a quantum system. It is by no means self-contained and contains a collection of results that were for one or the other reason appealing or of special interest to the author.

One of the most fascinating quantum effects is contained in the no-cloning Theorem of Wootters and Zurek [**245**]: a quantum system cannot be cloned.

This is a direct consequence of the linearity of quantum physics[2]. If cloning were possible, a qubit would be able to transmit an infinite amount of information: given enough copies of an identical quantum system, the Bloch vector associated to it could be determined to arbitrary high accuracy.

The no-cloning Theorem tells us something very deep about quantum measurements: a measurement disturbs the system, and any attempt to gain information about the system implies disturbance of the system in a stochastic irreversible way. In the next section we will investigate this question in the context of continuous measurements.

The fact that any quantum measurement disturbs the system can be exploited to construct cryptographic protocols that are provably secure [**27, 84, 25, 149**], unlike the current classical cryptographic protocols that are secure under the assumption of limited computational resources. The idea is the following: if somebody tries to eavesdrop, she will disturb the system, and this disturbance will reveal itself in the measurement statistics. Quantum cryptography is therefore very promising[3]. Note also that almost all the equations on which classical cryptographic schemes are based can be solved in polynomial time on a quantum computer: the most celebrated result is of course Shor's factoring algorithm [**193**], but for example the solution to the Pell equation can also be solved in polynomial time [**104**].

Let us now move to the following central question: given a sender that encodes information in an ensemble of quantum states $\mathcal{E} = \{p_i, \rho_i\}$, what is the optimal measurement strategy for the receiver such as to maximize the mutual information? The mutual information is defined as

$$I(X:Y) = H(\{p(i)\}) - \sum_\alpha p_\alpha H(\{p(i|\alpha)\}) = H(\{p(\alpha)\}) - \sum_i p_i H(\{p(\alpha|i)\}) \tag{254}$$

where $H(\{p_i\})$ is the Shannon entropy function, and is a measure of how much the ignorance of the sender about the states send labelled by $i$ decreases by performing the measurements labelled by $\alpha$ (note that $p(\alpha|i) = \mathrm{Tr}E_\alpha\rho_i$). More rigourously, it is the asymptotic amount of bits of information that can be decoded by the receiver if the sender used an appropriate code. The maximum of the mutual information over all measurement strategies is called the accessible information, and the maximum of the accessible information over all possible output ensembles of a channel is the capacity of a channel [**192**].

---

[2] The proof of this is simple: suppose a cloning machine existed; then it would bring $|0\rangle \to |0\rangle|0\rangle$, $|1\rangle \to |1\rangle|1\rangle$ and $a|0\rangle + b|1\rangle \to (a|0\rangle + b|1\rangle)(a|0\rangle + b|1\rangle)$. The last thing is clearly different from $a|0\rangle|0\rangle + b|1\rangle|1\rangle$, thus reaching a contradiction.

[3] In a realistic setup however, one also has to do authentication, and there seems to be no quantum trick that can be exploited here; this indicates that the combination of classical cryptography with quantum cryptography will be needed to obtain a very secure communication channel.

Following Davies [**62**], the problem of optimizing the measurement with relation to the mutual information can be brought down to a more tractable problem:

**Lemma 11.** *[Davies [**62**]] Let the input ensemble $\mathcal{E} = \{p_i, \rho_i\}$ be defined over a $n$-dimensional Hilbert space. Then the measurement maximizing the mutual information can be chosen such that all elements in the POVM are pure, and the number of elements $N$ in the optimal POVM can be bounded by $n \le N \le n^2$.*

A proof is given in the appendix.

This Theorem is surprising in that the upper bound on the number of elements in the POVM is independent of the number of elements in the quantum ensemble under investigation. Note also that a stronger upper bound can be obtained if for example all quantum states of the ensemble under investigation are real: then the POVM-elements can be chosen to be real and this will lead to the bound $n \le N \le n(n+1)/2$ (see also [**184**]).

An immediate consequence of the previous Theorem is the fact that the accessible information of an ensemble in a $n$-dimensional Hilbert space is bounded above by $2\log_2(n)$ bits of information. This bound however is not tight and can be strengthened considerably:

**Theorem 33** (Holevo [**112**]). *Given an ensemble of quantum states $\mathcal{E} = \{p_i, \rho_i\}$, then the accessible information is bounded above by the Holevo $\chi$ function:*

$$\chi(\mathcal{E}) = S(\sum_i p_i \rho_i) - \sum_i p_i S(\rho_i). \tag{255}$$

A very elegant proof of this important Theorem (that will not be given here) was found by Schumacher et al. [**190**] using the non-trivial property of strong subadditivity of the quantum entropy [**147**]:

$$S(\rho_{ABC}) + S(\rho_A) \le S(\rho_{AB}) + S(\rho_{AC}). \tag{256}$$

A direct consequence of this Theorem of Holevo is the fact that the accessible information of an ensemble in a $n$-dimensional Hilbert space is bounded above by $\log_2(n)$ bits of information. Consider for example a system of qubits: each qubit can at most reveal one bit of information. The optimal strategy for encoding bits in quantum bits is therefore obtained by preparing a "classical" quantum ensemble of orthogonal states. This may sound a bit disappointing, and indicates that the power of quantum information originates from another characteristic feature of quantum systems: this power mainly comes from the fact that quantum systems can be entangled.

More recently, a very interesting generalization of the previous Theorem has been obtained. It can be proven that the bound in the previous Theorem is tight under certain general conditions:

**Theorem 34** (Holevo [**114**], Schumacher and Westmoreland [**189**]). *Consider a large number of quantum states taken from an ensemble of eventually mixed states $\mathcal{E} = \{p_i, \rho_i\}$. If joint (entangled) measurements are allowed, then there exist a coding and a decoding strategy such that the mutual information approaches $\chi(\mathcal{E})$ in the limit of an infinite number of encoded quantum states.*

The proof is based on ideas of random coding and compression of quantum states. This generalization is of great practical value as typical transmitted states are mixed due to corruption by noise. The Theorem allows to calculate the classical capacity of an imperfect quantum channel using product input states. In the last chapter this will explicitly be done for some special kind of channels.

The classical capacity of a quantum channel can be obtained by maximizing the Holevo $\chi$-quantity over all probabilities $\{p_i\}$ and states $\{\rho_i\}$ of a convex ensemble (this convex ensemble is given by all possible outputs of the quantum channel). Similarly to Lemma 11, we can prove the following:

**Lemma 12.** *The capacity of a quantum channel acting on a $n$-dimensional Hilbert space can be achieved using at most $n^2$ pure input states.*

The proof is very similar to the proof of Lemma 11 and it also given in the appendix.

Before moving on to investigate continuous measurement strategies, let us hop to another interesting problem, extensively studied by Helstrom [**109**]: given one copy of a quantum state out of an ensemble of quantum states $\mathcal{E} = \{p_i, \rho_i\}$, what is the optimal measurement strategy such that the Bayes decision cost (related to the probability of making a wrong decision that $i'$ was prepared instead of $i$) is minimized? The Bayes decision cost is defined as

$$B(\mathcal{E}) = \sum_{i\alpha} p_i p(\alpha|i) C_{i\alpha}, \tag{257}$$

where the index $\alpha$ labels the possible measurement outcomes. A typical choice for the Bayes cost is $C_{i\alpha} = 1 - \delta_{i\alpha}$. If the ensemble consists of non-orthogonal (or non-commuting) states, the Bayes decision cost will typically be obtained by performing a POVM (and not a projective measurement) with elements $\{E_\alpha\}$ (the most celebrated example of this kind is the one involving the trine states [**113**]). The problem is therefore equivalent to: minimize

$$B = \sum_{i\alpha} p_i C_{i\alpha} \mathrm{Tr} \rho_i E_\alpha \tag{258}$$

subject to the constraints:

$$\forall \alpha : E_\alpha \geq 0 \qquad\qquad \sum_\alpha E_\alpha = I. \tag{259}$$

This is definitely a semidefinite program [**206**] and the exact solution can therefore efficiently be found using numerical algorithms that exhibit guaranteed convergence.

## 5.3. Quantum parameter estimation by Continuous Measurement

This section is a reprint of the article of Verstraete, Doherty and Mabuchi entitled "Sensitivity optimization in quantum parameter estimation"[**216**]. It deals with the following fundamental question: how can one devise a quantum measurement strategy such as to gain as much information as possible while trying to reduce the disturbance introduced by the measurement. This question lies at the heart of quantum information theory, and quantifies the trade-off between information gain and disturbance. In the case of finite dimensional systems, the problem has been extensively discussed by Fuchs and Peres [**91, 88**] and Banaszek [**13, 14**]. Here we consider a system in an infinite dimensional Hilbert space, and we will show how optimal measurement strategies can be found using techniques of continuous quantum measurement. An intriguing connection between Kalman filtering and the evolution of quantum systems will be revealed. Let us now reproduce the main part of this paper[4]:

In this work we present a general framework for sensitivity optimization in quantum parameter estimation schemes based on continuous (indirect) observation of a dynamical system. As an illustrative example, we analyze the canonical scenario of monitoring the position of a free mass or harmonic oscillator to detect weak classical forces. We show that our framework allows the consideration of *sensitivity scheduling* as well as estimation strategies for non-stationary signals, leading us to propose corresponding generalizations of the Standard Quantum Limit for force detection.

The primary motivation for work presented here has been to contribute to the continuing integration of quantum measurement theory with traditional (engineering) disciplines of measurement and control. Various researchers engaged in this endeavor have found that the concepts and methods of theoretical engineering provide a fresh perspective on how differences and relationships between quantum and classical metrology can be most cleanly understood. This approach has been especially fruitful in scenarios involving *continuous* measurement, for which a number of important physical insights and results of practical utility follow simply from the formal connections between quantum trajectory theory and Kalman filtering [**240, 159, 152, 153, 75, 74, 18**].

[4]In the original paper, a general formalism for quantum parameter estimation was also presented; as most of that work was done by A. Doherty, we do not include it in this thesis.

Here we describe a general formalism for parameter estimation via continuous quantum measurement, the equations of which are amenable to analytic and numerical optimization strategies. In addition to being useful for practical design of quantum measurements, we find that this approach sharpens our understanding of the significance and origin of Standard Quantum Limits (SQL's) in precision metrology. Following the basic notion that the "standard limit" for any measurement scenario should be derivable by optimization over some parametric family of "standard" measurement strategies, we present results that generalize the SQL for force estimation through continuous monitoring of the position of a test mass. Our analysis shows that the canonical expression for the force SQL in continuous position measurement stems from a rather arbitrary limitation of the set of allowable measurement strategies to those with constant sensitivity, and we find that a lower expression (by a factor of 3/4) can be obtained when time variations are allowed. It follows that further expansions of the optimization space (such as adaptive measurements with real-time feedback [**240**]) should be considered in order to arrive at an SQL that consistently accounts for a natural set of measurement strategies that are "practically equivalent" in terms of inherent experimental difficulty.

Very recently Gambetta and Wiseman have discussed a similar approach to parameter estimation for resonance fluorescence of a two-level atom paying particular attention to how information about the unknown parameter, and also about the quantum state, changes with different kinds of measurements [**93**].

### 5.3.1. Force estimation by continuous measurement of position

The aim of this section is to present a formalism for continuous parameter estimation in the specific context of a harmonic oscillator subject to an unknown force linear in the position $\hat{x}$. This section gives a rigorous and a more general treatment of the ideas previously worked out by Mabuchi [**153**]. We first derive the conditional evolution equations for the oscillator under continuous position measurement, then discuss their control-theoretic interpretation as Kalman filtering equations. We then show how a Bayesian parameter estimator can be obtained from the Kalman filter in this scenario.

5.3.1.1. *Conditional evolution equations.* We will derive the equations of motion of a continuously observed system conditioned on the measurement record. Our treatment is based on the model of continuous measurement of Caves and Milburn [**50**], which in turn was based on work of Barchielli *et al* [**15**]. Their derivation is solely based on the standard techniques of operations and effects in quantum mechanics which makes it very transparent. Similar results could have been obtained by making use of the quantum-stochastic calculus of Hudson [**126**] as was done by Belavkin and Staszewski [**19**].

In continuous measurement — often an accurate description of experimentally realizable measurements — the projective collapse of the wavefunction, and hence also the Zeno effect, can be avoided by continually performing infinitesimally weak measurements. A weak measurement consists of weakly coupling the system under interest to a (quantum-mechanical) meter, followed by a von Neumann measurement of the meter state. As there was only a weak coupling, only very little information about the system of interest is revealed and there will only be a limited amount of back-action. At first we will introduce the concept of weak measurements in the framework of position measurement. Then we will show how to derive the equations of motion for a quantum particle subject to a whole series of weak measurements. The treatment of continuous measurements will then be obtained by taking appropriate limits.

The aim of a weak position measurement is to get some information out of the system, although without disturbing it too much. This can be done by applying a selective POVM $\{\hat{A}_\xi(\hat{x})\}$ where there is a lot of overlap between the $\hat{A}_\xi(\hat{x})$ associated with different measurement results $\xi$. This overlap is proportional to the variance of the measurement outcome, but inversely proportional to the variance of the back-action noise. As shown by Braginsky and Khalili [**43**], the product of those variances always exceeds $\hbar^2/4$. Equality is achieved if and only if $\hat{A}_\xi(\hat{x})$ is Gaussian in $\hat{x}$. As we are interested in the ultimate limits imposed by quantum mechanics, we will assume our measurement device is optimally constructed so as to yield a Gaussian $\hat{A}_\xi(\hat{x})$:

$$\hat{A}_\xi(\hat{x}) = \frac{1}{(\pi D)^{1/4}} \exp\left(-\frac{(\xi-\hat{x})^2}{2D}\right)$$

This is equivalent to the model of Barchielli and also of Caves and Milburn [**50**] who obtained it by explicitly working out the case of linear coupling between a (Gaussian) meter and the particle followed by a von Neumann measurement on the meter.

We will now assume that the wavefunction of the observed particle is also Gaussian. This is a reasonable assumption as we will soon take the limit of many Gaussian measurements, each of which effects a Gaussian “conditioning” of the particle's wavefunction. Ultimately the wavefunction itself will become Gaussian, whatever its original shape (this is a consequence of the Central Limit Theorem). We furthermore assume that the Hamiltonian of the unobserved particle would be given by:

$$H_0 = \frac{\hat{p}^2}{2m} + \frac{m\omega^2}{2}\hat{x}^2 + \theta\hat{x}, \tag{260}$$

where $\theta$ is the (eventually time-dependent) force to be estimated. It will turn out to be very useful to parameterize the Gaussian wavefunction of the particle by a complex mean $\tilde{x} = \tilde{x}_r + i\tilde{x}_i$ and complex variance $\tilde{\sigma} = \tilde{\sigma}_r + i\tilde{\sigma}_i$ (throughout this section the notation $\sigma$ instead of $\sigma^2$ will be used to denote the variance):

$$
\begin{aligned}
|\psi\rangle &= |\tilde{x}(t), \tilde{\sigma}(t)\rangle \\
\langle x|\psi\rangle &= \left(\frac{\tilde{\sigma}_r}{\pi|\tilde{\sigma}|^2}\right)^{1/4} \exp\left(-\frac{(x-\tilde{x})^2}{2\tilde{\sigma}} - \frac{\tilde{x}_i^2}{2\tilde{\sigma}_r}\right) \\
\bar{x} &= \tilde{x}_r + \frac{\tilde{\sigma}_i}{\tilde{\sigma}_r}\tilde{x}_i \qquad \bar{p} = \hbar\frac{\tilde{x}_i}{\tilde{\sigma}_r}
\end{aligned}
$$

$$
\overline{\Delta x^2} = \frac{|\tilde{\sigma}|^2}{2\tilde{\sigma}_r} \qquad \overline{\Delta p^2} = \frac{\hbar^2}{2\tilde{\sigma}_r} \qquad \overline{\Delta x\Delta p + \Delta p\Delta x} = \frac{\hbar\tilde{\sigma}_i}{\tilde{\sigma}_r} \tag{261}
$$

The values of these quantities will in general depend on the value of $\theta$. In this subsection we will suppress this dependence but in the following we will denote the mean position conditioned on a particular value of $\theta$ by $\bar{x}_\theta$ and likewise for the other expectation values. We will now derive the dynamics of this state if a measurement takes place at time $\tau$. From time 0 to $\tau^-$, just before the measurement, the equations of motion are governed by the Schrödinger equation:

$$
\frac{d\tilde{\sigma}}{dt} = \frac{i\hbar}{m}\left(1 - \frac{m^2\omega^2}{\hbar^2}\tilde{\sigma}(t)^2\right) \qquad \frac{d\tilde{x}}{dt} = \frac{\tilde{\sigma}(t)}{i\hbar}\left(\theta + m\omega^2\tilde{x}\right) \tag{262}
$$

The corresponding $\bar{x}$, $\bar{p}$ and second order moments can easily be derived. The equation for $\tilde{\sigma}$ indicates the spreading and contracting of the wavepacket induced by the harmonic oscillation. At time $\tau$, the POVM $\{\hat{A}_\xi(\hat{x})\}$ is performed. $\xi$ will be a Gaussian distributed random variable with expectation value $\bar{x}(\tau^-)$ and variance $D + \overline{\Delta x^2}(\tau^-)$. Straightforward calculations show that the post-measurement wavefunction, conditioned on the result $\xi$, is parameterized by:

$$
\frac{1}{\tilde{\sigma}(\tau)} = \frac{1}{\tilde{\sigma}(\tau^-)} + \frac{1}{D} \qquad \tilde{x}_\xi(\tau) = \frac{\tilde{\sigma}(\tau^-)\xi + D\tilde{x}(\tau^-)}{\tilde{\sigma}(\tau^-) + D} \tag{263}
$$

The equation for $\tilde{\sigma}$ now indicates the contracting effect of the position measurement. The expectation values $\bar{x}$ and $\bar{p}$ become:

$$
\begin{aligned}
\bar{x}(\tau) &= \bar{x}(\tau^-) + \frac{|\tilde{\sigma}(\tau)|^2}{\tilde{\sigma}_r(\tau)D}\left(\xi - \bar{x}(\tau^-)\right) \\
\bar{p}(\tau) &= \bar{p}(\tau^-) + \frac{\hbar\tilde{\sigma}_i(\tau)}{D\tilde{\sigma}_r(\tau)}\left(\xi - \bar{x}(\tau^-)\right)
\end{aligned} \tag{264}
$$

Note that the back-action manifests itself by constantly introducing white noise, i.e. $\xi - \bar{x}(\tau^-)$, into the system.

It is trivial to write down the dynamical equations in the case of a finite number ($N$) of measurements: we just have to repeat the previous two-stage procedure $N$ times. However we are interested in taking the limit of infinitesimal time intervals $dt$ between two measurements. This will only make sense if at each infinitesimal time step the wavefunction is only subject to an infinitesimal disturbance. Referring to equation (263), this implies that the measurement accuracy $D$ has to scale as $1/dt$. Therefore we define the finite sensitivity $k$

by the relation $D = 1/(kdt)$, implying that only an infinitesimal amount of information is obtained in each measurement. In this limit, the random zero-mean variable $(\xi - \bar{x}(\tau^-))/D$ has a standard deviation given by $\sqrt{kdt/2}$. This is very convenient as a Gaussian random variable with zero mean and variance $\sqrt{dt}$ is by definition a Wiener increment, and therefore we can make use of the theory of Ito calculus. Defining $d\xi(t) = \xi_t dt$ as being the measurement record, and using the notation of Ito calculus, the complete equations of motion conditioned on the measurement result for a Gaussian particle subject to continuous observation of the position can be written down:

$$d\xi(t) = \bar{x}(t)dt + v_\xi(t)dW \tag{265}$$

$$d\bar{x}(t) = \frac{\bar{p}(t)}{m}dt + v_x(t)dW \tag{266}$$

$$d\bar{p}(t) = -m\omega^2\bar{x}(t)dt - \theta(t)dt + v_p(t)dW \tag{267}$$

$$\dot{\tilde{\sigma}}(t) = \frac{i\hbar}{m}\left(1 - \frac{m^2\omega^2}{\hbar^2}\tilde{\sigma}(t)^2\right) - k(t)\cdot\tilde{\sigma}(t)^2 \tag{268}$$

$$v_x(t) = \sqrt{\frac{k(t)}{2}}\frac{|\tilde{\sigma}(t)|^2}{\tilde{\sigma}_r(t)} \qquad v_p(t) = \sqrt{\frac{k(t)}{2}}\frac{\hbar\tilde{\sigma}_i(t)}{\tilde{\sigma}_r(t)} \qquad v_\xi(t) = \frac{1}{\sqrt{2k(t)}} \tag{269}$$

If the sensitivity $k$ is kept constant during the whole observation $(\forall t, k(t) = k(0))$, equation (268) can be solved exactly. Given initial condition $\tilde{\sigma}_0$, the solution is:

$$\tilde{\sigma}(t) = \tilde{\sigma}_\infty\left(\frac{\frac{\tilde{\sigma}_\infty+\tilde{\sigma}_0}{\tilde{\sigma}_\infty-\tilde{\sigma}_0}\exp(2i\Omega t) - 1}{\frac{\tilde{\sigma}_\infty+\tilde{\sigma}_0}{\tilde{\sigma}_\infty-\tilde{\sigma}_0}\exp(2i\Omega t) + 1}\right), \qquad \Omega = \sqrt{\omega^2 - \frac{i\hbar k}{m}}, \qquad \tilde{\sigma}_\infty = \frac{\hbar/m}{\Omega} \tag{270}$$

This shows that the position variance of the wavefunction evolves at least exponentially fast to a steady state. The damping is roughly proportional to the square root of the sensitivity, while the steady state solution has a variance inversely proportional to it. This result means that a continuously observed particle is localized, although not confined, in space. It is interesting to note that this localization increases with the mass of the particle, such that it is very difficult to localize a light particle. Indeed the steady state position variance can be understood from the point of view of Standard Quantum Limits for position measurement [**43**]. For example if $\omega^2 \gg \hbar k/m$ then $\overline{\Delta x^2}_\infty \simeq \hbar/2m\omega$. Similarly, if we take $t = 1/\mathcal{R}\rceil(\Omega)$ to be the time for an effectively complete measurement, then for a free particle $\overline{\Delta x^2}_\infty = \hbar t/m$ and so the steady state position variance is the same as the SQL for ideal position measurements separated by time intervals of length $1/\mathcal{R}\rceil(\Omega)$.

5.3.1.2. *Kalman filtering interpretation.* Let us now try to give a "signal processing" interpretation to equations (265-269). The Wiener increment was defined as the difference between the actual and the expected measurement result. As it is white noise, it is clear that the expected measurement result was actually the best possible guess for the result. This is reminiscent to the innovation process in classical control theory: the optimal filtering equations of a classical stochastic process can be obtained by imposing that the difference between the actual and expected (i.e. filtered) measurement be white noise. Indeed, in a previous paper [**75**], Doherty noticed that the equations (265-269) have exactly the structure of the Kalman filtering equations associated with a classical stochastic linear system. This is in complete accordance with the dynamical interpretation of quantum mechanics as describing the evolution of our knowledge about the system.

The classical stochastic system that has exactly the same filtering equations as our continuously observed quantum system is given by:

$$\begin{aligned} d\begin{pmatrix} x_\theta \\ p_\theta \end{pmatrix} &= \begin{pmatrix} 0 & \frac{1}{m} \\ -m\omega^2 & 0 \end{pmatrix}\begin{pmatrix} x_\theta \\ p_\theta \end{pmatrix} dt + \begin{pmatrix} 0 \\ 1 \end{pmatrix}\theta(t)dt + \begin{pmatrix} 0 \\ \hbar/2 \end{pmatrix}\sqrt{2k}dV_1 \\ d\xi &= \begin{pmatrix} 1 & 0 \end{pmatrix}\begin{pmatrix} x_\theta \\ p_\theta \end{pmatrix} dt + \frac{1}{\sqrt{2k}}dV_2 \end{aligned} \tag{271}$$

$dV_1$ and $dV_2$ are two independent Wiener increments and correspond to the process noise and measurement noise respectively. It is very enlightening to look at the corresponding weights of these noise processes: the higher the sensitivity, the more accurate the measurements, but the more noise is introduced into the system. Moreover measuring the position only introduces noise into the momentum. This clearly is a succinct manifestation of the Heisenberg uncertainty relation. Indeed, the product of the amplitude of the noise processes of measurement and back-action is independent on the sensitivity $k$ and exactly given by $\hbar/2$.

The equations for the means $\bar{x}_\theta$ and $\bar{p}_\theta$ are now given by the Kalman filter equations of this classical system, and the equations for the variances $\overline{\Delta x_\theta^2}, \overline{\Delta p_\theta^2}, \overline{\Delta x_\theta \Delta p_\theta + \Delta p_\theta \Delta x_\theta}$ are given by the associated Riccati equations. This is very convenient as this will allow us to use the convenient language of classical control theory to solve the estimation problem.

5.3.1.3. *Continuous Parameter Estimation.* Let us now consider the basic question of this section: how can we get the best estimates of the unknown force $\{\theta(t)\}$ acting on the system, given the measurement record $\{d\xi_t\}$? The natural way to attack this problem is the use of Bayes rule. As we have a linear system with $\{d\xi_t\}$ a linear function of $\{\theta(t)\}$, and the noise in the system is Gaussian, this will lead to a Gaussian distribution in $\{\theta(t)\}$. Moreover, due to the linearity, the second order moments of this distribution will be independent of the actual measurement record. Therefore the accuracy of our estimates will only be a function of the sensitivity chosen during the observation process and of the prior knowledge we have about the signal $\{\theta(t)\}$ (for example that it is constant) . This will allow us to devise optimal measurement strategies.

The formalism that we have developed is particularly useful in the case that we parameterize $\{\theta(t)\}$ as a linear combination of known time-dependent functions $\{f_i(t)\}$, but with unknown weights $\{\theta_i\}$:

$$\theta(t) = \sum_{i=1}^{n} \theta_i f_i(t)$$

The estimation, based on Bayes rule, will lead to a joint Gaussian distribution in the parameters $\{\theta_i\}$. Indeed, we have the relations:

$$\begin{aligned} p(\{\theta_i\}|\{\xi(t+dt)\}) &\sim p(d\xi(t)|\{\theta_i\},\{\xi(t)\})p(\{\theta_i\}|\{\xi(t)\}) \\ &\sim p\left(d\xi(t)|\bar{x}\left(t,\{\theta_i\},\{\xi(t)\}\right)\right)p(\{\theta_i\}|\{\xi(t)\}) \end{aligned} \quad (272)$$

In the last step we made use of the fact that the Kalman estimate $\bar{x}_{\{\theta_i\}}(t)$ is a sufficient statistic for $d\xi(t)$. Moreover all distributions are Gaussian, while $\bar{x}_{\{\theta_i\}}(t)$ is some linear function of $\{\theta_i\}$ due to the linear character of the Kalman filter:

$$\bar{x}_{\{\theta_i\}}(t) = \sum_i \theta_i \int_0^t dt' g(t,t') f_i(t')$$

The function $g(t,t')$ can easily be calculated using equations (265-269). To obtain the variance of the optimal estimates of $\{\theta_i\}$, formula (272) has to be applied recursively. By explicitly writing out the Gaussian distributions, and making use of the fact that the product of Gaussians is still a Gaussian, it is then easy to show that the variances at time $\tau$ are given by:

$$\frac{1}{\sigma_{\theta_i}} = \int_0^\tau \frac{dt}{v_\xi^2(t)} \left(\int_0^t dt' g(t,t') f_i(t')\right)^2 \quad (273)$$

A more intuitive way of obtaining the same optimal estimation, given a fixed measurement strategy, of $\{\theta_i\}$ can be obtained by a little trick: we can enlarge the state vector $(x_\theta, p_\theta)$ with the unknowns, and construct the Kalman filter and Riccati equation of the new enlarged system. $\bar{x}_\theta$ and $\bar{p}_\theta$, till now the expected values conditioned on a fixed value of the force, then get the meaning of the mean of these expected values over the probability distribution of the unknown force. In other words, the new $\bar{x}$ and $\bar{p}$ become the ensemble averages over the pure states labelled by a fixed force $\theta$. The new enlarged system, in the case of one unknown parameter $\theta$, reads:

$$\begin{aligned} d\begin{pmatrix} x \\ p \\ \theta \end{pmatrix} &= \underbrace{\begin{pmatrix} 0 & 1/m & 0 \\ -m\omega^2 & 0 & f(t) \\ 0 & 0 & 0 \end{pmatrix}}_{A(t)} \begin{pmatrix} x \\ p \\ \theta \end{pmatrix} dt + \underbrace{\begin{pmatrix} 0 \\ \hbar/2 \\ 0 \end{pmatrix}}_{B} \sqrt{2k(t)}dV_1 \\ d\xi &= \underbrace{\begin{pmatrix} 1 & 0 & 0 \end{pmatrix}}_{C} \begin{pmatrix} x \\ p \\ \theta \end{pmatrix} dt + \underbrace{\frac{1}{\sqrt{2k}}}_{D} dV_2 \end{aligned} \quad (274)$$

The Kalman filter equations will give us the best possible least-squares estimation of the vector $(x, p, \theta)$ at each time, while the Riccati equation determines the evolution of the covariance matrix $P$ of the estimation error:

$$\frac{d}{dt}\begin{pmatrix} \bar{x} \\ \bar{p} \\ \bar{\theta} \end{pmatrix} = A(t)\begin{pmatrix} \bar{x} \\ \bar{p} \\ \bar{\theta} \end{pmatrix} + 2k(t)P(t)C^T\left(d\xi(t) - C\begin{pmatrix} \bar{x} \\ \bar{p} \\ \bar{\theta} \end{pmatrix}\right) \quad (275)$$

$$\dot{P} = A(t)P + PA^T(t) - 2k(t)PC^TCP + 2k(t)BB^T \quad (276)$$

An optimal measurement strategy, dependent on the sensitivity, will then be this one that minimizes the $(3, 3)$ element in $P$ at time $t_{final}$. An analytic solution of this problem does not exist in general, as the Riccati equations are quadratic. However, in the case of constant $f(t) = f(0)$ and constant sensitivity $k(t) = k(0)$ analytical results will be derived.

Before proceeding however, it is interesting to do a dimensional analysis to see how the variances will scale. We begin by scaling $\tilde{t} = t/\tau$ with $\tau$ the duration of the complete measurement. Introducing the matrix

$$T = \begin{pmatrix} \sqrt{\frac{\hbar\tau}{2m}} & 0 & 0 \\ 0 & \sqrt{\frac{\hbar m}{2\tau}} & 0 \\ 0 & 0 & \sqrt{\frac{\hbar m}{2\tau^3}} \end{pmatrix}, \quad (277)$$

it can easily be checked that $\tilde{P} = T^{-1}PT^{-1}$ is dimensionless. If we then scale the sensitivity as $k(t) = \tilde{k}(\tilde{t})\hbar\tau^2/(2m)$, the force $\theta = \tilde{\theta}\sqrt{\hbar m/2\tau^3}$ and do the appropriate transformations $B \to \tilde{B}$ and $C \to \tilde{C}$, we get the equivalent state space model:

$$\tilde{A} = \begin{pmatrix} 0 & 1 & 0 \\ -\omega^2\tau^2 & 0 & f(t) \\ 0 & 0 & 0 \end{pmatrix} \qquad \tilde{B} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad \tilde{C} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \quad (278)$$

The new filter equations are still given by (275,276) with the substitution $(A, B, C, k(t)) \to (\tilde{A}, \tilde{B}, \tilde{C}, \tilde{k}(\tilde{t}))$. This observation has an immediate consequence if we are measuring the force acting on a free particle ($\omega = 0$): the standard deviation on our estimate will always scale like $\sqrt{\hbar m/2\tau^3}$, and the chosen sensitivity will only affect the accuracy by a multiplicative pre-factor.

### 5.3.2. Standard Quantum Limits

In this section we will derive the explicit equations of the variances on these estimates.

5.3.2.1. *Detection of stationary signals.* Let us first introduce the idea of the standard quantum limit in the context of von Neumann measurements. The idea is that a particle is prepared in some optimal way at time 0, such that at time $\tau$ a projective measurement is performed to determine the displacement associated with the force. The optimal preparation is crucial as it has to balance the position and the momentum uncertainty. The optimal preparation leads to the expression of the Standard Quantum Limit. Consider a free particle with a Gaussian wavefunction $\langle x|\psi\rangle$ and initial parameters $\tilde{x}(0), \tilde{\sigma}(0)$ (see equation (261)) and subject to an unknown force $\theta$. The integrated equations of motion (262) are given by:

$$\tilde{x}(t) = \tilde{x}_0 + \theta\left(t\tilde{\sigma}(0)/i\hbar + t^2/2m\right) \qquad \tilde{\sigma}(t) = \tilde{\sigma}(0) + i\frac{\hbar}{m}t$$

Suppose that at time $\tau$ we perform a von Neumann measurement of the position. The probability distribution associated with this measurement is given by:

$$p(x|\theta) \sim \exp\left(-\frac{\left(x - \frac{\theta t^2}{2m}\right)^2}{|\tilde{\sigma}|^2/\tilde{\sigma_r}}\right) \tag{279}$$

Using Bayes rule assuming a flat prior distribution for $\theta$ the variance on the estimate of $\theta$ given the measurement result $x$ can easily be derived:

$$\sigma_\theta = \frac{2m^2|\tilde{\sigma}(t)|^2}{\tilde{\sigma}_r(t)t^4} = \frac{2m^2\left(\tilde{\sigma}_r^2(0) + (\tilde{\sigma}_i(0) + \frac{\hbar t}{m})^2\right)}{\tilde{\sigma}_r(0)t^4} \tag{280}$$

This function is heavily dependent on the initial conditions of the wavefunction of the particle. The standard quantum limit can now be derived by choosing the initial conditions such that $\sigma_\theta$ is minimized. This variance can in principle go to zero if we allow $\langle\Delta x\Delta p\rangle$ to be negative, but we will not consider such "contractive" states [**247, 166**] here. We therefore impose the condition $\tilde{\sigma}_i(0) \geq 0$ in order to focus our attention on the specific issue of sensitivity optimization. The optimal $\tilde{\sigma}(0)$ is then given by $\tilde{\sigma}(0) = \hbar t/m$, and this leads to the expression of the Standard Quantum Limit:

$$\sigma_\theta = \frac{4\hbar m}{t^3} \tag{281}$$

It is clear that the square of the amplitude of a detectable force has to be bigger than the variance on its estimation to be detectable. Therefore the previous formula is the expression of the minimal force that can be detected by a free particle of mass $m$ over a time $t$. Note that the derived formula exceeds the normal equation of the SQL by a factor 8 as the standard equation is not derived in the context of parameter estimation.

We will now apply an analogous reasoning to a quantum particle subject to continuous measurement. The explicit expression of the variance on the estimated force was given by equation (273). As noted at the end of the first section, the resulting variance will be given by the standard quantum limit multiplied by a certain factor. From here on we will therefore work in the dimensionless picture as defined in (278). In general it is very hard to find the explicit expression for the autocorrelation function $g(t,t')$ in equation (273). Things get much more feasible if we do not vary the sensitivity during the measurement as the system then becomes stationary. It follows that we can assume that the values of the variances reached their steady state values given by equation (270). After some straightforward linear algebra, the explicit expression for $g(t,t')$ in the case of steady state is given by:

$$g(t,t') = \frac{1}{b}\exp(-a(t-t'))\sin(b(t-t')) \tag{282}$$

$$a = \omega\tau\sqrt{\frac{1}{2}\left(-1+\sqrt{1+\frac{(2k)^2}{(\omega\tau)^4}}\right)} \tag{283}$$

$$b = \omega\tau\sqrt{\frac{1}{2}\left(1+\sqrt{1+\frac{(2k)^2}{(\omega\tau)^4}}\right)} \tag{284}$$

Due to the stationarity of the variances, the autocorrelation function $g(t,t')$ is indeed only dependent on $(t-t')$, and from here on we will therefore use the notation $g(t,t') = g(t-t')$. The full expression of the variance on our estimate now becomes:

$$\frac{1}{\sigma_\theta} = 2k\int_0^1 dt\left(\int_0^t dt' g(t-t')f(t')\right)^2 \tag{285}$$

The force that acted on the system was assumed to be of the form $\theta(t) = \theta f(t)$ with $f(t)$ a known function. Note that this expression is dimensionless and has to be multiplied by $\frac{2\tau^3}{\hbar m}$. We next introduce $F(\omega)$ and $G(\omega)$ the Fourier transforms of the functions $f(t)\cdot u_{[0,1]}(t)$ and $g(t)\cdot u_{[0,1]}(t)$, where $u_{[0,1]}(t)$ is the window function over the interval $[0,1]$. The damping effect due to the back-action noise is responsible for broadening the spectrum of the harmonic oscillator with a width of approximately $k/(\omega\tau)$. Basic properties of Fourier transformations lead to the expression:

$$\frac{1}{\sigma_\theta} = \frac{2k}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\omega_1 d\omega_2 \exp\left(i\frac{\omega_1-\omega_2}{2}\right)\frac{\sin\left(\frac{\omega_1-\omega_2}{2}\right)}{\frac{\omega_1-\omega_2}{2}}G(\omega_1)G^*(\omega_2)F(\omega_1)F^*(\omega_2) \tag{286}$$

This formula clearly shows that only the frequencies of the signal $F(\beta)$ near to the natural frequencies of the oscillator $G(\beta)$ will be detectable.

Now we shall explicitly calculate the value of $\sigma_\theta$ in some different cases. Let us first of all assume that the spectrum $F(\beta)$ is almost constant for all values where $G(\beta)$ is substantially different from 0, i.e. around $\beta \simeq (\omega\tau)$. This is realistic in some scenarios of interest for the detection of gravitational waves [**43**]. Let us furthermore assume that $\omega\tau \gg 1$, which means that the period of the oscillator is much smaller than the observation time. Next we observe that we are allowed to approximate the $\text{sinc}((\omega_1-\omega_2)/2)$ function by a delta-Dirac function if the width of the spectrum $G(\beta)$, determined by the number $k/(\omega\tau)$, is much bigger than 1. This leads to the expression:

$$\begin{aligned} \frac{1}{\sigma_\theta} &\simeq \frac{k|F(\omega\tau)|^2}{2\pi}\int_{-\infty}^{\infty} d\omega |G(\omega)|^2 && (287)\\ &= \frac{k|F(\omega\tau)|^2}{2\pi}\int_{-\infty}^{\infty} d\omega \frac{1}{(a^2+b^2-\omega^2)^2+4a^2b^2} && (288)\\ &= \frac{|F(\omega\tau)|^2}{4\omega\tau}\chi\left(\frac{2k/(\omega\tau)^2}{\sqrt{1+(2k/(\omega\tau)^2)^2}}\right) && (289)\\ \chi(x) &= (1-x^2)^{1/4}\sqrt{\frac{1+\sqrt{1+x^2}}{2(1+x^2)}} && (290)\end{aligned}$$

The function introduced in the last line is only dependent on $2k/(\omega\tau)^2$, which can be tuned freely by changing the value of our sensitivity. The function $\chi(x)$ reaches its maximum value 1 for small values of $x$, meaning that optimal detection requires $k \ll (\omega\tau)^2$. The derivation however required that $1 \ll k/(\omega\tau)$. Therefore, the optimal choice of the sensitivity will be given by a value $(\omega\tau) \ll k \ll (\omega\tau)^2$, leading to the variance on the estimate:

$$\sigma_\theta^2 \simeq \frac{4\omega\tau}{|F(\omega\tau)|^2}\frac{\hbar m}{2\tau^3} = \frac{1}{|F(\omega\tau)|^2}\frac{2\hbar m\omega}{\tau^2} \qquad (291)$$

This corresponds exactly to the expression of the standard quantum limit for an oscillator [**43**]. A similar expression can be obtained by explicitly integrating (285) with $f(t) = \delta(t)$. The conditions under which this SQL can be reached are: 1. The total duration of the measurement is much bigger than the period of the oscillator; 2. The spectrum of the signal to be detected is flat around the natural frequencies of the observed oscillator.

We will now investigate what happens if this second condition is not fulfilled. In the extreme case, the force to be detected is constant, corresponding to a delta-Dirac function in the frequency domain. Again under the condition that $1 \ll \omega\tau \ll k/(\omega\tau)$, a good approximation of equation (286) becomes:

$$\frac{1}{\sigma_\theta} \simeq k|G(0)|^2 = \frac{1}{(\omega\tau)^2}\frac{2k/(\omega\tau)^2}{1+(2k/(\omega\tau)^2)^2} \qquad (292)$$

The optimal sensitivity is now given by $2k = (\omega\tau)^2$, indicating that one has to choose a much higher sensitivity to detect constant forces than resonant oscillating forces. The expression for the SQL for detecting constant forces with a harmonic oscillator therefore becomes:

$$\sigma_\theta \simeq 2(\omega\tau)^2\frac{\hbar m}{2\tau^3} = \frac{m\hbar\omega^2}{\tau} \qquad (293)$$

It is now natural to look what happens in the limit of $\omega \to 0$, it is if the observed particle is free and only subject to a constant force. In that case the explicit integration of (285) becomes possible, as $a$ and $b$ both become equal to the sensitivity $\sqrt{k}$. Straightforward but long integrations lead to:

$$\sigma_\theta = \frac{8k^{3/2}}{4\sqrt{k} - 5 + 8\exp(-\sqrt{k})\cos(\sqrt{k}) - \exp(-2\sqrt{k})\left(2 + \cos(2\sqrt{k}) + \sin(2\sqrt{k})\right)} \tag{294}$$

Minimization over the sensitivity leads to an expression for the SQL for the detection of a constant force with a free particle subject to continuous observation: the minimal value is obtained when $k \simeq 3.033$ (see figure 2) and leads to

$$\sigma_\theta \simeq 3\frac{4\hbar m}{\tau^3}. \tag{295}$$

Note that this expression differs from the corresponding one derived in [**153**], where calculations were done without properly accounting for the damping effect of measurement back-action. Comparing this result with (281), the variance of our estimate obtained by continuous measurement is 3 times bigger than if we were doing projective measurements. This is caused by two factors: at the end of the continuous measurement, there is still a lot of information encoded about the force in the wavefunction as the variance on the position at time $\tau$ is not at all equal to $\infty$. Secondly, the previous result was obtained by assuming that the variances of our Gaussian wavefunction were in steady state, and this is not necessarily the optimal initial condition. Indeed, it turns out that the optimal initial state (not considering contractive states) of the continuously observed particle is a Gaussian state with well defined momentum ($\langle \Delta p^2 \rangle \ll 1$) and therefore undefined position $\langle \Delta x^2 \rangle \gg 1$. This makes sense as the force to be detected can only be seen because it manifests itself through the momentum. The fact that the position uncertainty is very large is not so bad as the position is continuously observed such that it becomes well-defined very quickly. The expression for the variance on the force estimate using this optimally prepared initial state can now be calculated exactly by explicitly solving the Riccati equations (276):

$$\sigma_\theta = \frac{2k^{3/2}\left(\sinh(2\sqrt{k}) + \sin(2\sqrt{k})\right)}{\sqrt{k}(\sinh(2\sqrt{k}) + \sin(2\sqrt{k})) - (\cosh(2\sqrt{k}) - \cos(2\sqrt{k}))} \tag{296}$$

Optimization over the sensitivity (see figure 2) leads to an enhancement of ca. 2/3 in comparison with the steady state case (this minimal value is obtained for a sensitivity $k \simeq 2.834$).

An even bigger gain would have been obtained if a projective measurement at the end of the continuous observation were allowed. A realistic way to implement this would be to make the sensitivity very large at the end of the measurement. If the matrix $P(1)$ is the solution of the Riccati equation (276) at time $t = 1$, some straightforward calculations show that a projective position measurement reduces the estimator variance by $P^2_{(3,1)}/P_{(1,1)}$. The optimal initial conditions are still given by ($\langle \Delta p^2 \rangle \ll 1$) and $\langle \Delta x^2 \rangle \gg 1$. The exact

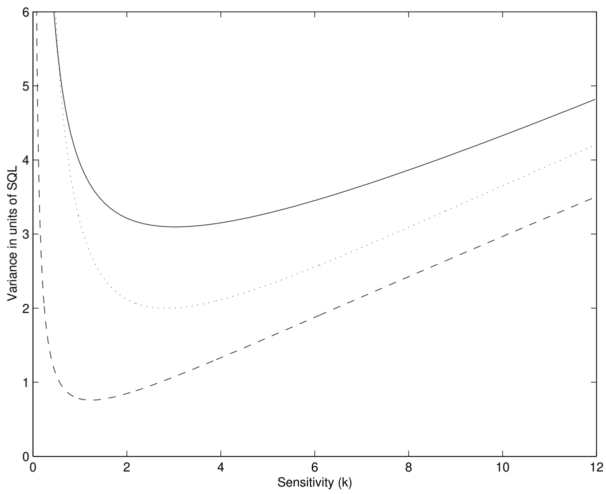



**Figure 2.** Variance on the estimator of a constant force in function of the sensitivity using three different setups: 1. (solid) initial steady state conditions and constant sensitivity k (eq. 294); 2. (dotted) optimal initial conditions and constant sensitivity k (eq. 296); 3. (dashed) optimal initial conditions, constant sensitivity k, followed by a final projective measurement (eq. 297)

expression of the variance on the estimate in function of the sensitivity $k$ is then given by:

$$\sigma_\theta = \frac{4k^{3/2}(\cosh(2\sqrt{k}) + \cos(2\sqrt{k}))}{2\sqrt{k}(\cosh(2\sqrt{k}) + \cos(2\sqrt{k})) - (\sinh(2\sqrt{k}) + \sin(2\sqrt{k}))} \tag{297}$$

Minimization over the sensitivity (see figure 2) leads to an optimal value $k \simeq 1.238$, yielding the equation:

$$\sigma_\theta \simeq 0.76\frac{4\hbar m}{t^3} \tag{298}$$

Therefore we have modestly beaten the usual standard quantum limit by optimally preparing the Gaussian wavepacket and doing a von Neumann measurement at the end of the continuous measurement. This shows that a continuous measurement together with a projective measurement at the end on a optimally prepared state can reveal more information than only projective measurements. In other words, the balance information gain versus disturbance is a little bit in favor of continuous measurement. Although noise is continuously fed into the system by the sensor, we can extract more information about the classical force.

An even better performance can be obtained if we vary the sensitivity continuously during the measurement (*sensitivity scheduling*). It is indeed the case that back-action noise introduced in the beginning of the measurement does more harm than back-action noise at the end of the measurement, as the random momentum kicks delivered at any given time corrupt all subsequent position readouts. In terms of systems theory, the optimal sensitivity as a function of time is simply an optimal control problem associated with Eq. (276). In this optimal control problem, the cost function is simply the value of the force estimator variance $P(3,3)$ at the final time. This is to be minimized by an appropriate choice of the time variation of the sensitivity. The optimal control can be determined by solving a Bellman equation using techniques of dynamic programming [**157**]. Due to the nonlinearity of the Riccati equation, this cannot be done analytically. The optimal sensitivity at time $\tau$, however, can easily be obtained: it tends to a Dirac- d function so as to mimic a projective position measurement. The variance on the estimator after such a projective measurement is reduced by $P^2_{(3,1)}/P_{(1,1)}$. In order to obtain a numerically tractable problem, we define the cost-function $K = P_{(3,3)}(\tau) - P^2_{(3,1)}(\tau)/P_{(1,1)}(\tau)$, the optimal control problem no longer contains a singularity and can be solved numerically. In this second problem, it is assumed that it is possible to make a projective measurement at the final time and the aim is to choose the sensitivity as a function of time such that the information gained during the continuous measurement and due to the projective measurement is maximized. Another way to regularize this problem would be to specify a maximum allowed sensitivity. We discretize the total time in, for example, 50 intervals, and in each interval we assume the sensitivity has a constant value $k_j$ . The solution can then be found by applying some kind of steepest descent algorithm over these 50 variables $\{k_j\}$. It turns out that the optimal $k(t)$ in the case of a free particle ($\omega = 0$) is a smooth monotonously but slowly increasing function of time. In this free particle case, the optimal time-varying sensitivity only leads to a marginal gain: the numerical optimization shows that the variance of the estimate becomes very nearly equal to a factor 3/4 of the usual standard quantum limit (281). Nevertheless, we can present this result as a generalization of the usual SQL to include strategies with sensitivity scheduling:

$$\sigma_\theta \simeq 3.000\frac{\hbar m}{t^3}. \tag{299}$$

Much greater improvements can be expected from the application of sensitivity scheduling to the case of a continuously observed harmonic oscillator. Indeed, the variance on the position of such a particle is small in the middle of the well and at the borders, while it is big elsewhere. Therefore the sensitivity should be varied in a sinusoidal manner, such as to measure more precisely at the positions where the variance is small. The optimal variation of sensitivity in time could be determined by solving a similar optimal control problem to the one explained in the previous paragraph. In the limit where projective measurements are allowed, one expects that the optimal variation of sensitivity should correspond to stroboscopic measurement [**43**], which is indeed well-known to beat the usual standard quantum limit.

5.3.2.2. *Detection of non-stationary signals.* The techniques introduced in the previous sections can also be used for the estimation of non-stationary signals, as one would have for example in the problem of gravitational wave detection when the arrival time of the signal is unknown. Suppose for example that we know that the signal to detect is of the form $\theta(t-t_1) = \theta_0 f(t-t_1)$ with $f(\tau)$ known but amplitude $\theta_0$ and arrival time $t_1$ unknown. An effective non-stationary measurement strategy can in fact be implemented by constructing a Kalman filter for system (271) assuming that $\theta = 0$ (assuming $f(\tau) = 0$ for $\tau < 0$). At times $t < t_1$, the quantity $d\xi - \bar{x}(t)dt$ is by construction white noise with variance $dt/2k(t)$. From time $t \geq t_1$ on however, the force will bias this white noise by an amount $\int_{t_1}^{t} dt' g(t,t')\theta(t')$ as the $\theta = 0$ Kalman filter models the wrong system. This bias will be detectable once it transcends the white noise at time $t_1 + \Delta t$:

$$\int_{t_1}^{t_1+\Delta t} dt \int_{t_1}^{t} dt' g(t,t')\theta_0 f(t'-t_1) \geq \sqrt{\int_{t_1}^{t_1+\Delta t} \frac{dt}{2k(t)}} \tag{300}$$

The goal is now to make this $\Delta t$ as small as possible. The previous equation can again be solved analytically if one has a constant sensitivity and steady state conditions. To make things easier we assume that the observed particle is free ($\omega = 0$), although all calculations can be performed in the more general case too. Let us first assume that the signal to detect is a kick at time $t_1$: $f(t-t_1) \simeq \delta(t-t_1)\tau$ with $\tau$ some measure of the duration of the kick [**43**]. Introducing the dimensionless parameter $\kappa = \Delta t\sqrt{\hbar k/2m}$, the previous inequality becomes:

$$\begin{aligned} \theta_0 &\geq \frac{1}{\tau}\sqrt{\frac{\hbar m}{\Delta t}}\frac{\kappa}{1-\exp(-\kappa)(\cos(\kappa)+\sin(\kappa))} && (301) \\ &\geq \frac{2}{\tau}\sqrt{\frac{\hbar m}{\Delta t}} && (302) \end{aligned}$$

In the last step the optimal $\kappa$, related to the optimal sensitivity $k$, was chosen. The meaning of this equation is clear: a kick with an amplitude $\theta_0$ will only be observed after a time span $\Delta t = 4\hbar m/\tau^2\theta_0^2$. Moreover, the sensitivity has to scale inversely with the square root of $\Delta t$.

An analogous treatment applies to the case of a constant force $f(t-t_1) = u_{[0,\infty]}(t-t_1)$. In this case inequality (300) becomes:

$$\begin{aligned} \theta_0 &\geq \sqrt{\frac{\hbar m}{\Delta t^3}}\frac{\kappa^2}{\exp(-\kappa)\cos(\kappa)+\kappa-1} && (303) \\ &\geq 4.25\sqrt{\frac{\hbar m}{\Delta t^3}} && (304) \end{aligned}$$

As expected, we recover the well known standard quantum limit, but now in a different set-up.

The previous arguments can be refined by using techniques of classical detection theory such as the concept of the matched filter. The results will however be qualitatively similar to the previous ones.

More advanced detection schemes can also be constructed by *adaptively* changing the sensitivity as a real-time function of the measurement record [**240**]. A possible application of this is a scheme for the detection of a signal with unknown arrival time: first one chooses the optimal sensitivity for estimating the arrival time, and from the moment on the signal is detected the sensitivity is brought to its optimal value for detecting the amplitude of the signal. More sophisticated versions of this adaptive measurement could be very useful in realistic stroboscopic measurements where the initial phase of the harmonic oscillator is unknown, as the measurement sensitivity could be made a real-time function of the estimated particle position.

CHAPTER 6

# Quantum Channels

The existence of non-local correlations or entanglement in multipartite quantum systems [**82, 186**] is one of the cornerstones on which the newly established field of quantum information theory is build. The main gain of quantum over classical information processing stems from the fact that we are allowed to perform operations on entangled states: through the quantum correlations, an operation on a part of the system affects the whole system. One of the most challenging open problems is to clarify and quantify how entanglement behaves when part of an entangled state is sent through a quantum channel.

Of central importance in the description of a quantum channel or completely positive map (CP-map) is the dual state associated to it. This state is defined over the tensor product of the Hilbert space itself (the input of the channel) with another one of the same dimension (the output of the channel). It is clear that there appears a natural tensor product structure, and indeed the notion of entanglement will be crucial in the description of quantum channels.

In a typical quantum information setting, Alice wants to send one qubit (eventually entangled with other qubits) to Bob through a quantum channel. The channel acts linearly on the input state, and the consistency of quantum mechanics dictates that this map be completely positive (CP) [**142**]. This implies that the map is of the form [**56**]

$$\Phi(\rho) = \sum_i A_i \rho A_i^\dagger.$$

Moreover the map is trace-preserving if no loss of the particle can occur. A natural way of describing the class of CP-maps is by using the duality between maps and states, first observed by Jamiolkowski [**129**] and since then rediscovered by many. We review some nice properties of CP-maps based on this dual description, and show how to obtain the extreme points of the convex set of trace-preserving CP-maps.

The dual state is defined on a Hilbert space that is the tensor product of two times the original Hilbert space on which the map acts, and is therefore naturally endowed with a notion of entanglement. Unitary evolution for example corresponds to maximal correlations between the in- and output state, and this kind of evolution leads to a dual state that is maximally entangled. We will show how normal forms derived for entangled states lead to interesting parameterizations of CP-maps, and will discuss some issues concerning the use of quantum channels to distribute entanglement.

It thus turns out that the techniques developed for describing entanglement can directly be applied for describing the evolution of a quantum system. Concepts as quantum steering and teleportation have a direct counterpart. A quantum channel for example will be useful for distributing entanglement if and only if the dual state associated to it is entangled, and optimal decompositions of states as derived in the case of entanglement of formation will yield very appealing parameterizations of quantum channels. Following Cirac et al. [**57**], it will moreover be shown how ideas of teleportation enable implementing global transformations on distributed quantum systems by means of local operations and a limited amount of entanglement. We end this chapter with a discussion of the optimal use of a given Hamiltonian to produce entanglement.

## 6.1. Characterization of CP-maps

The most general evolution of a quantum system is described by a linear CP-map [**142**]. In this section we will give a self-contained description of CP-maps or quantum channels. Most of the mathematics presented originate from the seminal papers of de Pillis [**65**] and Choi [**56**]. The fact that the evolution of quantum systems is described by linear completely positive maps is a consequence of the assumption of the linearity of the evolution (the complete positivity follows from consistency arguments once the linearity is accepted).

Let us now recall some notations and useful tricks. Consider a pure state $|\chi\rangle$ in a Hilbert space that is a tensor product of two Hilbert spaces of dimension $n$

$$|A\rangle = \sum_{ij}^{n} a_{ij}|i\rangle|j\rangle.$$

Define

$$|I\rangle = \sum_{i}^{n} |i\rangle|i\rangle$$

an unnormalized maximally entangled state and $A$ the operator with elements $\langle i|A|j\rangle = a_{ij}$, then

$$|A\rangle = A \otimes I_n|I\rangle.$$

Moreover it holds that

$$X \otimes Y|A\rangle = XA \otimes Y|I\rangle = XAY^T \otimes I_n|I\rangle = I_n \otimes YA^T X^T|I\rangle.$$

The symbol $|I\rangle$ will solely be used to denote the unnormalized maximally entangled state $|I\rangle = \sum_i |ii\rangle$. We are now ready for the following fundamental Theorem of de Pillis[**65**]:

**Theorem 35.** *A linear map $\Phi$ acting on a matrix $X$ is Hermitian-preserving if and only if there exist operators $\{A_i\}$ and real numbers $\lambda_i$ such that*

$$\Phi(X) = \sum_i \lambda_i A_i X A_i^\dagger$$

*Proof:* Suppose the map $\Phi$ acts on a $n \times n$ matrix. Then due to linearity, $\Phi$ is completely characterized if we know how it acts on a complete basis of $n \times n$ matrices, for example on all matrices $|e_i\rangle\langle e_j|$, $1 \le i, j \le n$ with $|e_i\rangle$ a complete orthonormal base in Hilbert space. Let us define the $n^2 \times n^2$ positive matrix

$$|I\rangle\langle I| = \begin{pmatrix} |e_1\rangle\langle e_1| & \cdots & |e_1\rangle\langle e_n| \\ \cdots & \cdots & \cdots \\ |e_n\rangle\langle e_1| & \cdots & |e_n\rangle\langle e_n| \end{pmatrix}, \tag{305}$$

being the matrix notation of a maximally entangled state in a $n \otimes n$ Hilbert space. It follows that all the information of a map $\Phi$ is encoded in the state

$$\rho_\Phi = I_n \otimes \Phi(|I\rangle\langle I|), \tag{306}$$

as the $n^2$ $n \times n$ blocks represent exactly the action of the map on the complete basis $|e_i\rangle\langle e_j|$. If $\Phi$ is Hermitian-preserving, then $\Phi(|e_i\rangle\langle e_j|)$ has to be equal to the Hermitian conjugate of $\Phi(|e_j\rangle\langle e_i|)$, and this implies that $\rho_\Phi$ is Hermitian. Let us therefore consider the eigenvalue decomposition of $\rho_\Phi = \sum_i \lambda_i |\chi_i\rangle\langle\chi_i|$. Using the trick $|A\rangle = (A \otimes I)|I\rangle$, we easily arrive at the conclusion that $\Phi(X) = \sum_i \lambda_i A_i X A_i^\dagger$, where $\{\lambda_i\}$ are the eigenvalues and where the operators $\{A_i^T\}$ are the reshaped versions of the eigenvectors of $\rho_\Phi$. □

A central ingredient in the proof was the introduction of the matrix

$$\rho_\Phi = I_n \otimes \Phi(|I\rangle\langle I|)$$

with $|I\rangle = \sum_i |i\rangle|i\rangle$ a maximally entangled state. We define this Hermitian matrix $\rho_\Phi$ as being the dual state corresponding to the map $\Phi$. It was already explained that it encodes all the information about the map, and its eigenvectors give rise to the operators $A_i$. The above lemma characterizes all possible Hermitian preserving maps, and therefore surely all positive and completely positive maps. For example, let us consider the positive map that corresponds to taking the transpose of the density operators of a qubit:

$$\lambda_1 = 1 \qquad A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{307}$$

$$\lambda_2 = 1 \qquad A_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{308}$$

$$\lambda_3 = 1 \qquad A_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} /\sqrt{2} \tag{309}$$

$$\lambda_4 = -1 \qquad A_4 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} /\sqrt{2} \tag{310}$$

Not all Hermitian-preserving maps are physical in quantum mechanics however: if a map acts on a subsystem, then it should conserve positivity of the complete density operator. This extra assumption leads to the condition of complete positivity, meaning that $I_m \otimes \Phi$ is positive for all $m$. Of course, this implies that the dual state $\rho_\Phi$ is not only Hermitian but also positive (i.e. all its eigenvalues are positive), as it is defined as the action of the map $I_n \otimes \Phi$ on a maximally entangled state. The positive eigenvalues can then be absorbed into the (Kraus) operators $\{A_i\}$, and we have therefore proven the Kraus representation Theorem (Choi[**56**]):

**Theorem 36.** *A linear map $\Phi$ acting on a density operator $\rho$ is completely positive if and only if there exist operators $\{A_i\}$ such that*

$$\Phi(\rho) = \sum_i A_i \rho A_i^\dagger.$$

Remarks:

- A CP-map is trace-preserving iff $\sum_i A_i^\dagger A_i = I_n$; this property is easily verified using the cyclicity of the trace. In terms of the (unique) dual state $\rho_\Phi$ associated to the map $\Phi$, this trace-preserving condition amounts to:
$$Tr_2(\rho_\Phi) = I_n.$$
Here the notation $Tr_2$ means the partial trace over the second subsystem. A CP-map is furthermore called bistochastic if also the condition
$$Tr_1(\rho_\Phi) = I_n$$
holds; this property is equivalent to the fact that the map is identity-preserving, i.e. $\Phi(I_n) = I_n$.
- The dual state $\rho_\Phi$ corresponding to a CP-map $\Phi$ is uniquely defined. The Kraus operators are obtained by considering the columns of a square root of $\rho_\Phi$ ($A_i$ is obtained by making a matrix out of the i'th column of a square root of $X$, with $\rho_\Phi = XX^\dagger$). As the square root of a matrix is not uniquely defined, the Kraus operators are not unique. Each different "square root" $X$ of $\rho_\Phi$ ($\rho_\Phi = XX^\dagger$) gives rise

to a different set of equivalent Kraus operators. This implies that all equivalent sets of Kraus operators are related by an isometry, and that the minimal number of Kraus operators is given by the rank of the density operator $\rho_\Phi$. Therefore we define the rank of a map to be the rank of the dual operator $\rho_\Phi$. This rank is bounded above by $n^2$ with $n$ the dimension of the Hilbert space. A unique Kraus representation can be obtained by for example enforcing the Kraus operators to be orthogonal, as these would correspond to the unique eigenvectors of $\rho_\Phi$. Note that a similar reasoning applies to all Hermitian preserving and all positive maps, although there an additional sign should be taken into account.

- By construction, we have proven that a map $\Phi$ acting on a n-dimensional Hilbert space is completely positive iff $I_n \otimes \Phi$ is positive: there is no need to consider auxiliary Hilbert spaces with dimension larger than the original one. The reasoning is as follows: if $I_n \otimes \Phi$ is positive, then $\rho_\Phi$ is positive, and therefore $\Phi$ has a Kraus representation, which implies complete positivity.
- Suppose $\Phi$ is positive but not completely positive. Then there exists a completely positive map $\tilde{\Phi}$ and a positive scalar $\epsilon$ such that

$$\Phi(\rho) = (1 + n\epsilon)\tilde{\Phi}(\rho) - \epsilon \mathrm{Tr}(\rho) I_n.$$

The proof of this fact is elementary: take $\epsilon$ to be the opposite of the smallest eigenvalue of $\rho_\Phi$ (this eigenvalue is negative as otherwise $\Phi$ would be completely positive), and define the CP-map $\tilde{\Phi}(\rho) = (\Phi(\rho)+n\epsilon\mathrm{Tr}(\rho)I/n)/(1+n\epsilon)$ (this map is completely positive because the dual state $A_{\tilde{\Phi}}$ associated to it is positive and has therefore a Kraus representation). Note that the whole reasoning is also valid for general Hermitian-preserving maps. As an example, consider again the transpose map on a qubit. Then it can be checked that the minimal value of $\epsilon$ is 1 (this is true for the PT operation in arbitrary dimensions) and that the Kraus operators corresponding to $\tilde{\Phi}$ become

$$\{A_i\} = \{\sqrt{\frac{2}{3}}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \sqrt{\frac{2}{3}}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \sqrt{\frac{1}{3}}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}/\sqrt{2}\}.$$

- To make the duality between maps and states more explicit, it is useful to consider the following identity:

$$\Phi(\rho) = Tr_2\left(\rho_\Phi^{T_1}(\rho \otimes I_n)\right), \tag{311}$$

where $T_1$ means partial transposition with relation to the first subsystem. This can be proven by explicitly writing the map $\Phi$ into Kraus operator form, and exploiting the cyclicity of the trace. Due to the partial transpose condition of Peres [**170**], it is clear that $\rho_\Phi^{T_1}$ will typically not longer be positive. This identity is very useful, and was used in the section on optimal teleportation with mixed states.

## 6.2. Extreme points of CP-maps

The set of completely positive maps is a convex set: indeed, if $\Phi_1$ and $\Phi_2$ are CP-maps, then so is $x\Phi_1 + (1-x)\Phi_2$. Due to the one to one correspondence between maps $\Phi$ and states $\rho_\Phi$, it is trivial to obtain the extreme points of the set of completely positive maps: these are the maps with one Kraus operator, corresponding to $\rho_\Phi$ having rank 1.

If however we consider the convex set of trace-preserving maps, the characterization of extreme points becomes more complicated. The knowledge of the set of extreme points of the trace-preserving CP-maps is very interesting from a physical perspective in the following way: suppose one has a multipartite state of qudits and one wants to maximize some convex functional of the state (e.g. the fidelity, ...) by performing local operations. Due to convexity, the optimal operation will correspond to an extreme point of the set of trace-preserving maps.

Let us now characterize all extremal trace-preserving maps:

**Theorem 37.** *Consider a TPCP-map $\Phi$ acting on a Hilbert space of dimension $n$ and of rank $m$. Consider the dual state $\rho_\Phi = XX^\dagger$ with $X$ a $n^2 \times m$ matrix, and the $n^2$ matrices $X_i = X^\dagger(\sigma_i \otimes I_n)X$ (the matrices $\{\sigma_i\}$ form a complete basis for the Hermitian $n \times n$ matrices). Then $\Phi$ is extremal if and only if $m \leq n$ and if the set of linear equations $\forall i : \mathrm{Tr}(QX_i) = 0$ has only the trivial solution $Q = 0$.*

*This condition is equivalent to the following one given by Choi[**56**]: given $m^2$ Kraus operators $\{A_i\}$ of a map $\Phi$, then the map is extremal iff the $m^2$ matrices $\{A_i^\dagger A_j\}$, $1 \leq i, j \leq m$ are linearly independent.*

*Proof:* The map $\Phi$ is extremal if and only if there does not exist a $R$ with the property that $RR^\dagger \neq I$ and such that $\mathrm{Tr}_2(XRR^\dagger X^\dagger) = I$. This condition is equivalent to the fact that the set of equations

$$\mathrm{Tr}\left( X \underbrace{(RR^\dagger - I)}_{Q} X^\dagger \sigma_i \otimes I \right) = 0$$

does only have the trivial solution $Q = 0$. As there are $n^2$ independent generators $\sigma_i$ and due to the fact that $Q$ has $m^2$ degrees of freedom, it is immediately clear that there will always be a non-trivial solution if $m > n$, ending the proof.

It remains to be proven that he condition obtained is equivalent to the one derived[1] by Choi [**56**]. This can be seen as follows: the condition $\mathrm{Tr}_2(XRR^\dagger X^\dagger) =$

[1]Actually, Choi derived the different problem of characterizing the extremal points of the (not necessarily trace-preserving) CP-maps that leave the identity unaffected, but his

$I$ is equivalent to the condition $\sum_{jk} A_k^\dagger A_j(\sum_i R_{ji}R_{ki}^* - \delta_{jk}) = 0$ (this is readily obtained using the trick $|A\rangle = A \otimes I|I\rangle$). Therefore a nontrivial solution of $Q$ is possible iff the set of matrices $\{A_i^\dagger A_j\}$, $1 \le i, j \le m$ are linearly dependent. □

Note that the given proof is constructive and can therefore be used for decomposing a given TPCP-map into a convex combination of extremal maps: once a non-trivial $Q$ and therefore $R$ is obtained, one can scale it such that $RR^\dagger \le I$, and define another $S = \sqrt{I - RR^\dagger}$. This $S$ is guaranteed to be another trace-preserving map up to a constant factor, and the original map is the sum of the maps parameterized by $XRR^\dagger X^\dagger$ and $XSS^\dagger X^\dagger$.

All TPCP maps $\Phi$ of rank 1 are of course extreme and correspond to unitary dynamics. One easily verifies that this implies that the dual $\rho_\Phi$ is a maximally entangled state. The intuition behind this is as follows: by equation (311), $\rho_\Phi$ characterizes the correlation between the output and the input of the channel. Maximal correlation happens iff the evolution occurs reversibly and thus unitarily, and therefore corresponds to maximal "entanglement"between in- and output. We will explore this connection between maps and entanglement more thoroughly in the following section.

One could go one step further, and try to characterize all extreme points of the convex set defined by all trace-preserving channels for which the extra condition holds that $\Phi(\rho_1) = \rho_2$ with $\rho_1$ and $\rho_2$ given density operators. (Note that $\rho_1$ and $\rho_2$ can be chosen completely arbitrary, as there will always exist at least one TPCP-map that transforms a given state into another given one: consider for example the map with its associated dual state $\rho_\Phi = I \otimes \rho_2$.) Bistochastic channels are a special subset of this convex set of maps (in that case $\rho_1 = \rho_2 \simeq I$). An adaption of Theorem 37 leads to the following:

**Theorem 38.** *Consider the convex set of trace-preserving CP-maps $\Phi$ for which $\Phi(\rho_1) = \rho_2$ with $\rho_1, \rho_2$ given. Suppose $\Phi$ is of rank $m$, its dual state is $\rho_\Phi = XX^\dagger$ with $X$ a $n \times n$ matrix, and that there are $m$ Kraus operators $\{A_i\}$. Then this map is extremal if and only if the set of $2m^2$ linear equations*

$$\mathrm{Tr}(QX^\dagger(\sigma_i \otimes I)X) = 0 \qquad \mathrm{Tr}(QX^\dagger(\rho_1^T \otimes \sigma_i)X) = 0 \tag{312}$$

*has only the trivial solution $Q = 0$, or equivalently if and only if the $m^2$ operators $\{A_i^\dagger A_j \oplus A_j \rho_1 A_i^\dagger\}$ $(1 \le i, j \le m)$ are linearly independent.*

Proof: The proof is completely analogous to the proof of Theorem 37, but here we have the extra condition

$$\mathrm{Tr}\left(X(RR^\dagger - I)X(\rho_1^T \otimes \sigma_i)\right) = 0.$$

arguments are readily translated to the present situation. Note also that his proof was much more involved.

In terms of Kraus operators, this additional condition becomes

$$\sum_{kj} A_j \rho_1 A_k^\dagger (\sum_i R_{ji} R_{ki}^* - \delta_{jk}) = 0$$

which ends the proof. □

A similar Theorem was stated by Landau and Streater [**144**] in the special case of bistochastic maps. In analogy with the conclusions of Theorem 37, we conclude that the number of Kraus operators in an extremal TPCP-map of the kind considered in the above Theorem is bounded by $\lfloor \sqrt{2n^2} \rfloor$.

Let us for example consider the case of qubits. Then the rank of an extremal $\Phi$ is bounded by 2, and extremal rank 2 TPCP-maps obeying the condition $\Phi(\rho_1) = \rho_2$ typically exist. There is however a notable exception if $\rho_1 = \rho_2 = I/2$ (i.e. when $\Phi$ is bistochastic): a bistochastic qubit map has a corresponding dual $\rho_\Phi$ that is Bell-diagonal. A Bell-diagonal state is a convex sum of maximally entangled states, and therefore a rank 2 bistochastic map cannot be extremal. Note however that this is an accident, and for Hilbert space dimensions larger than 2 there exist extremal bistochastic channels that are not unitary [**144**]. Sometimes the name "unital"is also used instead of "bistochastic". The foregoing argument however shows that this terminology is not completely justified.

One could now add more constraints $\Phi(\rho_{2i}) = \rho_{2i+1}$, and this would lead to similar conditions for extremality in terms of the Kraus operators. Note however that the $\rho_i$ appearing in the constraints cannot be chosen completely arbitrary, as in general non-compatible constraints can arise due to the complete positivity condition on the physical maps (Deciding whether a set of conditions $\Phi(\rho_{2i}) = \rho_{2i+1}$ is physical can be solved using the techniques of semidefinite programming [**206**]).

Let us now formulate another interesting Theorem:

**Theorem 39.** *Given a Hilbert space of dimension $n$ and a trace-preserving map $\Phi$ of rank $m \leq n$, then there exist pure states $|\psi\rangle$ such that $\Phi(|\psi\rangle\langle\psi|)$ are states of rank $m-1$.*

Proof: Let us first consider the case $m = n$, and define $m$ Kraus operators $\{A_i\}$ corresponding to $\Phi$. Given a pure state $|\psi\rangle$, then $\Phi$ maps this state to one that is not full rank iff there exists a pure state $|\chi\rangle$ such that

$$\langle\chi|\Phi(|\psi\rangle\langle\psi|)|\chi\rangle = 0 = \sum_i |\langle\chi|A_i|\psi\rangle|^2.$$

Writing $|\chi\rangle = \sum_i y_i|i\rangle$, $|\psi\rangle = \sum_i x_i|i\rangle$ and $\langle j|A_i|k\rangle = A^j_{ik}$, then the previous equation amounts to solving the following set of bilinear equations:

$$\forall i = 1 : n, \sum_{k=1}^{n} (\sum_{j=1}^{m=n} x_j A^j_{ik}) y_k = 0.$$

This set of equations always has a non-trivial solution. Indeed, the parameters $x_j$ can always be chosen such that the matrix $\tilde{A} = \sum_j x_j A^j_{ik}$ is singular (if all $A_i$ are full rank then this can be done by fixing all but one of them, and then choosing the remaining parameter such that the determinant vanishes; if one of the $A_i$ is rank deficient then the solution is of course direct). Then the parameters $y_k$ can be chosen such that the vector $y$ is in the right kernel of $\tilde{A}$ (the right kernel is not zero-dimensional as the dimension of the matrix $\tilde{A}$ is $n \times n$), and therefore $\Phi(|\psi\rangle\langle\psi|)$ is not full rank. If $m < n$, then the right kernel of $\tilde{A}$ is at least $n-m+1$ dimensional, such that $n-m+1$ linearly independent $|\chi\rangle$ can be found such that $\langle\chi|\Phi(|\psi\rangle\langle\psi|)|\chi\rangle = 0$, which ends the proof. □

In general , it is thus proven that one can always find states $|\psi\rangle$ such that the rank of $\Phi(|\psi\rangle\langle\psi|)$ is smaller than the rank of the map, which is surprising. Note that the bound in the Theorem is generically tight, i.e. the minimal rank of the output state will typically be $m-1$; this follows from the fact that decreasing the rank of the matrix $\tilde{A}$ with two units would need $n(n-1)/2$ independent degrees of freedom, while there are only $n-1$ available.

Note that extremal TPCP-maps always fulfil the conditions of the Theorem. In particular, extremal qubit channels are generically of rank 2, and the previous Theorem implies that there always exist pure states that remain pure after the action of a rank 2 extremal map (This was also observed by Ruskai et al.[**181**]).

The above Theorem has also some consequences for the study of entanglement. Applying the foregoing proof to the dual state $\rho_\Phi$, we can easily prove the following: if the rank of a mixed state $\rho$ defined in a $n \times n$ dimensional Hilbert space is given by $m \leq n$, then there always exist at least $(n-m+1)$ linearly independent product states orthogonal to it.

Let us now consider an example of the use of extremal maps. Suppose we want to characterize the optimal local trace-preserving operations that one has to apply locally to each of the qubits of a 2-qubit entangled mixed state, such as to maximize the fidelity (i.e. the overlap with a maximally entangled state). This problem is of interest in the context of teleportation [**28, 124**] as the fidelity of the state used to teleport is the standard measure of the quality of teleportation. Badziag and the Horodecki's [**12**] discovered the intriguing property that the fidelity of a mixed state can be enhanced by applying an amplitude damping channel to one of the qubits. This is due to the fact that the fidelity is both dependent on the quantum correlations and on the classical correlations, and enhancing the classical correlations by mixing (and hence losing quantum correlations) can sometimes lead to a higher fidelity.

With the help of the previous analysis of extremal maps, we are in the right position to find the optimal trace-preserving map that maximizes the fidelity. Indeed, the optimization problem is to find the trace-preserving CP-maps $\Phi_A, \Phi_B$ such as to maximize the fidelity $F$ defined as

$$F(\rho, \Phi_A, \Phi_B) = \langle\psi|\Phi_A \otimes \Phi_B(\rho)|\psi\rangle = \mathrm{Tr}\left\{\rho\left(\Phi_A^\dagger \otimes \Phi_B^\dagger(|\psi\rangle\langle\psi|)\right)\right\} \qquad (313)$$

with $|\psi\rangle$ the maximally entangled state. This problem is readily seen to be jointly convex in $\Phi_A$ and $\Phi_B$, and therefore the optimal strategy will certainly consist of applying extremal (rank 2) maps $\Phi_A, \Phi_B$. As we just have derived an easy parameterization of these maps, it is easy to devise a numerical algorithm that will yield the optimal solution.

Note that the problem, although convex in $\Phi_A$ and $\Phi_B$, is bilinear and therefore can have multiple (local) maxima. This problem disappears when only one party (Alice or Bob) applies a map (i.e. $\Phi_B = I$). This problem was studied in more detail by Rehacek et al.[**179**], where a heuristic algorithm was proposed to find the optimal local trace-preserving map to be applied by Bob. As the optimization problem is however convex, the powerful techniques of semidefinite programming [**206**] should be applied, for which an efficient algorithm exists that is assured to converge to the global optimum. Indeed, due to linearity the problem now consists of finding the 2-qubit state $\rho_{\Phi^\dagger} \geq 0$ with constraint $\mathrm{Tr_B}(\rho_{\Phi^\dagger}) = I$ such that the fidelity is maximized. As we already know, the algorithm will converge to a $\rho_\Phi$ of maximal rank 2 in the case of qubits. Exactly the same reasoning holds for systems in higher dimensional Hilbert spaces: if only one party is to apply a trace-preserving operation to enhance the fidelity, the above semidefinite program will produce the optimal local map that maximally enhances the fidelity.

Other situations in which extremal maps will be encountered are for example the problem of optimal cloning[**47, 51, 9**]: given an unknown input state $\rho$, one wants to construct the optimal trace-preserving CP-map such as to yield an output for which the fidelity with $\rho \otimes \rho$ is maximal. This can again be rephrased as a semidefinite program whose unique solution will be given by an extremal trace-preserving CP-map.

## 6.3. Quantum channels and entanglement

The physical interpretation of the dual state corresponding to a CP-map or quantum channel is straightforward. It is the density operator that corresponds to the state that can be made as follows: Alice prepares a maximally entangled state $|I\rangle$, and sends one half of it to Bob through the channel $\Phi$. This results into $\rho_\Phi$.

A perfect quantum channel is unitary and the corresponding state $\rho_\Phi$ is a maximally entangled state. This corresponds to the case of perfect transmission of qudits, and indeed a maximally entangled state is the state with perfect quantum correlations. Consider now a completely depolarizing channel. In that case it is possible to transmit a classical bit perfectly, and indeed $\rho_\Phi$ corresponds to a separable state with maximal classical correlations. As a third example, consider the complete amplitude damping channel. Then $\rho_\Phi$ is a separable pure state with no correlations whatever between Alice and Bob. It is therefore clear that the study of the character of correlation present in the quantum state $\rho_\Phi$ tells us a lot about the character of the quantum channel.

This way of looking at quantum channels gives a nice way of unifying statics and dynamics in one framework: the future is entangled (or at least correlated) with the past. Just as a measurement in the future gives us information about the prepared system (through the use of the quantum Bayes rule), a measurement on Bob's side enables Alice to refine her knowledge of her local system (through the use of the quantum steering Theorem)[2]. It is therefore clear that the description of entanglement will shed new light on the question of describing correlations between the states of the same system at two different instants of time, and vice-versa. Therefore we expect that many useful results concerning entanglement can directly be applied to quantum channels. On the other hand, a lot of work has been done concerning the quantification of the classical capacity of a quantum channel. These results offer a nice starting point for the study of classical correlations present in a quantum state.

### 6.3.1. Quantum capacity

The quantum capacity of a quantum channel is related to the asymptotic number of uses of the channel needed for obtaining states whose fidelity tends to one. To transmit quantum information with high fidelity, one indeed needs almost perfect singlets. It is immediately clear that ideas of entanglement distillation will be crucial: sending one part of an EPR through the channel will result in a mixed state, and these mixed states will have to be purified.

Let us first establish a result that was already intrinsically used by many [**34, 124, 176, 57**]:

[2] In some sense one could argue that this was expected due to the fact that space and time play analogous roles in the theory of relativity. It is very nice however that in the non-relativistic case considered here, the duality is already present. This gives hope that it should be possible to generalize the current findings to the relativistic case.

**Theorem 40.** *A quantum channel $\Phi$ can be used to distribute entanglement if and only if $\rho_\Phi$ is entangled. If $\rho_\Phi$ is separable, then the Kraus operators of the map $\Phi$ can be chosen to be projectors, and the map $\Phi$ is entanglement breaking.*

Proof: The if part is obvious, as $\rho_\Phi$ is the state obtained by sending one part of a maximally entangled state through the channel. To prove the only if part, assume that $\rho_\Phi$ is separable. Then all Kraus-operators can be chosen to be projectors (corresponding to the decomposition with separable pure states), destroying all entanglement. $\square$

It is also possible to make a quantitative statement:

**Theorem 41.** *Suppose we want to use the channel $\Phi$ to distribute entanglement by sending one part of an entangled state through the channel. The maximal attainable fidelity (i.e. overlap with a maximally entangled state) corresponds to the largest eigenvalue of $\rho_\Phi$. This maximal fidelity is obtained if Alice sends one half of the state described by the eigenvector of $\rho_\Phi$ corresponding to its largest eigenvalue.*

Proof: Suppose Alice prepares the entangled state $|\chi\rangle$ and sends the second part to Bob through the channel $\Phi$ with Kraus-operators $\{A_i\}$. We want to find the state $|\chi\rangle$ such that

$$\langle I|\sum_i I\otimes A_i|\chi\rangle\langle\chi|I\otimes A_i^\dagger|I\rangle = \langle\chi|\rho_\Phi|\chi\rangle \tag{314}$$

is maximized, which immediately gives the stated result. $\square$

The above result is amazing: it tells us that it is not always the best strategy to send one part of a maximally entangled state through the channel. It would be tempting to conjecture that the entanglement of distillation of the obtained state represents the quantum capacity of the given channel.

Note that the eigenvalues and eigenvectors of $\rho_\Phi$ got an appealing interpretation: these represent the fidelities that are obtained by sending one half of the eigenvectors through the channel. Note also that the reduction criterion [**119, 52**],

$$I\otimes \mathrm{Tr}_2(\rho_\Phi) - \rho_\Phi = \frac{\mathrm{I}}{\mathrm{n}} - \rho_\Phi$$

implies that $\rho_\Phi$ is entangled if its largest eigenvalue exceeds $1/n$. This is of course in complete accordance with the previous Theorem, as the maximal fidelity for a separable state is also given by $1/n$.

A more sophisticated treatment of the quantum capacity of a quantum channel would involve ideas of coding and of quantum error correction, although only partial results have been obtained yet; the following is an incomplete list of papers where interesting results have been obtained [**34, 191, 72, 239, 16, 106, 105**].

### 6.3.2. Classical Capacity

Let us now move towards the well-studied problem of classical capacity of a quantum channel. The central result is the Holevo- Schumacher- Westmoreland Theorem 34 [**114, 189**], which tells us that the classical product state capacity of a quantum channel $\Phi$ is given by

$$\chi(\Phi) = \max_{p_j,\rho_j} \left\{ S(\Phi(\sum_j p_j\rho_j)) - \sum_j p_j S(\Phi(\rho_j)) \right\}. \tag{315}$$

We have already discussed this Theorem in the last chapter. Let us now ask the following question: what would be the analogy and the interpretation of this formula in the dual picture of states $\rho_\Phi$? Using formula (311), it holds that

$$\Phi(\rho_j) = \mathrm{Tr}_1(\rho_\Phi(\rho_j^T \otimes I)).$$

Suppose Alice and Bob share the state $\rho_\Phi$. Then the above formula describes how Bob has to update his local density operator when Alice did a measurement with corresponding POVM-element $\rho_j^T$. Reasoning along the lines of the HSW-Theorem, the natural interpretation would now be that formula (315) will give us a measure of how much (secret) classical randomness Alice and Bob can create using the state $\rho_\Phi$: if Alice implements a POVM measurement with elements $\{p_j, \rho_j^T\}$, this drives the system at Bob's side into a particular direction, and a measurement of Bob will reveal some information about the (random) outcome of Alice. Note that we interpret the presence of a bipartite state as being a particular kind of quantum channel. Of course the depicted strategy of creating shared randomness is just another application of the quantum steering Theorem (see also [**200, 53**]).

The foregoing discussion suggests the following definition for the classical random correlations $C^{cl}$ present in a quantum state $\rho$:

$$\begin{aligned} C_B^{cl}(\rho_{AB}) &= \max_{\{E_j\}} S(\rho_B) - \sum_j p_j S(\rho_B^j) && (316)\\ p_j &= \mathrm{Tr}\rho(E_j \otimes I) && (317)\\ \rho_B^j &= \frac{1}{p_j}\mathrm{Tr}_1\left(\rho(E_j \otimes I)\right). && (318)\end{aligned}$$

Here $\{E_j\}$ presents the elements of the POVM implemented by Alice. Observe that there is an asymmetry in the definition, in that $C_A^{cl}$ is not necessarily equal to $C_B^{cl}$. This definition coincides with the one given by Henderson and Vedral [**110**], where they introduced this measure because it fulfilled the condition of monotonicity under local operations.

In general, the classical mutual information obtained by the actions of Alice and Bob to obtain classical randomness will be smaller than the derived quantity (316), as coding is needed to achieve the Shannon capacity. This coding could be implemented by doing joint measurements, but we do not expect that the upper bound is tight (private communication of Patrick Hayden); a better rate could be obtained if also public classical communication is allowed (A. Winter, unpublished).

## 6.4. One-qubit channels

In the case of qubit channels, much more explicit results can be obtained, due to the fact that we have a fairly good insight into the properties of mixed states of two qubits. In this section we highlight some questions about qubit channels that can be solved analytically.

Recall formula (311)

$$\Phi(\rho) = Tr_1 \left( \rho_\Phi^{T_1} (\rho \otimes I_n) \right) \tag{319}$$

which is almost exactly the same expression as if Alice were measuring the POVM-element $\rho$ on the joint state $\rho_\Phi$; the difference it that the partial transpose of this state has to be taken. It is now natural to look at the R-picture of the dual state $\rho_\Phi$ associated to the map (cfr. section 4.2.2). In the R-picture, a partial transpose corresponds to a multiplication of the third column or row with a minus sign. Let us therefore define $R_\Phi$ to be the parameterization of $\rho_\Phi^{T_1}$ in the $R$-picture, i.e. the R-picture of $\rho_\Phi$ in which the third row is multiplied by $-1$. Note that the first row of $R_\Phi$ is given by $[1; 0; 0; 0]$, as this corresponds to the trace-preserving condition.

If $x$ is the Bloch vector corresponding, then the action of the map with corresponding $\rho_\Phi^{T_1}$ or $R_\Phi$ is the following:

$$\begin{pmatrix} 1 \\ x' \end{pmatrix} = R_\Phi \begin{pmatrix} 1 \\ x \end{pmatrix} \tag{320}$$

. Exactly as in the case of quantum states, the image of the Bloch sphere yields an ellipsoid (see section 4.3). Here however, the situation is a bit simpler as the local density operator of Alice is always the maximally mixed state: the inner ellipsoid reduces to a point exactly in the middle of the ellipsoid. This implies that the knowledge of the ellipsoid corresponds to the complete knowledge of the quantum channel up to local unitaries at the input. (Recall that not all ellipsoids correspond to physical maps, but that there is some restriction on the ratio of the axis (see section 4.3)).

Let us now consider the analogue of LU and SLOCC equivalence classes derived for mixed states of two qubits. What we are looking for are normal forms $\Omega$ (where $\Omega$ is a map) such that $\Phi(\rho) = B\Omega(A\rho A^\dagger)B^\dagger$ with $A, B \in SU(2)$ or $\in SL(2, \mathcal{C})$.

The LU case is very easy: each $R_\Phi$ can be brought into the unique form

$$R_\Phi = \begin{pmatrix} 1 & 0 & 0 & 0 \\ x & \lambda_1 & 0 & 0 \\ y & 0 & \lambda_2 & 0 \\ z & 0 & 0 & \pm\lambda_3 \end{pmatrix}$$

by local unitary transformations, where $\lambda_1 \geq \lambda_2 \geq |\lambda_3|$ and $x, y \geq 0$; one just has to take the singular value decomposition of the lower $3 \times 3$ block of $R$, taking into account that the orthogonal matrices have determinant $+1$ (see also Fujiwara and Algoet [**92**] and King and Ruskai [**140**] for a different approach but with the same result).

Let us next move to SLOCC equivalence classes; it is clear that the Lorentz singular value decomposition is all we need:

**Theorem 42.** *Given a 1-qubit trace-preserving CP-map $\Phi$ and its dual $R_\Phi$. Then the SLOCC normal form $\Omega$ of $R_\Phi$ is proportional to one of the following unique normal forms:*

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & s_1 & 0 & 0 \\ 0 & 0 & s_2 & 0 \\ 0 & 0 & 0 & s_3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & x/\sqrt{3} & 0 & 0 \\ 0 & 0 & x/\sqrt{3} & 0 \\ 2/3 & 0 & 0 & 1/3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

*Here $1 \geq s_1 \geq s_2 \geq |s_3|$, $1 - s_1 - s_2 - s_3 \geq 0$ and $0 \leq x \leq 1$. For maps with a normal from of the first kind, one can choose the Kraus operators equal to*

$$\{A_i\} = \{p_0 A\sigma_0 B, p_1 A\sigma_1 B, p_2 A\sigma_2 B, p_3 A\sigma_3 B\} \tag{321}$$

*with $A, B$ complex $2 \times 2$ matrices and $p_i \geq 0$, related to the $\{s_i\}$ by the formula relating the eigenvalues of a Bell diagonal state to its Lorentz singular values. The Kraus operators of maps with a normal form of the second kind can be chosen to be of the form*

$$\{A_i\} = \{\sqrt{\frac{1+x}{2}} A \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\sqrt{3}} \end{pmatrix} B, \sqrt{\frac{1-x}{2}} A \begin{pmatrix} 1 & 0 \\ 0 & -\frac{1}{\sqrt{3}} \end{pmatrix} B, \sqrt{\frac{2}{3}} A \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} B\}, \tag{322}$$

*again with $A, B$ complex $2 \times 2$ matrices. In the third case, the map is trivial as it maps everything to the same point. $\{s_i\}, x, A, B, \{p_i\}$ can be calculated explicitly by calculating the Lorentz singular value decomposition of the state $\rho_\Phi$.*

*Proof:* The proof is immediate given the Lorentz singular value decomposition. The first case corresponds to a diagonalizable $R$, and a diagonal $R$ corresponds to a bistochastic channel. The second and third case correspond to non-diagonalizable cases (note that there are 2 normal forms in the case of states that do not apply here as they cannot lead to trace-preserving channels). $\square$

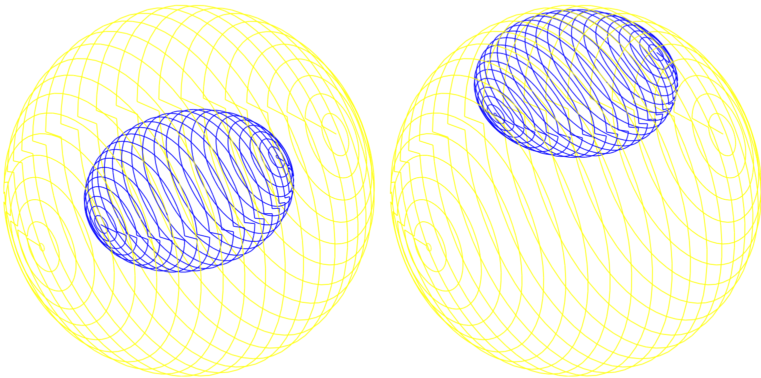


**Figure 1.** The image of a channel in generic normal form (left) or in non-generic normal form (right).

This gives a nice classification of all the classes of TPCP-maps on qubits: the generic class is the one that can be brought into unital form by adding appropriate filtering transformations $A, B$, i.e. the ellipsoid can be continuously deformed to an ellipsoid whose center is the maximally mixed state. The non-generic class however cannot be deformed in this way: it is easy to show that the ellipsoid corresponding to the normal form touches the Bloch sphere at one and only at one point; there is no filtering operation that can change this property. We conclude that the ellipsoids in the non-generic case are not (and cannot be made by filtering operations) symmetric around the origin and that they touch the Bloch sphere at exactly one point.

We depict both types of normal ellipsoids in figure 1. Note that this geometrical picture will be very useful in guessing input states that maximize the classical capacity of the state (see e.g. [**140**]).

### 6.4.1. Extremal maps for qubits

In the case of a qubit channel $\Phi$, the dual state $\rho_\Phi$ is a mixed state of two qubits. It is possible to obtain an explicit parameterization of all extremal qubit maps (see also Ruskai et al. [**181**] for a different approach):

**Theorem 43.** *The set of dual states $\rho_\Phi$ corresponding to extreme points of the set of completely positive trace preserving maps $\Phi$ on 1 qubit is given by the union of all maximally entangled pure states, and all rank 2 states $\rho$ for which $\mathrm{Tr}_2(\rho_\Phi)$ is equal and $\mathrm{Tr}_1(\rho_\Phi)$ is not equal to the identity. The Kraus operators corresponding to the rank 1 extreme points are unitary, while the ones corresponding to the rank 2 extreme points have a representation of the* *form:*

$$A_1 = U \begin{pmatrix} s_0 & 0 \\ 0 & s_1 \end{pmatrix} V^\dagger \qquad A_2 = U \begin{pmatrix} 0 & \sqrt{1-s_1^2} \\ \sqrt{1-s_0^2} & 0 \end{pmatrix} V^\dagger \tag{323}$$

*with $U, V$ unitary.*

Proof: We have already proven that extremal TPCP-maps have maximal rank 2. Due to the duality between maps and states, it is sufficient to consider rank 2 density operators of two qubits $\rho_\Phi$ for which $Tr_2(\rho_\Phi) = I_2$. A real parameterization of all 2-qubit density operators $\rho$ is given by the real $4 \times 4$ matrix $R$ with coefficients

$$R_{ij} = Tr\left(\rho \sigma_i \otimes \sigma_j\right) \tag{324}$$

where $0 \leq i, j \leq 3$. An appropriate choice of local unitary bases can always make the $R_{1:3,1:3}$ block diagonal, and the trace-preserving condition translates into $R_{0,1:3} = 0$. Therefore $R$ is given by:

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ t_1 & \lambda_1 & 0 & 0 \\ t_2 & 0 & \lambda_2 & 0 \\ t_3 & 0 & 0 & \lambda_3 \end{pmatrix}.$$

The corresponding $\rho$ is given by

$$\rho = \frac{1}{4} \begin{pmatrix} 1+t_3+\lambda_3 & 0 & t_1 - it_2 & \lambda_1 - \lambda_2 \\ 0 & 1+t_3-\lambda_3 & \lambda_1+\lambda_2 & t_1 - it_2 \\ t_1+it_2 & \lambda_1+\lambda_2 & 1-t_3-\lambda_3 & 0 \\ \lambda_1-\lambda_2 & t_1+it_2 & 0 & 1-t_3+\lambda_3 \end{pmatrix},$$

and the positivity of $\rho$ constrains the allowed range of the 6 parameters. Let us now impose that the rank of the corresponding $\rho$ is 2. This implies that linear combinations of $3 \times 3$ minors of $\rho$ be zero, and after some algebra one obtains the following conditions:

$$\begin{aligned} t_3(\lambda_3 + \lambda_1\lambda_2) &= 0 \\ t_2(\lambda_2 + \lambda_1\lambda_3) &= 0 \\ t_1(\lambda_1 + \lambda_2\lambda_3) &= 0 \end{aligned}$$

These equations, supplemented with the fact that diagonal elements of a positive semidefinite matrix are always bigger than the elements in the same column, lead to the conclusion that all $t_i$ but one have to be equal to zero if $\rho$ is rank 2. Without loss of generality, we can choose $t_1 = t_2 = 0$ and parameterize $\lambda_1 = \cos(\alpha)$, $\lambda_2 = \cos(\beta)$. We thus arrive at the canonical form

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\alpha) & 0 & 0 \\ 0 & 0 & \cos(\beta) & 0 \\ \sin(\alpha)\sin(\beta) & 0 & 0 & -\cos(\alpha)\cos(\beta) \end{pmatrix}. \tag{325}$$

Suppose that $\sin(\alpha)\sin(\beta) = 0$ (this condition is equivalent to $Tr_1(\rho_\Phi) = I/2$. Then the state corresponding to this $R$ is Bell-diagonal and thus a convex sum of two maximally entangled states, and therefore the map corresponding to this state cannot be extremal. In the other case, an extremal rank 2 TPCP-map is obtained, which can easily be shown to yield the given Kraus representation, where $s_0 = \sqrt{1-\cos(\alpha+\beta)/2}$ and $s_1 = \sqrt{1-\cos(\alpha-\beta)}/2$. $\square$

Note that the corresponding Theorem for bistochastic qubit channels is not very useful, as extremal TPCP qubit channels are always unitary. Theorem 39 however is very interesting, and indicates that there always exist pure states that remain pure after the action of the extremal qubit channel: indeed, if the basis vectors $\{|i\rangle\}$ are chosen according to the unitary $V$ in (323), then it is easily checked that the states $|\psi\rangle \simeq s_2\sqrt{1-s_2^2}|0\rangle \pm s_1\sqrt{1-s_1^2}$ remain pure by the action of the extremal map. Note that these two states are the only ones with this property, and note also that they are not orthogonal to each other.

### 6.4.2. Quantum capacity

Let us now move on to the relation between 1-qubit quantum channels and entanglement. We can now make use of the plethora of results derived for mixed states of two qubits. Let us first consider Theorem 40 about entanglement breaking channels. In the case of mixed states of two qubits, a state is entangled iff it violates the reduction criterion $I \otimes \rho_B - \rho \geq 0$. But in the case of the dual state $\rho_\Phi$, it holds that $\rho_B = I/2$, and therefore it holds that a quantum channel $\Phi$ can be used to distribute entanglement iff the maximal eigenvalue of $\rho_\Phi$ exceeds $1/2$ (this was first observed by Michael Horodecki[3]). In the light of Theorem 41, it follows that such a non-entanglement breaking channel can always be used to distribute an entangled state with fidelity larger than $1/2$, which implies on its turn that it can be used to distill entanglement[**34**].

Consider now an entanglement breaking channel, i.e. a channel for which $\rho_\Phi$ is separable. In this case all the Kraus operators can be chosen to be projectors. An explicit way of calculating this Kraus representation exists. Indeed, in the section about entanglement of formation of two qubits, a constructive way of decomposing a separable mixed state of two qubits as a convex combination of separable pure states was given. It was furthermore proven that a separable state of rank 2 or 4 can always be written as a convex combination of 2 respectively 4 separable pure states, thus giving rise to 2 respectively 4 rank one Kraus operators. Surprisingly, most separable rank 3 mixed states of two qubits can only be written as a convex combination of 4 separable pure states. This implies that a generic entanglement breaking channel of rank 3 needs 4 Kraus operators if these are to be chosen rank 1. Let us also mention

[3]Private communication

that the set of separable states is not of measure zero, implying that the set of entanglement breaking channels is also not of measure zero.

The results of Wootters [**244**] can of course also be applied to non-entanglement-breaking channels. A direct application of the formalism developed in section 4.4.1 yields the following Theorem:

**Theorem 44.** *Given a 1-qubit channel $\Phi$ and the state $\rho_\Phi$ associated to it. If $C$ is the concurrence of $\rho_\Phi$, then the channel has a Kraus representation of the form:*

$$\Phi(\rho) = \sum_i p_i(U_i\tilde{C}V_i)\rho(U_i\tilde{C}V_i)^\dagger \tag{326}$$

$$\tilde{C} = \frac{1}{2}\begin{pmatrix} \sqrt{1+C}+\sqrt{1-C} & 0 \\ 0 & \sqrt{1+C}-\sqrt{1-C} \end{pmatrix} \tag{327}$$

*where $U_i, V_i$ are unitary matrices.*

Proof: The Theorem is a direct consequence of the fact that a mixed state with concurrence $C$ can be written as a convex sum of pure states all with concurrence equal to $C$. □

The geometrical meaning in the context of channels is the following: each trace-preserving CP-map is a convex combination of contractive maps in unique different directions, where each contraction has the same magnitude.

Let us next address the question of calculating the quantum capacity of the one-qubit channel. Clearly, Theorem 41 tells us what states to send through the channel such as to maximize the fidelity of the shared entangled states. In general, the quantum capacity cannot be calculated as we even don't have a way of calculating the entanglement of distillation of mixed states of two qubits (which is a simpler problem).

In the case of unital channels of rank 2 however, the eigenvectors of $\rho_\Phi$ are maximally entangled and the quantum capacity can be calculated explicitly:

**Theorem 45.** *Consider a bistochastic qubit channel $\Phi$ of rank 2. Then its quantum capacity is given by $C_Q = 1-H(p)$, where $p$ is the maximal eigenvalue of $\rho_\Phi$ and $H(p) = -p\log_2(p)-(1-p)\log_2(1-p)$.*

Proof: A unital qubit channel exhibits the nice property that no loss whatever occurs by sending a maximally entangled state through the channel: it can easily be shown (see Bennett et al.[**34**]) that sending a quantum system through the channel is equivalent to using the standard teleportation channel induced by the (non-maximally entangled state) $\rho_\Phi$. Because we can use the state $\rho_\Phi$, obtained by sending a Bell state through the channel, to perfectly simulate the channel, this is clearly the optimal thing to do, and the quantum capacity of the channel is therefore equal to the distillable entanglement of $\rho_\Phi$. Now Rains

[**175**] has proven that the distillable entanglement of a Bell diagonal state of rank 2 is given by $E_{dist}(\rho) = 1 - S(\rho)$, which ends the proof of the Theorem. □

More general, the quantum capacity of bistochastic qubit channel is always equal to the entanglement of distillation of the corresponding dual states (due to the arguments in the previous proof).

As a last remark, we observe that the channels of the non-generic kind that touch the Bloch sphere at exactly one point are never entanglement-breaking: this follows from the fact that the concurrence of $\rho_\Phi$ always exceeds 0 in that case.

### 6.4.3. Classical capacity

Far more progress has been made concerning the classical capacity of quantum channels: it is known that the classical capacity using product inputs is given by the Holevo-$\chi$ quantity. Here the geometrical picture derived in section 6.4 can sharpen our intuition. Consider for example the case of a unital channel. It is immediately clear that Holevo-$\chi$ will be maximized by choosing a mixture of two states that lie on the opposite side of the major axis of the ellipsoid. This implies that the optimal input states are orthogonal. King and Ruskai [**140, 138**] even proved that entangled inputs cannot help in the case of unital channels, and we conclude that the classical capacity of the unital channels is completely understood.

Consider however a non-unital channel of the generic kind. As proven before, this channel can be interpreted as the succession of a filter, a unital channel, and another filter. The critical source of noise or decoherence and irreversibility in a channel is the mixing, and the previous analysis tells us that this mixing can always be interpreted to happen in a unital way, whereas the in- and output of the unital channel is reversibly but non-orthogonally filtered. It follows that orthogonal inputs will not appear orthogonally in the unital channel, and typically orthogonal inputs will not achieve capacity. This strange fact was indeed discovered by Fuchs [**89**], and it appears to be generic for non-unital channels.

Let us now have a look at the non-generic family of channels, whose ellipsoids touch the Bloch sphere at exactly one point. It happens that the so-called stretched channel belongs to this family, and this channel has the property that its (product) capacity is only achieved for an input ensemble with three states[**139**]. This is surprising but not too surprising given the geometrical picture, as one of the input states corresponds to the pure output state, while the other two ones are chosen to lie symmetric around the axis connecting the maximally entangled state with the pure output state. Note however that most of the non-generic states achieve capacity with 2 input states.

Let us now move to calculate the classical capacity of the extremal qubit channels. In the case of extremal qubit channels, it is possible to reduce the problem of calculating the classical (Holevo) capacity to an optimization problem over the ensemble average. The problem to be solved is as follows: find the optimal ensemble $\{\rho_i, p_i\}$ such that

$$S(\sum_i p_i \Phi(\rho_i)) - \sum_i p_i S(\Phi(\rho_i))$$

is maximized. We assume that $\Phi$ is rank 2 and therefore has a Kraus representation of the form (323). It is clear that only pure states $\{\rho_i\}$ have to be considered. It is easily seen that in the case of qubits, the entropy of a state is a convex monotonously increasing function of the determinant of the density operator: $S(\rho) = H(1/2(1 - \sqrt{1 - 4\det(\rho)^2}))$ with $H(p) = p\log(p) + (1-p)\log(1-p)$ the Shannon entropy function. Inspired by the analysis of 2-qubit channels by Uhlmann in terms of anti-linear operators [**203**], we make the following observation:

$$\det\left(A_1|\psi\rangle\langle\psi|A_1^\dagger + A_2|\psi\rangle\langle\psi|A_2^\dagger\right) = |\psi^T(A_1^T\sigma_y A_2 - A_2^T\sigma_y A_1)\psi|. \qquad (328)$$

Here $\psi$ is the vector notation (in the computational basis) of $|\psi\rangle$, and $\sigma_y$ is a Pauli matrix. Suppose now that we add an additional constraint to the problem, namely that the ensemble average $\rho$ is given. Taking a square root $X$ of $\rho = XX^\dagger$, all possible pure state decompositions can be written as $X' = XU$ with $U$ an arbitrary isometry (note that the columns of $XU$ represent all unnormalized pure states in the decomposition). With this additional constraint, the problem can be solved exactly as we solved the entanglement of formation problem. A constructive way of obtaining the optimal decomposition of $\rho$ is as follows: take a square root $X$ of $\rho$, and calculate the singular value decomposition of the symmetric matrix $X^T(A_1^T\sigma_y A_2 - A_2^T\sigma_y A_1)X = V\Sigma V^T$. Call $C = \sigma_1 - \sigma_2$ the concurrence with $\{\sigma_i\}$ the singular values of the above symmetric matrix. Then the optimal decomposition is obtained by choosing $U = V^*O$ with $O$ the real orthogonal matrix that is chosen such that the diagonal entries of the matrix $R = O^T(\mathrm{Diag}[\sigma_1, -\sigma_2] - C\rho)O)$ vanish (for a more elaborate discussion, we refer to section 4.4.1). For given ensemble average $\rho$, the classical capacity is therefore given by the following formula: $S(\Phi(\rho)) - f(C)$ (see also Uhlmann [**203**]).

To derive an explicit formula for the classical capacity of the extremal channels, we still have to do an optimization over all possible ensemble averages $\rho$. Note that the previous analysis already learned us that the capacity will always be reached with an ensemble of two input states. Both the terms $\Phi(\rho)$ and $C$ can easily be extremized separately, but unfortunately even if the eigenvalues of $\rho$ are fixed, the optimal eigenvectors for maximizing $S(\rho)$ and minimizing $C$ are not compatible. However, the capacity can easily be calculated numerically, as it just an optimization problem over three real parameters.

On the other hand, we have seen that the definition of the classical capacity had a direct counterpart in giving an appealing definition for the number of classical correlations present in a (mixed) bipartite state $C_{cl}$ (see 316). The techniques used in the foregoing paragraph are perfectly adequate to give an exact expression of this quantity if the shared quantum state is a rank 2 bipartite state $\rho$ of qubits. Indeed, a mixed bipartite state of two qubits can just be seen as a more general kind of quantum channel.

## 6.5. Maps on entangled systems

### 6.5.1. General Case

The formalism developed can readily be generalized to describe operations on entangled systems. The most general TPCP-map on two qubits for example is described by the $2 \times 2 \times 2 \times 2$ density operator obtained by letting the map act on two locally maximally entangled states. It is clear that the dual state corresponding to a unitary map is a pure maximally entangled state.

Following Rains [**176**], let us now try to get a description of maps that can be implemented locally, i.e. let us try to describe LOCC maps. It is clear that a necessary condition is the fact that the dual state corresponding to the map is separable (note however that this condition is not sufficient: in a remarkable paper [**32**] it was shown that there exist separable operations that cannot be implemented locally). But checking separability is extremely hard, and therefore Rains introduced the class of PPT-operations: this is the class of operations whose dual states have a positive partial transpose, and it thus strictly larger than the class of separable and LOCC operations. This was exactly the technique used in section 4.5, where the remarkable result was obtained that in the special case of optimal enhancement of the fidelity of two qubits, an optimization over all PPT-maps lead to a physically implementable LOCC map. This is supporting evidence for the fact that the class of PPT-operations is not much larger than the class of LOCC-operations.

Let us now move to a completely different topic, namely the implementation of non-local maps. When entanglement and classical communication are for free, then it is obvious that all non-local maps can easily be implemented by LOCC operations through teleportation (note that this fact is one of the reasons why teleportation is so important). But is it also possible to implement some operations with less entanglement? This was the question raised in a paper of Cirac et al.[**57, 77**]. Using the identity (see equation (311))

$$\rho_{A'}^T = \mathrm{Tr}_A \left( |I_{AA'}\rangle\langle I_{AA'}|(\rho_A \otimes I'_A) \right),$$

they proved the following very interesting relation (we represent the dual state to the map $\Phi$ by $X$):

$$
\begin{aligned}
\Phi(\rho)_{A_2B_2} &= \mathrm{Tr}_{A_1B_1}\left(X^{T_{A_2B_2}}_{A_1B_1A_2B_2}\rho_{A_1B_1}\right) \\
&= \mathrm{Tr}_{A_3B_3}\left(\underbrace{\mathrm{Tr}_{A_1B_1}\left\{X^{T_{A_2B_2}}_{A_1B_1A_2B_2}|I_{A_1A_3}\rangle\langle I_{A_1A_3}|\otimes|I_{B_1B_3}\rangle\langle I_{B_1B_3}|\right\}}_{\Psi_{A_2B_2A_3B_3}}\rho^T_{A_3B_3}\right) \\
&= \langle I_{A_3A_4}|\langle I_{B_3B_4}|\Psi_{A_2B_2A_3B_3}\otimes\rho_{A_4B_4}|I_{A_3A_4}\rangle|I_{B_3B_4}\rangle \qquad (329)
\end{aligned}
$$

The interpretation of $\Psi_{A_2B_2A_3B_3}$ is the following: it is the density operator obtained by applying the map $\Phi$ to the halves of two locally prepared maximally entangled states. The crucial point is the fact that $\Psi$ will not contain a lot of entanglement if the operation $\Phi$ is not able to do so. The interpretation of the last step in the derivation represents a local Bell measurement: a Bell measurement has to be implemented on the original state and the state $\Psi$. Note that this implies that this technique only allows to implement the map with a certain probability of success.

To summarize, the previous derivation indicates that every non-local operation can be implemented locally using a limited amount entanglement. Note that due to the Bell measurement this only works with a certain probability, and that classical communication is required as both parties have to agree on the measurement outcome. If the non-local map to be implemented is unitary, the whole process is reversible (even if the wrong Bell measurement outcome was obtained) and by repeating the procedure until the desired outcome is obtained, one can implement a unitary map with 100% probability (see [**57, 77**]). If a certain unitary operation has to be implemented that has not too much entangling capacity, then this procedure turns out to consume much less entanglement than the teleportation scheme.

The previous formalism can readily be extended to the multipartite case. It also tells us something very sensible about the class of PPT-operations: these can be implemented by making use of bound entangled states (this would maybe be the most interesting application of bound entangled states). Note also that there is no contradiction with the fact that not all separable operations can be implemented [**32**]: the present discussion is only probabilistic, and one can always implement a separable operation with a certain probability.

The previous formalism developed by Cirac et al. induces some new kind of duality between maps and entangled states: it is possible to “store”entangling operations in entangled states; it is possible to compress them using techniques of entanglement distillation; it is possible to teleport them, etc. It also allows to translate a lot of results obtained in the context of entanglement transformations into the context of maps. Questions like “Can I locally implement this specific operation using another one?”are readily translated (see [**78**]) into the equivalent question “Can I transform this entangled state into another one?”. This is of great practical interest, as the kind of interactions that one can typically implement experimentally are fixed.

This implies that all the results on LU,LOCC and SLOCC equivalence classes derived in the first part of this thesis are of direct use in the context of simulation of one map in terms of another one. Especially the central Theorem 6 of section 3.1 is interesting in this context: due to the unicity of the normal form, to check whether one unitary can simulate another one by a certain probability amounts to checking if their corresponding SLOCC-normal form is equivalent up to local unitary operations.

### 6.5.2. Entanglement Capability of non-local Hamiltonians

Let us finally move to another related topic: given a Hamiltonian that can couple two systems, what states have to be prepared such as to maximize the entanglement capability of this Hamiltonian [**79, 248**]? This question turns out to be equivalent to finding the optimal way of making use of a given Hamiltonian to exchange classical information [**35**]. This section is a reprint of a recent article of Childs, Leung, Verstraete and Vidal [**55**] under the title *Asymptotic entanglement capacity of the Ising and anisotropic Heisenberg interactions*:

The fundamental resource for quantum information processing is an interaction between two quantum systems. Any Hamiltonian $H_{AB} \neq H_A + H_B$ that is not a sum of local terms couples the systems $A$ and $B$. Together with local operations, the coupling can be used to generate entanglement [**79, 35, 248, 141**], to transmit classical and quantum information [**35, 107, 36**], and more generally, to simulate the bipartite dynamics of some other Hamiltonian $H'_{AB}$ and thus to perform arbitrary unitary gates on the composite space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ [**73, 242, 137, 224, 243, 163, 54, 227, 156, 31**].

Much experimental effort has been devoted to creating entangled states of quantum systems, including those in quantum optics, nuclear magnetic resonance, and condensed matter physics [**128**]. Determining the ability of a system to create entangled states provides a benchmark of the “quantumness” of the system. Furthermore, such states could ultimately be put to practical use in various quantum information processing tasks, such as superdense coding [**23**] or quantum teleportation [**28**].

The theory of optimal entanglement generation can be approached in different ways. For example, Ref. [**79**] considers *single-shot* capacities. In the case of two-qubit interactions, and assuming that ancillary systems are not available, Ref. [**79**] presents a closed expression for the entanglement capacity and optimal protocols by which it can be achieved. In contrast, Ref. [**35**] considers the *asymptotic* entanglement capacity, allowing the use of ancillary systems, and shows that when ancillas are allowed, the single-shot and asymptotic capacities are in fact the same. However, such capacities can be difficult to calculate because the ancillary systems may be arbitrarily large.

In this section, we calculate the asymptotic entanglement capacity of any two-qubit interaction that is locally equivalent to $\mu_x\, \sigma_x \otimes \sigma_x + \mu_y\, \sigma_y \otimes \sigma_y$, and thus present a connection between the results of Refs. [**79**] and [**35**]. We consider the use of ancillary systems, and show that they do not increase the entanglement capacity of these interactions. Thus in these cases, the asymptotic capacity discussed in Ref. [**35**] is in fact given by the expression presented in Ref. [**79**]. As an application of this result, we present an explicit ensemble for entanglement assisted classical communication [**35**], implicitly found in Ref. [**36**], at a rate equal to the entanglement capacity. We also give an alternative ensemble achieving the same rate. Finally, we conclude by presenting some numerical data on the entanglement capacity of general two-qubit interactions.

We begin by reviewing some definitions and known results. Let $|\psi\rangle$ be a state of the systems $A$ and $B$. This state can always be written using the Schmidt decomposition [**169**],

$$|\psi\rangle := \sum_i \sqrt{\lambda_i}\ |\phi_i\rangle_A \otimes |\eta_i\rangle_B\,, \tag{330}$$

where $\{|\phi_i\rangle\}$ and $\{|\eta_i\rangle\}$ are orthonormal sets of states, and $\lambda_i > 0$ with $\sum_i \lambda_i = 1$. The entanglement between $A$ and $B$ is defined as

$$E(|\psi\rangle) := -\sum_i \lambda_i \log \lambda_i\,. \tag{331}$$

(Throughout this section, log is base 2.)

Reference [**79**] considers maximizing the rate of increase of entanglement when a pure state is acted on by $e^{-iHt}$, the evolution according to a Hamiltonian $H$ (we set $\hbar = 1$ throughout this section). We refer to this maximal rate as the *single-shot* entanglement capacity. When no ancilla's are used, this is given by

$$E_H^{(1*)} := \max_{|\psi\rangle\in\mathcal{H}_{AB}} \lim_{t\to 0} \frac{E(e^{-iHt}|\psi\rangle) - E(|\psi\rangle)}{t}\,. \tag{332}$$

Here the rate of increasing entanglement is optimized over all possible pure initial states of $\mathcal{H}_{AB}$ without ancillary systems. In fact, the single-shot capacity may be higher if ancillary systems $A'$ and $B'$, not acted on by $H$, are used. For this reason, we may consider the alternative single-shot entanglement capacity

$$E_H^{(1)} := \sup_{|\psi\rangle\in\mathcal{H}_{AA'BB'}} \lim_{t\to 0} \frac{E(e^{-iHt}|\psi\rangle) - E(|\psi\rangle)}{t}\,. \tag{333}$$

For any two-qubit Hamiltonian $H$, Ref. [**79**] shows that it is locally equivalent to a *normal form*

$$\sum_{i=x,y,z} \mu_i\, \sigma_i \otimes \sigma_i\,, \quad \mu_x \geq \mu_y \geq |\mu_z|\,. \tag{334}$$

In terms of this normal form, the optimal single-shot entanglement capacity of any two-qubit interaction without ancillas is given by

$$E_H^{(1*)} \;=\; \alpha(\mu_x + \mu_y)\,, \tag{335}$$

$$\alpha \;:=\; 2 \max_x \sqrt{x\,(1-x)}\, \log\left(\frac{x}{1-x}\right) \approx 1.9123\,, \tag{336}$$

where the maximum is obtained at $x_0 \approx 0.9168$. In addition, $E_H^{(1)}$ may be strictly larger than $E_H^{(1*)}$ when $|\mu_z| > 0$ [**79**].

Reference [**35**] considers the *asymptotic* entanglement capacity $E_H$ for an arbitrary Hamiltonian $H$. $E_H$ is defined as the maximum rate at which entanglement can be produced by using many interacting pairs of systems, in parallel or sequentially. These systems may be acted on by arbitrary collective local operations (attaching or discarding ancillary systems, unitary transformations, and measurements). Furthermore, classical communication between $A$ and $B$ and possibly mixed initial states are allowed. Reference [**35**] proves that the asymptotic entanglement capacity in this general setting turns out to be just the single-shot capacity in Ref. [**79**], $E_H = E_H^{(1)}$, for all $H$, so

$$E_H = \sup_{|\psi\rangle \in \mathcal{H}_{AA'BB'}} \lim_{t\to 0} \frac{E(e^{-iHt}|\psi\rangle) - E(|\psi\rangle)}{t}\,. \tag{337}$$

Let $|\psi\rangle$ be the optimal input in Eqs. (333) or (337). When $|\psi\rangle$ is finite dimensional, the entanglement capacity can be achieved [**79, 35**] by first inefficiently generating some EPR pairs, and repeating the following three steps: $(i)$ transform $nE(|\psi\rangle)$ EPR pairs into $|\psi\rangle^{\otimes n}$, $(ii)$ evolve each $|\psi\rangle$ according to $H$ for a short time $\delta t$, and $(iii)$ concentrate the entanglement into $n(E(|\psi\rangle) + \delta t E_H)$ EPR pairs.

In this section, we show that $E_K^{(1)} = E_K^{(1*)}$ for any two-qubit Hamiltonian with normal form

$$K := \mu_x\, \sigma_x \otimes \sigma_x + \mu_y\, \sigma_y \otimes \sigma_y\,, \quad \mu_x \geq \mu_y \geq 0\,, \tag{338}$$

so that all three entanglement capacities are equal:

$$E_K = E_K^{(1)} = E_K^{(1*)}\,. \tag{339}$$

The optimal input is therefore a two-qubit state, and the optimal protocol applies. In particular, for these Hamiltonians, which include the Ising interaction $\sigma_z \otimes \sigma_z$ and the anisotropic Heisenberg interaction $\sigma_x \otimes \sigma_x + \sigma_y \otimes \sigma_y$, entanglement can be optimally generated from a 2-qubit initial state $|\psi\rangle$ without ancillary systems $A'B'$. As mentioned above, this result is not generic, since ancillas increase the amount of entanglement generated by some two-qubit interactions, such as the isotropic Heisenberg interaction $\sigma_x\otimes\sigma_x+\sigma_y\otimes\sigma_y+\sigma_z\otimes\sigma_z$ [**79**].

In the following, we will focus on computing the asymptotic entanglement capacity of the interaction

$$K_{xx} := \sigma_x \otimes \sigma_x \,. \tag{340}$$

One way to see that this is sufficient to determine the asymptotic entanglement capacity of $K$ in Eq. (338) is to note that $K$ is *asymptotically equivalent* to

$$K' := (\mu_x + \mu_y)\sigma_x \otimes \sigma_x \tag{341}$$

(notice that $E_{tH} = |t|E_H$ for two-qubit Hamiltonians) [**225**]. This equivalence is based on the following two facts: $(i)$ $K'$ and fast local unitary transformations on qubits $A$ and $B$ can simulate $K$ [**31**]; conversely, $(ii)$ the Hamiltonian $K$ can be used to simulate $K'$ given a *catalytic* maximally entangled state, without consuming the entanglement of $A'B'$, which subsequently can be reused [**225**]. Therefore, Hamiltonians $K$ and $K'$ are asymptotically equivalent resources given local operations and an asymptotically vanishing amount of entanglement. Thus *any* asymptotic capacity must be equal for $K$ and $K'$, and in particular, $E_K = E_{K'}$. For the specific case of entanglement capacity, a simpler proof is available. The simulation $(i)$, which does not require a catalyst, demonstrates $E_K \leq E_{K'}$. After computing $E_{K'}$, we will see that the protocol of Ref. [**79**] saturates this bound, so in fact $E_K = E_{K'}$ with no need for ancillas to achieve either capacity.

We now present the optimization of Eq. (337) for $K_{xx}$. We suppose that in addition to the qubits $A$ and $B$ on which $K_{xx}$ acts, $d$-dimensional ancillas $A'$ and $B'$ are used, where $d$ is arbitrary. We can always write the Schmidt-decomposed initial state $|\psi\rangle$ as

$$\begin{aligned} |\psi\rangle &= \sum_i \sqrt{\lambda_i}\, |\phi_i\rangle_{AA'} \otimes |\eta_i\rangle_{BB'} && (342)\\ &= (U \otimes V)(\sqrt{\Lambda} \otimes I_{BB'})|\Phi\rangle && (343)\\ &= U\sqrt{\Lambda}V^T \otimes I_{BB'}|\Phi\rangle \,, && (344) \end{aligned}$$

where $U$ and $V$ are unitary matrices on $\mathcal{H}_{AA'}$ and $\mathcal{H}_{BB'}$, $\Lambda$ is a diagonal matrix with diagonal elements $\Lambda_{ii} = \lambda_i$, $|\Phi\rangle = \sum_i |i\rangle_{AA'} \otimes |i\rangle_{BB'}$, and we have used

$$(I \otimes M)|\Phi\rangle = (M^T \otimes I)|\Phi\rangle \tag{345}$$

for any operator $M$. Defining $\rho := \mathrm{Tr}_{BB'}|\psi\rangle\langle\psi|$, the entanglement capacity of $K_{xx}$ is

$$\begin{aligned} E_{K_{xx}} &= \max_{|\psi\rangle} \mathrm{Tr}\left(-\frac{\mathrm{d}\rho}{\mathrm{d}t}\log\rho - \rho\frac{\mathrm{d}\log\rho}{\mathrm{d}t}\right)\\ &= \max_{|\psi\rangle} \mathrm{Tr}\left(-\frac{\mathrm{d}\rho}{\mathrm{d}t}\log\rho\right) \,. && (346) \end{aligned}$$

The variation of $\rho$ can be computed using perturbation theory [**79**]:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -i\mathrm{Tr}_{BB'}[K_{xx}, |\psi\rangle\langle\psi|] = 2\,\mathrm{Im}\,\mathrm{Tr}_{BB'}(K_{xx}|\psi\rangle\langle\psi|). \tag{347}$$

Letting $R = U\sqrt{\Lambda}V^T$, we have

$$\begin{aligned} & \mathrm{Tr}_{BB'}\left(K_{xx}|\psi\rangle\langle\psi|\right) \\ = \; & \mathrm{Tr}_{BB'}\left[(X\otimes X)(R\otimes I_{BB'})|\Phi\rangle\langle\Phi|(R^\dagger\otimes I_{BB'})\right] \\ = \; & \mathrm{Tr}_{BB'}\left[(XRX^T\otimes I_{BB'})|\Phi\rangle\langle\Phi|(R^\dagger\otimes I_{BB'})\right] \\ = \; & XRX^TR^\dagger\,, \end{aligned} \tag{348}$$

where we have introduced $X := \sigma_x \otimes I$, with the identity operator acting on the ancilla. The first equality follows simply from substitution of $K_{xx}$ and $|\psi\rangle$ by their expressions in Eqs. (340) and (344); the second uses Eq. (345); and the third employs the fact that for any operators $M_1, M_2$,

$$\mathrm{Tr}_{BB'}[(M_1\otimes I_{BB'})|\Phi\rangle\langle\Phi|(M_2\otimes I_{BB'})] = M_1M_2\,. \tag{349}$$

Since $\rho = U\Lambda U^\dagger$, we have

$$E_{K_{xx}} = \max_{|\psi\rangle}\mathrm{Tr}\left(-U^\dagger\frac{\mathrm{d}\rho}{\mathrm{d}t}U\log\Lambda\right)\,. \tag{350}$$

Using Eqs. (347) and (348), and introducing the Hermitian operators $X_U = U^\dagger XU$ and $X_V = V^\dagger XV$, we have

$$U^\dagger\frac{\mathrm{d}\rho}{\mathrm{d}t}U = 2\,\mathrm{Im}\,X_U\sqrt{\Lambda}X_V^T\sqrt{\Lambda}\,. \tag{351}$$

Letting $U, V, \Lambda$ obtain the max in Eq. (350), we find

$$\begin{aligned} E_{K_{xx}} & = -2\,\mathrm{Im}\,\mathrm{Tr}\left(X_U\sqrt{\Lambda}X_V^T\sqrt{\Lambda}\log\Lambda\right) \\ & \leq i\mathrm{Tr}\left[(X_U\sqrt{\Lambda}X_V^T - X_V^T\sqrt{\Lambda}X_U)\sqrt{\Lambda}\log\Lambda\right] \\ & = i\mathrm{Tr}\left[M(X_U\circ X_V)\right]\,, \end{aligned} \tag{352}$$

where we have introduced the real, skew-symmetric matrix

$$M_{ij} := \sqrt{\lambda_i\lambda_j}\log(\lambda_j/\lambda_i)\,, \tag{353}$$

and the symbol $\circ$ denotes the Hadamard (or element-wise) product of matrices. In the second line of Eq. (352) we have used

$$\mathrm{Im}\,\mathrm{Tr}A = \mathrm{Tr}(A - A^\dagger)/2i \tag{354}$$

and the fact that $\Lambda$, $X_U$, and $X_V$ are Hermitian. The last line can be checked by explicitly writing the trace in terms of matrix elements.

From Eq. (352) we obtain the following upper bound for $E_{K_{xx}}$ (here $\wr A \wr$ denotes the element-wise absolute value, i.e., $\wr A \wr_{ij} = |A_{ij}|$):

$$\begin{aligned} E_{K_{xx}} &\leq \mathrm{Tr}(\wr M \wr \, \wr X_U \circ X_V \wr) \\ &\leq \max_P \mathrm{Tr}(\wr M \wr P) \\ &\leq 2 \max_x \sqrt{x(1-x)} \log[x/(1-x)] \\ &= \alpha \approx 1.9123 \,, \end{aligned} \tag{355}$$

where $P$ is a permutation operator and $x \in [0,1]$. The first line uses the triangle inequality. The second inequality follows from noticing that $\wr X_U \circ X_V \wr$ is a doubly substochastic matrix [**37**]. Indeed, for any two complex numbers $v$ and $w$ one has that $2|vw| \leq |v| + |w|$, and consequently, for any two unitary matrices $V$ and $W$,

$$\begin{aligned} \sum_i |V_{ij} W_{ij}| &\leq \sum_i (|V_{ij}|^2 + |W_{ij}|^2)/2 = 1, \\ \sum_j |V_{ij} W_{ij}| &\leq \sum_j (|V_{ij}|^2 + |W_{ij}|^2)/2 = 1, \end{aligned} \tag{356}$$

which implies that the matrix $\wr V \circ W \wr$, with entries $|V_{ij} W_{ij}|$, is doubly substochastic. Therefore a doubly stochastic matrix $Q$ exists such that $|X_U \circ X_V|_{ij} \leq Q_{ij}$ for all $i$ and $j$ [**37**], so that $\mathrm{Tr}(\wr M \wr \, \wr X_U \circ X_V \wr) \leq \mathrm{Tr}(\wr M \wr Q)$. But $Q$ is a convex combination of permutation operators $P_k$, $Q = \sum_k p_k P_k$, which implies that $\mathrm{Tr} \wr M \wr Q \leq \max_P \mathrm{Tr}(\wr M \wr P)$. Finally, the third inequality in Eq. (355) follows from noticing that

$$\begin{aligned} & |M|_{ij} = \sqrt{\lambda_i \lambda_j} |\log(\lambda_j/\lambda_i)| \\ =\ & (\lambda_i + \lambda_j) \sqrt{\frac{\lambda_i}{\lambda_i + \lambda_j} \frac{\lambda_j}{\lambda_i + \lambda_j}} |\log(\lambda_j/\lambda_i)| \\ \leq\ & (\lambda_i + \lambda_j) \max_x \sqrt{x(1-x)} \log[x/(1-x)] \\ =\ & (\lambda_i + \lambda_j) \alpha/2 \,, \end{aligned} \tag{357}$$

and that

$$\mathrm{Tr}(\wr M \wr P) \leq \frac{\alpha}{2} \sum_{ij} (\lambda_i + \lambda_j) P_{ij} = \alpha \sum_i \lambda_i = \alpha \,, \tag{358}$$

where we have used that $P$ is a permutation matrix and that $\sum_i \lambda_i = 1$. Comparison of Eqs. (336) and (355) shows that, indeed, $E_{K_{xx}} = E^{(1*)}_{K_{xx}}$, completing the proof.

We have shown that ancillary systems are not needed when optimizing entanglement generation by any two-qubit Hamiltonian with normal form given by Eq. (338)[4]. More specifically, there is a universal optimal two-qubit initial state

[4]It is interesting to note that this Hamiltonian is also special in that the ground state of a chain of particles interacting by this Hamiltonian can be found analytically[**6**]; this emerges out of the fact that this Hamiltonian is very special and that the calculation of the

given by [**79**]

$$|\psi_{\max}\rangle := \sqrt{x_0}|0\rangle_A \otimes |1\rangle_B - i\sqrt{1-x_0}|1\rangle_A \otimes |0\rangle_B\,. \tag{359}$$

As an application of the above, we discuss how to use Hamiltonian $K$ to enable classical communication between Alice and Bob. This has been studied in [**35**], and the entanglement-assisted forward classical capacity, $C^E_{\rightarrow}$ (maximum rate for the Hamiltonian $H$ to communicate from Alice to Bob when unlimited free entanglement is available), is shown to be

$$C^E_{\rightarrow}(H) = \sup_{\mathcal{E}} \left[ \lim_{t\rightarrow 0} \frac{\chi(\mathrm{Tr}_{AA'} e^{-iHt}\mathcal{E}) - \chi(\mathrm{Tr}_{AA'}\mathcal{E})}{t} \right], \tag{360}$$

where $\mathcal{E} = \{p_i, |\psi_i\rangle\}$ is an ensemble of bipartite states, $e^{-iHt}\mathcal{E}$ and $\mathrm{Tr}_{AA'}\mathcal{E}$ denote the respective transformed ensembles $\{p_i, e^{-iHt}|\psi_i\rangle\}$ and $\{p_i, \mathrm{Tr}_{AA'}|\psi_i\rangle\langle\psi_i|\}$, and

$$\chi(\{p_i,\rho_i\}) := S\left(\sum_i p_i\rho_i\right) - \sum_i p_i S(\rho_i) \tag{361}$$

is the Holevo information of the ensemble $\{p_i,\rho_i\}$ and $S$ is the von Neumann entropy. Reference [**35**] also describes a protocol to achieve the rate in the bracket of Eq. (360) for any ensemble $\mathcal{E}$.

For any two-qubit Hamiltonian $H$, Ref. [**36**] constructs an ensemble with communication rate $E_H$, so that $C^E_{\rightarrow}(H) \geq E_H$. This ensemble, which is not necessarily optimal, is defined in terms of an optimal state for entanglement generation. This ensemble $\mathcal{E}_1$ can now be made more explicit for Hamiltonian $K$ in light of our findings:

$$\begin{aligned} p_1 &:= \frac{1}{2},\ |\psi_1\rangle := \sqrt{x_0}|0\rangle_A\otimes|1\rangle_B + i\sqrt{1-x_0}|1\rangle_A\otimes|0\rangle_B, \\ p_2 &:= \frac{1}{2},\ |\psi_2\rangle := \sqrt{x_0}|0\rangle_A\otimes|0\rangle_B - i\sqrt{1-x_0}|1\rangle_A\otimes|1\rangle_B, \end{aligned}$$

where $x_0$ is defined after Eq. (336). For ensemble $\mathcal{E}_1$ we find

$$\begin{aligned} &\chi(\mathrm{Tr}_A\mathcal{E}_1) = S(I/2) - S(\mathrm{Tr}_A|\psi_1\rangle\langle\psi_1|) = 1 - E(|\psi_{\max}\rangle) \\ &\chi(\mathrm{Tr}_A(e^{-i\delta t K}\mathcal{E}_1)) = 1 - (E(|\psi_{\max}\rangle) - \delta t E_K) \end{aligned} \tag{362}$$

and therefore the net rate at which classical bits are transmitted is indeed $\Delta\chi/\delta t = E_K$.

ground state can be transformed to an exactly solvable problem with fermions. It would be interesting to compare this connection with the one discovered by Schliemann et al. [**185**], where there turns out to be an equivalence between states of two qubits and two fermions described in a 6-dimensional space.

Next we present an alternative ensemble $\mathcal{E}_2$ of product states with the same communication rate:

$$\begin{aligned} p_1 &:= \frac{1}{2},\ |\psi_1\rangle := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}_A \otimes \begin{pmatrix} \sqrt{x_0} \\ -i\sqrt{1-x_0} \end{pmatrix}_B , \\ p_2 &:= \frac{1}{2},\ |\psi_2\rangle := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}_A \otimes \begin{pmatrix} \sqrt{x_0} \\ i\sqrt{1-x_0} \end{pmatrix}_B , \end{aligned}$$

Here, we use $K$ to simulate $K'$ [**225**], under which the ensemble evolves. For ensemble $\mathcal{E}_2$, $S(\mathrm{Tr}_A|\psi_i\rangle\langle\psi_i|) = 0$,

$$\begin{aligned} \chi(\mathrm{Tr}_A\mathcal{E}_2) &= H_2(x_0)\,, \text{ and} \\ \chi(\mathrm{Tr}_A(e^{-i\delta t K}\mathcal{E}_2)) &= H_2(x_0 - 2\delta t\sqrt{x_0}\sqrt{1-x_0}) \\ &= H_2(x_0) + E_K \delta t \end{aligned} \tag{363}$$

(where $H_2$ is the binary entropy), so that the communication rate is again $\Delta\chi/\delta t = E_K$.

The main difference between these two ensembles is that the states in ensemble $\mathcal{E}_1$ are entangled while states in ensemble $\mathcal{E}_2$ are not. In the first case the interaction $K$ is used to decrease the degree of entanglement between Alice and Bob or, equivalently, to make states of Bob's ensemble $\mathrm{Tr}_A\mathcal{E}_1$ less mixed and thus more distinguishable. The same increase of distinguishability for the pure states of Bob's ensemble $\mathrm{Tr}_A\mathcal{E}_2$ is achieved by conditionally rotating them with $K$, in a way that they become more orthogonal to each other. We note, in addition, that ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$ can be prepared using different remote state preparation techniques[5].

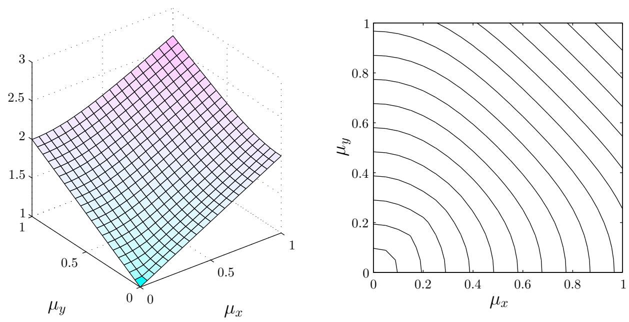



**Figure 2.** Numerically optimized entanglement capacity of the two-qubit Hamiltonian $\mu_x\ \sigma_x \otimes \sigma_x + \mu_y\ \sigma_y \otimes \sigma_y + \sigma_z \otimes \sigma_z$. The vertical axis in the left figure is in units of $\alpha$.

In conclusion, we have computed the asymptotic entanglement capacity of all two-qubit Hamiltonians that are locally equivalent to $\mu_x\ \sigma_x \otimes \sigma_x + \mu_y\ \sigma_y \otimes \sigma_y$. Though we do not yet have a closed form expression for general two-qubit

[5] Reference [**35**] cites a method due to P. Shor that can be used to prepare $\mathcal{E}_1$, and $\mathcal{E}_2$ can be prepared by a technique due to P. Shor and A. Winter (in preparation).

Hamiltonians, we can present some preliminary results in this direction. The numerically optimized entanglement capacity is shown in Fig. 2. As discussed above, achieving the capacity in general requires the use of ancillas. Numerically, we find that the optimum can be achieved with single-qubit ancillas on both sides.

## 6.6. Conclusion

We have shown that the natural description of quantum channels or positive linear maps is given by a dual quantum state associated to the map. This dual state is defined over a Hilbert space that is naturally endowed with a tensor product structure of the in- and output of the channel. We showed that the techniques developed in the context of entanglement are of direct use in describing positive maps. We derived a characterization of the extreme points of the convex set of trace-preserving completely positive maps, and gave some generalizations. We discussed some new results about the classical and quantum capacity of a quantum channel, and in the case of one-qubit channels we showed how to exploit the duality between qubit channels and mixed states of two qubits to obtain useful parameterizations. Finally we gave a short review of some recent results of how to implement non-local operations using entanglement, and showed how a given non-local Hamiltonian can optimally be used to create entanglement or to transmit classical information.

## CHAPTER 7

# Conclusion

As Schrödinger already pointed out, entanglement is the main characteristic trait of quantum mechanics. In a first part of this thesis, we aimed at giving a systematic and unified analysis of entanglement in multipartite quantum systems. The following list is a rough sketch of the main own contributions in this part. The citations at the beginning of each item refer to the papers in which the results appeared.

- [**217**] The quantum steering Theorem of Schrödinger was reformulated and identified as the essential ingredient in describing entanglement in pure bipartite quantum systems. We argued how quantum teleportation and the results on entanglement monotones and entanglement transformations involving majorization follow almost trivially from the quantum steering theorem.
- [**211**] We presented a unified way of classifying pure multipartite quantum states by means of local equivalence classes. A constructive way of generalizing the singular value decomposition to tensors was derived, yielding local unitary equivalence classes, and we argued why this generalization failed to be unique. Next we looked for another normal form by enlarging the class of local operations from local unitary operations ($SU(n)$) to local filtering operations ($SL(n,\mathcal{C})$). A much more appealing normal form was obtained, and we presented supporting evidence for the uniqueness of the normal forms obtained. The introduced formalism led to a natural way of defining entanglement measures, and we identified the normal form as the state in the SLOCC-equivalence class with the maximal amount of entanglement. We also managed to generalize all these findings to the case of mixed multipartite states.
- [**161, 213**] The first complete characterization was given of all pure $2 \times 2 \times N$-systems. The analysis revealed that there exist 9 different ways in which $2 \times 2 \times N$ systems can be entangled. Furthermore we derived the optimal way to distill a $GHZ$-state from a state belonging to its equivalence class.

- [**215**] A complete characterization was obtained of all LU and all SLOCC equivalence classes in the case of pure states of four qubits. This was possible by formulating a highly non-trivial generalization of the singular value decomposition to the case of complex orthogonal equivalence. The analysis revealed the existence of nine different families of entangled states, where each family contains a continuum of equivalence classes labelled by 8 real parameters. We also indicated how to generalize these results to higher dimensional systems.

- [**212, 213**] The analysis of entanglement in mixed states is much more involved. In the case of mixed states of two qubits however, we were able to effectively separate the local from the non-local characteristics of the density operator, yielding a very convenient parameterization of all density operators of two qubits. The central result was the existence of the Lorentz singular value decomposition, which is a generalization of the singular value decomposition but with Lorentz instead of unitary transformations. In analogy with the singular value decomposition, we obtained a nice variational characterization of the Lorentz singular values. This enabled to construct entanglement measures for mixed states, and we derived the optimal filtering procedure to maximize them.

- [**217**] As a first application of the Lorentz singular value decomposition, we generalized the quantum steering theorem to mixed states of two qubits. An appealing geometrical picture of all states of two qubits was obtained, sharpening our intuition about entanglement considerably. Moreover, the quantum steering theorem was generalized to mixed bipartite states of arbitrary dimensions.

- [**10, 212**] We presented a novel and constructive derivation of how to calculate the entanglement of formation of mixed states of two qubits. We made the connection with the Lorentz singular value decomposition, and showed how this naturally lead to a proof of the celebrated partial transpose condition for entanglement.

- [**209, 221, 220**] We made a comparison of the entanglement measures concurrence (or entanglement of formation), negativity, relative entropy of entanglement, fidelity (i.e. maximal singlet fraction) and violation of the Bell-CHSH inequalities for mixed states of two qubits. Tight upper and lower bounds were derived for one measure in function of another one, where extensive use was made of the Lorentz singular value decomposition. We proved that the local filtering operations maximizing all these entanglement measures correspond to the ones bringing a state into Bell-diagonal normal form. This resolved a long standing open question of identifying all quantum states that violate Bell inequalities after an appropriate filtering operation.

- [**219**] We considered the problem of optimal teleportation with a mixed state of two qubits. We optimized the achieved fidelity allowing all possible (unphysical) PPT-operations, the class of which is strictly larger than the class of LOCC operations, and we arrived at the surprising result that the optimal solution corresponded to a physically implementable 1-LOCC protocol involving local filtering and classical communication. The geometrical quantum steering picture gave a nice illustration of the effect of these optimal operations.
- [**64**] A general formalism for constructing entanglement distillation protocols was created, and we illustrated this by introducing the best known distillation protocols.
- [**210, 234**] We introduced the concept of maximally entangled mixed states, defined as states whose entanglement cannot be enlarged by arbitrary global unitary operations. We proved their uniqueness in the case of two qubits. This enabled to give a complete characterization of the maximal ball of separable states surrounding the identity and to give a complete characterization of the states with maximal amount of entropy for a given amount of entanglement (or vice-versa).
- [**214**] We showed that the problem of finding the closest PPT-state to an entangled one can easily be solved if the Hilbert-Schmidt norm is used. This yields a geometrical interpretation of negativity, and we illustrated this by drawing contour plots of states with constant negativity. This gives some intuition about the shape of the convex set of separable states, and enabled non-trivial proofs such that the mixed set of $W$-states is not of measure zero.
- [**146**] A mixed state of two qubits can also be interpreted as the partial trace of a pure tri- or fourpartite system. This leads to the concept of entanglement of assistance. We devised optimal strategies for maximizing the entanglement shared by two parties by measurements of the other remaining parties on the shared tri- or fourpartite state.

The second part treated selected topics in quantum information theory. The following list contains the main contributions:

- [**216**] We developed a general scheme for quantum parameter estimation in the context of continuous quantum measurement. Exploiting an intriguing connection between quantum evolution and Kalman filtering, we devised optimal detection strategies that could beat the standard quantum limit.
- [**218**] We reinvented the duality between completely positive maps and entangled quantum systems. This led to a unified way of describing quantum channels and their classical and quantum capacity. On the other hand, this gave some insight in how to quantify classical

correlations in entangled quantum systems. We showed how to characterize the extreme points of completely positive maps, described entanglement breaking channels and discussed which states have to be send through a channel such as to maximize the quantum capacity. In the case of qubit channels, we made use of the plethora of results derived in the first part about mixed states of two qubits to give a complete classification of all qubit channels, and introduced interesting normal forms. Furthermore, we indicated how the quantum steering theorem translates into the language of quantum channels, and made some progress in calculating the Holevo capacity of all extremal qubit channels.

- [**55**] We investigated the problem of how to maximize the entanglement capacity of a given two-qubit Hamiltonian, and obtained the exact result in the case of the Ising interaction. This lead to the first calculation of the asymptotic entanglement capacity of a Hamiltonian.

Fortunately, the work of the previous years raised a lot of open questions. The following is a biased list of open problems in the field of quantum information theory:

- Find an information theoretic explanation for the fundamental postulates of quantum theory. It is indeed not very appealing that a concept such as a Hilbert space has to be invoked to justify a physical theory. Note that the work presented in section 5.1 was motivated by such considerations.
- Generalize the asymptotic results on pure bipartite entanglement transformations to the multipartite case. The quest of local filtering equivalence classes revealed the fact that many different kinds of entanglement exist. Does the same classification still holds in the asymptotic case?
- Construct more sophisticated versions of the hashing protocol to distill singlets that work also for barely entangled states on. The general formalism developed in section 4.6 gives a first hint of how to achieve this. This should ultimately lead to the calculation of the entanglement of distillation.
- Find a useful parameterization of all possible LOCC transformations (local quantum operations assisted by classical communication) on distributed mixed quantum systems. It would be very nice if one could for example prove that the number of classical communication rounds can always be bounded by a number depending on the dimension of the subsystems. The work on quantum steering presented in this thesis was motivated by this program.
- Find an efficient algorithm for checking whether a mixed density operator is separable, and characterize all states that cannot be distilled

(it seems very reasonable to conjecture that the class of undistillable states is strictly larger than the class of PPT-states).

- Devise new applications of multipartite quantum systems.
- One of the fundamental problems of quantum information theory is the calculation of the quantum capacity of a quantum channel. The solution will depend on our ability of calculating the entanglement of distillation, and of solving the following problem: which states do I have to send through a quantum channel such as to maximize the quantum capacity? Note that we touched upon these problems in section 4.6 and 6.3.1 & 6.4.2.
- Can the classical capacity of a quantum channel be increased by allowing entangled inputs? This problem is strongly related to the following open problem: is the entanglement of formation additive?
- Given some bipartite quantum state. What is the maximal amount of secret classical correlations that both parties can create out of it using local operations (and eventually classical public communication). More generally, what kind of irreversibility occurs when quantum correlations are partially transformed into classical correlations? Is it possible to obtain a new conservation law? The work presented in section 6.3.2 & 6.4.3 is a nice starting point for investigating this issue.
- What is the minimal amount of entanglement needed to implement a global unitary operation? (cfr. section 6.5.1)
- Given a specific Hamiltonian coupling distributed quantum systems. Devise an optimal way of using this Hamiltonian to exchange classical or quantum information (we refer to section 6.5.2 for the first results).
- Construct a general theory of quantum feedback control. How can the results of classical bilinear control theory be translated into the quantum language? How should one implement detection procedures such as to minimize the quantum back action of the system under observation? These questions motivated the research presented in section 5.3.
- Can the present understanding of entanglement reveal some new fundamental insights into other branches of physics (e.g. in the context of phase transitions)? Are there some interesting macroscopic effects originating from quantum entanglement?
- How can one translate the present results in quantum information theory into a covariant (quantum field) theory?
- $\cdots$

APPENDIX A

# Basic Concepts of Quantum Mechanics

In this appendix some basic mathematical machinery of quantum mechanics is collected from the perspective of linear algebra; as argued by Bart De Moor, this section is intended for *nitwits*.

Three different but related concepts are needed to describe a quantum system: quantum states, maps, and observables. Quantum states are somehow related to distributions while observables to random variables. Both can be represented as linear operators on the complex Hilbert space $\mathcal{H}$, and we will only consider Hilbert spaces of finite dimensions.

- ***States:*** Quantum states are defined as bounded linear operators (i.e. matrices) on the positive cone, and have trace equal to 1. A state of maximal knowledge $\rho$ corresponds to a positive operator of rank 1, i.e. a pure state

$$\rho = |\psi\rangle\langle\psi|, \tag{364}$$

while the most general state is a convex sum of pure states, i.e. a positive (semi-)definite operator with trace equal to 1.

Let us for example consider the simplest of all quantum systems, i.e. a qubit, described in a Hilbert space of dimension 2. Then a quantum state $\rho$ can be represented by a $2 \times 2$ positive matrix with trace 1, that has three degrees of freedom. The Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{365}$$

supplemented by $\sigma_0 = I_2$ form a complete orthogonal basis for all Hermitian $2 \times 2$ matrices (i.e. $\mathrm{Tr}\sigma_i\sigma_j = 2\delta_{ij}$), and it follows that a density operator $\rho$ has a unique decomposition of the form

$$\rho = \frac{1}{2}\left(I_2 + \sum_{i=1}^{3} x_i\sigma_i\right) \tag{366}$$

$$x_i = \mathrm{Tr}(\sigma_i\rho) \tag{367}$$

with the coefficients $x_i$ real. This defines the Bloch vector $x = (x_1, x_2, x_3)$. The condition of positivity corresponds to the fact that the Bloch vector obeys the relation

$$\|x\|_2 \leq 1,$$

and therefore a state is uniquely parameterized by a point inside a sphere of radius 1. This sphere is called the Bloch sphere. Note that pure states lie on the boundary, and orthogonal pure states have the opposite coordinates.

The von-Neumann entropy is a measure of the mixedness of a state and is defined as

$$S(\rho) = -\mathrm{Tr}\left(\rho \log_2(\rho)\right). \tag{368}$$

It is solely function of the eigenvalues of a matrix and can be given an information theoretic meaning à la Shannon if the mixed state originates from a source that stochastically emits orthogonal pure states.

If two quantum systems are coupled together, then the new Hilbert space arising is described by the tensor product of the two original ones. This is the origin of the superposition principle and of the more fantastic concept of entanglement.

- ***Maps:*** The evolution of quantum systems is described by a linear map that maps states to states. The map should therefore conserve positivity, and even conserve positivity if applied on a subsystem (i.e. on one part in the tensor product). The most general map $\Phi$ is therefore given by a completely positive trace-preserving linear map. Following a theorem of Choi [**56**], this implies that the map is completely specified by a set of so-called Kraus operators $\{A_i\}$:

$$\Phi(\rho) = \sum_i A_i \rho A_i^\dagger. \tag{369}$$

The trace condition translates into $\sum_i A_i^\dagger A_i = I$ with $I$ the identity operator in Hilbert space. In the case of a closed unobserved system, the evolution is unitary (i.e. there is only one Kraus operator). More generally, the case of more then 1 Kraus operator arises e.g. when the system couples to the unaccessible degrees of freedom of the environment.

- ***Observables:*** An observable in its full generality is called a *positive operator valued measure* , corresponding to a positive operator $E$ bounded above by the identity $E \leq I$ (this notation means that $I - E \geq 0$, i.e. positive). An observable $E$ together with a state $\rho$ yield a probability measure $P_E(\rho)$ via the formula

$$P_E(\rho) = \mathrm{Tr}\left(\rho E\right). \tag{370}$$

The most general measurement corresponds to a set of positive operator valued measures $\{E_i\}$ such that $\sum_i E_i = I$, and the measurement itself is generally called a POVM-measurement. Its physical meaning is as follows: a POVM-measurement $\{E_i\}; i = 1..m$ corresponds to a measurement with $m$ different outcomes[1], each outcome occurring with probability $p_i = \mathrm{Tr}\,(\rho E_i)$. In some sense a POVM-measurement can always be interpreted as a von-Neumann measurements (where all the elements of the set $\{E_i\}$ are orthogonal projectors) on a Hilbert space that is enlarged by an auxiliary system (called an ancilla).

A measurement reveals information about a quantum system and our knowledge of the quantum system should be updated correspondingly. The POVM-elements $\{E_i\}$ however do not contain enough information to do this in an unique way; this has to do with the fact that the evolution should be described by a completely positive map and that there are many inequivalent ways of choosing Kraus operators associated to a POVM-element. The description of a POVM as a set of positive operators $\{E_i\}$ is therefore incomplete, and a more complete description is obtained by specifying a set of Kraus operators $\{A_i\}$ with $A_i^\dagger A_i = E_i$. Suppose a POVM-measurement with Kraus operators $\{A_i\}$ has been performed with measurement outcome $i$, then the density operator should, in analogy with Bayes' rule, be updated as

$$\rho_i = \frac{A_i \rho A_i^\dagger}{\mathrm{Tr}\left(A_i \rho A_i^\dagger\right)}. \tag{371}$$

It is now clear how to interpret the most general type of quantum evolution as described by a completely positive map: it is as the environment is performing measurements, whose outcomes are inaccessible, and therefore the convex sum should be taken with associated probabilities.

Let us now concentrate on a quantum system that exhibits a tensor product structure. The simplest case arises when two coupled qubits are described, on which we will mostly focus from now on. We will restrict the following discussion to the case of pure states. A complete set of orthogonal basis vectors for a state $|\psi\rangle$ of 2 qubits is given by

$$|00\rangle, |01\rangle, |10\rangle, |11\rangle, \tag{372}$$

for which it is often convenient to choose an explicit vector representation as

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{373}$$

[1] A similar reasoning holds in the case of a non-countable set as needed in section 5.3.

The vector representation of a state

$$|\psi\rangle = x_{00}|00\rangle + x_{01}|01\rangle + x_{10}|10\rangle + x_{11}|11\rangle$$

is therefore given by

$$\psi = \begin{pmatrix} x_{00} \\ x_{01} \\ x_{10} \\ x_{11} \end{pmatrix}.$$

The notation already suggests that pure states defined on a Hilbert space endowed with a tensor product structure can be represented as matrices:

$$\Psi = \begin{pmatrix} x_{00} & x_{01} \\ x_{10} & x_{11} \end{pmatrix}.$$

A local change of basis (i.e. passively choosing another set of orthonormal basis vectors or actively rotating the qubit) amounts to multiplying the vector representation with a tensor product of unitary matrices $U \otimes V$, and to multiplying the matrix representation left and right by unitaries:

$$\Psi' = U\Psi V^T.$$

Using the singular value decomposition, it is immediately clear that there always exist a local basis such that $\Psi$ is diagonal, which is also called the Schmidt decomposition of a state. The Schmidt coefficients of a pure state correspond to the singular values of this matrix.

A pure state $|\psi\rangle$ is separable if and only its associated matrix representation $\Psi$ is rank 1, and is called entangled iff it is not separable.

If an observer has only access to one of the two subsystems arising in a tensor product, and performs a POVM-measurement $\{E_i\}$ on it, then his measurement outcomes obey the statistical rule:

$$p_i = \mathrm{Tr}\left(\rho(E_i \otimes I_2)\right) = \mathrm{Tr}\left(E_i \mathrm{Tr}_2(\rho)\right). \tag{374}$$

Here we used the notation $\mathrm{Tr}_2(\rho)$, denoting the partial trace operation. It maps a density operator defined in a $n \times m$ dimensional Hilbert space on an operator in a $n$ dimensional Hilbert space, and is defined as the unique operator with the following characteristic

$$\forall \rho, i, j: \quad \langle i|\mathrm{Tr}_2(\rho)|j\rangle = \sum_{k=1}^{m} \langle ik|\rho|jk\rangle.$$

The partial trace of a pure density operator can readily be expressed in terms of the matrix representation $\Psi$:

$$\mathrm{Tr}_2(|\psi\rangle\langle\psi|) = \Psi\Psi^\dagger, \qquad \mathrm{Tr}_1(|\psi\rangle\langle\psi|) = \Psi^\dagger\Psi.$$

The fact that a state is entangled therefore corresponds to the fact that its local density operator is not pure. This was first observed by E. Schrödinger, which let him to the following sentence: "the best possible knowledge of a *whole* does not necessarily include the best possible knowledge of all its *parts*": it is not because a state is pure that its local density operators are pure.

The more mixed a local density operator of a pure state is, the more the state is entangled. Maximally entangled pure states are those for which the entropy of the local density operators are maximal (i.e. proportional to the identity). A complete orthonormal basis of maximally entangled states is given by the Bell states (sometimes also called EPR-pairs):

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right) \quad |\Phi^-\rangle = \tfrac{1}{\sqrt{2}}\left(|00\rangle - |11\rangle\right)$$
$$|\Psi^+\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle + |10\rangle\right) \quad |\Psi^-\rangle = \tfrac{1}{\sqrt{2}}\left(|01\rangle - |10\rangle\right)$$

These states can be e.g. be used to do dense coding and perfect quantum teleportation, and are therefore the basic resources for most of the quantum information theoretic applications.

Entanglement gives rise to strange non-local effects without classical analogue. Two quantum systems can only be entangled if they interacted in one or the other way with each other. Local operations and classical communication (i.e. sending bits of information) cannot create entanglement between two distant parties, and these kind of operations are called LOCC operations (local operations assisted by classical communication). An entanglement monotone is defined as a quantity that does not increase under LOCC operations; this is a natural requirement for all sensible entanglement measures.

A related class of operations is called the class of SLOCC operations (Stochastic Local Operations assisted by Classical Communication). These operations can only be implemented with a certain probability, and correspond to the situation in which all parties implement local POVM's $\{E_i^\alpha\}$, and only keep the state if they all get a prespecified measurement result. With these type of operations, it is possible to transform a pure state with a limited amount of entanglement to one with more entanglement, but of course this can only occur with a small probability. This type of operations play a crucial role in this thesis.

As an example, consider a pure bipartite state with matrix representation $\Psi$, and two POVM's with Kraus operators $\{A_i\}$, $\{B_j\}$ implemented by both parties (Alice and Bob). If Alice gets outcome $i$ and Bob $j$, then the state becomes

$$\frac{A_i \Psi B_j^T}{\|A_i \Psi B_j^T\|_F}.$$

The probability by which these outcomes were obtained is given by $\|A_i \Psi B_j^T\|_F$. The previous formula indicates that $A_i$ and $B_j$ can always be chosen such that

$A_i \Psi B_j^T$ is proportional to a unitary matrix iff $\Psi$ was full rank, yielding a maximally entangled state. This type of transformation is a SLOCC transformation, and enables to transform barely entangled states to maximally entangled states as long as their matrix representation is full rank (recall that this does not apply to separable states as they have a rank 1 matrix representation).

In the case of pure multipartite states with more then 2 parties, a similar discussion leads to the decomposition of higher dimensional tensor objects. Much of the difficulties encountered in this study arise from the fact that no nice analogue of the singular value decomposition exists in that case, and chapter 3 is devoted to this problem.

Armed with the previous list of basic notions, one should now be prepared to devour the main results of this thesis.

# APPENDIX B

# Miscellaneous Proofs

**Theorem 17** The 4x4 matrix $R$ with elements $R_{ij} = \mathrm{Tr}(\rho\sigma_i \otimes \sigma_j)$ can be decomposed as

$$R = L_1 \Sigma L_2^T \tag{375}$$

with $L_1, L_2$ proper orthochronous Lorentz transformations, and $\Sigma$ either of diagonal form $\Sigma = \mathrm{diag}[s_0, s_1, s_2, s_3]$ with $s_0 \geq s_1 \geq s_2 \geq |s_3|$, either of the form

$$\Sigma = \begin{pmatrix} a & 0 & 0 & b \\ 0 & d & 0 & 0 \\ 0 & 0 & -d & 0 \\ c & 0 & 0 & a+c-b \end{pmatrix} \tag{376}$$

with $a, b, c, d$ real.

*Proof:* The proof of this theorem heavily relies on results on matrix decompositions in spaces with indefinite metric [**100**]. We first introduce the matrix $C = MRMR^T$ which is $M$-selfadjoint. Using theorem (5.3) in [**100**], it follows that there exist matrices $X$ and $J$ with $C = X^{-1}JX$, $J$ consisting of a direct sum of real Jordan blocks and $XMX^T = N_J$ with $N_J$ a direct sum of symmetric nxn matrices of the form $[S_{ij}] = \pm[\delta_{i+j,n+1}]$ with $n$ the size of the corresponding Jordan block. Using Sylvester's law of inertia, there exists orthogonal $O_J$ such that $N_J = O_J^T M O_J$. It is then easy to check that $O_J X = L_1^T$ is a Lorentz transformation. Therefore the relations $C = MRMR^T = ML_1MO_JJO_J^TL_1^T$ hold. Multiplying left by $M$, Sylvester's law of inertia implies that there exist a matrix $\Sigma$ with the same rank as $J$ such that $MO_JJO_J^T = \Sigma M \Sigma^T$. Therefore we have the relation $RMR^T = L_1 \Sigma M \Sigma^T L_1^T$. If $R$ has the same rank as $RMR^T$, this relation implies that there exists a Lorentz transformation $L_2$ such that $R = L_1 \Sigma L_2^T$.

Let us now investigate the possible forms of $\Sigma$. As $N_J = O_J^T M O_J$ has signature $(+ - --)$, $J$ can only be a direct sum of the following form: 4 1x1 blocks; 1 orthogonal 2x2 block and 2 1x1 blocks; 1 2x2 Jordan block and 2 1x1 blocks;

1 3x3 Jordan block and 1 1x1 block. Noting the eigenvalues of $C$ as $\{\lambda_i\}$, it is easy to verify that a "square root" $\Sigma$ in the four cases is respectively given by

(1) $\Sigma = \mathrm{diag}[\sqrt{|\lambda_0|}, \sqrt{|\lambda_1|}, \sqrt{|\lambda_2|}, \sqrt{|\lambda_3|}]P$ with $P$ a permutation matrix permutating the first column with one other column;

(2) $\Sigma = \mathrm{diag}\left[\sqrt{|\lambda_0|}\begin{pmatrix} \cos(\phi) & \sin(\phi) \\ \sin(\phi) & -\cos(\phi) \end{pmatrix}, \sqrt{|\lambda_2|}, \sqrt{|\lambda_3|}\right]$;

(3) $\Sigma = \mathrm{diag}\left[\begin{pmatrix} a & b \\ c & a+c-b \end{pmatrix}, \sqrt{|\lambda_2|}, \sqrt{|\lambda_3|}\right]$;

(4) $\Sigma = \mathrm{diag}\left[\begin{pmatrix} a & 0 & 0 \\ b & \sqrt{a^2+b^2} & 0 \\ 0 & \frac{-ab}{\sqrt{a^2+b^2}} & \frac{a^2}{\sqrt{a^2+b^2}} \end{pmatrix}, \sqrt{|\lambda_3|}\right]$ with $a = \sqrt{|\lambda_0|}$ and $b = -1/\sqrt{2|\lambda_0|}$.

Now we go back to the relation $R = L_1^T \Sigma L_2$. $L_1$ and $L_2$ can be made proper and orthochronous by absorbing factors $-1$ into the rows and columns of $\Sigma$ yielding $\Sigma'$. Theorem (2) now implies that this $\Sigma'$ corresponds to an unnormalized physical state, which means that $\rho'$ corresponding to $\Sigma'$ has no negative eigenvalues. It is easy to show that this requirement excludes cases 2 and 4 of the possible forms of $\Sigma$. The third case corresponds to (376). Furthermore in the first case the permutation matrix has to be the identity and $|\lambda_0| \geq \max(|\lambda_1|, |\lambda_2|, |\lambda_3|)$. Multiplying left and right by proper orthochronous Lorentz transformations, the elements $\{s_i\}$ of this diagonal $\Sigma$ can always be ordered as $s_0 \geq s_1 \geq s_2 \geq |s_3|$.

The case where the rank of $C$ is lower then the rank of $R$ still has to be considered. This is only possible if the rowspace of $R$ has an isotropic subspace $Q$ for which $QMQ^T = 0$. Some straightforward calculations reveal that the only physical states for which this hold have normal form (376) with $a = b = c$ and $d = 0$ or $a = b$ and $c = d = 0$. This completes the proof. □

**Lemma 11** [Davies [**62**]] *Let the input ensemble $\mathcal{E} = \{p_i, \rho_i\}$ be defined over a $n$-dimensional Hilbert space. Then the measurement maximizing the mutual information can be chosen such that all elements in the POVM are pure, and the number of elements $N$ in the optimal POVM can be bounded by $n \leq N \leq n^2$.*

Proof: The following proof is much more direct than the original one of Davies [**62**] (based on a theorem of Caratheodory on convex sets) and is constructive. First of all we note that the set of POVM's forms a convex set and that the mutual information is convex in the POVM performed; this follows from the fact that the mutual information is a convex function of $p(\alpha|i)$ for fixed $p(i)$ [**61**]. Suppose now that an element $E_\alpha$ of the POVM is not pure but has rank $\geq 2$: $E_\alpha = E_\alpha^1 + E_\alpha^2$. Define a first POVM by substituting the element $E_\alpha$ by $E_\alpha^1$ and adding an extra element to the POVM $E^2_\alpha$, and a second POVM by substituting the element $E_\alpha$ by $E^2_\alpha$ and adding an extra element to the POVM $E^1_\alpha$. Then the original POVM yields exactly the same mutual information as the equally weighted convex combination of these new POVM's. Due to convexity, the new POVM's with purer elements give rise to at least the same mutual information, and by repeating this argument recursively it is proven that all elements in the POVM can be chosen to be pure. The POVM elements are represented by $n \times n$ hermitian matrices summing up to the identity. Suppose that the POVM has $n^2+1$ elements. The system of equations $\sum_\alpha x_\alpha E_\alpha = I$ in the unknowns $\{x_\alpha\}$ has an infinite number of solutions as there are more unknowns than equations and as the set of equations has at least one solution (indeed, $\forall \alpha, x_\alpha = 1$ is a solution). Take such a solution $\{x_\alpha\}$ for which the minimal element $x_{\min}$ is negative. It follows that the maximal element $x_{\max} > 1$ as the sum of positive definite matrices has to remain the same as if all coefficients were equal to 1. Define now two new POVM's $\{E^1_\alpha = (|x_{\min}| + x_\alpha)/(|x_{\min}| + 1)E_\alpha\}$ and $\{E^1_\alpha = (x_{\max} - x_\alpha)/(x_{\max} - 1)E_\alpha\}$. Both have one (different) element equal to zero and are therefore POVM's with $n^2$ elements. Moreover, the original POVM is the convex sum of both with weights respectively given by $(|x_{\min}| + 1)/(|x_{\min}| + x_{\max})$ and $(x_{\max} - 1)/(|x_{\min}| + x_{\max})$, and due to the convexity of the mutual information these new POVM's must yield a larger value of the mutual information. If the original POVM has more than $n^2+1$ elements, exactly the same reasoning applies. It remains to be proven that $N \geq n$. This follows trivially from the fact that one needs at least $n$ pure states to sum up to the identity. □

**Lemma 12** *The capacity of a quantum channel acting on a $n$-dimensional Hilbert space can be achieved using at most $n^2$ pure input states.*

Proof: Suppose the input ensemble has $n^2+1$ states and has ensemble average $\rho_0 = \sum_i p_i \rho_i$. Repeating the arguments of the proof of lemma 11, it follows that there exist two ensembles $\{p^1_i, \rho_i\}$ and $\{p^2_i, \rho_i\}$ having ensemble average $\rho_0$, each having one (different) probability equal to zero and such that the original ensemble is a convex sum of both with weights $q$ and $1-q$. It follows that

$$\chi(\{p_i, \rho_i\}) = q\chi(\{p^1_i, \rho_i\}) + (1-q)\chi(\{p^2_i, \rho_i\})$$

such that one of the two ensembles (having $n^2$ non-zero elements) has a larger $\chi$. If the input ensemble has more than $n^2+1$ elements, the proof can be completed by repeating this argument recursively . (Observe that the proof is constructive). □

APPENDIX C

# Some Matlab Code

The first routine calculates the (non-unique) generalization of the SVD to tensors as described in theorem 5. The input is a $n_1 \cdot n_2 \cdots n_N$ (complex) vector, representing a state in a $n_1 \otimes \cdots \otimes n_N$ dimensional Hilbert space. $m$ is a vector labelling the dimensions of the tensor: $m = [n_1, n_2, \cdots, n_N]$. The output $x$ is the normal form and $A(1 : n_k, 1 : n_k, k)$ represent the local unitaries needed to transform the original $x$ to normal form (note that the program only places zeros at the right places without taking into account the adjustment of the phases).

```
function [x,A]=tnormalU(x,m);
%result: look at reshape(y,[max(size(m)):-1:1])

mn=max(size(m));m0=m; A=zeros(max(m),max(m),mn); for
k=1:mn,A(1:m(k),1:m(k),k)=eye(m(k));end;

xt=x; for p=0:min(m)-1, if p>0, q=1;for
kk=mn:-1:1,q=round(log(kron(exp(pro(kk)*[1:(m(kk)-1)]),exp(q))));end;
m=m-1;xt=xt(q);end; mn=max(size(m));mtot=prod(m); for k=1:mn-1,
pro(k)=prod(m(k+1:mn));end;pro(mn)=1; for k=1:mn,
ind(1:m(k),k)=[1:pro(k):(m(k)-1)*pro(k)+1]';end; for k=1:mn,
proo(k)=mtot/pro(k)/m(k);end;

tr=1; tel=0;while (tr>10^(-13))*(tel<200), tr=0;tel=tel+1; for
k=1:mn,if m(k)>1;
    [u,s,v]=svd(xt(ind(1:m(k),k)));tr=tr+norm(u(2:m(k),1));
    A((1:m(k))+p,1:m0(k),k)=u'*A((1:m(k))+p,1:m0(k),k);
    xt=kron(eye(proo(k)),kron(u',eye(pro(k))))*xt;
end; end; end; end;

mn=max(size(m0));AA=1;for k=1:mn,
AA=kron(AA,A(1:m0(k),1:m0(k),k));end; x=AA*x;
```

The following program calculates the normal SLOCC form for pure multipartite states as defined in theorem 6 that is unique up to LU. The input is a complex vector $x$ of dimension $n_1 \cdot n_2 \cdots n_N$ (i.e. the coefficients of the pure state), a vector $mm$ specifying the dimensions $[n_1, n_2, \cdots, n_N]$, and an optional number nmax controlling the maximal number of iterations. The output is the normal form $X$, the SLOCC operations $A(:,:,k)$ needed to bring the state into normal form, and a number indicating the number of iterations done.

```
function [X,A,kk]=normalp(x,mm,nmax);

if nargin==2,nmax=200;end;trnorm=norm(x);x=x/trnorm;
m=max(size(mm));mmm=mm;ma=max(mm);tot=prod(mm);mmm(m+1)=1;

for k=1:m,
 P=prod(mmm(m-k+2:m+1));nn=[];
 for kk=1:P,nn=[nn;kk+[0:P:P*(mmm(m-k+1)-1)]'];end;
 P=round(prod(mmm)/prod(mmm(m-k+1:m+1)));nn1=nn;PP=max(size(nn));
 for kk=2:P,nn1=[nn1;nn+(kk-1)*PP];end;
nnn(:,k)=nn1; end;

A=zeros(ma,ma,m);for k=1:m,si=mm(k);A(1:si,1:si,k)=eye(si);end;
tr=1;tr0=2;kk=0;
while (tr0-tr>10^(-12))*(kk<nmax),kk=kk+1;
tr0=tr; for k=1:m,
   ss=mm(k);
   X=reshape(x(nnn(:,m+1-k)),ss,tot/ss);
   [U,S,V]=svd(X);
   A(1:ss,1:ss,k)=A(1:ss,1:ss,k)*U*S(1:ss,1:ss)/(det(U*S(1:ss,1:ss))^(1/ss)+eps);
   X=inv(S(1:ss,1:ss)+eps)*(eps+det(U*S(1:ss,1:ss))^(1/ss))*U'*X;
   x(nnn(:,m+1-k))=reshape(X,tot,1);
end; tr=norm(X,'fro'); end;
X=x*trnorm;
```

The last program calculates the normal SLOCC form in the case of mixed states. The only difference in the assignment of the variables is now that $\rho$ is a density operator (i.e. a matrix) instead of a vector (note that this program also works for pure states $\rho = |\psi\rangle\langle\psi|$, although this is neither efficient nor good for precision). Note also that this program can be used to calculate the Lorentz singular value decomposition of generic mixed states of two qubits in an efficient way and to a high accuracy.

```
function [X,A,kk]=normal(rho,mm,nmax);

if nargin==2,nmax=200;end;
m=max(size(mm));mmm=mm;ma=max(mm);tot=prod(mm);mmm(m+1)=1; for
k=1:m,
 P=prod(mmm(m-k+2:m+1));nn=[];
 for kk=1:P,nn=[nn;kk+[0:P:P*(mmm(m-k+1)-1)]'];end;
 P=round(prod(mmm)/prod(mmm(m-k+1:m+1)));nn1=nn;PP=max(size(nn));
 for kk=2:P,nn1=[nn1;nn+(kk-1)*PP];end;
```

```
nnn(:,k)=nn1; end;

A=zeros(ma,ma,m);for k=1:m,si=mm(k);A(1:si,1:si,k)=eye(si);end;
X=rho;tr=1;tr0=2;kk=0;
while (tr0-tr>10^(-16))*(kk<nmax),kk=kk+1;
tr0=tr; for tel=1:m,
   si=mm(m+1-tel);XX=zeros(si);
   for k=1:tot/si,
      XX=XX+X(nnn([1:si]+si*(k-1),tel),nnn([1:si]+si*(k-1),tel));
   end;
   [u,s,v]=svd(XX);C=inv(s+eps)^(1/2)*u'*A(1:si,1:si,m+1-tel);
   C=C/(det(C))^(1/si);A(1:si,1:si,m+1-tel)=C;
   AA=A(1:mm(1),1:mm(1),1);
   for kkk=2:m, AA=kron(AA,A(1:mm(kkk),1:mm(kkk),kkk));end;
   X=AA*rho*AA';
 end;
tr=real(trace(X)); end;

for k=1:m,A(1:mm(k),1:mm(k),k)=inv(A(1:mm(k),1:mm(k),k));end;
```

# Publications


(1) F. Verstraete, J. Dehaene and B. De Moor, “Local filtering operations on two qubits”, *Phys. Rev. A* **64**, 1, 010101(R) (2001).

(2) F. Verstraete, K. Audenaert and B. De Moor, “Maximally entangled mixed states of two qubits”, *Phys. Rev. A* **64**, 1, 012316 (2001).

(3) F. Verstraete, A. Doherty and H. Mabuchi, “Sensitivity optimization in quantum parameter estimation”, *Phys. Rev. A* **64**, 032111 (2001).

(4) K. Audenaert, F. Verstraete and B. De Moor, “Variational characterizations of separability and entanglement of formation”, *Phys. Rev. A* **64**, 5, 052304 (2001).

(5) F. Verstraete, K. Audenaert, J. Dehaene and B. De Moor, “A comparison of the entanglement measures negativity and concurrence”, *J. Phys. A* **34**, 10327 (2001).

(6) F. Verstraete, J. Dehaene and B. De Moor, “Lorentz singular-value decomposition and its applications to pure states of three qubits”, *Phys. Rev. A* **65**, 3, 032308 (2002).

(7) F. Verstraete, J. Dehaene, B. De Moor and H. Verschelde, “Four qubits can be entangled in nine different ways”, *Phys. Rev. A* **65**, 052112 (2002).

(8) F. Verstraete, J. Dehaene and B. De Moor, “On the geometry of entangled states”, *J. Mod. Opt.* **49**, 1277 (2002).

(9) F. Verstraete and H. Verschelde, “On the fidelity of mixed states of two qubits”, *Phys. Rev. A* **66**, 022307 (2002).

(10) F. Verstraete and M. M. Wolf, “Entanglement versus Bell violations and their behaviour under local filtering operations”, to appear in Phys. Rev. Lett.

(11) F. Verstraete, J. Dehaene and B. De Moor, “Normal forms, entanglement monotones and optimal filtration of multipartite quantum systems”, quant-ph/0105090.

(12) F. Verstraete and H. Verschelde, “On quantum channels”, quant-ph/0202124.

(13) F. Verstraete and H. Verschelde, “Optimal teleportation with a mixed state of two qubits”, quant-ph/0203073.

(14) T. Laustsen, F. Verstraete and S. van Enk, “Local vs. Joint Measurements for the Entanglement of Assistance”, quant-ph/0206192.

(15) A. Childs, D. Leung, F. Verstraete and G. Vidal, “Asymptotic Entanglement Capability of Ising and anisotropic Heisenberg interactions”, quant-ph/0207052.

(16) J. Dehaene, M. Van den Nest, F. Verstraete and B. De Moor, "Local Permutations of Products of Bell states and Entanglement Distillation", quant-ph/0207154

(17) T.C. Wei, K. Nemoto, P.M. Goldbart, P.G. Kwiat, W.J. Munro and F. Verstraete, "Maximal entanglement versus entropy for mixed quantum states", quant-ph/0208138.

# Bibliography


1. A. Acín, A. Andrianov, L. Costa, E. Jané, J. I. Latorre, and R. Tarrach. Schmidt decomposition and classification of three-quantum-bit states. *Phys. Rev. Lett.*, 85:1560, 2000.
2. A. Acín, D. Bruß, M. Lewenstein, and A. Sanpera. Classification of mixed three-qubit states. *Phys. Rev. Lett.*, 87:040401, 2001.
3. A. Acín, E. Jané, W. Dür, and G. Vidal. Optimal distillation of a ghz state. *Phys. Rev. Lett.*, 85:4811, 2000.
4. A. Acín, G. Vidal, and J.I. Cirac. On the structure of a reversible entanglement generating set for three-partite states. quant-ph/0202056.
5. D. Aerts and I. Daubechies. Physical justification for using the tensor product to describe two quantum systems as one joint system. *Helv. Phys. Acta*, 51:661, 1978.
6. M. Araki and T. Matsui. Ground state of the xy-model. *Comm. Math. Phys.*, 101:213, 1985. (see also Lett. Math. Phys., 11:87, 1986).
7. A. Aspect, J. Dalibard, and G. Roger. Experimental test of bell's inequalities using time-varying analyzers. *Phys. Rev. Lett.*, 49:1804, 1982.
8. K. Audenaert, J. Eisert, E. Jané, M.B. Plenio, S. Virmani, and B. De Moor. The assymptotic relative entropy of entanglement. *Phys. Rev. Lett.*, 87:217902, 2001.
9. K. Audenaert and B. De Moor. Optimizing completely positive maps using semidefinite programming. *Phys. Rev. A*, 65:030302, 2002.
10. K. Audenaert, F. Verstraete, and B. De Moor. Variational characterizations of separability and entanglement of formation. *Phys. Rev. A*, 64:052304, 2001.
11. P. Badziąg, P. Deuar, M. Horodecki, P. Horodecki, and R. Horodecki. Concurrence in arbitrary dimensions. *J. Mod. Opt.*, 49:1289, 2002.
12. P. Badziąg, M. Horodecki, P. Horodecki, and R. Horodecki. Local environment can enhance fidelity of quantum teleportation. *Phys. Rev. A*, 62:012311, 2000.
13. K. Banaszek. Fidelity balance in quantum operations. *Phys. Rev. Lett.*, 86:1366, 2001.

14. K. Banaszek and I. Devetak. Fidelity trade-off for finite ensembles of identically prepared qubits. *Phys. Rev. A*, 64:052307, 2001.
15. A. Barchielli, L.Lanz, and G.M.Prosperi. A model for macroscopic description and continuous observations in quantum mechanics. *Nuovo Cimento B*, 72:79, 1982.
16. H. Barnum, E. Knill, and M. A. Nielsen. On quantum fidelities and channel capacities. *IEEE Trans. Inf. Theory*, 46:1317, 2000.
17. T. Bayes. *Phil. Trans. Roy. Soc.*, 53:370, 1763. reprinted in *Biometrika* 45 (1958) 293.
18. V. P. Belavkin. Measurement, filtering and control in quantum open dynamical systems. *Rep. Math. Phys.*, 43:405, 1999. see also V. P. Belavkin, in *Modelling and Control of Systems*, edited by A. Blaquiere (Springer-Verlag, Berlin, 1989).
19. V. P. Belavkin and P. Staszewski. Nondemolition observation of a free quantum particle. *Phys. Rev. A*, 45:1347, 1992.
20. J.S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200, 1964.
21. J.S. Bell. On the problem of hidden variables in quantum theory. *Rev. of Mod. Phys.*, 38:447–452, 1966.
22. C. H. Bennett, S. Popescu, D. Rohrlich, J. A. Smolin, and A. V. Thapliyal. Exact and asymptotic measures of multipartite pure state entanglement. *Phys. Rev. A*, 63:012307, 2000.
23. C. H. Bennett and S. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Phys. Rev. Lett.*, 69:2881–2884, 1992.
24. C.H. Bennett. Logical reversibility of computation. *IBM J. Res. Develop*, 17:525–532, 1973.
25. C.H. Bennett. Quantum cryptography using any two nonorthogonal states. *Phys. Rev. Lett.*, 68:3121, 1992.
26. C.H. Bennett, H.J. Bernstein, S. Popescu, and B. Schumacher. Concentrating partial entanglement by local operations. *Phys. Rev. A*, 53:2046–2052, 1996.
27. C.H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. Proc. of the IEEE Int. Conf. on Computers, Systems and Signal Processing, pages 175–179, 1984.
28. 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.
29. C.H. Bennett, G. Brassard, C. Crépeau, and U. Mauer. Generalized privacy amplification. *IEEE Trans. on Inf. Theory*, 41:1915, 1995.

30. 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.
31. C.H. Bennett, J.I. Cirac, M. S. Leifer, D. W. Leung, N. Linden, S. Popescu, and G. Vidal. Optimal simulation of two-qubit hamiltonians using general local operations. *Phys. Rev. A*, 66:012305, 2002.
32. C.H. Bennett, D.P. DiVincenzo, C.A. Fuchs, T. Mor, E.M. Rains, P.W. Shor, J.A. Smolin, and W.K. Wootters. Quantum nonlocality without entanglement. *Phys. Rev. A*, 59:1070–1091, 1999.
33. C.H. Bennett, D.P. DiVincenzo, J.A. Smolin, B.M. Terhal, and W.K. Wootters. Remote state preparation. *Phys. Rev. Lett.*, 87:077902, 2001.
34. 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.
35. C.H. Bennett, A. Harrow, D. W. Leung, and J. A. Smolin. On the capacities of bipartite hamiltonians and unitary gates. quant-ph/0205057.
36. D.W. Berry and B.C. Sanders. Classical communication capacity is equal to the entanglement capability for two-qubit operations. quant-ph/0205181.
37. R. Bhatia. *Matrix Analysis*. Springer-Verlag, New York, 1997.
38. D. Bohm. *Quantum Theory*. Prentice-Hall, New York, 1951.
39. N. Bohr. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 48:696, 1935.
40. A.W. Bojancayk, R. Onn, and A.O. Steinhardt. Existence of the hyperbolic singular value decomposition. *Lin. Alg. Appl.*, 185:21, 1993.
41. Y. Bolshakov and B. Reichstein. Unitary equivalence in an indefinite scalar product: An analogue of singular-value decomposition. *Lin. Alg. Appl.*, 222:155–226, 1995.
42. G. Bowen and S. Bose. Teleportation as a depolarizing quantum channel, relative entropy and classical capacity. *Phys. Rev. Lett.*, 87:267901, 2001.
43. V.B. Braginsky and F.Y. Khalili. *Quantum Measurement*. Cambridge University Press, 1992.
44. S. Braunstein, C.M. Caves, R. Jozsa, N. Linden, and S. Popescu. Separability of very noisy mixed states and implications for nmr quantum computing. *Phys. Rev. Lett.*, 83:1054, 1999.
45. H.J. Briegel and R. Raussendorf. Persistent entanglement in arrays of interacting particles. *Phys. Rev. Lett.*, 86:910, 2001.
46. H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *Proceedings of 30th STOC*, pages 63–68, 1998.
47. V. Bužek and M. Hillery. Quantum copying: Beyond the no-cloning theorem. *Phys. Rev. A*, 54:1844, 1996.

48. A.R. Calderbank, E.M. Rains, P.W. Shor, and N.J.A. Sloane. Quantum error correction and orthogonal geometry. *Phys. Rev. Lett.*, 78:405, 1997.
49. H.A. Carteret, A. Higuchi, and A. Sudbery. Multipartite generalization of the schmidt decomposition. *J. Math. Phys.*, 41:7932, 2000.
50. C.M. Caves and G.J. Milburn. Quantum-mechanical model for continuous position measurements. *Phys. Rev. A*, 36:5543, 1987.
51. N. Cerf. Pauli cloning of a quantum bit. *Phys. Rev. Lett.*, 84:4497, 2000.
52. N. J. Cerf, C. Adami, and R. M. Gingrich. Quantum conditional operator and a criterion for separability. *Phys. Rev. A*, 60:893–898, 1999.
53. N.J. Cerf, S.Massar, and S. Schneider. Multipartite classical and quantum secrecy monotones. quant-ph/0202103.
54. H. Chen. Necessary conditions for the efficient simulation of hamiltonians using local unitary transformations. quant-ph/0109115.
55. A. Childs, D. Leung, F. Verstraete, and G. Vidal. Asymptotic entanglement capability of ising and anisotropic heisenberg interactions. quant-ph/0207052.
56. M.-D. Choi. Completely positive linear maps on complex matrices. *Linear Algebra and Its Applications*, 10:285–290, 1975.
57. I. Cirac, W. Dür, B. Kraus, and M. Lewenstein. Entangling operations and their implementation using a small amount of entanglement. *Phys. Rev. Lett.*, 86:544, 2001.
58. J.F. Clauser, M.A. Horne, A. Shimony, and R.A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880, 1969.
59. V. Coffman, J. Kundu, and W.K. Wootters. Distributed entanglement. *Phys. Rev. A*, 61:052306, 2000.
60. O. Cohen. Unlocking hidden entanglement with classical information. *Phys. Rev. Lett.*, 80:2493, 1998.
61. T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 1991.
62. E.B. Davies. Information and quantum measurement. *IEEE Trans. on Inf. Theory*, 24(5):596, 1978.
63. J. Dehaene. *Continuous-time matrix algorithms, systolic algorithms and adaptive neural networks*. PhD thesis, SISTA/ESAT, KULeuven, 1995.
64. J. Dehaene, M. Van den Nest, B. De Moor, and F. Verstraete. Local permutations of products of bell states and entanglement distillation. 2002. quant-ph/0207154.
65. J. dePillis. Linear transformations which preserve hermitian and positive semidefinite operators. *Pacific J. Math.*, 23:129, 1967.
66. D. Deutsch. Quantum theory, the Church-Turing principle, and the universal quantum Turing machine. In *Proceedings of the Royal Society of London*, volume A400, pages 97–117, 1985.

67. 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, 1996.

68. D. Deutsch and R. Jozsa. Rapid solution of problems by quantum computation. In *Proceedings of the Royal Society of London*, volume A439, pages 553–558, 1992.

69. P.A.M. Dirac. *The principles of quantum mechanics.* Oxford University Press, Oxford, 1930 (1st edition), 1958 (4th edition).

70. D. P. DiVincenzo and D.W. Leung B.M. Terhal. Quantum data hiding. quant-ph/0103098.

71. D.P. DiVincenzo, C.A. Fuchs, H. Mabuchi, J.A. Smolin, A. Thapliyal, and A. Uhlmann. *The entanglement of assistance*, pages 247–257. Springer-Verlag, Berlin, 1999.

72. D.P. DiVincenzo, P.W. Shor, and J.A. Smolin. Quantum-channel capacity of very noisy channels. *Phys. Rev. A*, 57:830, 1998.

73. J. L. Dodd, M. A. Nielsen, M. J. Bremner, and R. T. Thew. Universal quantum computation and simulation using any entangling hamiltonian and local unitaries. *Phys. Rev. A*, 65:040301(R), 2002.

74. A. C. Doherty and K. Jacobs. Feedback control of quantum systems using continuous state estimation. *Phys. Rev. A*, 60:2700, 1999.

75. A. C. Doherty, S. M. Tan, A. S. Parkins, and D. F.Walls. State determination in continuous measurement. *Phys. Rev. A*, 60:2380, 1999.

76. A.C. Doherty, P.A. Parrilo, and F.M. Spedalieri. Distinguishing separable and entangled states. *Phys. Rev. Lett.*, 88:187904, 2002.

77. W. Dür and J.I. Cirac. Non-local operations: Purification, storage, compression, tomography, and probabilistic implementation. *Phys. Rev. A*, 64:012317, 2001.

78. W. Dür and J.I. Cirac. Equivalence classes of non-local unitary operations. *Quantum Information and Computation*, 2:240, 2002.

79. W. Dür, G. Vidal, J. I. Cirac, N. Linden, and S. Popescu. Entanglement capabilities of non-local hamiltonians. *Phys. Rev. Lett.*, 87:137901, 2001.

80. W. Dür, G. Vidal, and J.I. Cirac. Three qubits can be entangled in two inequivalent ways. *Phys. Rev. A*, 62:062314, 2000.

81. T. Eggeling and R. Werner. Hiding classical data in multi-partite quantum states. quant-ph/0203004.

82. A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev. A*, 47:777–780, 1935.

83. J. Eisert and M. Plenio. A comparison of entanglement measures. *Journal of Modern Optics*, 46:145, 1999.

84. A. Ekert. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.*, 67:661–663, 1991.

85. R. Feynman. Simulating physics with computers. *Int. Jour. of Theor. Phys.*, 21(6/7):467–488, 1982.
86. R. Feynman. Quantum mechanical computers. *Optics News*, 11:11–20, 1985.
87. C.A. Fuchs. *Distinguishability and Accessible Information in Quantum Theory*. PhD thesis, University of New Mexico, Albuquerque, 1995.
88. C.A. Fuchs. Information gain vs. disturbance in quantum theory. *Physcomp96*, 1996. quant-ph/9611010.
89. C.A. Fuchs. Nonorthogonal quantum states maximize classical information capacity. *Phys. Rev. Lett.*, 79:1162, 1997.
90. C.A. Fuchs. Quantum mechanics as quantum information (and a little more). 2002. quant-ph/0205039.
91. C.A. Fuchs and A. Peres. Quantum state disturbance vs. information gain: Uncertainty relations for quantum information. *Phys. Rev. A*, 53:2038, 1996.
92. A. Fujiwara and P. Algoet. Affine parameterization of completely positive maps on a matrix algebra. *Phys. Rev. A*, 59:3290, 1999.
93. J. Gambetta and H.M. Wiseman. State and dynamical parameter estimation for open quantum systems. quant-ph/0103032.
94. F.R. Gantmacher. *The theory of matrices*. Chelsea Publishing Company, New York, 1959.
95. I.M. Gelfand, M.M. Kapranov, and A.V. Zelevinsky. *Discriminants, Resultants and Multidimensional Determinants*. Birkhäuser, Boston, 1994.
96. N. Gisin. Bell's inequality hold for all non-product states. *Phys. Lett. A*, 154:6, 1991.
97. N. Gisin. Hidden quantum nonlocality revealed by local filters. *Phys. Lett. A*, 210:151, 1996.
98. N. Gisin and H. Bechmann-Pasquinucci. Bell inequality, bell states and maximally entangled states for n qubits. *Phys. Lett. A*, 246:1, 1998.
99. A.M. Gleason. Measures on the closed subspaces of a hilbert space. *J. Math. Mech.*, 6:885, 1957.
100. I. Gohberg, P. Lancaster, and L. Rodman. *Matrices and indefinite scalar products*. Birkhauser Verlag, 1983.
101. D. Gottesman. Class of quantum error-correcting codes saturating the quantum hamming bound. *Phys. Rev. A*, 54:1862, 1996.
102. D.M. Greenberger, M. Horne, and A. Zeilinger. *Bell's Theorem, Quantum theory and Conceptions of the Universe*, chapter Going beyond Bell's theorem, page 69. Kluwer, Dordrecht, 1989.
103. L. Gurvits. Quantum matching theory. quant-ph/0201022.
104. S. Hallgren. Polynomial-time quantum algorithms for pell's equation and the principal ideal problem. STOC 2002.

105. M. Hamada. A lower bound on the quantum capacity of channels with correlated errors. quant-ph/0201056.

106. M. Hamada. Lower bounds on the quantum capacity and error exponent of general memoryless channels. quant-ph/0112103.

107. K. Hammerer, G. Vidal, and J. I. Cirac. Characterization of non-local gates. quant-ph/ 0205100.

108. J.B. Hartle. Quantum mechanics of individual systems. *Am. J. Phys.*, 36:704, 1968.

109. C.W. Helstrom. *Quantum Detection and Estimation Theory.* Academic Press, New York, 1976.

110. L. Henderson and V. Vedral. Classical, quantum and total correlation. *Jour. of Phys. A: Math. and Gen.*, 34(35):6899–6905, 2001.

111. S. Hill and W.K. Wootters. Entanglement of a pair of quantum bits. *Phys. Rev. Lett.*, 78:5022, 1997.

112. A. S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problemy Peredachi Informatsii*, 9(3):3–11, 1973. English translation in *Problems of Information Transmission*, 9:177–183, 1973.

113. A.S. Holevo. On the capacity of quantum communication channel. *Probl. Peredachi Inform.*, 15(4):3, 1979.

114. A.S. Holevo. The capacity of quantum channel with general signal states. *IEEE Trans. on Inf. Theory*, 44:269, 1998.

115. Yoopyo Hong. A canonical form for hermitean matrices under complex orthogonal congruence. *SIAM J. Matrix Anal. Appl.*, 10(2):233–243, April 1989.

116. R.A. Horn and C.R. Johnson. *Matrix Analysis.* Cambridge University Press, 1985.

117. R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, 1991.

118. R.A. Horn and D.I. Merino. Contragredient equivalence - a canonical form and some applications. *Lin. Alg. Appl.*, 214:43–92, 1995.

119. M. Horodecki and P. Horodecki. Reduction criterion of separability and limits for a class of distillation protocols. *Phys. Rev. A*, 59:4206, 1999.

120. M. Horodecki, P. Horodecki, and R. Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223:1–8, 1996.

121. M. Horodecki, P. Horodecki, and R. Horodecki. Inseparable two spin 1/2 density matrices can be distilled to a singlet form. *Phys. Rev. Lett.*, 78:574–577, 1997.

122. M. Horodecki, P. Horodecki, and R. Horodecki. Entanglement and thermodynamical analogies. *Acta Phys. Slov.*, 48:141, 1998.

123. 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.

124. M. Horodecki, P. Horodecki, and R. Horodecki. General teleportation channel, singlet fraction, and quasidistillation. *Phys. Rev. A*, 60:1888–1898, 1999.

125. R. Horodecki, P. Horodecki, and M. Horodecki. Violating bell inequality by mixed spin-1/2 state: necessary and sufficient condition. *Phys. Lett. A*, 200:340, 1995.

126. R. L. Hudson and K. R. Parthasarathy. Quantum ito's formula and stochastic evolutions. *Comm. Math. Phys.*, 93:301, 1984.

127. L.P. Hughston, R. Jozsa, and W.K. Wootters. A complete classification of quantum ensembles having a given density operator. *Phys. Lett. A*, 183:14, 1993.

128. Special issue. *Fortschr. Phys.*, 48(9-11), 2000.

129. A. Jamiolkowski. Linear transformations which preserve trace and positive semidefiniteness of operators. *Rev. of Mod. Phys.*, 3:275–278, 1972.

130. E. Jané. Purification of two-qubit mixed states. *Quant. Inf. Comp.*, 2:348, 2002.

131. E.T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106:620–630, 1957.

132. E.T. Jaynes. Information theory and statistical mechanics. ii. *Physical Review*, 108:171–190, 1957.

133. D. Jonathan and M. Plenio. Entanglement-assisted local manipulation of pure quantum states. *Phys. Rev. Lett.*, 83:3566, 1999.

134. D. Jonathan and M. Plenio. Minimal conditions for local pure-state entanglement manipulation. *Phys. Rev. Lett.*, 83:1455, 1999.

135. A. Kent. Entangled mixed states and local purification. *Phys. Rev. Lett.*, 81:2839, 1998.

136. A. Kent, N. Linden, and S. Massar. Optimal entanglement enhancement for mixed states. *Phys. Rev. Lett.*, 83:2656, 1999.

137. N. Khaneja, R. Brockett, and S. J. Glaser. Time optimal control in spin systems. *Phys. Rev. A*, 63:032308, 2O01.

138. C. King. Additivity for a class of unital qubit channels. quant-ph/0103156.

139. C. King, M. Nathanson, and M.B. Ruskai. Qubit channels can require more than two inputs to achieve capacity. *Phys. Rev. Lett.*, 88:057901, 2002.

140. C. King and M.-B. Ruskai. Minimal entropy of states emerging from noisy quantum channels. *IEEE Trans. on Inf. Theory*, 47:192–209, 2001.

141. B. Kraus and J. I. Cirac. Optimal creation of entanglement using a two-qubit gate. *Phys. Rev. A*, 63:062309, 2001.

142. K. Kraus. *States, Effects and Operations: Fundamental Notions of Quantum Theory.* Springer-Verlag, 1983.

143. A. Lamas-Linares, J. C. Howell, and D. Bouwmeester. Stimulated emission of polarization-entangled photons. *Nature*, 412:887, 2001.

144. L.J. Landau and R.F. Streater. On birkhoff's theorem for doubly stochastic completely positive maps of matrix algebras. *Lin. Alg. Appl.*, 193:107, 1993.

145. R. Landauer. Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.*, 5:183, 1961.

146. T. Laustsen, F. Verstraete, and S. van Enk. Local vs. joint measurements for the entanglement of assistance. quant-ph/0206192.

147. E. Lieb and M.B. Ruskai. A fundamental property of quantum-mechanical entropy. *Phys. Rev. Lett.*, 30:434, 1973.

148. N. Linden, S. Massar, and S. Popescu. Purifying noisy entanglement requires collective measurements. *Phys. Rev. Lett.*, 81:3279, 1998.

149. H. Lo and H.F. Chau. Unconditionl security of quantum key distribution over arbitrarily long distances. *Science*, 283:2050, 1999.

150. H.-K. Lo and S. Popescu. The classical communication cost of entanglement manipulation: Is entanglement an inter-convertible resource? *Phys. Rev. Lett.*, 83:1459–1462, 1999.

151. H.-K. Lo and S. Popescu. Concentrating entanglement by local actions—beyond mean values. *Phys. Rev. A*, 63:022301/1–16, 2001.

152. H. Mabuchi. Dynamical identification of open quantum systems. *Quantum Semiclass. Opt.*, 8:1103, 1996.

153. H. Mabuchi. Standard quantum limits for broadband position measurement. *Phys. Rev. A*, 58:123, 1998.

154. C. Macchiavello. On the analytical convergence of the qpa procedure. *Phys. Lett. A*, 247:385, 1998.

155. A.W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and Its Applications.* Acedemic Press, 1979.

156. L. Masanes, G. Vidal, and J.I. Latorre. Time-optimal hamiltonian simulation and gate synthesis using homogeneous local unitaries. quant-ph/0202042.

157. P.S. Maybeck. *Stochastic Models, Estimation and Control.* Academic Press, 1982.

158. N.D. Mermin. What's wrong with these elements of reality? *Physics Today*, 43:9, 1990.

159. G. J. Milburn. Classical and quantum conditional statistical dynamics. *Quantum Semiclass. Opt.*, 8:269, 1996.

160. A. Miyake. Topological classification of multipartite entangled states by the hyperdeterminant. 2002. quant-ph/0206111.

161. A. Miyake and F. Verstraete. A complete classification of all $2 \times 2 \times n$ quantum systems. (in preparation).

162. M.A. Nielsen. Conditions for a class of entanglement transformations. *Phys. Rev. Lett.*, 83:436–439, 1999.

163. M.A. Nielsen, M.J. Bremner, J.L. Dodd, A.M. Childs, and C.M. Dawson. Universal simulation of hamiltonian dynamics for qudits. quant-ph/0109064.

164. M.A. Nielsen and J. Kempe. Separable states are more disordered globally than locally. *Phys. Rev. Lett.*, 86:5184–5187.

165. T. Ogawa and H. Nagaoka. Strong converse and stein's lemma in quantum hypothesis testing. *IEEE Trans. Inf. Theor.*, 46:2428, 2000.

166. M. Ozawa. Measurement breaking the standard quantum limit for free-mass position. *Phys. Rev. Lett.*, 60:385, 1988.

167. M. Ozawa. Entanglement measures and the hilbert-schmidt distance. *Phys. Lett. A*, 268:158, 2000.

168. J. Pan, M. Daniell, S. Gasparoni, G. Weihs, and A. Zeilinger. Experimental demonstration of four-photon entanglement and high-fidelity teleportation. *Phys. Rev. Lett.*, 86:4435, 2001.

169. A. Peres. *Quantum Theory: Concepts and Methods*. Kluwer Academic Publishers, 1993.

170. A. Peres. Separability criterion for density matrices. *Phys. Rev. Lett.*, 77:1413–1415, 1996.

171. A. Peres and D.R. Terno. Convex probability domain of generalized quantum measurements. *J. Phys. A*, 31:671, 1998.

172. M.B. Plenio. The holevo bound and landauer's principle. *Phys. Lett. A*, 263:4, 1999.

173. M.B. Plenio, S. Virmani, and P. Papadopoulos. Operator monotones, the reduction criterion and the relative entropy. *J. Phys. A*, 33:193, 2000.

174. S. Popescu. Bell's inequalities and density matrices: revealing hidden nonlocality. *Phys. Rev. Lett.*, 74:2619–2622, 1995.

175. E. M. Rains. Rigorous treatment of distillable entanglement. *Phys. Rev. A*, 60:173, 179, 1999.

176. E.M. Rains. A semidefinite program for distillable entanglement. *IEEE Trans. on Inf. Theory*, 47:2921, 2001.

177. A.V.G. Rao and K.S. Mallesh. On the algebraic characterization of a mueller matrix in polarization optics- i: Identifying a mueller matrix from its n-matrix; ii: Necessary and sufficient conditions for jones-derived mueller matrices. *J. Mod. Opt.*, 45:955 − 989, 1998.

178. R. Raz. Exponential separation of quantum and classical communication complexity. In *Proceedings of 31th STOC*, pages 358–367, 1999.

179. J. Rehácek, Z. Hradil, J. Fiurásek, and C. Brukner. Designing optimum cp maps for quantum teleportation. *Phys. Rev. A*, 64:060301, 2001.

180. P. Rungta, V. Bužek, C.M. Caves, M. Hillery, and G.J. Milburn. Universal state inversion and concurrence in arbitrary dimensions. *Phys. Rev. A*, 64:042315, 2001.

181. M.B. Ruskai, S. Szarek, and E. Werner. An analysis of completely-positive trace-preserving maps on 2x2 matrices. *Lin. Alg. Appl.*, 347:159–187, 2002.

182. C.A. Sackett, D. Kielpinski, B.E. King, C. Langer, V. Meyer, C.J. Myatt, M. Rowe, Q.A. Turchette, W.M. Itano, D.J. Wineland, and C. Monroe. Experimental entanglement of four particles. *Nature*, 404:256–259, 2000.

183. A. Sanpera, R. Tarrach, and G. Vidal. Local description of quantum inseparability. *Phys. Rev. A*, 58:826, 1998.

184. M. Sasaki, S.M. Barnett, R. Jozsa ans M. Osaki, and O. Hirota. Accessible information and optimal strategies for real symmetrical quantum sources. *Phys. Rev. A*, 59:3325, 1999.

185. J. Schliemann, J.I. Cirac, M. Kuś, M. Lewenstein, and D. Loss. Quantum correlations in two-fermion systems. *Phys. Rev. A*, 64:022303, 2001.

186. E. Schrödinger. Discussion of probability distributions between separated systems. *Proc. Camb. Phil. Soc.*, 31:555, 1935.

187. E. Schrödinger. Probability relations between separated systems. *Proc. Camb. Phil. Soc.*, 32:446, 1936.

188. B. Schumacher. Quantum coding. *Phys. Rev. A*, 51:2738, 1995.

189. B. Schumacher and M. Westmoreland. Sending classical information via noisy quantum channels. *Phys. Rev. A*, 56:131–138, 1997.

190. B. Schumacher, M. Westmoreland, and W.K. Wootters. Limitation on the amount of accessible information in a quantum channel. *Phys. Rev. Lett.*, 76:3453, 1996.

191. B. W. Schumacher. Sending entanglement through noisy quantum channels. *Phys. Rev. A*, 54:2614, 1996.

192. C.E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, 1948.

193. P. W. Shor. Algorithms for quantum computation: discrete logarithms and factoring. In Proceedings, 35th Annual Symposium on Foundations of Computer Science, IEEE Press, Los Alamitos, CA, 1994.

194. P. W. Shor. Scheme for reducing decoherence in quantum memory. *Phys. Rev. A*, 52:2493, 1995.

195. D. Simon. On the power of quantum computation. In Proceedings, 35th Annual Symposium on Foundations of Computer Science, IEEE Press, Los Alamitos, CA, 1994.

196. A.M. Steane. Error correcting codes in quantum theory. *Phys. Rev. Lett.*, 77:793, 1996.

197. A. Sudbery. On local invariants of pure three-qubit states. *J. Phys. A*, 34:643, 2001.

198. L. Szilard. Uber die entropieverminderung in einen thermodynamischen system bei eingriffen intelligenter wesen. *Z. Phys.*, 53:840, 1929.

199. G. 't Hooft. Equivalence relations between deterministic and quantummechanical systems. *J. Stat. Phys.*, 53:323, 1988.

200. B.M. Terhal, M. Horodecki, D.W. Leung, and D.P. DiVincenzo. The entanglement of purification. quant-ph/0202044.

201. R.C. Thompson. Singular values and diagonal elements of complex symmetric matrices. *Lin. Alg. Appl.*, 26:65, 1979.

202. A. Uhlmann. *Rep. Math. Phys.*, 1:147, 1970.

203. A. Uhlmann. On 1-qubit channels. *J. Phys. A*, 34:7047, 2001.

204. W. van Dam and P. Hayden. Embezzling entangled quantum states. quant-ph/0201041.

205. C.V.M. van der Mee. An eigenvalue criterion for matrices transforming stokes parameters. *J. Math. Phys.*, 34:5072, 1993.

206. L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38:49, 1996.

207. V. Vedral and M.B. Plenio. Entanglement measures and purification procedures. *Phys. Rev. A*, 57:1619, 1998.

208. V. Vedral, M.B. Plenio, M.A. Rippin, and P.L. Knight. Quantifying entanglement. *Phys. Rev. Lett.*, 78:2275, 1997.

209. F. Verstraete, K. Audenaert, J. Dehaene, and B. De Moor. A comparison of the entanglement measures negativity and concurrence. *J. Phys. A*, 34:10327, 2001.

210. F. Verstraete, K. Audenaert, and B. De Moor. Maximally entangled mixed states of two qubits. *Phys. Rev. A*, 64:012316, 2001.

211. F. Verstraete, J. Dehaene, and B. De Moor. Normal forms and entanglement monotones for multipartite quantum systems. quant-ph/0105090.

212. F. Verstraete, J. Dehaene, and B. De Moor. Local filtering operations on two qubits. *Phys. Rev. A*, 64:010101(R), 2001.

213. F. Verstraete, J. Dehaene, and B. De Moor. The lorentz singular value decomposition and its applications to pure states of three qubits. *Phys. Rev. A*, 65:032308, 2002.

214. F. Verstraete, J. Dehaene, and B. De Moor. On the geometry of entangled states. *J. Mod. Opt.*, 49:1277, 2002.

215. F. Verstraete, J. Dehaene, B. De Moor, and H. Verschelde. Four qubits can be entangled in nine different ways. *Phys. Rev. A*, 65:052112, 2002.

216. F. Verstraete, A. Doherty, and H. Mabuchi. Sensitivity optimization in quantum parameter estimation. *Phys. Rev. A*, 64:032111, 2001.

217. F. Verstraete and T. Rudolph. Quantum steering. (in preparation).

218. F. Verstraete and H. Verschelde. On quantum channels. quant-ph/0202124.

219. F. Verstraete and H. Verschelde. Optimal teleportation with a mixed state of two qubits. quant-ph/0203073.
220. F. Verstraete and H. Verschelde. On the fidelity of mixed states of two qubits. *Phys. Rev. A*, 66:022307, 2002.
221. F. Verstraete and M.M. Wolf. Entanglement versus bell violations and their behaviour under local filtering operations. *Phys. Rev. Lett.*, 89:XXXXX, 2002.
222. G. Vidal. Entanglement of pure states for a single copy. *Phys. Rev. Lett.*, 83:1046, 1999.
223. G. Vidal. Entanglement monotones. *Jour. of Modern Optics*, 47(2/3):355–376, 2000.
224. G. Vidal and J. Cirac. Optimal simulation of nonlocal hamiltonians using local operations and classical communication. quant-ph/0108076.
225. G. Vidal and J.I. Cirac. Catalysis in non–local quantum operations. *Phys. Rev. Lett.*, 88:167903, 2002.
226. G. Vidal, W. Dür, and I. Cirac. Entanglement cost of antisymmetric states. *Phys. Rev. Lett.*, 2002.
227. G. Vidal, K. Hammerer, and J.I. Cirac. Interaction cost of non-local gates. *Phys. Rev. Lett.*, 88:237902, 2002.
228. G. Vidal and R. Tarrach. Robustness of entanglement. *Phys. Rev. A*, 59:141, 1999.
229. G. Vidal and R. Werner. Computable measure of entanglement. *Phys. Rev. A*, 65:032314, 2002.
230. J. von Neumann. *Mathematische Grundlagen der Quantenmechanik*. Springer, Berlin, 1932. English transl.: Princeton Univ. Press, Princeton (1955).
231. J. von Neumann. Some matrix-inequalities and metrization of matric-space. *Tomsk Univ. Rev.*, 1:205–218, 1937.
232. J. Walgate, A.J. Short, L. Hardy, and V. Vedral. Local distinguishability of multipartite orthogonal quantum states. *Phys. Rev. Lett.*, 85:4972–4975, 2000.
233. B.-Y. Wang and B.-Y. Xi. Some inequalities for singular values of matrix products. *Lin. Alg. Appl.*, 264:109, 1997.
234. T.C. Wei, K. Nemoto, P.M. Goldbart, P.G. Kwiat, W.J. Munro, and F. Verstraete. Maximal entanglement versus entropy for mixed quantum states. quant-ph/0208138.
235. H. Weinfurter and M. Zukowski. Four-photon entanglement from down-conversion. *Phys. Rev. A*, 62:010102(R), 2001.
236. R.F. Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Phys. Rev. A*, 40:4277–4281, 1989.
237. R.F. Werner and M.M. Wolf. Bell inequalities and entanglement. *Quant. Inf. Comp.*, 1, 2001.

238. J.A. Wheeler and W.H. Zurek (eds.). *Quantum theory and measurement.* Princeton University Press, Princeton, 1983.

239. A. Winter. *Coding theorems of quantum information theory.* PhD thesis, Bielefeld University, 1999. quant-ph/9907077.

240. H.M. Wiseman. Adaptive phase measurements of optical modes: Going beyond the marginal q distribution. *Phys. Rev. Lett.*, 75:4587, 1995.

241. C. Witte and M. Trucks. A new entanglement measure induced by the hilbert-schmidt norm. *Phys.Lett. A*, 257:14, 1999.

242. P. Wocjan, D. Janzing, and Th. Beth. Simulating arbitrary pair-interactions by a given hamiltonian: Graph-theoretical bounds on the time complexity. quant-ph/0106077.

243. P. Wocjan, M. Roetteler, D. Janzing, and Th. Beth. Universal simulation of hamiltonians using a finite set of control operations. quant-ph/0109063,0109088.

244. W.K. Wootters. Entanglement of formation of an arbitrary state of two qubits. *Phys. Rev. Lett.*, 80:2245, 1998.

245. W.K. Wootters and W.H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982.

246. S. L. Woronowicz. Positive maps of low dimensional matrix algebras. *Rep. Math. Phys.*, 10:165–183, 1976.

247. H.P. Yuen. Contractive states and the standard quantum limit for monitoring free-mass positions. *Phys. Rev. Lett.*, 51:719, 1983.

248. P. Zanardi, C. Zalka, and L. Faoro. Entangling power of quantum evolutions. *Phys. Rev. A*, 62:030301, 2000.

249. K. Życzkowski. Volume of the set of separable states. ii. *Phys. Rev. A*, 60:3496, 1999.

250. K. Życzkowski, P. Horodecki, A. Sanpera, and M. Lewenstein. On the volume of the set of separable states. *Phys. Rev. A*, 58:883, 1998.