Information-theoretic Studies and Capacity Bounds: Group Network Codes and Energy Harvesting Communication Systems

Thesis by

Wei Mao

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2015

(Defended April 10, 2015)

To those who appreciate the beauty of mathematics

Acknowledgments

I am deeply grateful to my advisor, Prof. Babak Hassibi, who provided me endless guidance and support, both in research and life, during my Ph.D. studies at Caltech. It is him who not only conveyed to me the idea of getting enough insight into the problem before delving into the details, but also coached me looking at the big picture and background of a problem before concentrating on the specifics. It is also him who entrenched in my mind that, effective communication of one's ideas/research to the colleagues, the academic community, or the public, both in oral or written form, is almost as important as the ideas/research themselves. It is still him, who, during my hard times in life, offered me great emotional and mental support and guidance without any reservation. To him, my gratitude is beyond words.

I am also deeply grateful to Prof. Michael Aschbacher and Prof. Palghat P. Vaidyanathan. Prof. Aschbacher is among the most important math mentors in my life who helped shaping my understanding of mathematics. In his lectures and the collaboration with me, the rigorous math training I received and the influence of the axiomatic system perspective on me will continuously benefit my academic career. Prof. Vaidyanathan is another mentor who guided and supported me both in my research and life. His enlightening conversations with me, mostly happening in the corridor and coffee room, endowed me great wisdom and strength, especially when I was baffled during my difficult times. He also served on my Ph.D. examination committee and provided invaluable feedback for my research.

My sincere appreciation goes to Prof. Jehoshua Bruck, Prof. Victoria Kostina, Prof. Adam Wierman, and Prof. Amin Shokrollahi, who served on my Ph.D. examination committee/candidacy committee and provided invaluable feedback for my

work. Some of them are also good friends of mine, I am very thankful for their friendship and helpful academic discussions.

My warmest thanks go to my colleagues and friends, who made my life more enjoyable during my Ph.D. studies. In particular, I would like to thank my current and former group-mates at Caltech, especially Weiyu Xu, Ravi Teja Sukhavasi, Matthew Thill, and also our group secretary, Shirley Slattery. My gratitude also goes to my friends Zhaoyan Zhu, Zhiying Wang, Na Li, Xiangyu Wei, and Su-Peng Yu. I am greatly indebted to my dearest friend, Ran Pang, for the unlimited friendship and support she brought to me, especially during the period when I am writing my thesis.

And finally, my deepest gratitude extends to my family, for their endless love and care. My mother, Xianglian Xie, continuously supported and encouraged me during the difficult times of my life, without any reservations. My father, Bingsheng Mao, and elder brother, Yu Mao, helped educating and supporting me as I grew up. To my family, I shall always be grateful.

Abstract

Network information theory and channels with memory are two important but difficult frontiers of information theory. In this two-parted dissertation, we study these two areas, each comprising one part. For the first area we study the so-called entropy vectors via finite group theory, and the network codes constructed from finite groups. In particular, we identify the smallest finite group that violates the Ingleton inequality, an inequality respected by all linear network codes, but not satisfied by all entropy vectors. Based on the analysis of this group we generalize it to several families of Ingleton-violating groups, which may be used to design good network codes. Regarding that aspect, we study the network codes constructed with finite groups, and especially show that linear network codes are embedded in the group network codes constructed with these Ingleton-violating families. Furthermore, such codes are strictly more powerful than linear network codes, as they are able to violate the Ingleton inequality while linear network codes cannot. For the second area, we study the impact of memory to the channel capacity through a novel communication system: the energy harvesting channel. Different from traditional communication systems, the transmitter of an energy harvesting channel is powered by an exogenous energy harvesting device and a finite-sized battery. As a consequence, each time the system can only transmit a symbol whose energy consumption is no more than the energy currently available. This new type of power supply introduces an unprecedented input constraint for the channel, which is random, instantaneous, and has memory. Furthermore, naturally, the energy harvesting process is observed causally at the transmitter, but no such information is provided to the receiver. Both of these features pose great challenges for the analysis of the channel capacity. In this work we use

techniques from channels with side information, and finite state channels, to obtain lower and upper bounds of the energy harvesting channel. In particular, we study the stationarity and ergodicity conditions of a surrogate channel to compute and optimize the achievable rates for the original channel. In addition, for practical code design of the system we study the pairwise error probabilities of the input sequences.

Contents

List of Figures

List of Tables

Chapter 1

Summary

In information theory, communication networks and channels with memory are among the most important and intriguing directions that attract generations of researchers. Generalizing the idea of single user discrete memoryless channels, they provide models for more practical communication systems. Despite their importance, these two frontiers are difficult fields of study. With continuing effort of investigation, researchers have made numerous contributions and advances in these fields; however, many major problems still remain unsolved, especially in the first area. In this two-parted dissertation, we study some aspects of these two areas, each of which comprises one part.

In the first part we study the so-called entropy vectors via finite group theory, and group network codes, which are network codes constructed from finite groups. An entropy vector for a set of n jointly distributed discrete random variables is the tuple of $2^n - 1$ (joint) entropies for these variables. The collection of all such entropy vectors comprises an entropy region in the $2^n - 1$ dimensional real space, which plays an important role in determining the capacity region of a multi-source multi-sink wired network. It is shown that such an entropy region can be represented using finite groups and their subgroups. This connection is potentially useful for designing good network codes, since in principle from finite groups one can construct network codes that are able to violate the Ingleton inequality, which is an inequality respected by all linear network codes, but not satisfied by all entropy vectors. However, finding a "meaningful" finite group that violates the Ingleton inequality is not a trivial task:

from any known examples of Ingleton-violating entropy vectors one always obtains huge permutation groups, which reveal very little information on the group structure. In this work, we use computer search to find the smallest finite group that violates the Ingleton inequality, and then extend it to a family of violations using the tool of abstract group presentations. This family has a very clear structure, both in the sense of group theory and in terms of the representing matrices. Based on the analysis of these groups we generalize them to several families of Ingleton-violating groups, both exploiting their matrix structure and using the theory of group actions. As mentioned earlier these families are good candidates for constructing network codes. We study various aspects of such group network codes, and especially show that linear network codes are embedded in the group network codes derived from these Ingleton-violating families. Such codes are strictly more powerful than linear network codes, as they are able to violate the Ingleton inequality, while linear network codes cannot.

In the second part we investigate the impact of a new form of channel memory, the memory in the input constraints, to the channel capacity, through the study of a novel communication system—the energy harvesting channel. Different from traditional communication systems, the transmitter of an energy harvesting channel is powered by an exogenous energy harvesting device and a finite-sized battery. As a consequence, each time the system can only transmit a symbol whose energy consumption is no more than the energy currently available. This new type of power supply introduces an unprecedented input constraint for the channel, which is random, instantaneous, and has memory. In addition, the energy harvesting process is observed causally at the transmitter naturally, but no energy information is provided to the receiver. Both of these features pose great challenges for the analysis of the channel capacity. In this work we first use techniques from channels with causal transmitter side information to transform the original channel to an equivalent channel, which has no input constraint or side information, but still has memory. The capacity formula of such a channel is given by the Verdu-Han formula [1], which is not computable in general. However, by imposing some restrictions on the input alphabet, we obtain a surrogate channel, which is a finite state channel [2] in many cases.

For this simpler model we study the required stationarity and ergodicity conditions on the channel and source to apply the Shannon-McMillan-Breiman theorem to the joint input-output process, and then use stochastic methods to compute and optimize the information rates, which are achievable rates for the original channel. Such rates serve as lower bounds for the capacity of the energy harvesting channel. For the upper bounds, we use Gallager's techniques for finite state channels [2] to derive a series of capacity upper bounds in terms of finite block length mutual information. As these bounds are highly computationally expensive (the complexity of computation is double exponential), we relax them further to achieve linear computational complexity. In addition, we also study the pairwise error probabilities of the input sequences, which are useful for practical code design of the energy harvesting system.

Part I

Finite Groups &
Network Information Theory

Chapter 2

Ingleton-violating Finite Groups

In this chapter we study the problem of using finite groups to violate the Ingleton Inequality, which is organized as follows. Section 2.1 provides a detailed description of the background. Section 2.2 introduces the necessary notations. Section 2.3 describes the computer search of Ingleton-violating groups and proves several conditions that help in pruning the search. Having found the smallest violation instance, Section 2.4 studies its structure using group presentations. Section 2.5 then generalizes the instance to an Ingleton-violating family in $PGL(2, p)$, and then to $PGL(2, q)$, through explicitly constructing the subgroups in the format of matrices. Furthermore, the preimage group $GL(2, q)$ is also examined and 15 new families of Ingleton violating subgroups are identified, in Section 2.6. The original family has a deep relation to the theory of group actions, as disclosed in the more abstract Section 2.7, which leads to several new violation constructions in this framework.

2.1 Introduction

Let $\mathcal{N} = \{1, 2, \ldots, n\}$, and let $X_1, X_2, \ldots, X_n$ be n jointly distributed discrete random variables. For any nonempty set $\alpha \subseteq \mathcal{N}$, let X_α denote the collection of random variables $\{X_i : i \in \alpha\}$, with joint entropy $h_\alpha \triangleq H(X_\alpha) = H(X_i; i \in \alpha)$. We call the ordered real $(2^n - 1)$-tuple $(h_\alpha : \emptyset \neq \alpha \subseteq \mathcal{N}) \in \mathbb{R}^{2^n-1}$ an *entropy vector*. The set of all entropy vectors derived from n jointly distributed discrete random variables is denoted by Γ_n^*. It is not too difficult to show that the closure of this set, i.e., $\overline{\Gamma_n^*}$,

is a *convex cone* [3].

The set $\overline{\Gamma_n^*}$ figures prominently in information theory since it describes the possible values that the joint entropies of a collection of n discrete random variables can obtain. From a practical point of view, it is of importance since it can be shown that the capacity region of any arbitrary multi-source multi-sink *wired* network, whose graph is acyclic and whose links are discrete memoryless channels, can be obtained by optimizing a linear function of the entropy vector over the convex cone $\overline{\Gamma_n^*}$ and a set of linear constraints (defined by the network) [4,5]. Despite this importance, the entropy region $\overline{\Gamma_n^*}$ is only known for $n = 2, 3$ random variables and remains unknown for $n \geq 4$ random variables. Nonetheless, there are important connections known between $\overline{\Gamma_n^*}$ and matroid theory (since entropy is a submodular[1] function) [6], determinantal inequalities (through the connection with Gaussian random variables) [7], and quasi-uniform arrays [8]. However, perhaps most intriguing is the connection to finite groups, which we briefly elaborate on below.

2.1.1 Groups and Entropy

Throughout this work we use the group theory notation defined in Section 2.2. Let G be a finite group, and let $G_1, G_2, \ldots, G_n$ be n of its subgroups. For any nonempty set $\alpha \subseteq \mathcal{N}$, the group $G_\alpha \triangleq \bigcap_{i\in\alpha} G_i$ is a subgroup of G. Define

$$g_\alpha = \log \frac{|G|}{|G_\alpha|}.$$

We call the ordered real $(2^n - 1)$-tuple $(g_\alpha : \emptyset \neq \alpha \subseteq \mathcal{N}) \in \mathbb{R}^{2^n-1}$ a (finite) *group characterizable vector*. Let Υ_n be the set of all group characterizable vectors derived from n subgroups of a finite group.

The major result shown by Chan and Yeung in [9] is that $\overline{\Gamma_n^*} = \overline{\text{cone}(\Upsilon_n)}$, i.e., the closure of Γ_n^*, is the same as the closure of the cone generated by Υ_n. Specifically, every group characterizable vector is an entropy vector, whereas every entropy vector is arbitrarily close to a scaled version of some group characterizable vector.

[1] A set function f on the subsets of $\mathcal{N}$ is *submodular* iff $f_\alpha + f_\beta - f_{\alpha\cap\beta} - f_{\alpha\cup\beta} \geq 0$ for all $\alpha, \beta \subseteq \mathcal{N}$.

To show the first part of this statement, let Λ be a random variable uniformly distributed on the elements of G and for $i = 1, \ldots, n$ define

$$X_i = \Lambda G_i,$$

i.e., the left coset of G_i in G with representative Λ. Then X_i is uniformly distributed on G/G_i and $H(X_i) = \log \frac{|G|}{|G_i|}$. To calculate the joint entropy $h_\alpha = H(X_\alpha)$ for a nonempty subset $\alpha \subseteq \mathcal{N}$, let $\mathcal{X}_\alpha$ denote the set of all coset tuples

$$\{ (xG_i : i \in \alpha) \mid x \in G \}.$$

Consider the intersection mapping $\Theta_\alpha : \mathcal{X}_\alpha \to G/G_\alpha$, where for all $x \in G$,

$$\Theta_\alpha : (xG_i : i \in \alpha) \mapsto \bigcap_{i \in \alpha} xG_i = xG_\alpha. \tag{2.1}$$

Θ_α is a well defined onto function on $\mathcal{X}_\alpha$, and it is one-to-one since if $(xG_i : i \in \alpha)$ and $(x'G_i : i \in \alpha)$ are mapped to the same coset $xG_\alpha = x'G_\alpha$, then $x^{-1}x' \in G_\alpha$ and so $x^{-1}x' \in G_i$ for all i, which implies

$$(xG_i : i \in \alpha) = (x'G_i : i \in \alpha).$$

So $H(X_\alpha) = H(\Theta_\alpha(X_\alpha))$, and as $\Theta_\alpha(X_\alpha) = \Lambda G_\alpha$, we have

$$h_\alpha = H(\Lambda G_\alpha)) = \log \frac{|G|}{|G_\alpha|} = g_\alpha.$$

Thus every group-characterizable vector is indeed an entropy vector. Showing the other direction, i.e., that every entropy vector can be approximated by a scaled group-characterizable vector, is more tricky (the interested reader may consult [9] for the details). Here we shall briefly describe the intuition.

Consider a random variable X_1 with alphabet size N and probability mass function $\{p_i, i = 1, \ldots, N\}$. Now, if we make T copies of this random variable to make

sequences of length T, the entropy of X_1 is roughly equal to the logarithm of the number of strongly typical sequences divided by T. These are sequences where X_1 takes its first value roughly Tp_1 times, its second value roughly Tp_2 times, and so on. Therefore, assuming that T is large enough so that the Tp_i are close to integers (otherwise, we have to round things) we may roughly write

$$H(X_1) \approx \frac{1}{T} \log \left(\begin{array}{ccccc} & & T & & \\ Tp_1 & Tp_2 & \ldots & Tp_{N-1} & Tp_N \end{array} \right),$$

where the argument inside the log is the usual multinomial coefficient. Written in terms of factorials this is

$$H(X_1) \approx \frac{1}{T} \log \frac{T!}{(Tp_1)!(Tp_2)! \ldots (Tp_N)!}. \tag{2.2}$$

If we consider the group G to be the symmetric group S_T, i.e., the group of permutations among T objects, then clearly $|G| = T!$. Now partition the T objects into N sets each with Tp_1 to Tp_N elements, respectively, and define the group G_1 to be the subgroup of S_T that permutes these objects *while respecting the partition*. Clearly, $|G_1| = (Tp_1)!(Tp_2)! \ldots (Tp_N)!$, which is the denominator in (2.2). Thus, $H(X_1) \approx \frac{1}{T} \log \frac{|G|}{|G_1|}$, so that the entropy $h_{\{1\}}$ is a scaled version of the group-characterizable $g_{\{1\}}$. This argument can be made more precise and extended to n random variables—see [9] for the details. We note, in passing, that this construction often needs T to be very large, so that the group G and the subgroups G_i are huge.

2.1.2 The Ingleton Inequality

As mentioned earlier, entropy satisfies submodularity and is connected to the notion of matroids. A matroid is defined by a ground set S and a rank function r (written as $r(\{\cdot\})$ or $r_{\{\cdot\}}$) defined over subsets of S, that satisfy the following axioms:

1) r is always a non-negative integer, and $r(U) \leq |U|$, $\forall U \subseteq S$.

2) r is monotonic: if $U \subseteq W \subseteq S$, then $r(U) \leq r(W)$.

3) r is submodular.

Axioms 2) and 3), together with positiveness, are called the *Shannon inequalities* for a set function. A matroid is defined in a way to extend the notion of a collection of vectors (in some vector space) along with the usual definition of the rank. It is called *representable* if its ground set can be represented as a collection of vectors (defined over some finite field) along with the usual rank function. Determining whether a matroid is representable or not is, in general, an open problem.

In 1971 Ingleton showed that for $n = 4$, the rank function r of any representable matriod must satisfy the inequality [10]

$$r_{12} + r_{13} + r_{14} + r_{23} + r_{24} \geq r_1 + r_2 + r_{34} + r_{123} + r_{124}$$

(where for simplicity we write r_{ij} and r_{ijk} for $r_{\{i,j\}}$ and $r_{\{i,j,k\}}$, respectively). In fact, these *Ingleton inequalities*, together with the Shannon inequalities and their combinations, are the only inequalities that the rank function of a representable matroid needs to satisfy (which are called linear rank inequalities) when $n = 4$ (see [11]). Furthermore, [11] shows that the rank function of any representable matroid is necessarily an entropy vector, but not every linear rank inequality is respected by a general entropy vector. For example, there are entropy vectors that violate the Ingleton inequality (e.g., [11, 12]), so that entropy is generally not a representable matroid. Using non-representable matroids, [13] constructs network coding problems that cannot be solved by linear network codes (since linear network codes are, by definition, representable).

When $n \geq 5$, there are many more linear rank inequalities besides the Shannon ones. But since the focus of this paper is the simplest case where $n = 4$ with only one such inequality, we refer the interested readers to the works of Kinser [14], Dougherty *et al.* [15–17], and Chan *et al.* [18] for recent development in this area.

From this point on we shall only study the Ingleton inequality, with $n = 4$. In the case of entropy vectors, it is written as

$$h_{12} + h_{13} + h_{14} + h_{23} + h_{24} \geq h_1 + h_2 + h_{34} + h_{123} + h_{124}. \tag{2.3}$$

The following sufficient condition is proposed in [11] for four general random variables X_1, X_2, X_3, and X_4 to satisfy (2.3):

Lemma 2.1.1. *If there exists a random variable Z that is a* common information *for X_1 and X_2, i.e., $H(Z|X_1) = H(Z|X_2) = 0$ while $H(Z) = I(X_1; X_2)$, then* (2.3) *is satisfied.*

In general, common information does not exist for two arbitrary random variables, but when the entropies correspond to ranks of vector subspaces, their common information does exist [11] and that is why representable matroids respect Ingleton. In Section 2.3 we will prove a similar condition for groups to satisfy Ingleton by constructing a common information.

As $\overline{\Gamma_n^*} = \overline{\text{cone}(\Upsilon_n)}$, we know there must exist finite groups and corresponding subgroups, such that their induced group-characterizable vectors violate the Ingleton inequality. In [19] it was shown that abelian groups cannot violate the Ingleton inequality, thereby giving an alternative proof as to why linear network codes (and even the more general abelian group network codes (defined below)) cannot achieve capacity on arbitrary networks, as the underlying groups for linear network codes are abelian. So we need to focus on non-abelian groups and their connections to nonlinear codes. Note that in the context of finite groups, the Ingleton inequality can be rewritten as

$$|G_1||G_2||G_{34}||G_{123}||G_{124}| \geq |G_{12}||G_{13}||G_{14}||G_{23}||G_{24}|. \tag{2.4}$$

2.1.3 Discussion

Since we know of distributions whose entropy vector violates the Ingleton inequality, we can, in principle, construct finite groups whose group-characterizable vectors violate Ingleton. Two such distributions are Example 1 in [12], where the underlying distribution is uniform over 7 points and the random variables correspond to different partitions of these seven points, and the example on page 1445 of [20], constructed from finite projective geometry and where the underlying distribution is uniform over $12 \times 13 = 156$ points. Unfortunately, constructing groups and subgroups for these distributions using the recipe of Section 2.1.1 results in $T = 29 \times 7 = 203$ and $T = 23 \times 156 = 3588$, which results in groups of size 203! and 3588!, which are too huge to give us any insight whatsoever.

These discussions lead us to the following questions.

1) Could the connection between entropy and groups be a red herring? Are the interesting groups too large to give any insight into the problem (e.g., the conditions for the Ingleton inequality to be violated)?
2) What is the smallest group with subgroups that violates the Ingleton inequality? Does it have any special structure?
3) Can one construct good network codes from such Ingleton-violating groups?

In this work we address the first two questions, and try to lay some groundwork for answering the third. We identify the smallest group that violates the Ingleton inequality—i.e., the symmetric group S_5, with 120 elements. Through a thorough investigation of the structure of its subgroups we conclude that it belongs to the family of groups $PGL(2, q)$, with $q \geq 5$ being a power of a prime. ($PGL(2, 5)$ is isomorphic to S_5.) We therefore believe that the connection to groups is not a red herring and that there may be some benefit to it.

Having a "recipe" for Ingleton violations, we generalize the family in two directions. Since $PGL(2, q)$ is the quotient group of $GL(2, q)$ modulo the scalar matrices, we explore the subgroups in $GL(2, q)$ and discover several new families of Ingleton violations. On the other hand, the projective general linear group $PGL(n, q)$ can be

viewed as the image of a permutation representation induced by the action of the general linear group $GL(n, q)$ on its projective geometry. It turns out that in this context, the Ingleton-violating subgroups of the family $PGL(2, q)$ all have nice interpretations: each of them is the stabilizer for a set of points in the projective geometry. Based on this viewpoint we obtain a few new families of Ingleton violations, including the groups $PGL(n, q)$ and $GL(n, q)$, and further give an abstract construction in general 2-transitive groups.

We can use these Ingleton-violating groups to contruct network codes, which have the potential of performing better than linear network codes, since the former can violate the Ingleton inequality while the latter cannot. We defer the detailed discussion for group network codes to Chapter 3.

Before we proceed to present the details of our results, we would like to mention some recent developments after our first paper [21] on this subject. In [22], Boston and Nan mainly study symmetric groups and discover many new Ingleton violations in the related groups. Furthermore, using the same group action theoretic approach as above (specifically, designing the subgroups to be the stabilizers of certain sets of points[2]), they systematically construct subgroups of a symmetric group to violate Ingleton. Many of these new violations are quite effective (see Section 2.5.3 for more discussions). Also, while all the Ingleton-violating groups in this work are non-solvable, [22] shows that there do exist solvable groups that violate Ingleton. Paajanen [23], however, focuses on the subclasses p-groups and nilpotent groups and shows that with some technical conditions they satisfy Ingleton. Recall that we have the hierarchy of finite groups

$$\textbf{Cyclic groups} \subset \textbf{Abelian groups} \subset \textbf{Nilpotent groups}$$
$$\subset \textbf{Solvable groups} \subset \textbf{All groups}$$

and that every nilpotent group is a direct product of groups, each of which is a p-

[2]In fact, in the original paper of Chan and Yeung [9] the same type of subgroups are also used to show that every entropy vector can be approximated by a scaled group-characterizable vector.

group for a distinct p. Now, roughly speaking we have a guideline for what class of groups one needs to explore to violate Ingleton. For linear rank inequalities in higher dimensions, [24] considers the case $n = 5$ and obtains some results on the groups that satisfy/violate some of these inequalities.

2.2 Notation

We use the following abstract algebra notations, in this and the following chapters. These are fairly standard (and follow Dummitt and Foote [25]). The interested reader who may not be familiar with all the concepts below may refer to [25], or any other standard abstract algebra textbook.

$\|G\|$	the order (cardinality) of the set/group G.
$\|g\|$	the order of element g = the smallest positive integer m s.t. $g^m = 1$.
x^g	the conjugate of element x by element g: $x^g = g^{-1}xg$. (No confusion with the powers of x as g is an element of G.)
X^g	the conjugate of subset X by element g: $X^g = \{x^g : x \in X\}$.
$G \cong H$	the group G is isomorphic to the group H.
$H \leq G$	H is a subgroup of G.
$H \trianglelefteq G$	H is a normal subgroup of G, i.e., $H^g = H$, $\forall g \in G$.
gH	the left coset of the subgroup H in G with representative g.
G/H	the set of all left cosets of subgroup H in G. When $H \trianglelefteq G$, G/H is a group, called the factor group or quotient group.
HK or $H \cdot K$	the "set product" of $H, K \subseteq G$: $HK = \{hk : h \in H, k \in K\}$.
$H \times K$	the direct product of groups H and K. The elements are the pairs $\{(h,k) : h \in H, k \in K\}$ and $(h_1,k_1)(h_2,k_2) = (h_1h_2, k_1k_2)$.
G^n	the direct product of n copies of the group G.
$H \rtimes K$	the semidirect product of groups H and K. The elements are the same as $H \times K$, but $(h_1,k_1)\cdot(h_2,k_2) = (h_1 \cdot \varphi(k_1)(h_2), k_1k_2)$

	where φ is a homomorphism of K into the automorphism group of H.
$\langle g_1, \ldots, g_m \rangle$, $\langle S \rangle$	the group generated by the elements $g_1, \ldots, g_m$, and by set S.
$G = \langle S \mid R \rangle$	$\langle S \mid R \rangle$ is a presentation of G. S is a set of generators of G, while R is a set of relations G should satisfy. See Definition 2.4.1.
1	the natural number "1", identity element of a group, or the trivial group. The meaning should be clear in different contexts with no confusion.
$\mathbb{Z}_n$	the integers modulo $n \cong$ the cyclic group of order n.
S_n	the symmetric group of degree n, consisting of all permutations on n points.
D_{2n}	the dihedral group of order $2n$.
$\mathbb{F}_q$	the finite field of q elements.
$\mathbb{Z}_n^\times$, $\mathbb{F}_q^\times$	the multiplicative group of units of $\mathbb{Z}_n$, and of $\mathbb{F}_q$, both consisting of all invertible elements under multiplication. $\mathbb{F}_q^\times =$ all nonzero elements of $\mathbb{F}_q$.
$GL(n, q)$	the general linear group of degree n on $\mathbb{F}_q$, which consists of all invertible $n \times n$ matrices with entries from $\mathbb{F}_q$. The identity element for $GL(n, q)$ is usually denoted by $I =$ identity matrix. $\lvert GL(n, q)\rvert = (q^n - 1)(q^n - q)(q^n - q^2) \cdots (q^n - q^{n-1})$.
V_q	the center of $GL(n, q)$, consisting of the collection of matrices that commute with every matrix in $GL(n, q) =$ all nonzero scalar matrices $= \{\alpha I : \alpha \in \mathbb{F}_q^\times\}$.
$PGL(n, q)$	the projective general linear group $= GL(n, q)/V_q$. $\lvert PGL(n, q)\rvert = \lvert GL(n, q)\rvert / \lvert V_q\rvert = \lvert GL(n, q)\rvert/(q - 1)$. In other words, it is the group of all invertible $n \times n$ matrices with entries from $\mathbb{F}_q$, where matrices that are proportional are considered the same group element.
$SL(2, q)$	the special linear group = all matrices in $GL(2, q)$ with deter-

	minant 1. $\lvert SL(2,q)\rvert = \lvert PGL(2,q)\rvert$.
$PSL(2,q)$	the projective special linear group $= SL(2,q)/\langle -I\rangle$. $\lvert PSL(2,q)\rvert$ $= \lvert SL(2,q)\rvert/2$.
$f \circ g$	the composition of two mappings f and g.
$U \oplus V$	the direct sum of vector spaces U and V.

To simplify expressions in later sections, let $\mathcal{K}_n \triangleq \{0, 1, \ldots, n-2\}$ for integers $n \geq 2$.

2.3 Ingleton Violation: Computer Search and Some Conditions

Since the Ingleton inequality (2.4) involves four subgroups of a finite group and their various intersections, designing a small admissible structure is very difficult without an existing example. So we use computer programs to search for a small instance. Specifically, we use the GAP system [26] to search its "Small Group" library, which contains all finite groups of order less than or equal to 2000, except those of 1024. We pick a group in this library (starting from the smallest, of course), find all its subgroups, then test the Ingleton inequality for all 4-combinations of these subgroups. This is a tremendous task, as there are already more than 1000 groups (up to isomorphism) of order less than or equal to 100, each of which might have hundreds of subgroups, or even more.

It is therefore extremely critical to prune our search. In fact, we used the following conditions to exclude groups or subgroups in the search, each of which guarantees that Ingleton is satisfied.

Condition 2.3.1. *G is abelian. [19]*

Condition 2.3.2. *$G_i \trianglelefteq G$, $\forall i$. [27]*

Condition 2.3.3. *$G_1G_2 = G_2G_1$, or equivalently $G_1G_2 \leq G$.*

Condition 2.3.4. *$G_i = 1$ or G, for some i.*

Condition 2.3.5. *$G_i = G_j$ for some distinct i and j.*

Condition 2.3.6. *$G_{12} = 1$.*

Condition 2.3.7. *$G_i \leq G_j$ for some distinct i and j.*

Note that Condition 2.3.2 subsumes Condition 2.3.1, while Condition 2.3.3 subsumes Condition 2.3.2. Also Conditions 2.3.4 and 2.3.5 are contained in Condition 2.3.7. Nevertheless, we still list these more restrictive conditions as they are easier to check using computer programs. In addition, Conditions 2.3.1, 2.3.3, and 2.3.6 are crucial in our program, as they appear in the outer loops and can save a large amount of search work.

For the above reasons we only list the proofs for Conditions 2.3.3, 2.3.6, and 2.3.7 below:

Proof 2.3.3: Construct random variables X_i's from uniformly distributed Λ on G as in Section 2.1.1. As $G_{1;2} \triangleq G_1 G_2 \leq G$, we can similarly construct random variable $Z = \Lambda G_{1;2}$. In fact, Z is a common information for X_1 and X_2: since $|G_{1;2}| = |G_1||G_2|/|G_{12}|$,

$$H(Z) = H(X_1) + H(X_2) - H(X_1, X_2) = I(X_1; X_2).$$

Also, $H(Z|X_1) = H(Z|X_2) = 0$ as $G_1, G_2 \leq G_{1;2}$. Thus Ingleton is satisfied by Lemma 2.1.1. □

In the proof above we used the group-entropy correspondence in Section 2.1.1 to translate the problem to the entropy domain. Henceforth, in order to show that a group satisfies Ingleton, we shall either prove (2.4) directly, or equivalently prove (2.3) using this correspondence. Furthermore, observe that the Ingleton inequality has symmetries between subscripts 1 and 2 and between 3 and 4, i.e., if we interchange the subscripts 1 and 2, or 3 and 4, the inequality stays the same. Thus if we prove conditions for some $i \in \{1, 2\}$ and $j \in \{3, 4\}$, we automatically get conditions for all $(i, j) \in \{1, 2\} \times \{3, 4\}$. So without loss of generality, we will just prove conditions for the case $(i, j) = (1, 3)$ when these symmetries apply.

Proof 2.3.6: Realize that (2.3) can be rewritten as

$$\delta_{13,14} + \delta_{23,24} + \delta_{134,234} - \delta_{123,124} \geq 0, \tag{2.5}$$

where for $\emptyset \neq \alpha, \beta \subseteq \mathcal{N}$,

$$\delta_{\alpha,\beta} \triangleq h_\alpha + h_\beta - h_{\alpha\cap\beta} - h_{\alpha\cup\beta}.$$

For example, $\delta_{134,234} = h_{134} + h_{234} - h_{34} - h_{1234}$. By submodularity of entropies, all $\delta_{\alpha,\beta} \geq 0$. If $G_{12} = 1$, then $\delta_{123,124} = 0$ and (2.5) holds. □

Proof 2.3.7: $(i,j) = (1,2)$ implies Condition 2.3.3. $(i,j) = (1,3)$ implies $\delta_{123,124} = 0$ in (2.5). $(i,j) = (3,1)$ implies $\delta_{123,234} = 0$ and so $\delta_{123,234} \leq \delta_{12,24}$, which further transforms to $\delta_{123,124} \leq \delta_{23,24}$, thus (2.5) holds. For $(i,j) = (3,4)$, (2.4) becomes

$$|G_1||G_2||G_3||G_{123}||G_{124}| \geq |G_{12}||G_{13}||G_{14}||G_{23}||G_{24}|,$$

which is true as $G_2 \geq G_{24}$ and by submodularity, $|G_1||G_{124}| \geq |G_{12}||G_{14}|$ and $|G_3||G_{123}| \geq |G_{13}||G_{23}|$. □

2.4 The Smallest Violation Instance and the Group Presentation

Using GAP we found that the smallest group that violates Ingleton is $G = S_5$, which has 60 sets of violating subgroups up to subscript symmetries. Further examination shows that these 60 sets of subgroups are in fact all conjugates of each other, and are thus virtually the same in terms of group structure. We list below some

information from GAP about one representative:[3]

$$
\begin{array}{lll}
G_1 = \big\langle (3,4,5), (1,2)(4,5) \big\rangle & \cong S_3 \cong D_6 & |G_1| = 6 \\
G_2 = \big\langle (1,2,3,4,5), (1,4,3,5) \big\rangle & \cong \mathbb{Z}_5 \rtimes \mathbb{Z}_4 & |G_2| = 20 \\
G_3 = \big\langle (2,3), (1,3,4,2) \big\rangle & \cong D_8 & |G_3| = 8 \\
G_4 = \big\langle (2,4), (1,2,5,4) \big\rangle & \cong D_8 & |G_4| = 8 \\
G_{12} = \big\langle (1,2)(3,5) \big\rangle & \cong \mathbb{Z}_2 & |G_{12}| = 2 \\
G_{13} = \big\langle (1,2)(3,4) \big\rangle & \cong \mathbb{Z}_2 & |G_{13}| = 2 \\
G_{14} = \big\langle (1,2)(4,5) \big\rangle & \cong \mathbb{Z}_2 & |G_{14}| = 2 \\
G_{23} = \big\langle (1,3,4,2) \big\rangle & \cong \mathbb{Z}_4 & |G_{23}| = 4 \\
G_{24} = \big\langle (1,2,5,4) \big\rangle & \cong \mathbb{Z}_4 & |G_{24}| = 4 \\
G_{34} = 1 & & |G_{34}| = 1 \\
G_{123} = 1 & & |G_{123}| = 1 \\
G_{124} = 1 & & |G_{124}| = 1.
\end{array}
$$

Simple calculation shows that

$$|G_1||G_2||G_{34}||G_{123}||G_{124}| = 120 < 128 = |G_{12}||G_{13}||G_{14}||G_{23}||G_{24}|,$$

so Ingleton is violated. Also we can check that G_1–G_4 indeed generate G.

To illustrate the structure of these subgroups, we use the group cycle graph. See Figure 2.1, where the dash-dotted lines denote the pairwise intersections of subgroups excluding identity. From the cycle graph we can obtain more structural information that GAP does not show us directly. First, not only is G_2 a semidirect product of two cyclic groups $\big\langle (1,2,3,4,5) \big\rangle \cong \mathbb{Z}_5$ and $\big\langle (1,4,3,5) \big\rangle \cong \mathbb{Z}_4$, but also $\big(G_2 \setminus \big\langle (1,2,3,4,5) \big\rangle\big) \bigcup \{1\}$ is the union of subgroups which are all isomorphic to (in fact, conjugate to) $\big\langle (1,4,3,5) \big\rangle$ and have trivial pairwise intersections. We say G_2 has a "flower" structure in this case. Second, G_4 is the conjugate of G_3 by $(3,4,5)$.

[3]The permutations are written in cycle notation, e.g., $(1,2)(3,4,5)$ is the permutation on the set $\{1,2,3,4,5\}$ that makes the following mapping: $1 \mapsto 2,\ 2 \mapsto 1,\ 3 \mapsto 4,\ 4 \mapsto 5,\ 5 \mapsto 3$. Also GAP's convention for permutations is used throughout this paper, i.e., permutations are applied to an element from the right.

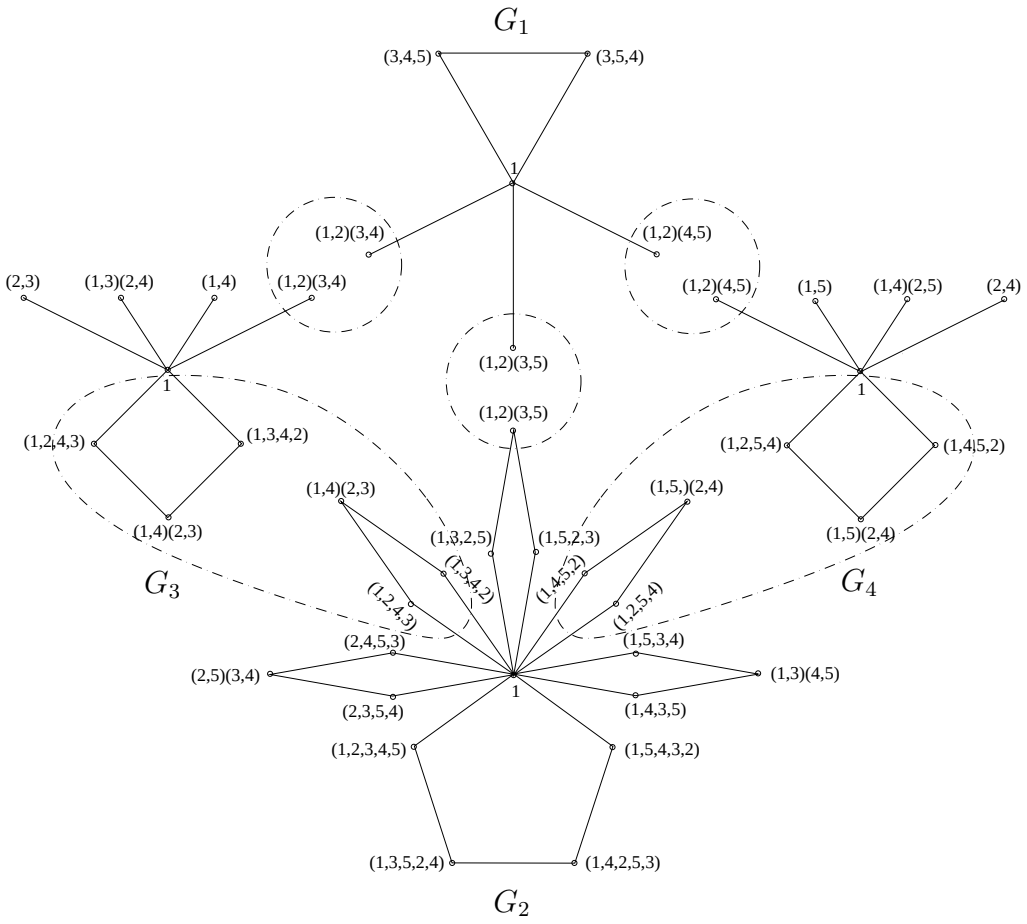


Figure 2.1: Cycle graph of the Ingleton violating subgroups of S_5

In particular, there is a conjugacy relation between the order-4 generators of G_3 and G_4: $(1,3,4,2)^{(3,4,5)} = (1,4,5,2) = (1,2,5,4)^{-1}$.

In order to generalize these subgroups to a family of violations, we seek a parameterized group presentation for G that retains the above structures. Although these group presentations are abstract, each of them can be input to GAP to yield an isomorphic concrete group, and Ingleton inequality can be checked against the corresponding subgroups. Observing that $|G_{23}|$ and $|G_{24}|$ (both equal to 4) contribute most to the right-hand side (RHS) of (2.4), we may try to let the "petals" of G_2 (conjugates of $\big\langle(1,4,3,5)\big\rangle$) grow while keeping other structures fixed.[4] In the rest of this section, we start from a presentation of G_2 and then extend it to the whole

[4]This approach is a little conservative, but it is the only successful extension according to our GAP trials. For example, one may try to expand G_1 at the same time, but the structures of G_3 and G_4 usually collapse.

group G.

Let us first define a presentation of a group. For a precise definition one needs to introduce the concept of free groups, which we will skip. The interested readers may consult abstract algebra textbooks, e.g., [25, 28]. Here we only give an informal but useful definition.

Definition 2.4.1 (Group Presentation)**.** A set S of *generators* of a group G is a subset of G, such that every group element can be written as a finite product of elements of S and their inverses. An equation satisfied in G involving only $S \cup \{1\}$ is called a *relation* in G among S. Let R be a set of such relations. We say G has a *presentation*

$$\langle\, S \mid R \,\rangle$$

if G is the largest ("freest") group generated by S subject only to the relations R. (Formally, the group G is said to have the above presentation if it is isomorphic to the quotient of a free group F on S by the normal subgroup of F generated by the relations R.)

For example, consider a presentation $\langle x \mid x^n = 1\rangle$. Any group generated by x contains only the powers of x, but by the relation $x^n = 1$ the order of such a group cannot exceed n. Among these groups the cyclic group $\mathbb{Z}_n$ has the maximum order, and hence has the above group presentation.

2.4.1 Presentation of G_2

Let G_2 be generated by two elements a and b, with a normal subgroup $N = \langle a\rangle \cong \mathbb{Z}_n$ and another subgroup $H = \langle b\rangle \cong \mathbb{Z}_m$, for some integers m, n. This gives us a presentation

$$G_2 = \left\langle\, a, b \,\middle|\, a^n = b^m = 1,\, a^b = a^s \,\right\rangle \tag{2.6}$$

for some $0 < s < n$. In order to violate Ingleton as much as possible, we may wish for n to be small while m is large. However, the flower structure of G_2 may limit the choices of n and m. First of all, for this presentation to be a semidirect product, we

need $s^m \equiv 1 \pmod{n}$ (see [28, Sec 5.4]), i.e.,

$$s \in \mathbb{Z}_n^{\times}, \quad |s| \,\Big|\, m, \tag{2.7}$$

where $|s|$ denotes the order of s in the multiplicative group $\mathbb{Z}_n^{\times}$. As a consequence, $|G_2| = mn$, $H \bigcap N = 1$, and by the relations in (2.6) we also have

$$(a^i)^{b^k} = a^{is^k}, \quad \forall i, k \in \mathbb{Z}. \tag{2.8}$$

Moreover, we need $(G_2 \setminus N) \bigcup \{1\}$ to be the union of subgroups which are all isomorphic to H with trivial pairwise intersections.

One possible way to achieve this is to restrict $H^{g_1} \bigcap H^{g_2} = 1, \ \forall g_1 \neq g_2 \in N$, as in our original example. This is equivalent to $H^g \bigcap H = 1, \forall g \in N \setminus \{1\}$. If this is the case, then there will be $|N| = n$ "petals" of size m in G_2, and the total number of nonidentity elements will equal $n(m-1) = nm - n = |G_2 \setminus N|$, and then indeed the flower structure would be achieved. Pick two nonidentity elements $h_1 = b^l \in H$, $h_2 = (b^k)^{a^i} \in H^{a^i}$ for some $0 < k, l < m$ and some $0 < i < n$. Then

$$h_1 = h_2 \ \Leftrightarrow \ a^{-i}b^k a^i = b^l \ \Leftrightarrow \ a^{-i}(a^i)^{b^{-k}}b^k = b^l \ \Leftrightarrow \ a^{-i}a^{is^{-k}} = b^{l-k} \ \Leftrightarrow \ a^{(s^{-k}-1)i} = b^{l-k}.$$

In the last equation, $LHS \in N$ and $RHS \in H$. But $H \bigcap N = 1$ forces that $a^{(s^{-k}-1)i} = b^{l-k} = 1$, i.e., $l = k$ and $n \,\Big|\, (s^{-k}-1)i$.

To guarantee that $H^{a^i} \bigcap H = 1$, we must have $m \leq |s|$. Otherwise if we let $0 < k = |s| < m$, then $s^{-k} \equiv 1 \pmod{n}$ and so $n \,\Big|\, (s^{-k}-1)i$ is satisfied. This means that by choosing $k = l = |s|$, we have found a nonidentity element $h_2 = (b^k)^{a^i} = b^l = h_1$ in $H^{a^i} \bigcap H$. Therefore $m \leq |s|$ and as $|s| \,\Big|\, m$ by (2.7), $m = |s|$. In particular, $m \leq |\mathbb{Z}_n^{\times}| \leq n - 1$.

For m to be as large as possible, s should be a primitive root modulo n, which makes $m = |\mathbb{Z}_n^{\times}|$. Pick $n = p$ for some prime p, then $m = |\mathbb{Z}_p^{\times}| = p - 1$ achieves the upper bound $m \leq n-1$. Also in this case, if $0 < k < m = |s|$ and $0 < i < n = p$, then $n \,\Big|\, (s^{-k}-1)i$ requires $p \,\Big|\, i$ or $p \,\Big|\, (s^{-k}-1)$. Since $p > i$, the latter must be true, which

Table 2.1: Correspondence of group elements

a	b	c	b_1	b_3	b_4
$(1,2,3,4,5)$	$(1,4,3,5)$	$(3,4,5)$	$(1,2)(3,5)$	$(1,3,4,2)$	$(1,2,5,4)$

implies that $|s| \,\big|\, k$. But this is a contradiction since $0 < k < |s|$. So indeed we have $H^g \bigcap H = 1$, $\forall g \in N$, and the flower structure is realized. Furthermore, to make H nontrivial, we need $p > 2$. Thus with such a choice of parameters, the presentation of G_2 becomes

$$G_2 = \big\langle\, a, b \,\big|\, a^p = b^{p-1} = 1,\, a^b = a^s \,\big\rangle, \tag{2.9}$$

where $p > 2$ is a prime and s is a primitive root modulo p.

2.4.2 Presentation of G

The next step is to extend the presentation (2.9) to the whole group G generated by G_1–G_4, with the structure in Figure 2.1. Consider the dihedral groups G_3 and G_4. The subgroups of rotations are just H^{a_3} and H^{a_4}, respectively, for some $a_3 = a^{k_3}, a_4 = a^{k_4} \in N$. Also G_3 and G_4 each shares one element of reflection with the dihedral group G_1, while the remaining reflection of G_1 is just $(b^{\frac{p-1}{2}})^{a_1}$ in G_2, for some $a_1 = a^{k_1} \in N$. Thus if we can determine the generator of the subgroup of rotations of G_1, then all elements of G_1–G_4 are determined. In other words, if we introduce an element c as the generator of rotations of G_1, then all elements from G_1–G_4 can be expressed as products of a, b, c, and their inverses. Define

$$b_1 = (b^{\frac{p-1}{2}})^{a^{k_1}}, \quad b_3 = b^{a^{k_3}}, \quad b_4 = b^{a^{k_4}} \tag{2.10}$$

for some integers k_1, k_3, k_4. If in Figure 2.1 we let a, b, c, b_1, b_3, b_4 correspond with the elements specified in Table 2.1, then the subgroups and the whole group in our presentation should be

$$G_1 = \big\langle c, b_1 \big\rangle, \;\; G_2 = \big\langle a, b \big\rangle, \;\; G_3 = \big\langle b_1 c^2, b_3 \big\rangle, \;\; G_4 = \big\langle b_1 c, b_4 \big\rangle, \;\; G = \big\langle a, b, c \big\rangle. \tag{2.11}$$

As $G_1 \cong D_6$, we should have the relation

$$c^3 = (cb_1)^2 = 1.$$

Furthermore, for G_3 and G_4 to be dihedral groups, we need

$$(b_3 \cdot b_1c^2)^2 = (b_4 \cdot b_1c)^2 = 1.$$

At this point we can try to plug in the presentation with these relations to GAP to find a concrete group. But still there are too many possible parameter values to choose. Especially, when p is large, the choices of k_1, k_3, k_4 are numerous. Also, for a fixed p not many such combinations yield successful Ingleton violations, according to our GAP trials. Therefore we need to ultilize more structural information from Figure 2.1 to obtain more restrictions on k_1, k_3, and k_4.

Observe that in the original violation G_4 is the conjugate of G_3 by $(3, 4, 5)$, and $(1, 3, 4, 2)^{(3,4,5)} = (1, 2, 5, 4)^{-1}$. In our presentation this translates to $b_3^c = b_4^{-1}$, according to Table 2.1. With this new relation, we claim that $(b_3 \cdot b_1c^2)^2 = (b_4 \cdot b_1c)^2 = 1$ is satisfied if and only if

$$k_3 - k_1 \equiv k_1 - k_4 \pmod{p}.$$

In fact, as $|b_1| = 2$, $c^3 = (cb_1)^2 = 1$, we have $cb_1 = b_1c^2$ and $b_1c = c^2b_1$. Using these relations we can establish the following equalities:

$$(b_3 \cdot b_1c^2)^2 = b_3b_1c^{-1}b_3cb_1 = b_3b_1b_4^{-1}b_1,$$

$$(b_4 \cdot b_1c)^2 = b_4b_1cb_4c^{-1}b_1 = b_4b_1b_3^{-1}b_1 = \left((b_3b_1b_4^{-1}b_1)^{-1}\right)^{b_1}.$$

So $(b_3 \cdot b_1c^2)^2 = 1$ if and only if $(b_4 \cdot b_1c)^2 = 1$. Using (2.8) and the fact that $b^{\frac{p-1}{2}} = (b^{\frac{p-1}{2}})^{-1}$ and plugging (2.10) in, we have

$$\begin{aligned} b_3b_1b_4^{-1}b_1 &= b^{a^{k_3}}(b^{\frac{p-1}{2}})^{a^{k_1}}(b^{-1})^{a^{k_4}}(b^{\frac{p-1}{2}})^{a^{k_1}} \\ &= a^{-k_3}ba^{k_3-k_1}b^{\frac{p-1}{2}}a^{k_1-k_4}b^{-1}a^{k_4-k_1}b^{\frac{p-1}{2}}a^{k_1} \end{aligned}$$

$$
\begin{aligned}
&= a^{-k_3} \cdot ba^{k_3-k_1}b^{-1} \cdot b^{\frac{p-1}{2}} \cdot ba^{k_1-k_4}b^{-1} \cdot a^{k_4-k_1}b^{\frac{p-1}{2}}a^{k_1} \\
&= a^{-k_3} \cdot a^{(k_3-k_1)s^{-1}} \cdot b^{\frac{p-1}{2}} \cdot a^{(k_1-k_4)s^{-1}} \cdot a^{k_4-k_1}b^{\frac{p-1}{2}}a^{k_1} \\
&= a^{(k_3-k_1)s^{-1}-k_3} \cdot (b^{\frac{p-1}{2}})^{-1}a^{(k_1-k_4)(s^{-1}-1)}b^{\frac{p-1}{2}} \cdot a^{k_1} \\
&= a^{(k_3-k_1)s^{-1}-k_3} \cdot a^{(k_1-k_4)(s^{-1}-1)s^{(p-1)/2}} \cdot a^{k_1} \\
&= a^{[(k_3-k_1)+(k_1-k_4)s^{(p-1)/2}](s^{-1}-1)}.
\end{aligned}
$$

Since s is a primitive root modulo p, $|s^{(p-1)/2}| = 2$. As $\mathbb{Z}_p^\times$ is cyclic of an even order $p-1$, it is clear that there is a unique element of order 2. But -1 has order 2 in $\mathbb{Z}_p^\times$, so $s^{(p-1)/2} \equiv -1 \pmod{p}$ and

$$(b_3 \cdot b_1c^2)^2 = b_3b_1b_4^{-1}b_1 = a^{[(k_3-k_1)-(k_1-k_4)](s^{-1}-1)}.$$

Now $p \nmid (s^{-1}-1)$ as $s \neq 1$, which implies

$$(b_3 \cdot b_1c^2)^2 = 1 \;\Leftrightarrow\; p \,\big|\, [(k_3-k_1)-(k_1-k_4)] \;\Leftrightarrow\; k_3-k_1 \equiv k_1-k_4 \pmod{p}.$$

This condition on k_1, k_3, and k_4 tells us that the petals G_{23} and G_{24} of G_2 should be symmetric (modulo p) w.r.t. G_{12}, i.e., G_{23}, G_{12}, and G_{24} should be equally spaced.[5]

In sum, our analysis leads to the following presentation:

$$G = \left\langle\, a, b, c \,\middle|\, a^p = b^{p-1} = c^3 = 1,\, a^b = a^s,\, (cb_1)^2 = b_3^c b_4 = 1 \,\right\rangle, \tag{2.12}$$

where p is an odd prime and s is a primitive root modulo p, $k_3-k_1 \equiv k_1-k_4 \pmod{p}$. If our extension of the subgroup structures succeeds, then the orders of subgroups and intersections would be:

$$|G_1| = 6, \quad |G_2| = p(p-1), \quad |G_3| = |G_4| = 2(p-1),$$

$$|G_{12}| = |G_{13}| = |G_{14}| = 2, \quad |G_{23}| = |G_{24}| = p-1, \quad |G_{34}| = |G_{123}| = |G_{124}| = 1.$$

[5] With this symmetry it is very easy for GAP to produce the desired structures, even with arbitrary choices of k_1 and k_3.

Hence in (2.4) $LHS = 6p(p-1)$ while $RHS = 8(p-1)^2$, and so when $p \geq 5$, Ingleton should be violated.

2.5 Explicit Violation Construction with $PGL(2,p)$ and $PGL(2,q)$

Feeding the above presentation to GAP, we find that for $p = 5, 7, \ldots, 23$ the outcome is a finite group that violates the Ingleton inequality.[6] Moreover, with GAP we verified for the first few primes (up to $p = 11$) that this group is isomorphic to the projective general linear group $PGL(2,p)$. This leads us to conjecturing that $PGL(2,p)$ is a family of Ingleton-violating groups. In fact, with an explicit identification of the generators in (2.12) with matrices in $PGL(2,p)$, we prove that $PGL(2,p)$ is indeed a family of Ingleton-violating groups for primes $p \geq 5$ by directly constructing their violating subgroups in (2.11) in the form of matrices. These matrix subgroups all have clear interpretations. Furthermore, once we have the formats of these subroups, we extend them to the Ingleton-violating family $PGL(2,q)$ for all finite field order $q \geq 5$.

2.5.1 The Family $PGL(2,p)$

First we introduce some necessary notations. Let p be an odd prime. For $A \in GL(2,p)$, let $\overline{A}$ denote the left coset of A in $GL(2,p)$ with respect to the center $V_p = \{\alpha I : \alpha \in \mathbb{F}_p^\times\}$. Thus $\overline{A} = \overline{B}$ if and only if each entry of A is a nonzero constant multiple of the corresponding entry of B. A^T denotes the transpose of A as usual. We denote the elements of $\mathbb{F}_p$ by ordinary integers, but the addition and multiplication, as well as equality, are modulo p. Furthermore, $-k$ and k^{-1} denotes the additive and multiplicative inverses of k in $\mathbb{F}_p$, respectively. If $s \in \mathbb{F}_p$, and A has multiplicative order p, then A^s simply indicates the s-th power of A, where s is viewed as an integer.

[6] The capability of the testing program is primarily limited by hardware. When p is too large the program runs out of memory.

This would not cause any confusion, as we only use elements from $\mathbb{F}_p$ for the entries of matrices.

We start by identifying the generators in $PGL(2,p)$ that correspond to presentation (2.12). Consider the following matrices in $GL(2,p)$:

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & t \end{bmatrix}, \qquad C = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix},$$

where t is a primitive root modulo p, i.e., a generator of $\mathbb{F}_p^\times$. Our guess is that $\overline{A}, \overline{B}, \overline{C}$ correspond to the generators a, b, c in (2.12), respectively. The powers of these matrices are:

$$A^k = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}, \quad B^k = \begin{bmatrix} 1 & 0 \\ 0 & t^k \end{bmatrix}, \quad C^2 = \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix}, \quad C^3 = I$$

for any integer k. Thus $\left|\overline{A}\right| = p$, $\left|\overline{B}\right| = p-1$, and $\left|\overline{C}\right| = 3$. Also,

$$A^B = B^{-1}AB = \begin{bmatrix} 1 & 0 \\ t^{-1} & 1 \end{bmatrix} = A^s,$$

where $s = t^{-1}$ is also a primitive root modulo p. So $\overline{A}^{\overline{B}} = \overline{A}^s$. Next we let

$$B_1 = (B^{\frac{p-1}{2}})^{A^{k_1}} = \begin{bmatrix} 1 & 0 \\ -k_1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ k_1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -2k_1 & -1 \end{bmatrix},$$

where we calculated $t^{\frac{p-1}{2}} = -1$, as it is the unique element of order 2 in $\mathbb{F}_p^\times$. Now check

$$CB_1 = \begin{bmatrix} -2k_1 & -1 \\ 2k_1 - 1 & 1 \end{bmatrix}, \quad (CB_1)^2 = \begin{bmatrix} 4k_1^2 - 2k_1 + 1 & 2k_1 - 1 \\ -(2k_1-1)^2 & 2 - 2k_1 \end{bmatrix}.$$

Thus if we want $\left(\overline{CB_1}\right)^2 = \overline{I}$, k_1 must be $2^{-1} = \frac{p+1}{2}$. In this case,

$$B_1 = \begin{bmatrix} 1 & 0 \\ -1 & -1 \end{bmatrix}, \quad CB_1 = \begin{bmatrix} -1 & -1 \\ 0 & 1 \end{bmatrix}, \quad \left(\overline{CB_1}\right)^2 = \overline{I}.$$

Let $B_3 = B^{A^{k_3}}$, $B_4 = B^{A^{k_4}}$. As $k_3 - k_1 = k_1 - k_4$, we have $k_3 = 1 - k_4$.

$$B^{A^k} = \begin{bmatrix} 1 & 0 \\ k(t-1) & t \end{bmatrix}, \quad B_3 C \cdot B_4 = \begin{bmatrix} 0 & 1 \\ -t & k_3(t-1) - t \end{bmatrix} \begin{bmatrix} 1 & 0 \\ k_4(t-1) & t \end{bmatrix},$$

whose $(1,1)$-entry is $k_4(t-1)$. If we want $\overline{B_3}\,\overline{C}\,\overline{B_4} = \overline{I}$, i.e., $\overline{B_3 C B_4} = \overline{C}$, k_4 must be 0, since the $(1,1)$-entry of C is 0 and $t \neq 1$. So $k_3 = 1 - k_4 = 1$,

$$B_3 = \begin{bmatrix} 1 & 0 \\ t-1 & t \end{bmatrix}, \quad B_4 = \begin{bmatrix} 1 & 0 \\ 0 & t \end{bmatrix} = B,$$

$$\overline{B_3 C B_4} = \overline{\begin{bmatrix} 0 & 1 \\ -t & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & t \end{bmatrix}} = \overline{\begin{bmatrix} 0 & t \\ -t & -t \end{bmatrix}} = \overline{C}.$$

So far for $\overline{A}, \overline{B}, \overline{C}$ we have verified all the relations in (2.12). We can also prove that they are actually a set of generators for $PGL(2,p)$. Observe that each matrix in $GL(2,p)$ can be written as a product of some elementary matrices, which are

$$\begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & \beta \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & t^i \end{bmatrix}, \quad \begin{bmatrix} t^j & 0 \\ 0 & 1 \end{bmatrix},$$

where $\alpha, \beta \in \mathbb{F}_p$ and $i, j \in \mathcal{K}_p$. They are generated by A, A^T, B and $t^{-1}B$, respectively. So $PGL(2,p)$ is generated by $\overline{A}, \overline{A^T}$ and $\overline{B}$. Now check

$$B_1 C = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad A^{B_1 C} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = A^T.$$

Thus $\overline{A}, \overline{B}$, and $\overline{C}$ generate $PGL(2,p)$, and hence setting

$$s = t^{-1}, \quad k_1 = \frac{p+1}{2}, \quad k_3 = 1, \quad k_4 = 0,$$

we see that $PGL(2,p)$ is a quotient of the group G in (2.12), whose generators $\overline{A}, \overline{B}$, and $\overline{C}$ correspond precisely to the generators a, b, and c of G.

Remark 2.5.1. Note that we have not proved that (2.12) is a presentation of $PGL(2,p)$. To do that, one must show that the order of the group generated by a, b, c in (2.12) is no more than $|PGL(2,p)| = (p-1)p(p+1)$, which we have not yet been able to prove. However, identifying possible corresponding generators still gives us a way to explicitly construct the subgroups to violate Ingleton.

Now we can write out the subgroups in $PGL(2,p)$ corresponding to subgroups in (2.11).

$G_1 = \langle \overline{C}, \overline{B_1} \rangle$. Note that $\left|\overline{C}\right| = 3$, $\left|\overline{B_1}\right| = 2$, and $\left(\overline{CB_1}\right)^2 = \overline{I}$, so $\overline{CB_1} = \overline{B_1}\left(\overline{C}\right)^2$ and G_1 has at most 6 elements $\left\{ \left(\overline{B_1}\right)^i \left(\overline{C}\right)^j : 0 \le i < 2,\ 0 \le j < 3 \right\}$. Calculating these elements we can see that $|G_1| = 6$ exactly and thus indeed $G_1 \cong D_6 \cong S_3$:

$$G_1 = \left\{ \overline{I}, \quad \overline{\begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}}, \quad \overline{\begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix}}, \quad \overline{\begin{bmatrix} 1 & 0 \\ -1 & -1 \end{bmatrix}}, \quad \overline{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}, \quad \overline{\begin{bmatrix} -1 & -1 \\ 0 & 1 \end{bmatrix}} \right\}.$$

$G_2 = \langle \overline{A}, \overline{B} \rangle$. We claim that G_2 is the subgroup of lower triangular matrices[7] in $GL(2,p)$ modulo V_p, i.e.,

$$G_2 = \left\{ \overline{\begin{bmatrix} 1 & 0 \\ \alpha & \beta \end{bmatrix}} \;\middle|\; \alpha \in \mathbb{F}_p,\ \beta \in \mathbb{F}_p^\times \right\}.$$

As A, B are lower triangular, any element in G_2 is a lower triangular matrix modulo

[7]We would end up with upper triangular matrices for G_2 if A^T were used in place of A, but the two resulting groups are actually conjugate to each other, e.g., consider conjugating by B_1C.

V_p. On the other hand, $\forall \alpha \in \mathbb{F}_p,\ \beta \in \mathbb{F}_p^{\times}$, then $\beta = t^l$ for some integer l. So

$$\begin{bmatrix} 1 & 0 \\ \alpha & \beta \end{bmatrix} = A^{\alpha} B^{l} \Rightarrow \overline{\begin{bmatrix} 1 & 0 \\ \alpha & \beta \end{bmatrix}} = \overline{A}^{\alpha} \overline{B}^{l} \in G_2.$$

Thus $|G_2| = p(p-1)$ and G_2 has presentation (2.9). Therefore, as proved in Section 2.4.1, $G_2 \cong \mathbb{Z}_p \rtimes \mathbb{Z}_{p-1}$ and it achieves the desired flower structure.

$G_3 = \langle \overline{B_1}(\overline{C})^2, \overline{B_3} \rangle = \langle \overline{CB_1}, \overline{B_3} \rangle$. Note that $|\overline{CB_1}| = 2$, $|\overline{B_3}| = |\overline{B}| = p-1$, also

$$B_3^k = \begin{bmatrix} 1 & 0 \\ t^k - 1 & t^k \end{bmatrix}, \quad B_3^{-1} = \begin{bmatrix} 1 & 0 \\ t^{-1} - 1 & t^{-1} \end{bmatrix},$$

$$\overline{B_3} \cdot \overline{CB_1} = \overline{\begin{bmatrix} -1 & -1 \\ 1-t & 1 \end{bmatrix}} = \overline{\begin{bmatrix} -t^{-1} & -t^{-1} \\ t^{-1} - 1 & t^{-1} \end{bmatrix}} = \overline{CB_1}(\overline{B_3})^{-1},$$

so G_3 has at most $2(p-1)$ elements $\left\{ (\overline{CB_1})^i (\overline{B_3})^j : 0 \le i < 2,\ 0 \le j < p-1 \right\}$. Calculating these elements we can see that $|G_3| = 2(p-1)$ exactly and so $G_3 \cong D_{2(p-1)}$:

$$G_3 = \left\{ (\overline{B_3})^k = \overline{\begin{bmatrix} 1 & 0 \\ t^k - 1 & t^k \end{bmatrix}}, \quad \overline{CB_1}(\overline{B_3})^k = \overline{\begin{bmatrix} -1 & -1 \\ 1 - t^{-k} & 1 \end{bmatrix}} \;\middle|\; k \in \mathcal{K}_p \right\}.$$

$G_4 = \langle \overline{B_1 C}, \overline{B_4} \rangle$. Note that $|\overline{B_1 C}| = 2$, $|\overline{B_4}| = |\overline{B}| = p-1$. Moreover,

$$\overline{B_4} \cdot \overline{B_1 C} = \overline{\begin{bmatrix} 0 & 1 \\ t & 0 \end{bmatrix}} = \overline{\begin{bmatrix} 0 & t^{-1} \\ 1 & 0 \end{bmatrix}} = \overline{B_1 C}(\overline{B_4})^{-1},$$

so G_4 has at most $2(p-1)$ elements $\left\{ (\overline{B_1 C})^i (\overline{B_4})^j : 0 \le i < 2,\ 0 \le j < p-1 \right\}$. Calculating these elements we can see that $|G_4| = 2(p-1)$ exactly and so $G_4 \cong D_{2(p-1)}$:

$$G_4 = \left\{ (\overline{B_4})^k = \overline{\begin{bmatrix} 1 & 0 \\ 0 & t^k \end{bmatrix}}, \quad \overline{B_1 C}(\overline{B_4})^k = \overline{\begin{bmatrix} 0 & t^k \\ 1 & 0 \end{bmatrix}} \;\middle|\; k \in \mathcal{K}_p \right\}.$$

These are all diagonal and anti-diagonal matrices in $GL(2,p)$ modulo V_p. Note that we have already verified that $\left(\overline{B_3}\right)^{\overline{C}} = \overline{B_4}^{-1}$, and also that $\left(\overline{CB_1}\right)^{\overline{C}} = \overline{B_1 C}$, and thus indeed $G_4 = G_3^{\overline{C}}$ as in the original instance (Figure 2.1).

With all four subgroups explicitly written, we can easily write down the intersections:

$$G_{12} = \left\langle \overline{B_1} \right\rangle = \left\{ \overline{I}, \quad \overline{\begin{bmatrix} 1 & 0 \\ -1 & -1 \end{bmatrix}} \right\} \cong \mathbb{Z}_2,$$

$$G_{13} = \left\langle \overline{CB_1} \right\rangle = \left\{ \overline{I}, \quad \overline{\begin{bmatrix} -1 & -1 \\ 0 & 1 \end{bmatrix}} \right\} \cong \mathbb{Z}_2,$$

$$G_{14} = \left\langle \overline{B_1 C} \right\rangle = \left\{ \overline{I}, \quad \overline{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}} \right\} \cong \mathbb{Z}_2,$$

$$G_{23} = \left\langle \overline{B_3} \right\rangle = \left\{ \left. \overline{\begin{bmatrix} 1 & 0 \\ t^k - 1 & t^k \end{bmatrix}} \right| k \in \mathcal{K}_p \right\} \cong \mathbb{Z}_{p-1},$$

$$G_{24} = \left\langle \overline{B_4} \right\rangle = \left\{ \left. \overline{\begin{bmatrix} 1 & 0 \\ 0 & t^k \end{bmatrix}} \right| k \in \mathcal{K}_p \right\} \cong \mathbb{Z}_{p-1},$$

$$G_{34} = G_{123} = G_{124} = 1.$$

$$|G_{12}| = |G_{13}| = |G_{14}| = 2,$$

$$|G_{23}| = |G_{24}| = p - 1.$$

So in (2.4), indeed

$$LHS = |G_1||G_2||G_{34}||G_{123}||G_{124}| = 6p(p-1),$$

$$RHS = |G_{12}||G_{13}||G_{14}||G_{23}||G_{24}| = 8(p-1)^2,$$

$$LHS - RHS = 2(p-1)(4-p).$$

Thus Ingleton is violated when $p \geq 5$, and the subgroup structures of $S_5 \cong PGL(2,5)$ are exactly reproduced.

2.5.2 The Family $PGL(2,q)$

With the explicit matrix forms of the Ingleton-violating subgroups, we can further extend the above violation to $PGL(2,q)$ for all finite field order $q \geq 5$. For a finite field $\mathbb{F}_q$, we know that $q = p^m$ for some prime p (the characteristic of $\mathbb{F}_q$) and some integer m. Since $\mathbb{F}_p$ is the prime subfield of $\mathbb{F}_q$, $GL(2,p)$ is a subgroup of $GL(2,q)$, which induces an isomorphic copy of $PGL(2,p)$ as a subgroup of $PGL(2,q)$. Therefore, using the same subgroups of $PGL(2,p)$ as in the previous section, we obtain a trivial Ingleton violation in $PGL(2,q)$ whenever the characteristic $p \geq 5$. Nevertheless, by extending the interpretations of these subgroups to $PGL(2,q)$, we can obtain a more general (nontrivial) violation for each finite field order $q \geq 5$.

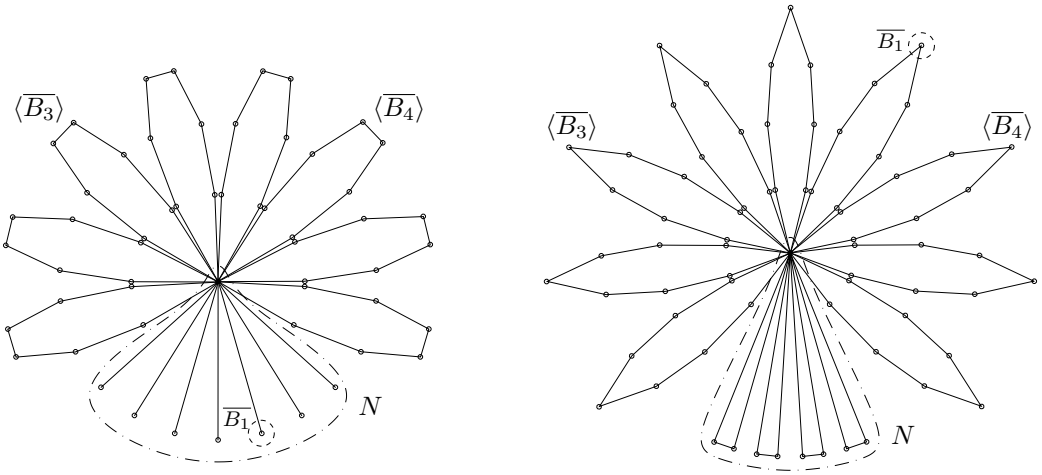
In the field $\mathbb{F}_q$, we continue to use the ordinary integers with modular arithmetic to represent the prime subfield $\mathbb{F}_p$. With this convention, all the matrices and subgroups in Section 2.5.1 are well defined[8], although now the cosets are taken with respect to V_q rather than V_p. These subgroups constitute a trivial embedding of our previous violation in $PGL(2,q)$. However, in $PGL(2,q)$, the previous sets of generators do not guarantee that G_2 is the full subgroup of all lower triangular matrices, nor that G_4 contains all the diagonal and anti-diagonal matrices.

To preserve these interpretations of the subgroups, we need to make some adjustment to the generators of G_2. Redefine t to be a primitive element of $\mathbb{F}_q$, i.e., t generates $\mathbb{F}_q^\times$. Then $\left|\overline{B}\right| = q-1$. Also, instead of a single A, we need to introduce more matrices to generate the subgroup $N \triangleq \left\{ \overline{A_\alpha} \,\middle|\, \alpha \in \mathbb{F}_q \right\}$, where for each $\alpha \in \mathbb{F}_q$ we define

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

Clearly $A_\alpha A_\beta = A_{\alpha+\beta}$, and $A_\alpha^k = A_{k\alpha}$ for each integer k. Thus $\left|\overline{A_\alpha}\right| = p$ for each

[8]The only problem that may arise is that when $p = 2$, $B_1 = (B^{\frac{p-1}{2}})^{A^{k_1}}$ is not well defined. But we can circumvent that by directly working with the final matrix form of B_1.



G_2 in $PGL(2,8)$ G_2 in $PGL(2,9)$

Figure 2.2: Generalized flower structures

$\alpha \in \mathbb{F}_q^{\times}$. Observe that $\mathbb{F}_q$ is an m-dimensional vector space over $\mathbb{F}_p$, so we pick a basis $(\xi_1, \xi_2, \ldots, \xi_m)$. Then for all $\alpha \in \mathbb{F}_q$, $\alpha = \sum_{i=1}^{m} k_i \xi_i$ for some $k_1, k_2, \ldots, k_m \in \mathbb{F}_p$ and $A_\alpha = \prod_{i=1}^{m} A_{\xi_i}^{k_i}$. Also $\langle\overline{A_{\xi_i}}\rangle \bigcap \langle\overline{A_{\xi_j}}\rangle = 1$ for distinct i and j. Thus

$$N = \langle\ \overline{A_{\xi_1}},\ \overline{A_{\xi_2}},\ \ldots,\ \overline{A_{\xi_m}}\ \rangle \cong \langle\ \overline{A_{\xi_1}}\ \rangle \times \langle\ \overline{A_{\xi_2}}\ \rangle \times \ldots \times \langle\ \overline{A_{\xi_m}}\ \rangle \cong \mathbb{Z}_p^m.$$

Actually, N is isomorphic to the additive group of the vector space $\mathbb{F}_q$ over $\mathbb{F}_p$ (cf. Section 3.2.1).

Let $G_2 = \langle\overline{A_{\xi_1}}, \overline{A_{\xi_2}}, \ldots, \overline{A_{\xi_m}}, \overline{B}\rangle = \langle N, \overline{B}\rangle$. Similar to the previous section, it is easy to show that now G_2 is indeed the subgroup of all lower triangular matrices modulo V_q. Furthermore, for any $\alpha \in \mathbb{F}_q$, we have $\overline{A_\alpha}^{\overline{B}} = \overline{A_{t^{-1}\alpha}}$, so $N \trianglelefteq G_2$ and $G_2 = NH$, where $H \triangleq \langle\overline{B}\rangle$. Also $N \bigcap H = 1$, and thus $G_2 \cong N \rtimes H \cong \mathbb{Z}_p^m \rtimes \mathbb{Z}_{q-1}$. Although in general G_2 does not have presentation (2.6) or (2.9) anymore, since N is not necessarily cyclic, we can prove that it does have a "generalized flower structure" when $q > 2$, i.e., $(G_2 \setminus N) \bigcup \{\overline{I}\}$ is the union of subgroups which are all isomorphic to H with trivial pairwise intersections. Similar to the analysis of the G_2 in Section 2.4.1, it suffices to show that $H^{\overline{A_\alpha}} \bigcap H = 1$, $\forall \overline{A_\alpha} \in N \setminus \{\overline{I}\}$. But this is true since for each

$\alpha \in \mathbb{F}_q^\times$ and some integers $k, l \in \mathcal{K}_q$,

$$(\overline{B}^k)^{\overline{A_\alpha}} = \overline{B}^l \iff \overline{B}^k \cdot \overline{A_\alpha} = \overline{A_\alpha} \cdot \overline{B}^l \iff \overline{\begin{bmatrix} 1 & 0 \\ t^k\alpha & t^k \end{bmatrix}} = \overline{\begin{bmatrix} 1 & 0 \\ \alpha & t^l \end{bmatrix}} \iff k = l = 0.$$

Figure 2.2 shows two representative generalized flower structures of G_2, for $q = 8$ and $q = 9$. The center point of each cycle graph denotes the identity element. For each G_2, there are $|N| = q$ petals and one "root system" (encircled by the dash-dotted line), which is the normal subgroup N. Every petal is a conjugate of H and has size $q - 1$. Since N has $q - 1$ nonidentity elements that each has order p, the root system consists of $(q - 1)/(p - 1)$ trivially intersecting "roots/tubers", each of which is a p-cycle. Note that when $m = 1$, there is only one root/tuber, as in the original flower structure in Figure 2.1.

Now using the same matrices,

$$C = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 1 & 0 \\ -1 & -1 \end{bmatrix},$$

$$B_3 = B^{A_1} = \begin{bmatrix} 1 & 0 \\ t-1 & t \end{bmatrix}, \quad B_4 = B = \begin{bmatrix} 1 & 0 \\ 0 & t \end{bmatrix},$$

as in Section 2.5.1 (except that t now generates $\mathbb{F}_q^\times$ instead of $\mathbb{F}_p^\times$), we write down the following subgroups:

$G_1 = \langle \overline{C}, \overline{B_1} \rangle \cong D_6 \cong S_3$. (Same as in Section 2.5.1.)

$G_2 = \langle \overline{A_{\xi_1}}, \overline{A_{\xi_2}}, \ldots, \overline{A_{\xi_m}}, \overline{B} \rangle = \langle N, \overline{B} \rangle \cong \mathbb{Z}_p^m \rtimes \mathbb{Z}_{q-1}$, which consists of all lower triangular matrices in $GL(2, q)$ modulo V_q.

$G_3 = \langle \overline{B_1}(\overline{C})^2, \overline{B_3} \rangle = \langle \overline{CB_1}, \overline{B_3} \rangle$. Now $|\overline{B_3}| = q - 1$, and we still have $\overline{B_3} \cdot \overline{CB_1} = \overline{CB_1}(\overline{B_3})^{-1}$, so

$$G_3 = \left\{ (\overline{B_3})^k = \overline{\begin{bmatrix} 1 & 0 \\ t^k - 1 & t^k \end{bmatrix}}, \quad \overline{CB_1}(\overline{B_3})^k = \overline{\begin{bmatrix} -1 & -1 \\ 1 - t^{-k} & 1 \end{bmatrix}} \;\middle|\; k \in \mathcal{K}_q \right\} \cong D_{2(q-1)}.$$

$G_4 = \langle \overline{B_1C}, \overline{B_4} \rangle$. Now $\left|\overline{B_4}\right| = q-1$ and $\overline{B_4} \cdot \overline{B_1C} = \overline{B_1C}\left(\overline{B_4}\right)^{-1}$, so

$$G_4 = \left\{ \left(\overline{B_4}\right)^k = \overline{\begin{bmatrix} 1 & 0 \\ 0 & t^k \end{bmatrix}}, \quad \overline{B_1C}\left(\overline{B_4}\right)^k = \overline{\begin{bmatrix} 0 & t^k \\ 1 & 0 \end{bmatrix}} \;\middle|\; k \in \mathcal{K}_q \right\} \cong D_{2(q-1)},$$

which comprises all diagonal and anti-diagonal matrices in $GL(2,q)$ modulo V_q.

Next we find the intersections $G_{12} = \langle \overline{B_1} \rangle$, $G_{13} = \langle \overline{CB_1} \rangle$, and $G_{14} = \langle \overline{B_1C} \rangle$, which are all isomorphic to $\mathbb{Z}_2$; $G_{23} = \langle \overline{B_3} \rangle$ and $G_{24} = \langle \overline{B_4} \rangle$, both of which are isomorphic to $\mathbb{Z}_{q-1}$; and $G_{34} = G_{123} = G_{124} = 1$.

The orders of the four subgroups and the intersections are

$$|G_1| = 6, \quad |G_2| = q(q-1), \quad |G_3| = |G_4| = 2(q-1),$$

$$|G_{12}| = |G_{13}| = |G_{14}| = 2, \quad |G_{23}| = |G_{24}| = q-1, \quad |G_{34}| = |G_{123}| = |G_{124}| = 1.$$

So in (2.4),

$$LHS = |G_1||G_2||G_{34}||G_{123}||G_{124}| = 6q(q-1),$$

$$RHS = |G_{12}||G_{13}||G_{14}||G_{23}||G_{24}| = 8(q-1)^2,$$

$$LHS - RHS = 2(q-1)(4-q).$$

Thus Ingleton is violated when $q \geq 5$.

Remark 2.5.2. Depending on the characteristic p of $\mathbb{F}_q$, the intersection $G_{12} = \langle \overline{B_1} \rangle$ might lie in either the petals or the roots of G_2, as depicted by the dashed circles in Figure 2.2. If $p \neq 2$, then q is odd and

$$\overline{B_1} = \left(\overline{B}^{\frac{q-1}{2}}\right)^{\overline{A_{k_1}}},$$

where $k_1 = 2^{-1} = \frac{p+1}{2}$, so G_{12} is on the petal $H^{\overline{A_{k_1}}}$; whereas if $p = 2$, then $-1 = 1$ and $\overline{B_1} = \overline{A_1} \in N$, so G_{12} becomes a root. Note that the patterns of the other intersections are not changed for different q.

Remark 2.5.3. We can also show that $\overline{A_{\xi_1}}, \overline{A_{\xi_2}}, \ldots, \overline{A_{\xi_m}}, \overline{B}$ and $\overline{C}$ generate $PGL(2,q)$,

using the same argument as in the previous section. The only difference is that the elementary matrices of $GL(2,q)$ are now generated by $A_{\xi_1}, A_{\xi_1}^T, \ldots, A_{\xi_m}, A_{\xi_m}^T, B$ and $t^{-1}B$. But as $A_\alpha^{B_1 C} = A_\alpha^T$, $\forall \alpha \in \mathbb{F}_q$, we see that $PGL(2,q)$ is indeed generated by the desired elements.

In Section 2.7, we will see that these subgroups have more fundamental interpretations in the framework of group actions and groups of Lie type: each subgroup is the stabilizer for a special set of points in the underlying projective geometry of $PGL(2,q)$.

2.5.3 Discussion

To measure "how much" the Ingleton inequality is violated, or how effective a set of subgroups is in terms of violating Ingleton, we need to compare the difference of the two sides of (2.3) for the corresponding entropy vector, i.e.,

$$\Delta_h \triangleq h_1 + h_2 + h_{34} + h_{123} + h_{124} - (h_{12} + h_{13} + h_{14} + h_{23} + h_{24}).$$

Translating to the finite group context, it equals $\log \frac{RHS}{LHS}$ of (2.4). Thus we can make the following definition to measure the extent to which Ingleton is violated.

Definition 2.5.1. For a 4-tuple of subgroups $\tau = (G_i : 1 \leq i \leq 4)$, we define the *Ingleton ratio* to be

$$r(\tau) = \frac{|G_{12}||G_{13}||G_{14}||G_{23}||G_{24}|}{|G_1||G_2||G_{34}||G_{123}||G_{124}|}. \tag{2.13}$$

Clearly $\Delta_h = \log r$ and Ingleton is violated iff $r > 1$. The family $PGL(2,q)$ have the Ingleton ratio

$$r = \frac{4(q-1)}{3q},$$

which approaches $4/3$ when q is large.

However, the Ingleton ratio is not precise enough to characterize the effectiveness of an Ingleton violation instance. Observe that $\overline{\Gamma_n^*}$ is a cone, and in fact, as remarked in [22], adding an entropy vector to itself yields another entropy vector. Thus we can

arbitrarily increase the Ingleton ratio by joining copies of a violation instance. For example, if $\tau = (G_i : 1 \leq i \leq 4)$ is such an instance, for each integer N let

$$G' = \times_{k=1}^{N} G \triangleq G \times \cdots \times G$$

be the direct product of N copies of G and define $\tau' = (G'_i : 1 \leq i \leq 4)$ with $G'_i = \times_{k=1}^{N} G_i$ for each i. Then the Ingleton ratio $r(\tau') = [r(\tau)]^N$, which grows unbounded when $N \to \infty$.

Therefore we need to consider the scaled version of Δ_h to be able to measure the effectiveness of an Ingleton violation. In [29] Dougherty *et al.* use the full joint entropy h_{1234} as a scaling factor to avoid the problem above:

Definition 2.5.2. For an entropy vector $h = (h_\alpha : \emptyset \neq \alpha \subseteq \{1, 2, 3, 4\})$, define the *Ingleton score* to be

$$\sigma(h) = -\frac{\Delta_h}{h_{1234}}.$$

In the context of groups, the Ingleton score of a 4-tuple τ of subgroups of G is

$$\sigma(\tau) = \frac{-\log r(\tau)}{\log(|G|/|G_{1234}|)}.$$

Note that Ingleton fails iff $\sigma < 0$, and a lower score means a larger violation. Essentially this definition forms a ray starting from the origin and passing through the point in $\mathbb{R}^{2^4-1}$ corresponding to an entropy vector, then finds its intersection with the hyperplane $h_{1234} = 1$ and computes $-\Delta_h$ for that point to measure the Ingleton violation. The best Ingleton score in the family $PGL(2, q)$ is attained when $q = 13$, with $\sigma = -0.0270$. In [22] many violations obtained have lower Ingleton scores, and hence are more effective than $PGL(2, q)$. In [29] a conjecture concerning the lowest Ingleton score attainable by an arbitrary entropy vector is proposed, but has been refuted recently by Matúš and Csirmaz [30].

A perhaps more geometrically meaningful scaling factor is the 2-norm of the entropy vector, as proposed in [31]:

Definition 2.5.3. For an entropy vector $h = (h_\alpha : \emptyset \neq \alpha \subseteq \{1, 2, 3, 4\})$, define the

Ingleton violation index to be

$$\iota(h) = \frac{\Delta_h}{\|h\|_2} = \frac{\Delta_h}{\sqrt{h^T h}}.$$

Essentially this definition measures the "sine" of the angle between an entropy vector and the Ingleton hyperplane $\Delta_h = 0$. The Ingleton inequality fails iff $\iota > 0$, and a larger index means a larger violation. Note that two entropy vectors might have the same violation index but different Ingleton scores, and vice versa. The best Ingleton violation index in the family $PGL(2, q)$ is again attained when $q = 13$, with $\iota = 0.0082$, whereas for an arbitrary entropy vector the best ι found in literature is 0.0276 using quasi-uniform distributions [32].

Next we discuss two directions for generalizing the above Ingleton-violating family and finding new violations. On the one hand, $PGL(2, q)$ is the quotient group of $GL(2, q)$, so supposedly $GL(2, q)$ should have a richer choice of subgroups violating Ingleton inequality. This approach is explored in the next section. On the other hand, since the subgroups in the $PGL(2, q)$ family have simple but fundamental interpretations in terms of group actions, we can generalize them in this framework. In particular, we obtain two new families of violations in $PGL(n, q)$ for general n, and further generalize to an abstract construction using 2-transitive groups. This approach is explored in Section 2.7. Note that it is more abstract than the previous approach and requires more background knowledge.

2.6 Ingleton Violations in $GL(2, q)$

As $PGL(2, q)$ is the quotient group of $GL(2, q)$ modulo the subgroup V_q of scalar matrices, naturally one may ask if the general linear groups also violate Ingleton. In fact, the following lemma shows that there is at least one set of subgroups in $GL(2, q)$ that violates Ingleton for all finite field orders $q \geq 5$:

Lemma 2.6.1. *If G is a finite group with a normal subgroup N such that $H \triangleq G/N$*

has a set of Ingleton-violating subgroups, then the preimages of these subgroups under the natural homomorphism $g \mapsto gN$ are subgroups of G that also violate Ingleton.

Proof. Let $(H_i : 1 \leq i \leq 4)$ be a set of Ingleton-violating subgroups in H. Define G_i to be the preimage of H_i under the natural homomorphism, and G_i is then a group containing N for each i. By the Lattice Isomorphism Theorem (see, e.g., [25]), for any nonempty subset $\alpha \subseteq \{1, 2, 3, 4\}$, $G_\alpha/N = H_\alpha$, and so $|G_\alpha| = |H_\alpha| \cdot |N|$. Thus by checking the orders in (2.4), $(G_i : 1 \leq i \leq 4)$ also violate Ingleton. □

Searching with GAP, we find $GL(2, 5)$ to be the smallest general linear group that violates Ingleton. Up to subscript symmetries and conjugations, it has 15 sets of Ingleton-violating subgroups. We would like to analyze their structures and generalize them for $q \geq 5$ if possible.

Throughout this section, we always assume q is a finite field order, and p is the characteristic of $\mathbb{F}_q$. We begin our analysis by identifying the preimages of the Ingleton-violating subgroups in the previous section under the natural homomorphism

$$\pi : GL(2, q) \to GL(2, q)/V_q = PGL(2, q),$$

according to Lemma 2.6.1. With no surprise, when $q = 5$ these correspond to one of the 15 violation instances in $GL(2, 5)$, and they take on nice matrix structures similar to the subgroups in Section 2.5. Based on this set of subgroups we have 10 other instances, all of which are essentially its variants: each instance differs from the preimages at exactly one subgroup (either G_1 or G_2). These 11 violation instances can be easily extended to families of Ingleton-violating subgroups in $GL(2, q)$ for $q \geq 5$, sometimes with an extra condition. The remaining 4 instances cannot be derived directly from the preimages; however, they are interrelated and all their subgroups are equal or conjugate to some known subgroups from the previous instances. They also generalize to Ingleton-violating families in $GL(2, q)$ with some extra conditions.

Table 2.2 summarizes how the generalization of these instances depends on the values of p and q. We can see that when $p = 2$, these 15 instances collapse to only

Table 2.2: (a) Identical instances when $p = 2$ (b) Cases when Ingleton is not violated

Instance No.	Identical Instance(s)
1	5
2	3, 4
6	8, 10
7	9, 11
12	13
14	15

Instance No.	$p = 3$	$p \neq 2$, $\frac{q-1}{2}$ odd
8, 9		×
12, 14	×	
13, 15	×	×

6 dinstinct ones; also some instances need specific conditions on p and q to violate Ingelton.

Table 2.3: Orders of subgroups and intersections

Ins. No.	$\lvert G_1\rvert$	$\lvert G_2\rvert$	$\lvert G_3\rvert$	$\lvert G_{34}\rvert$	$\lvert G_{123}\rvert$	$\lvert G_{12}\rvert$	$\lvert G_{13}\rvert$	$\lvert G_{23}\rvert$	$LHS - RHS$ in (2.4)
0	6	$q(q-1)$	$2(q-1)$	1	1	2	2	$q-1$	$2(q-1)(4-q)$
1	$6(q-1)$	$q(q-1)^2$	$2(q-1)^2$	$q-1$	$q-1$	$2(q-1)$	$2(q-1)$	$(q-1)^2$	$2(q-1)^6(4-q)$
2,4	6	$q(q-1)^2$	$2(q-1)^2$	$q-1$	1	2	2	$(q-1)^2$	$2(q-1)^3(4-q)$
3	12	$q(q-1)^2$	$2(q-1)^2$	$q-1$	2	4	4	$(q-1)^2$	$16(q-1)^3(4-q)$
5	$3(q-1)$	$q(q-1)^2$	$2(q-1)^2$	$q-1$	$\frac{q-1}{2}$	$q-1$	$q-1$	$(q-1)^2$	$\frac{1}{4}(q-1)^6(4-q)$
6–9	$6(q-1)$	$q(q-1)$	$2(q-1)^2$	$q-1$	1	2	$2(q-1)$	$q-1$	$2(q-1)^3(4-q)$
10,11	$6(q-1)$	$2q(q-1)$	$2(q-1)^2$	$q-1$	2	4	$2(q-1)$	$2(q-1)$	$16(q-1)^3(4-q)$
12–15	6	$q(q-1)$	$q(q-1)$	1	1	2	2	$q-1$	$2(q-1)(4-q)$
8',9'	$6(q-1)$	$q(q-1)$	$2(q-1)^2$	$q-1$	2	2	$2(q-1)$	$q-1$	$8(q-1)^3(2q+1)$
13',15'	6	$q(q-1)$	$q(q-1)$	2	1	1	1	$q-1$	$(q-1)(11q+1)$

In Table 2.3, the orders of the subgroups for the cases we have explored in $PGL(2,q)$ and $GL(2,q)$ are listed. No. 0 denotes the instance in $PGL(2,q)$, and No. 1–15 denote the generalizations of the 15 violation instances in $GL(2,5)$ to $GL(2,q)$. Since all instances have the subgroup order symmetries

$$|G_3| = |G_4|, \quad |G_{123}| = |G_{124}|, \quad |G_{13}| = |G_{14}|, \quad |G_{23}| = |G_{24}|,$$

only one of each pair of orders is listed. Note that when $p = 2$, there are only 6 such dinstinct generalizations, which are Instances 1, 2, 6, 7, 12, and 14. Thus for the order calculation of all other instances in $GL(2,q)$ assume $p \neq 2$. Moreover, No. 8', 9', 13', and 15' correspond to Instances 8, 9, 13, and 15 when $p \neq 2$ but $\frac{q-1}{2}$ is odd, in which case Ingleton is satisfied. Finally, the order calculation for Instances 12–15 only works for $p \neq 3$. From Table 2.3, we can calculate that all violation instances in the table have the same Ingleton ratio $r = 4(q-1)/(3q)$, which is the same as the family $PGL(2,q)$. But because the scaling factors for both the Ingleton score and the violation index are no larger than $PGL(2,q)$ in these instances, they are no more effective.

In the following, we present all of these extended violation families, with Section 2.6.1 being the set of preimage subgroups, Sections 2.6.2 and 2.6.3 its 10 variants, and Section 2.6.4 the remaining 4 instances. We continue to use the notations from Section 2.5 with t being a primitive element of $\mathbb{F}_q$, but we redefine

$$N = \{\, A_\alpha |\, \alpha \in \mathbb{F}_q\} = \langle A_{\xi_1}, A_{\xi_2}, \ldots, A_{\xi_m}\rangle \cong \langle A_{\xi_1}\rangle \times \langle A_{\xi_2}\rangle \times \ldots \times \langle A_{\xi_m}\rangle \cong \mathbb{Z}_p^m.$$

In addition, we introduce the following matrices and subgroups in $GL(2,q)$ to facilitate our presentation:

$$B' = \begin{bmatrix} -1 & 0 \\ 0 & t \end{bmatrix}, \qquad P = \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix}, \qquad P' = \begin{bmatrix} t & 0 \\ 0 & -1 \end{bmatrix},$$

$$M = \langle C, B_1 \rangle = \left\{ I, \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}, \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ -1 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} -1 & -1 \\ 0 & 1 \end{bmatrix} \right\},$$

$$K = \langle N, B \rangle = \left\{ \begin{bmatrix} 1 & 0 \\ \alpha & \beta \end{bmatrix} \middle| \begin{array}{l} \alpha \in \mathbb{F}_q, \\ \beta \in \mathbb{F}_q^\times \end{array} \right\},$$

$$K' = \langle N, B' \rangle = \left\{ \begin{bmatrix} (-1)^k & 0 \\ \alpha & t^k \end{bmatrix} \middle| \begin{array}{l} \alpha \in \mathbb{F}_q, \\ k \in \mathcal{K}_q \end{array} \right\},$$

$$J = \langle N, P \rangle = \left\{ \begin{bmatrix} \beta & 0 \\ \alpha & 1 \end{bmatrix} \middle| \begin{array}{l} \alpha \in \mathbb{F}_q, \\ \beta \in \mathbb{F}_q^\times \end{array} \right\},$$

$$J' = \langle N, P' \rangle = \left\{ \begin{bmatrix} t^k & 0 \\ \alpha & (-1)^k \end{bmatrix} \middle| \begin{array}{l} \alpha \in \mathbb{F}_q, \\ k \in \mathcal{K}_q \end{array} \right\}.$$

Note that when $p = 2$, we have $-1 = 1$, so $B' = B$, $P' = P$, and $K' = K$, $J' = J$. Also note that M and K precisely correspond to G_1 and G_2 in Section 2.5, respectively. The group M is isomorphic to $D_6 \cong S_3$, while the other four groups are all semidirect products $\mathbb{Z}_p^m \rtimes \mathbb{Z}_{q-1}$, with $K \cong J$ and $K' \cong J'$. Moreover, K and J have generalized flower structures for all $q > 2$. However, if $p \neq 2$, K' and J' only have flower structures when $\frac{q-1}{2}$ is even, in which case they are also isomorphic to K. (See Section A.1.1 in Appendices for proofs.) This turns out to be a necessary condition to violate Ingleton in all the instances where K' and J' are involved.

2.6.1 Instance 1: The Preimage Subgroups

To obtain the preimage H_0 of a subgroup $H \leq PGL(2, q)$ under π, we can generate H_0 in $GL(2, q)$ with the generators of H (without overlines) and tI, since $V_q = \langle tI \rangle \cong \mathbb{Z}_{q-1}$.

$G_1 = \langle tI, C, B_1 \rangle = \langle V_q, M \rangle$. Since V_q is the center of $GL(2, q)$ and intersects M trivially, G_1 is a direct product:

$$G_1 = \{ t^k X \mid X \in M, k \in \mathcal{K}_q \} \cong V_q \times M \cong \mathbb{Z}_{q-1} \times S_3.$$

$G_2 = \langle tI, A_{\xi_1}, A_{\xi_2}, \ldots, A_{\xi_m}, B\rangle = \langle tI, N, B\rangle = \langle V_q, K\rangle$. G_2 is the subgroups of all lower triangular matrices in $GL(2,q)$, and as $V_q \bigcap K = 1$, we have

$$G_2 \cong V_q \times K \cong \mathbb{Z}_{q-1} \times (\mathbb{Z}_p^m \rtimes \mathbb{Z}_{q-1}).$$

$G_3 = \langle tI, B_1C^2, B_3\rangle = \langle tI, CB_1, B_3\rangle = \langle CB_1, T\rangle$, where $T = \langle tI, B_3\rangle$. As $V_q \bigcap \langle B_3\rangle = 1$, we have

$$T = \{\, t^k B_3^m \,|\, k, m \in \mathcal{K}_q \,\} \cong V_q \times \langle B_3\rangle \cong \mathbb{Z}_{q-1} \times \mathbb{Z}_{q-1}.$$

It is easy to check that $(t^k B_3^m)^{CB_1} = t^{k+m} B_3^{-m} \in T$, so $G_3 = \langle CB_1\rangle \cdot T$ and $T \trianglelefteq G_3$. Furthermore, $|CB_1| = 2$ and $T \bigcap \langle CB_1\rangle = 1$, thus

$$G_3 \cong T \rtimes \langle CB_1\rangle \cong (\mathbb{Z}_{q-1} \times \mathbb{Z}_{q-1}) \rtimes \mathbb{Z}_2,$$

$$G_3 = \left\{ t^k \begin{bmatrix} 1 & 0 \\ t^m - 1 & t^m \end{bmatrix}, \quad t^{k+m} \begin{bmatrix} -1 & -1 \\ 1 - t^{-m} & 1 \end{bmatrix} \middle| \, k, m \in \mathcal{K}_q \right\}.$$

$G_4 = \langle tI, B_1C, B_4\rangle = \langle tI, B_1C, B\rangle = \langle B_1C, D\rangle$, where $D = \langle tI, B\rangle$. Since $V_q \bigcap \langle B\rangle = 1$, we have

$$D = \{\, t^k B^m \,|\, k, m \in \mathcal{K}_q \,\} = \cong V_q \times \langle B\rangle \cong \mathbb{Z}_{q-1} \times \mathbb{Z}_{q-1},$$

which consists of all diagonal matrices in $GL(2,q)$. Note that

$$\begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}^{B_1C} = \begin{bmatrix} \beta & 0 \\ 0 & \alpha \end{bmatrix} \in D,$$

so $G_4 = \langle B_1C\rangle \cdot D$ and $D \trianglelefteq G_4$. Since $|B_1C| = 2$ and $D \bigcap \langle B_1C\rangle = 1$,

$$G_4 \cong D \rtimes \langle B_1C\rangle \cong (\mathbb{Z}_{q-1} \times \mathbb{Z}_{q-1}) \rtimes \mathbb{Z}_2.$$

Table 2.4: G_1 for Instances 2–5

Ins. No.	2	3	4	5
G_1	$\langle C, B_1\rangle$	$\langle -C, B_1\rangle$	$\langle C, -B_1\rangle$	$\langle C, tB_1\rangle$

Actually G_4 is the subgroups of all diagonal and anti-diagonal matrices in $GL(2,q)$:

$$G_4 = \left\{ \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}, \begin{bmatrix} 0 & \beta \\ \alpha & 0 \end{bmatrix} \middle| \, \alpha, \beta \in \mathbb{F}_q^\times \right\}.$$

Calculating the intersections, we have

$$G_{12} = \langle tI, B_1\rangle \cong V_q \times \langle B_1\rangle \cong \mathbb{Z}_{q-1} \times \mathbb{Z}_2,$$

$$G_{13} = \langle tI, CB_1\rangle \cong V_q \times \langle CB_1\rangle \cong \mathbb{Z}_{q-1} \times \mathbb{Z}_2,$$

$$G_{14} = \langle tI, B_1C\rangle \cong V_q \times \langle B_1C\rangle \cong \mathbb{Z}_{q-1} \times \mathbb{Z}_2,$$

$$G_{23} = T, \quad G_{24} = D, \quad G_{34} = G_{123} = G_{124} = \langle tI\rangle = V_q.$$

From the calculation in Table 2.3, Ingleton is violated when $q \geq 5$.

2.6.2 Instances 2–5: Variants with Different G_1's

In all the instances in this section, only G_1 is different from Instance 1; it is now a *proper* subgroup of $\langle tI, C, B_1\rangle$ (see Table 2.4, where the generator-form for these groups is used to better demonstrate the subgroup relations). When $p \neq 2$, these instances are all distinct; however, when $p = 2$, clearly Instances 3 and 4 collapse to Instance 2, while Instance 5 becomes Instance 1. From Table 2.3, we can see that they all violate Ingleton when $q \geq 5$.

2.6.2.1 Instance 2

$G_1 = M$.

$G_{12} = \langle B_1\rangle$, $G_{13} = \langle CB_1\rangle$ and $G_{14} = \langle B_1C\rangle$ are all isomorphic to $\mathbb{Z}_2$, and

$G_{123} = G_{124} = 1$.

2.6.2.2 Instance 3

$G_1 = \langle -C, B_1 \rangle$.

We only consider the case $p \neq 2$, since otherwise this is the same as Instance 2. As $|C| = 3$, we have $(-C)^3 = -I$ and $(-C)^4 = C$. Thus

$$G_1 = \langle -I, C, B_1 \rangle = \langle -I, M \rangle \cong \langle -I \rangle \times M \cong \mathbb{Z}_2 \times S_3 \cong D_{12},$$

since $\langle -I \rangle$ is a subgroup of V_q and intersects M trivially. So $G_1 = \{\pm X \mid X \in M\}$,

$$G_{12} = \langle -I, B_1 \rangle \cong \langle -I \rangle \times \langle B_1 \rangle \cong \mathbb{Z}_2 \times \mathbb{Z}_2,$$

$$G_{13} = \langle -I, CB_1 \rangle \cong \langle -I \rangle \times \langle CB_1 \rangle \cong \mathbb{Z}_2 \times \mathbb{Z}_2,$$

$$G_{14} = \langle -I, B_1C \rangle \cong \langle -I \rangle \times \langle B_1C \rangle \cong \mathbb{Z}_2 \times \mathbb{Z}_2,$$

$$G_{123} = G_{124} = \langle -I \rangle \cong \mathbb{Z}_2.$$

2.6.2.3 Instance 4

$G_1 = \langle C, -B_1 \rangle$.

Here we also need only consider the case $p \neq 2$. Observe that $|C| = 3$, $|{-}B_1| = 2$, and $(C \cdot (-B_1))^2 = (CB_1)^2 = I$. This gives us

$$G_1 = \left\{ I, C, C^2, -B_1, -B_1C, -CB_1 \right\},$$

so $G_1 \cong D_6 \cong S_3$.

For the intersections, we have $G_{12} = \langle -B_1 \rangle$, $G_{13} = \langle -CB_1 \rangle$, and $G_{14} = \langle -B_1C \rangle$ all isomorphic to $\mathbb{Z}_2$, and $G_{123} = G_{124} = 1$.

2.6.2.4 Instance 5

$G_1 = \langle C, tB_1 \rangle$.

Table 2.5: G_2 for Instances 6–11

Ins. No.	6	7	8	9	10	11
G_2	$\langle N, B\rangle$	$\langle N, P\rangle$	$\langle N, B'\rangle$	$\langle N, P'\rangle$	$\langle -I, N, B\rangle$	$\langle -I, N, P\rangle$

When $p = 2$, q is even. Since $|B_1| = 2$ and $|t| = q - 1$, we have $(tB_1)^q = tI$ and $(tB_1)^{q-1} = B_1$. Thus $G_1 = \langle tI, C, B_1\rangle$ and this instance is the same as Instance 1.

Now assume $p \neq 2$. As q is odd, $|tB_1| = q - 1$. When k is even, $(tB_1)^k = t^k I$ and so $C^{(tB_1)^k} = C$. Otherwise $(tB_1)^k = t^k B_1$, then $C^{(tB_1)^k} = B_1 C B_1 = C^{-1}$ since $(CB_1)^2 = I$. So $G_1 = \langle tB_1\rangle \cdot \langle C\rangle$ and $\langle C\rangle \trianglelefteq G_1$. Furthermore, $\langle tB_1\rangle \bigcap \langle C\rangle = 1$ and $|C| = 3$, thus

$$G_1 = \left\{ t^k I,\, t^k C,\, t^k C^2 \;\middle|\; k \text{ even},\, k \in \mathcal{K}_q \right\} \bigcup \left\{ t^k B_1,\, t^k B_1 C,\, t^k C B_1 \;\middle|\; k \text{ odd},\, k \in \mathcal{K}_q \right\},$$

$$G_1 \cong \langle C\rangle \rtimes \langle tB_1\rangle \cong \mathbb{Z}_3 \rtimes \mathbb{Z}_{q-1}.$$

The intersections are $G_{12} = \langle tB_1\rangle$, $G_{13} = \langle tCB_1\rangle$, and $G_{14} = \langle tB_1C\rangle$, which are all isomorphic to $\mathbb{Z}_{q-1}$, and $G_{123} = G_{124} = \langle t^2 I\rangle \cong \mathbb{Z}_{\frac{q-1}{2}}$.

2.6.3 Instances 6–11: Variants with Different G_2's

In all the instances in this section, only G_2 is different from Instance 1; it is now a *proper* subgroup of $\langle tI, N, B\rangle$ (see Table 2.5). It is easy to see that these instances are distinct when $p \neq 2$; otherwise Instances 8 and 10 collapse to Instance 6, while Instances 9 and 11 become Instance 7. Thus in the analysis of Instances 8–11, we assume $p \neq 2$. From Table 2.3, Instances 6, 7, 10, and 11 violate Ingleton whenever $q \geq 5$; however, if $p \neq 2$, Instances 8 and 9 only violate Ingleton when in addition $\frac{q-1}{2}$ is even. Please refer to Section A.1.2 in Appendices for the calculation of subgroup intersections in Instances 8 and 9.

2.6.3.1 Instance 6

$G_2 = K$.

In this case, $G_{12} = \langle B_1 \rangle \cong \mathbb{Z}_2$ and $G_{123} = G_{124} = 1$. Also $G_{23} = \langle B_3 \rangle$ and $G_{24} = \langle B \rangle$, both of which are isomorphic to $\mathbb{Z}_{q-1}$.

2.6.3.2 Instance 7

$G_2 = J$.

Here $G_{12} = \langle -B_1 \rangle \cong \mathbb{Z}_2$, $G_{123} = G_{124} = 1$. Also, $G_{23} = \langle t^{-1}B_3 \rangle$ and $G_{24} = \langle P \rangle$, and are both isomorphic to $\mathbb{Z}_{q-1}$.

2.6.3.3 Instance 8

$G_2 = K'$.

We have

$$G_{12} = \begin{cases} \langle B_1 \rangle \cong \mathbb{Z}_2 & \text{if } \frac{q-1}{2} \text{ is even} \\ \langle -I \rangle \cong \mathbb{Z}_2 & \text{otherwise} \end{cases},$$

$$G_{123} = G_{124} = \begin{cases} 1 & \text{if } \frac{q-1}{2} \text{ is even} \\ \langle -I \rangle \cong \mathbb{Z}_2 & \text{otherwise} \end{cases}.$$

In this case, $G_{23} = \langle -B_3^{\frac{q+1}{2}} \rangle$ and $G_{24} = \langle B' \rangle$ are both isomorphic to $\mathbb{Z}_{q-1}$.

2.6.3.4 Instance 9

$G_2 = J'$.

We have

$$G_{12} = \begin{cases} \langle -B_1 \rangle \cong \mathbb{Z}_2 & \text{if } \frac{q-1}{2} \text{ is even} \\ \langle -I \rangle \cong \mathbb{Z}_2 & \text{otherwise} \end{cases},$$

$$G_{123} = G_{124} = \begin{cases} 1 & \text{if } \frac{q-1}{2} \text{ is even} \\ \langle -I \rangle \cong \mathbb{Z}_2 & \text{otherwise} \end{cases}.$$

Here $G_{23} = \langle tB_3^{\frac{q-3}{2}} \rangle$ and $G_{24} = \langle P' \rangle$ are isomorphic to $\mathbb{Z}_{q-1}$.

2.6.3.5 Instance 10

$G_2 = \langle -I, N, B \rangle$.

Now we have

$$G_2 = \langle -I, K \rangle \cong \langle -I \rangle \times K \cong \mathbb{Z}_2 \times (\mathbb{Z}_p^m \rtimes \mathbb{Z}_{q-1}),$$

since $\langle -I \rangle \bigcap K = 1$. Thus $G_2 = \{ \pm X \mid X \in K \}$.

For the intersections, we have

$$G_{12} = \langle -I, B_1 \rangle \cong \mathbb{Z}_2 \times \mathbb{Z}_2,$$

$$G_{123} = G_{124} = \langle -I \rangle \cong \mathbb{Z}_2,$$

$$G_{23} = \langle -I, B_3 \rangle \cong \langle -I \rangle \times \langle B_3 \rangle \cong \mathbb{Z}_2 \times \mathbb{Z}_{q-1},$$

$$G_{24} = \langle -I, B \rangle \cong \langle -I \rangle \times \langle B \rangle \cong \mathbb{Z}_2 \times \mathbb{Z}_{q-1}.$$

2.6.3.6 Instance 11

$G_2 = \langle -I, N, P \rangle$.

Here

$$G_2 = \langle -I, J \rangle \cong \langle -I \rangle \times J \cong \mathbb{Z}_2 \times (\mathbb{Z}_p^m \rtimes \mathbb{Z}_{q-1}),$$

since $\langle -I \rangle \bigcap J = 1$. Thus $G_2 = \{ \pm X \mid X \in J \}$.

Moreover,

$$G_{12} = \langle -I, -B_1 \rangle = \langle -I, B_1 \rangle \cong \mathbb{Z}_2 \times \mathbb{Z}_2,$$

$$G_{123} = G_{124} = \langle -I \rangle \cong \mathbb{Z}_2,$$

$$G_{23} = \langle -I, t^{-1}B_3 \rangle \cong \langle -I \rangle \times \langle t^{-1}B_3 \rangle \cong \mathbb{Z}_2 \times \mathbb{Z}_{q-1},$$

$$G_{24} = \langle -I, P \rangle \cong \langle -I \rangle \times \langle P \rangle \cong \mathbb{Z}_2 \times \mathbb{Z}_{q-1}.$$

2.6.4 Instances 12–15

For these last four instances, G_1 is always M, and G_2–G_4 are equal or conjugate to one of K, K', J, and J', as listed in Table 2.6. Thus G_2–G_4 are all semidirect

Table 2.6: Subgroups for Instances 12–15

Ins. No.	G_1	G_2	G_3	G_4
12	M	$\langle N, B\rangle$	$\langle N, P\rangle^E$	$\langle N, P\rangle^Q$
13	M	$\langle N, B'\rangle$	$\langle N, P'\rangle^E$	$\langle N, P'\rangle^Q$
14	M	$\langle N, P\rangle^E$	$\langle N, B\rangle$	$\langle N, B\rangle^W$
15	M	$\langle N, P'\rangle^E$	$\langle N, B'\rangle$	$\langle N, B'\rangle^W$

products $\mathbb{Z}_p^m \rtimes \mathbb{Z}_{q-1}$ and the structures of G_3 and G_4 are different from all previous instances. The conjugators E, Q, and W and the elements of new subgroups are listed as follows.

$$E = \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix}, \qquad Q = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}, \qquad W = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix}.$$

$$J^E = \langle N, P\rangle^E = \left\{ \begin{bmatrix} 1-v & v \\ 1-u-v & u+v \end{bmatrix} \middle| \begin{array}{l} u \in \mathbb{F}_q^\times, \\ v \in \mathbb{F}_q \end{array} \right\},$$

$$(J')^E = \langle N, P'\rangle^E = \left\{ \begin{bmatrix} (-1)^j - \alpha & \alpha \\ (-1)^j - t^j - \alpha & t^j + \alpha \end{bmatrix} \middle| \begin{array}{l} \alpha \in \mathbb{F}_q, \\ j \in \mathcal{K}_q \end{array} \right\},$$

$$J^Q = \langle N, P\rangle^Q = \left\{ \begin{bmatrix} 1+2y & y \\ 2(x-2y-1) & x-2y \end{bmatrix} \middle| \begin{array}{l} x \in \mathbb{F}_q^\times, \\ y \in \mathbb{F}_q \end{array} \right\},$$

$$(J')^Q = \langle N, P'\rangle^Q = \left\{ \begin{bmatrix} (-1)^i + 2\beta & \beta \\ 2\left(t^i - 2\beta - (-1)^i\right) & t^i - 2\beta \end{bmatrix} \middle| \begin{array}{l} \beta \in \mathbb{F}_q, \\ i \in \mathcal{K}_q \end{array} \right\},$$

$$K^W = \langle N, B\rangle^W = \left\{ \begin{bmatrix} x & y \\ 0 & 1 \end{bmatrix} \middle| \begin{array}{l} x \in \mathbb{F}_q^\times, \\ y \in \mathbb{F}_q \end{array} \right\} = \left\{ X^T \mid X \in J \right\},$$

$$(K')^W = \langle N, B'\rangle^W = \left\{ \begin{bmatrix} t^i & \beta \\ 0 & (-1)^i \end{bmatrix} \middle| \begin{array}{l} \beta \in \mathbb{F}_q, \\ i \in \mathcal{K}_q \end{array} \right\} = \left\{ X^T \mid X \in J' \right\}.$$

As mentioned in Table 2.2, Instances 12–15 do not violate Ingleton when $p = 3$. The reasons are as follows. If $p = 3$, then $2 = -1$, so $E = Q$ and $M \leq J^E$. Thus in Instance 12 we have $G_3 = G_4$ and $G_1 \leq G_3$, while in Instances 13 and

14 we have $G_3 = G_4$ and $G_1 \leq G_2$ respectively. So these three instances satisfy Conditions 2.3.5 and/or 2.3.7. Instance 15, however, satisfies Condition 2.3.3 in this case (see Section A.1.3 in Appendices).

Also, we need $p \neq 2$ to make Instances 13 and 15 distinct, otherwise they collapse to Instances 12 and 14, respectively. Thus in the rest of this section we always assume $p \neq 3$, while for Instances 13 and 15 we assume $p > 3$. From Table 2.3, Instances 12 and 14 violate Ingleton when $q \geq 5$ (and of course, $p \neq 3$), while if $p \neq 2$, Instances 13 and 15 only violate Ingleton when in addition $\frac{q-1}{2}$ is even. Please refer to Section A.1.4 in Appendices for the intersection calculations.

2.6.4.1 Instance 12

$G_2 = K, G_3 = J^E, G_4 = J^Q$.

We have $G_{12} = \langle B_1 \rangle$, $G_{13} = \langle B_1 C \rangle$, and $G_{14} = \langle C B_1 \rangle$, all isomorphic to $\mathbb{Z}_2$, and $G_{34} = G_{123} = G_{124} = 1$. Furthermore,

$$G_{23} = \left\{ \begin{bmatrix} 1 & 0 \\ 1 - t^j & t^j \end{bmatrix} \middle| \, j \in \mathcal{K}_q \right\} = \langle P \rangle^E,$$

$$G_{24} = \left\{ \begin{bmatrix} 1 & 0 \\ 2(t^i - 1) & t^i \end{bmatrix} \middle| \, i \in \mathcal{K}_q \right\} = \langle P \rangle^Q,$$

both of which are isomorphic to $\mathbb{Z}_{q-1}$.

2.6.4.2 Instance 13

$G_2 = K', G_3 = (J')^E, G_4 = (J')^Q$.

When $\frac{q-1}{2}$ is even, G_{12}, G_{13}, G_{14}, and G_{34} are the same as in Instance 12. Otherwise $G_{12} = G_{13} = G_{14} = 1$ and $G_{34} = \langle -I \rangle \cong \mathbb{Z}_2$. G_{123} and G_{124} are always trivial. Also,

$$G_{23} = \left\{ \begin{bmatrix} (-1)^j & 0 \\ (-1)^j - t^j & t^j \end{bmatrix} \middle| \, j \in \mathcal{K}_q \right\} = \langle P' \rangle^E,$$

$$G_{24} = \left\{ \begin{bmatrix} (-1)^i & 0 \\ 2\left(t^i - (-1)^i\right) & t^i \end{bmatrix} \middle| \, i \in \mathcal{K}_q \right\} = \langle P' \rangle^Q,$$

both of which are isomorphic to $\mathbb{Z}_{q-1}$.

2.6.4.3 Instance 14

$G_2 = J^E, G_3 = K, G_4 = K^W$.

Observe that G_2 and G_3 are obtained from swapping the corresponding subgroups from Instance 12. Therefore G_{12} and G_{13} are also swapped while G_{23} remains the same. It turns out that G_{14}, G_{34}, G_{123}, and G_{124} are also the same as in Instance 12. Furthermore,

$$G_{24} = \left\{ \begin{bmatrix} t^i & 1 - t^i \\ 0 & 1 \end{bmatrix} \middle| \, i \in \mathcal{K}_q \right\} = \langle B \rangle^W \cong \mathbb{Z}_{q-1}.$$

2.6.4.4 Instance 15

$G_2 = (J')^E, G_3 = K', G_4 = (K')^W$.

In this case, G_2 and G_3 from Instance 13 are swapped to yield the corresponding subgroups here. So G_{12} and G_{13} are also swapped while G_{23} stays the same. Moreover, G_{14}, G_{34}, G_{123}, and G_{124} are the same as in Instance 13, both when $\frac{q-1}{2}$ is even and otherwise. Finally,

$$G_{24} = \left\{ \begin{bmatrix} t^i & (-1)^i - t^i \\ 0 & (-1)^i \end{bmatrix} \middle| \, i \in \mathcal{K}_q \right\} = \langle B' \rangle^W \cong \mathbb{Z}_{q-1}.$$

2.7 Interpretation and Generalizations of Violation in $PGL(2, q)$ using Theory of Group Actions

Instead of invertible matrices, we can also regard a general linear group as the group of all invertible linear transformations on a vector space. In this section, we

take this point of view and consider the actions of linear groups on their corresponding projective geometries. Such actions induce a permutation representation for each general linear group on its projective geometry, and the projective linear groups are naturally defined in this framework. Using the theory of group actions, we show that the Ingleton violation in $PGL(2, q)$ from Section 2.5 has a nice interpretation: each subgroup is some sort of stabilizer for a set of points in the projective geometry. Furthermore, based on this understanding, we generalize the construction in $PGL(2, q)$ to two new families of Ingleton violations in $PGL(n, q)$ for a general n.[9] Finally, we provide an abstract construction in 2-transitive groups generalizing these ideas.

Throughout this section we assume basic knowledge in the theory of group actions, which can be found in standard group theory textbooks. In particular, we make extensive use of the orbit-stabilizer theorem, which says that the order of the orbit of an element is equal to the index of it stabilizer (see, e.g., [25, Sec. 4.1, Prop. 2]). Most notations are standard abstract algebra notations (see, e.g., [25]); the rest are introduced when they first appear. Note that this section is more abstract than the others and assumes more background knowledge in abstract algebra.

This section is mostly based on Prof. M. Aschbacher's correspondences with us. We have furnished various details and explanations for clarity.

2.7.1 Preliminaries for Linear Groups

Let V be an n-dimensional vector space over a field F. Recall that $GL(V)$ and $SL(V)$ are the general linear group and special linear group on V, respectively. They are examples of groups of Lie type, a notion which is not totally well defined.

Each group G of Lie type possesses a *building*, a simplicial complex on which G is represented as a group of automorphisms. A (abstract) *simplicial complex* consists of a set X of *vertices* together with a collection of nonempty subsets of X called *simplices*; the only axiom says that each nonempty subset of a simplex is a simplex.

Example 2.7.1. Let X be a partially ordered set. The *order complex* of X is the

[9] Note that with Lemma 2.6.1, the families in $PGL(n, q)$ can also be easily extended to families of violations in $GL(n, q)$.

simplicial complex with vertex set X and with the simplices the nonempty chains in the poset.

Example 2.7.2. The *projective geometry* $PG(V)$ of V is the poset of nonzero proper subspaces of V, partially ordered by inclusion. The building of $GL(V)$ and $SL(V)$ is the order complex of this poset. Of course $GL(V)$ permutes the subspaces of V, supplying a representation of $GL(V)$ on $PG(V)$ whose kernel is the subgroup of scalar maps. The images of $GL(V)$ and $SL(V)$ in the automorphism group $Aut(PG(V))$ are the *projective general linear group* $PGL(V)$ and *projective special linear group* $PSL(V)$. Write $GL(n, F)$, $SL(n, F)$, $PGL(n, F)$, $PSL(n, F)$ for the corresponding group when $\dim(V) = n$ and the field is F.

Example 2.7.3. Specialize to the case $n = 2$. Then $PG(V)$ consists of the *points* of V; i.e., the 1-dimensional subspaces of V. This is the so-called *projective line*. Let $\mathcal{X} = \{x_1, x_2\}$ be a basis of V. We regard the projective line as $\Omega = F \cup \{\infty\}$, where ∞ denotes Fx_1 and for $e \in F$, e denotes $F(ex_1 + x_2)$. Then given an invertible matrix

$$M(a, b, c, d) = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

in $GL(V)$, one can check that, subject to the identification of $PG(V)$ with Ω, $M(a, b, c, d)$ acts on Ω via

$$M(a, b, c, d) : x \mapsto \frac{ax + b}{cx + d},$$

where arithmetic involving ∞ is suitably interpreted; e.g., $(a\infty + b)/(c\infty + d) = a/c$ if $c \neq 0$ and ∞ if $c = 0$. So we can regard $PGL(V) = PGL(2, F)$ as the group of these projective linear maps $M(a, b, c, d)$, $ad - bc \neq 0$ on the projective line Ω.

The following result is well known and easy to prove:

Lemma 2.7.1. *$PGL(2, F)$ is sharply 3-transitive on the projective line $PG(V)$. That is, $PGL(V)$ is transitive on ordered 3-tuples of distinct points, and only the identity fixes three points.*

Next we introduce several types of subgroups for these linear groups.

A *Borel subgroup* of a group G of Lie type is the stabilizer of a maximal simplex in its building.

Example 2.7.4. A maximal simplex in $PG(V)$ is a flag

$$\tau = (0 < V_1 < \cdots < V_{n-1} < V),$$

where $\dim(V_k) = k$. If we pick a basis $\mathcal{X} = \{x_1, ..., x_n\}$ for V such that

$$V_k = \langle x_i : 1 \leq i \leq k \rangle,$$

then the Borel subgroup stabilizing τ is the subgroup whose matrices with respect to $\mathcal{X}$ are the upper triangular invertible matrices.

Let $G = PGL(2, F)$. By definition, the stabilizers $G_{Fx_1} = G_\infty$ and $G_{Fx_2} = G_0$ are both Borel subgroups of G. The matrices of these subgroups are upper triangular and lower triangular, respectively. As G is transitive on Ω, for each of $u = \infty, 0$ we have the bijection $gG_u \mapsto g(u)$ of the coset space G/G_u with Ω (by orbit-stabilizer theorem).

Buildings have certain special subcomplexes called *apartments*. For a group G of Lie type, the pointwise stabilizer of an apartment is called a *Cartan subgroup* of G.

Example 2.7.5. In the projective geometry, the apartments are of the form $\Sigma(\mathcal{X})$ for $\mathcal{X} = \{x_1, \cdots, x_n\}$, a basis for V, where $\Sigma(\mathcal{X})$ consists of the subspaces spanned by nonempty proper subsets of $\mathcal{X}$. The matrices in the Cartan subgroup stabilizing $\Sigma(\mathcal{X})$ are the diagonal matrices.

Suppose $n = 2$. Then $\Sigma(\mathcal{X}) = \{Fx_1, Fx_2\} = \{\infty, 0\}$ is just a pair of points. The *global stabilizer* $G(u, v)$ of a pair of points is the subgroup of G permuting the 2-subset $\{u, v\}$. In $G = PGL(2, F)$ it is (usually) the normalizer of the Cartan subgroup and dihedral. Furthermore, $G_0 \cap G(0, \infty) = G_{0,\infty}$ is a Cartan subgroup isomorphic to the multiplicative group $F^\times$ of F.

Let G be $GL(V)$ or $PGL(V)$ in the rest of this section.

An element of $GL(V)$ is *unipotent* if all its eigenvalues are 1. A subgroup of $GL(V)$ is *unipotent* if all its elements are unipotent. The *unipotent radical* $Q(H)$ of a subgroup H of $GL(V)$ is the largest normal unipotent subgroup of H. For example if F is finite of characteristic p, then $Q(H)$ is the largest normal p-subgroup of H. Passing to images in $PGL(V)$, we have the corresponding notions in that group also.

A subgroup H of G is a *parabolic* if H is the stabilizer of a simplex in the projective geometry $PG(V)$. Thus, for example, Borel subgroups are parabolics, and indeed the parabolics are the overgroups of the Borel subgroups.

Example 2.7.6. Let $F = \mathbb{F}_q$, U be an m-dimensional subspace of V with $0 < m < n$, $G = GL(V)$, and $H = N_G(U)$ be the (global) stabilizer of U in G. As $\{U\}$ is a simplex in $PG(V)$, H is a parabolic. Pick a complement W to U in V, and let $\mathcal{X}_1$ and $\mathcal{X}_2$ be bases for U and W, respectively. Then the matrices of H with respect to $\mathcal{X}_1 \cup \mathcal{X}_2$ have the form $\begin{bmatrix} K & L \\ 0 & R \end{bmatrix}$ with K and R invertible. Define

$$q_n = q^{n(n-1)/2}, \quad M_k = \prod_{i=1}^{k} (q^i - 1)$$

for $1 \leq k \leq n$, then

$$|GL(k,q)| = q_k M_k,$$

$$|H| = |GL(m,q)| \cdot |GL(n-m,q)| \cdot q^{m(n-m)} = q_n M_m M_{n-m}.$$

Furthermore, in $PGL(V)$ the image of H has order $q_n M_m M_{n-m}/(q-1)$.

2.7.2 Interpretation of the Ingleton Violation in $PGL(2,q)$

Let $F = \mathbb{F}_q$ and $G = PGL(2,q) = PGL(2,\mathbb{F}_q)$. In the Ingleton violation construction in Section 2.5 we have a 4-tuple of subgroups $\rho = (G_i : 1 \leq i \leq 4)$ of G. The group $G_2 = G_{Fx_2} = G_0$ is a Borel subgroup. The subgroups G_3 and G_4 are isomorphic to the dihedral group $D_{2(q-1)}$ of order $2(q-1)$, and their intersection G_{2i} with G_2 is cyclic of order $q-1$ and with G_{34} of order 1. This forces G_{2i}, $i = 3, 4$,

to be distinct Cartan subgroups of G_2, and hence $G_i = G(0, e_i)$ for some $e_i \in F$. In fact, from the forms of the matrices in G_3 and G_4 it is easy to check that $e_3 = -1$ and $e_4 = \infty$.

Finally, $G_1 \cong S_3$ with G_{1i} being the three subgroups of G_1 of order 2 for $2 \leq i \leq 4$. For $2 \leq i \leq 4$ let $G_{1i} = \langle t_i \rangle$, and for $1 \leq j \leq 4$ let Δ_j be the orbit of G_j on Ω containing 0. Then $|\Delta_j| = |G_j : G_{2j}| = n_j$, where $n_3 = n_4 = 2$ and $n_1 = 3$. Indeed, $\Delta_i = \{0, t_i(0)\}$ for $i = 3, 4$, with $\Delta_3 = \{0, -1\}$ and $\Delta_4 = \{0, \infty\}$. Then as $G_1 = \langle t_3, t_4 \rangle$ and $n_1 = 3$, $\Delta = \Delta_1 = \{0, -1, \infty\}$. But as G is sharply 3-transitive, the global stabilizer $G(\Delta)$ is isomorphic to S_3. Hence $G_1 = G(\Delta)$, and is determined by G_2, G_3, and G_4.

Thus the 4-tuple ρ is determined by the ordered triple $(0, -1, \infty)$ with the four subgroups being various (global) stabilizers on it. Furthermore, given an arbitrary ordered triple (α, β, γ) of distinct points in Ω, we can construct a 4-tuple ρ' in the same fashion, where

$$G_2 = G_\alpha, \quad G_3 = G(\alpha, \beta), \quad G_4 = G(\alpha, \gamma), \quad G_1 = G(\alpha, \beta, \gamma).$$

Since G is 3-transitive on Ω, by the same element in G all four subgroups in ρ' are conjugate to their counterparts in ρ. In particular, the new tuple ρ' also violates Ingleton.

With respect to the "flower structure" of $G_2 = G_0$, this follows from the fact that G_0 is a Frobenius group on $\Omega' = \Omega - \{0\}$. That is, G_0 is a transitive permutation group on Ω' in which the maximum number of fixed points of a nonidentity element is 1. (This is guaranteed by the sharp 3-transitivity of G.) Then by a theorem of Frobenius, the identity 1 of G_0, together with the set of elements with no fixed points, forms a normal subgroup K called the *Frobenius kernel* of the Frobenius group. In our case, K is the subgroup N in Sections 2.4 and 2.5, which is the unipotent radical of the Borel subgroup G_0 and is isomorphic to the additive group of the field F. Also $G_0 - K$ is partitioned by the sets $G_{0,a} - \{1\}, a \in \Omega'$; these are the $|\Omega'| = q$ petals in the flower. The subgroups $G_{0,a}$ are the q Cartan subgroups contained in G_0, and

each is isomorphic to $F^{\times}$.

2.7.3 Generalizations in $PGL(n, q)$

Let $\tau = (G_i : 1 \leq i \leq 4)$ be a family of subgroups of a finite group G. The Ingleton inequality (2.4) fails iff

$$|G_1 G_2| < \frac{|G_{13} G_{23}||G_{14} G_{24}|}{|G_{34}|}.$$

In all constructions we will consider in this section, $G_i = G_{1i} G_{2i}$ for $i = 3, 4$ and $|G_3| = |G_4|$. Also $|G_1 G_2| = |G_1 : G_{12}||G_2|$. Hence in such constructions Ingleton is violated iff

$$|G_1 : G_{12}||G_2| < \frac{|G_3|^2}{|G_{34}|}, \tag{2.14}$$

and the Ingleton ratio (2.13) becomes

$$r(\tau) = \frac{|G_3|^2}{|G_1 : G_{12}||G_2||G_{34}|}.$$

Now we explore three different approaches, all trying to extend the $PGL(2, q)$ family of violations ρ to $PGL(n, q)$.

2.7.3.1 Generalization 1

Let $G = PGL(n, q)$ with $n \geq 3$. It is easy to see that G is doubly transitive on the points of $PG(V)$ and transitive on triples of independent points. Let P_i, $2 \leq i \leq 4$, be independent points in V, $\Delta_i = \{P_2, P_i\}$ for $i = 3, 4$, and $\Delta = \{P_2, P_3, P_4\}$. Set $G_2 = N_G(P_2)$, $G_i = N_G(\Delta_i)$, $i = 3, 4$, and $G_1 = N_G(\Delta)$. Let $\tau = (G_i : 1 \leq i \leq 4)$.

Now G_2 is a parabolic and by Example 2.7.6,

$$|G_2| = q_n M_{n-1}. \tag{2.15}$$

Next $D = P_2 + P_3 + P_4$ is a 3-dimensional subspace of V, so by Example 2.7.6 again, $|N_G(D)| = q_n M_3 M_{n-3}/(q-1)$. Further, through calculation of the preimages in

$GL(n,q)$ we have

$$|N_G(D) : G_1| = \frac{|GL(3,q)|}{6(q-1)^3} = \frac{q^3 M_3}{6(q-1)^3},$$

since G_1 acts as the symmetric group on Δ of order 3, and for each pair of points there are $q-1$ different choices of mappings. So

$$|G_1| = \frac{|N_G(D)| \cdot 6(q-1)^3}{q^3 M_3} = \frac{6q_n M_{n-3}(q-1)^2}{q^3}. \tag{2.16}$$

As G_1 is transitive on Δ of order 3, $|G_1 : G_{12}| = 3$. Therefore

$$|G_1 : G_{12}||G_2| = 3|G_2| = 3q_n M_{n-1}. \tag{2.17}$$

Also for $i = 3, 4$, G_i and G_{1i} are both transitive on Δ_i of order 2, so $|G_i : G_{2i}| = |G_{1i} : G_{12i}| = 2$. Thus $|G_{1i}G_{2i}| = |G_{1i} : G_{12i}||G_{2i}| = |G_i|$ and $G_i = G_{1i}G_{2i}$ for $i = 3, 4$. Since G is doubly transitive on the points, G_3 is conjugate to G_4 and so $|G_3| = |G_4|$. Further $U = P_2 + P_3$ is a 2-dimensional subspace of V, so by Example 2.7.6, $|N_G(U)| = q_n M_2 M_{n-2}/(q-1)$. Also by calculating the preimages

$$|N_G(U) : G_3| = \frac{|GL(2,q)|}{2(q-1)^2} = \frac{qM_2}{2(q-1)^2},$$

$$|G_3| = \frac{|N_G(U)| \cdot 2(q-1)^2}{qM_2} = \frac{2q_n M_{n-2}(q-1)}{q}. \tag{2.18}$$

Finally, $G_{34} = G_\Delta$ is the pointwise stabilizer of Δ. Since G_1 is 3-transitive on Δ, $|G_1 : G_{34}| = 3! = 6$. So by (2.16):

$$|G_{34}| = \frac{q_n M_{n-3}(q-1)^2}{q^3}. \tag{2.19}$$

It follows from (2.17), (2.18), and (2.19) that (2.14) is satisfied iff

$$3q_n M_{n-1} < \frac{4q_n^2 M_{n-2}^2 (q-1)^2 \cdot q^3}{q^2 \cdot q_n M_{n-3}(q-1)^2} = 4q_n q M_{n-2}(q^{n-2} - 1),$$

which holds iff $3(q^{n-1}-1) < 4q(q^{n-2}-1)$ iff

$$q^{n-1} - 4q + 3 > 0. \tag{2.20}$$

This inequality holds when $n \geq 4$ or $n = 3$ and $q \geq 4$.

Since G is transitive on all triples of independent points, all 4-tuples in this generalization are conjugate to each other.

The Ingleton ratio is

$$r(\tau) = \frac{4q_n^2 M_{n-2}^2 (q-1)^2 \cdot q^3}{q^2 \cdot 3q_n M_{n-1} \cdot q_n M_{n-3}(q-1)^2} = \frac{4q(q^{n-2}-1)}{3(q^{n-1}-1)},$$

which approaches $4/3$ for large q or n; whereas in the original instance ρ, $r(\rho) = 4(q-1)/(3q)$, which has the same asymptotics. But the scaling factors for both the Ingleton score and the violation index are usually larger than $PGL(2,q)$, so in general τ is less effective in violating Ingleton.

2.7.3.2 Generalization 2

As usual let $F = \mathbb{F}_q$ and $G = PGL(n,q)$, with $n \geq 2$. Let P_i, $2 \leq i \leq 4$, be distinct but dependent points in V. Thus $P_i = Fx_i$, $i = 2,3$, for two independent vectors $x_2, x_3 \in V$, and $P_4 = Fx_4$, where $x_4 = ex_2 + x_3$ for some $e \in F$. Let U, Δ, Δ_i, $i = 3,4$, and G_i, $1 \leq i \leq 4$, be defined the same as in Generalization 1. Note that when $n = 2$ this is our original construction ρ.

From Generalization 1, $|G_2| = q_n M_{n-1}$ and $|N_G(U)| = q_n M_2 M_{n-2}/(q-1)$. Since U is a 2-dimensional subspace of V, $PGL(U)$ is sharply 3-transitive on the points of U by Lemma 2.7.1. Now as Δ is a set of three distinct points in U, its global stabilizer in $PGL(U)$ is isomorphic to S_3. Thus G_1 is 3-transitive on Δ. Observe that each vector in $\{x_i : 2 \leq i \leq 4\}$ is a unique linear combination of the other two, with both coefficients nonzero. Then, fixing a permutation of $\{P_i : 2 \leq i \leq 4\}$, there are only

$q-1$ linear transformations in $GL(U)$ that respect this permutation. Hence

$$|N_G(U) : G_1| = \frac{|GL(2,q)|}{6(q-1)} = \frac{qM_2}{6(q-1)},$$

$$|G_1| = \frac{|N_G(U)| \cdot 6(q-1)}{qM_2} = \frac{6q_nM_{n-2}}{q}. \tag{2.21}$$

G_1 is transitive on Δ, while for $i = 3, 4$, G_i and G_{1i} are both transitive on Δ_i. G is doubly transitive on the points of $PG(V)$. Thus from arguments in Generalization 1 we have $|G_1 : G_{12}||G_2| = 3q_nM_{n-1}$, $G_i = G_{1i}G_{2i}$ for $i = 3, 4$, and $|G_3| = |G_4|$. Also $|G_3| = 2q_nM_{n-2}(q-1)/q$. Since $G_{34} = G_\Delta$ is of index 6 in G_1, by (2.21):

$$|G_{34}| = \frac{q_nM_{n-2}}{q}.$$

Thus (2.14) is satisfied iff

$$3q_nM_{n-1} < \frac{4q_n^2M_{n-2}^2(q-1)^2 \cdot q}{q^2 \cdot q_nM_{n-2}} = \frac{4q_nM_{n-2}(q-1)^2}{q},$$

which holds iff $3q(q^{n-1}-1) < 4(q-1)^2$ iff

$$3q\sum_{i=0}^{n-2} q^i - 4q + 4 < 0. \tag{2.22}$$

When $n = 2$, this inequality holds iff $q > 4$. When $n > 2$, however, it always fails because $3q^2 - q + 4 > 0$ for all q.

Therefore, the original instance ρ is the only successful case in this construction, with Ingleton ratio $r(\rho) = 4(q-1)/(3q)$.

2.7.3.3 Generalization 3

Again take $G = PGL(n,q)$ with $n \geq 3$. Let U_2 be a point of V, U_i, $i = 3, 4$, distinct 2-dimensional subspaces of V with $U_3 \cap U_4 = U_2$, and $U_1 = U_3 + U_4$ the 3-dimensional subspace of V generated by U_3 and U_4. Set $G_i = N_G(U_i)$ for $1 \leq i \leq 4$, and $\lambda = (G_i : 1 \leq i \leq 4)$. Then all the G_i are parabolics with $|G_2| = q_nM_{n-1}$ from

(2.15),

$$|G_3| = |G_4| = \frac{q_n M_2 M_{n-2}}{q-1}, \qquad |G_1| = \frac{q_n M_3 M_{n-3}}{q-1}.$$

As G_1 is transitive on the $(q^3-1)/(q-1) = q^2+q+1$ points in U_1, so

$$|G_1 : G_{12}| = q^2+q+1, \qquad |G_1 : G_{12}||G_2| = (q^2+q+1)q_n M_{n-1}.$$

For $i = 3, 4$, G_i and G_{1i} are both transitive on the $(q^2-1)/(q-1) = q+1$ points in U_i, so $G_i = G_{1i}G_{2i}$ for $i = 3, 4$. Also G_{34} is the subgroup of G fixing U_2 and the points U_3/U_2 and U_4/U_2 of the quotient space U_1/U_2; in particular it is a subgroup of G_1. If we pick a basis $\mathcal{X}_1 = \{x_3, x_2, x_4\}$ for U_1 such that $U_2 = \langle x_2 \rangle$ and $U_i = \langle x_2, x_i \rangle$ for $i = 3, 4$, then elements of G_{34} correspond to the linear transformations in $GL(U_1)$ whose matrices with respect to $\mathcal{X}_1$ take the form

$$\begin{bmatrix} a & 0 & 0 \\ x & b & y \\ 0 & 0 & c \end{bmatrix},$$

where a, b, and c are nonzero. So

$$|G_1 : G_{34}| = \frac{|GL(3,q)|}{q^2(q-1)^3} = \frac{qM_3}{(q-1)^3},$$

$$|G_{34}| = \frac{|G_1|}{qM_3/(q-1)^3} = \frac{q_n M_3 M_{n-3} \cdot (q-1)^3}{(q-1) \cdot qM_3} = \frac{q_n M_{n-3}(q-1)^2}{q}.$$

It follows that (2.14) is satisfied iff

$$(q^2+q+1)q_n M_{n-1} < \frac{q_n^2 M_2^2 M_{n-2}^2 \cdot q}{(q-1)^2 \cdot q_n M_{n-3}(q-1)^2} = q_n q(q+1)^2(q^{n-2}-1)M_{n-2},$$

which holds iff $(q^2+q+1)(q^{n-1}-1) < q(q+1)^2(q^{n-2}-1)$ iff

$$q^n - q^3 - q^2 + 1 > 0,$$

which holds iff $n \geq 4$.

The Ingleton ratio is

$$r(\lambda) = \frac{q_n^2 M_2^2 M_{n-2}^2 \cdot q}{(q-1)^2 \cdot (q^2+q+1) q_n M_{n-1} \cdot q_n M_{n-3} (q-1)^2} = \frac{q(q+1)^2(q^{n-2}-1)}{(q^2+q+1)(q^{n-1}-1)},$$

which approaches 1 for large q and $(q+1)^2/(q^2+q+1)$ (which is smaller than $4/3$) for large n. So this generalization seems less effective than the other two.

2.7.4 Generalizations in General 2-transitive Groups

In the following we generalize the Ingleton violation ρ in $PGL(2,q)$ to a more abstract construction, which includes Generalizations 1 and 2 as special cases.

Let G be a doubly transitive group on a set Ω of order $l \geq 3$, let α and β be distinct points in Ω, and assume $\gamma \in \Omega - \{\alpha, \beta\}$ such that the global stabilizer $G(\Delta)$ of $\Delta = \{\alpha, \beta, \gamma\}$ acts as the symmetric group on Δ (which is clearly the case when G is 3-transitive). Let

$$G_2 = G_\alpha, \quad G_3 = G(\alpha, \beta), \quad G_4 = G(\alpha, \gamma), \quad G_1 = G(\Delta).$$

Set $\mu = (G_i : 1 \leq i \leq 4)$.

Let $k = \left|G_{\alpha,\beta}\right|$, $d = |G_\Delta|$, Γ the orbit of γ under the action of $G_{\alpha,\beta}$, and $c = |\Gamma|$. Observe that $c = |G_{\alpha,\beta} : G_\Delta| = k/d$ and $c \leq l-2$ as $\Gamma \subseteq \Omega - \{\alpha, \beta\}$. Further $c = l-2$ iff G is 3-transitive.

Since G is 2-transitive on Ω, G_2 is transitive on $\Omega - \{\alpha\}$, and so $|G_2 : G_{\alpha,\beta}| = l-1$. Also $|G_1 : G_{12}| = 3$ as G_1 is transitive on Δ, thus

$$|G_1 : G_{12}||G_2| = 3|G_2| = 3(l-1)k.$$

Next, G_3 is conjugate to G_4 by 2-transitivity of G and for $i = 3, 4$, G_i and G_{1i} are both transitive on Δ_i of order 2, so $G_{1i}G_{2i} = G_i$ and $|G_i| = 2k$ for $i = 3, 4$. Finally $G_{34} = G_\Delta$ is of order d. Thus

$$|G_3|^2/|G_{34}| = 4k^2/d = 4kc,$$

so condition (2.14) is satisfied iff $3(l-1)k < 4kc$ iff

$$3(l-1) < 4c. \tag{2.23}$$

Further the Ingleton ratio $r(\mu) = 4c/(3(l-1))$.

If G is 3-transitive then $c = l-2$, so $3(l-1) < 4c = 4(l-2)$ iff $l > 5$. Further $r(\mu) = 4(l-2)/(3(l-1))$.

Both Generalization 1 and 2 fit in this construction, with ρ being the only 3-transitive case. In Generalization 1, $l = (q^n-1)/(q-1)$ and by independence of points in Δ,

$$c = \frac{(q^n-1)-(q^2-1)}{q-1} = \frac{q^2(q^{n-2}-1)}{q-1},$$

so by (2.23), (2.14) is satisfied iff

$$3(\frac{q^n-1}{q-1}-1) < \frac{4q^2(q^{n-2}-1)}{q-1},$$

which gives (2.20). In Generalization 2, l has the same value, but since $GL(U)$ is 3-transitive on the $(q^2-1)/(q-1) = q+1$ points of U, $c = q+1-2 = q-1$. Then by (2.23), (2.14) is satisfied iff

$$3(\frac{q^n-1}{q-1}-1) < 4(q-1),$$

which gives (2.22).

We see that the 3-transitive groups give rise to simple and effective Ingleton violation constructions. This category of groups include the alternating and symmetric groups, the groups $PGL(2,q)$ with $l = q+1$, the Mathieu groups, the affine groups of degree 2^e (which are the semidirect product of an e-dimensional vector space E over $\mathbb{F}_2$ by $GL(E)$), and the subgroup of the affine group for $e = 4$, where the complement is A_7 rather than $GL(4,2) \cong A_8$.

Chapter 3

Group Network Codes

We can use the Ingleton-violating groups obtained in the previous chapter to construct network codes that have the potential to have a better performance than linear network codes. However, we will see that designing such a desirable code is not a trivial task, since, in order to construct a network code from a given group, the choice of the subgroups is subject to certain constraints. In this chapter we first give the definitions of general network codes and group network codes, together with some discussion of these concepts, and then study some aspects of the network code construction for our Ingleton-violating groups.

3.1 Definitions

A communication network is usually represented by a directed acyclic graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the node set $\mathcal{V}$ and the edge set $\mathcal{E}$ model the communication nodes and channels, respectively. Let $\mathcal{S} \subset \mathcal{V}$ be the set of source nodes and $\mathcal{D}(s)$ be the set of sink nodes demanding source s for each $s \in \mathcal{S}$. For any node v and any edge e, $\mathcal{I}(v)$ and $\mathcal{I}(e)$ denote the sets of incoming edges to v and to the tail node of e, respectively.

A network code should include

1) the assignment of a symbol Y_s from some alphabet $\mathcal{Y}_s$ for a source message at each source s;
2) the encoding of a symbol Y_e in some alphabet $\mathcal{Y}_e$ at each edge e, from the symbols on $\mathcal{I}(e)$. Namely, $Y_e = \phi_e\left(Y_f : f \in \mathcal{I}(e)\right)$ for some deterministic encoding

function ϕ_e;

3) the decoding of the symbol Y_s at each $u \in \mathcal{D}(s)$ for all sources s, i.e., Y_s is uniquely determined from the symbols on $\mathcal{I}(u)$: $Y_s = \phi_{u,s}(Y_f : f \in \mathcal{I}(u))$ for some decoding function $\phi_{u,s}$.

It is clear that at each edge e the symbol Y_e is a deterministic function of the source symbols $\{Y_s : s \in \mathcal{S}\}$, which is denoted by φ_e and is called the *global mapping* at e. Also the source random variables $\{Y_s : s \in \mathcal{S}\}$ are usually assumed to be independent and uniform on their respective alphabets.

For example, a linear network code is defined as follows: 1) for each $t \in \mathcal{S} \cup \mathcal{E}$, the alphabet $\mathcal{Y}_t$ is a vector space F^{d_t} over a finite field F with some finite dimension d_t; 2) all encoding/decoding functions are linear: if t is an edge or a sink node, then the encoding/decoding function ϕ_t at t can be written as

$$\phi_t\left(Y_f : f \in \mathcal{I}(t)\right) = \sum_{f \in \mathcal{I}(t)} M_{t,f} Y_f$$

for some matrices $M_{t,f} \in F^{d_t} \times F^{d_f}$. Thus the global mappings at the edges are also linear.

Group network codes were first proposed by Chan in [33, 34], where the author considered the fact that finite groups can generate the whole entropy region, and noted that linear network codes are included as a special case. Suppose G is a finite group, and $\{G_e : e \in \mathcal{E}\}$ and $\{G_s : s \in \mathcal{S}\}$ are some of its subgroups. One can construct a network code with $\mathcal{Y}_t = G/G_t$ for each $t \in \mathcal{S} \cup \mathcal{E}$ if the following requirements are met:

(R1) *Source independence*: $H(Y_{\mathcal{S}}) = \sum_{s \in \mathcal{S}} H(Y_s)$, which means that the cardinalities of $G/G_{\mathcal{S}}$ and $\prod_{s \in \mathcal{S}} \mathcal{Y}_s$ (the Cartesian product of the source alphabets) are equal, where $G_{\mathcal{S}} \triangleq \bigcap_{s \in \mathcal{S}} G_s$. This is equivalent to

$$\prod_{s \in \mathcal{S}} |G_s| = |G|^{|\mathcal{S}|-1} |G_{\mathcal{S}}|.$$

(R2) *Encoding*: $\forall e \in \mathcal{E}$, $\bigcap_{f \in \mathcal{I}(e)} G_f \leq G_e$.

(R3) *Decoding*: $\forall s \in \mathcal{S}$, $\bigcap_{f \in \mathcal{I}(u)} G_f \leq G_s$ for each $u \in \mathcal{D}(s)$.

Note that the source and edge symbols for the group network code are (left) cosets. The encoding and decoding operations are as follows: at an edge or a sink node t, the encoding/decoding function takes an input coset tuple $(Y_f : f \in \mathcal{I}(t))$ and first forms the intersection of them, which is a coset of $G_{\mathcal{I}(t)}$, then maps this coset to the unique coset of G_e (or G_s, whichever is appropriate) that contains it. For a rigorous justification of the validity of such operations, and of the code being a valid network code, see Section A.2.1 in Appendices.

We also make two observations regarding group network codes. First, the entropy vector for the network symbols $\{Y_t : t \in \mathcal{S} \cup \mathcal{E}\}$ is characterizable by the group G and its subgroups $\{G_t : t \in \mathcal{S} \cup \mathcal{E}\}$. Second, linear network codes are a special case of group network codes; in particular, for each linear network code we can construct an equivalent group network code. These observations are elaborated in Sections A.2.2 and A.2.3 in Appendices, respectively.

3.2 Considerations for Constructing Group Network Codes with Ingleton-violating Groups

We can use our Ingleton-violating groups to build group network codes. From the first observation above, the resulting entropy vectors are characterizable by the subgroups used, and are thus capable of violating the Ingleton inequality. In contrast, the entropy vectors of linear network codes always respect Ingleton. Furthermore, let G be any of $PGL(n,p)$, $PGL(n,q)$, $GL(n,p)$, or $GL(n,q)$. We will show in the following that linear network codes can be embedded in the group network codes constructed with direct products of copies of G. Apparently a direct product of any copies of an Ingleton-violating group still violates Ingleton, and thus such classes of group network codes are strictly more powerful than linear network codes.

To construct a group network code, the choices of subgroups are not arbitrary: they should meet requirements (R1)–(R3). In particular, (R1) limits what subgroups

can be associated with the sources: they need to satisfy

$$\prod_{s\in\mathcal{S}} |G_s| = |G|^{|\mathcal{S}|-1}|G_{\mathcal{S}}|. \tag{3.1}$$

When this is the case, we simply say the subgroups $\{G_s : s \in \mathcal{S}\}$ are independent in G. We will study the constructions of independent source subgroups in the context of $PGL(2,q)$ and $GL(2,q)$ (since they have simpler structures than the other higher-degree linear groups), and also provide a universal source subgroup construction for direct products of groups.

3.2.1 Embeddings of Linear Network Codes

As observed above and elaborated in Section A.2.3 in Appendices, linear network codes are a special type of group network codes. In particular, they are determined by the underlying additive group structure. The direct sum V of source vector spaces can be called the *ambient vector space* of a linear network code. Let $(V,+)$ denote the additive group of V. If we can find a finite group G such that $(V,+) \leq G$, then the linear network code is said to be *embedded* in the group network codes using G, since we can use subgroups of G to construct an equivalent group network code.

Consider a linear network code with ambient vector space $V = \mathbb{F}_q^d$ for some d and q, where $q = p^m$ for some prime p and some integer m. Observing that $\mathbb{F}_q$ is an m-dimensional vector space over $\mathbb{F}_p$, we can establish the following facts:

i) $(\mathbb{F}_p,+) \cong \mathbb{Z}_p$,

ii) $(\mathbb{F}_q,+) \cong (\mathbb{F}_p,+)^m \cong \mathbb{Z}_p^m$,

iii) $(V,+) \cong (\mathbb{F}_q,+)^d \cong \mathbb{Z}_p^{md}$.

Thus $(V,+)$ is embedded in the direct product of $m \cdot d$ copies of a group G, provided that G contains an element of order p—by Cauchy's theorem, this condition is equivalent to p divides $|G|$. It then follows that linear network codes over $\mathbb{F}_q$ are embedded in the group network codes using direct products of copies of G^m. In particular, let G be any of the linear groups $PGL(2,p)$, $PGL(2,q)$, $GL(2,p)$, or $GL(2,q)$. We have

the following embeddings in these groups, using properties of the matrix A and the subgroup N:

1) In $PGL(2,p)$, $\left|\overline{A}\right| = p$. So $(V,+) \cong \left\langle \overline{A} \right\rangle^{md} \leq PGL(2,p)^{md}$.
2) In $GL(2,p)$, $|A| = p$. So $(V,+) \cong \langle A \rangle^{md} \leq GL(2,p)^{md}$.
3) In $PGL(2,q)$, $N = \left\{ \overline{A_\alpha} \,\middle|\, \alpha \in \mathbb{F}_q \right\} \cong \mathbb{Z}_p^m$. So $(V,+) \cong N^d \leq PGL(2,q)^d$.
4) In $GL(2,q)$, $N = \{ A_\alpha | \, \alpha \in \mathbb{F}_q \} \cong \mathbb{Z}_p^m$. So $(V,+) \cong N^d \leq GL(2,q)^d$.

Therefore, we also have the corresponding network code embeddings. Furthermore, these results for the degree-2 linear groups are easily extended to degree n, since the former are subgroups of the latter.

3.2.2 Sources Independence Requirement Considerations

If we want to utilize the Ingleton-violating groups $PGL(2,q)$ and $GL(2,q)$ to construct network codes, we need to find their independent subgroups. GAP searching shows that up to conjugation, $PGL(2,5)$ has 16 independent pairs of subgroups, 1 triple, and no quadruple. For $GL(2,5)$, the numbers are 86, 14, and 0, respectively. It might be desirable to use some of the Ingleton-violating subgroups as sources, but we find no independent pairs in any violation instance in either $PGL(2,5)$ or $GL(2,5)$. Furthermore, we can prove the following negative results:

Lemma 3.2.1. *Let $i,j \in \{1,2,3,4\}$ and $(i,j) \neq (3,4)$. For four random variables X_1, X_2, X_3, and X_4, if X_i and X_j are independent, then the Ingelton inequality* (2.3) *is satisfied.*

Proof. By symmetry of (2.3), we only need to prove the result for when $(i,j) = (1,2)$ or $(1,3)$. In the first case, $h_{12} = h_1 + h_2$, so

$$\begin{aligned} h_{12} + h_{13} + h_{14} + h_{23} + h_{24} &\geq h_1 + h_2 + h_3 + h_{123} + h_4 + h_{124} \\ &\geq h_1 + h_2 + h_{34} + h_{123} + h_{124}, \end{aligned}$$

where we used $h_{13} + h_{23} \geq h_3 + h_{123}$ and $h_{14} + h_{24} \geq h_4 + h_{124}$ by submodularity of entropy. The second case is similar. □

Corollary 3.2.1. *There is no independent triple or quadruple in a set of four subgroups that violates* (2.4)*.*

On another note, if we want to use the Ingleton-violating subgroups in the network, Proposition A.2.2 in Appendix A.2 tells us that their intersection should contain the intersection of all the source subgroups. Since in $PGL(2,q)$ the intersection of the Ingleton-violating subgroups is trivial, we need to find trivially intersecting independent subgroups to serve as sources. In $PGL(2,5)$, there are 4 such pairs and no such triples. At least one of these pairs also extends to a general family:

Proposition 3.2.1. *Let* $U = \begin{bmatrix} 0 & -1 \\ t & 0 \end{bmatrix} \in GL(2,q)$*, where* t *is a primitive element in* $\mathbb{F}_q$*. Let* H *be the image of* $SL(2,q)$ *in* $PGL(2,q)$ *under the natural homomorphism, which is isomorphic to* $PSL(2,q)$*. When* $p \neq 2$*,* H *and* $\langle \overline{U} \rangle$ *are independent in* $PGL(2,q)$ *with trivial intersection.*

Proof. It is easy to see that $\left|\overline{U}\right| = 2$, $\det U = t$. The determinant of any matrix representing an element in H takes the form $t^{2k} \in \langle t^2 \rangle$, for some k. But $t \notin \langle t^2 \rangle$ as $q-1$ is even, so $H \bigcap \langle \overline{U} \rangle = 1$. Also

$$\left|\langle \overline{U} \rangle\right| \cdot |H| = 2 \cdot |SL(2,q)|/2 = |SL(2,q)| = |PGL(2,q)|,$$

and thus (3.1) holds. □

In $GL(2,q)$ there are more Ingleton-violating instances, which have various intersections, so the requirement on the sources is not so strict and we have a richer class of subgroups to work with. As in $PGL(2,q)$, there exist trivially intersecting independent pairs, for example:

Proposition 3.2.2. *In* $GL(2,q)$*,* $SL(2,q)$ *and* $\langle B \rangle$ *(or* $\langle P \rangle$*) are independent with trivial intersection.*

Proof. Obviously $\det B^k = 1$ iff $B^k = I$, so $SL(2,q)$ and $\langle B \rangle$ have trivial intersection. Also

$$|B| \cdot |SL(2,q)| = (q-1) \cdot |GL(2,q)|/(q-1) = |GL(2,q)|,$$

and thus (3.1) is satisfied. The proof for $\langle P \rangle$ is similar. □

In general it is not easy to find many independent subgroups in a group. If the group is a direct product of n of its subgroups, however, it admits a natural construction of n independent subgroups:

Proposition 3.2.3. *If $G = G_1 \times G_2 \times \cdots \times G_n$, then*

$$1 \times G_2 \times \cdots \times G_n, \qquad G_1 \times 1 \times \cdots \times G_n, \qquad \ldots, \qquad G_1 \times G_2 \times \cdots \times 1$$

are n trivially intersecting independent subgroups in G.

Proof. Trivial intersection is obvious, and it is easy to check that both sides of (3.1) are equal to $\prod_{i=1}^{n} |G_i|^{n-1}$. □

This construction is the generalization of the source construction for linear network codes, in which case the subgroup at source s is the W_s defined in Appendix A.2.3. Also we see that by using direct products we can obtain independent subgroups for an arbitrary number of sources, but the group order also grows.

If we further require the sources to be of the same alphabet size, then the independent subgroups must have the same order. In the above proposition, this can be simply achieved by choosing G_i to be the same subgroup for each i. Additionally, for an arbitrary pair of independent subgroups, we have the following proposition:

Proposition 3.2.4. *If G_s and G_r are independent in G, then $G_s \times G_r$ and $G_r \times G_s$ are independent in G^2 with the same order.*

Proof. G_s and G_r satisfy $|G_s||G_r| = |G||G_s \bigcap G_r|$. Thus for the direct product construction, the LHS and RHS of (3.1) are $|G_s|^2|G_r|^2$ and $|G|^2|G_s \bigcap G_r|^2$, respectively, which are equal. □

Part II

Energy Harvesting Systems & Channels with Causal CSIT

Chapter 4

Energy Harvesting Channels and FSC-X

4.1 Introduction

In many future wireless systems, such as low-power wireless sensor networks, one may encounter transmitters that harvest and store energy for transmission. Such communication systems have recently been introduced and studied by Ulukus et al. [35, 36]. In particular [37] shows that with unlimited battery the entire capacity of an additive white gaussian noise (AWGN) channel can be achieved. Using Shannon's method [38], [39] tries to analyze the AWGN channel capacity in the zero battery case; however, the proof is incomplete.[1] Nevertheless, treating such a case with a discrete channel is elementary. The intermediate case, i.e., the case with a finite nonzero battery, is first considered in [40], where the optimum offline transmission policy for an energy harvesting node is obtained. However, in general, determining the channel capacity in such a case remains open. For the simplest case of a unit-sized battery, [41] assumes that the transmitter only uses the causal battery state information (see Section 4.2.1) and transforms the system of a binary energy harvesting transmitter connected to a noiseless discrete channel into a timing channel. Using related results, a capacity formula involving an auxiliary random variable is derived and upper and lower bounds are obtained. [42] explores the continuous case with an AWGN channel

[1] Two major problems of the proof in [39] are that (15) cannot be analytically extended to $\mathbb{C}^2$, while (18) cannot be implied by the identity theorem on $\mathbb{C}^2$.

and provides upper and lower bounds that have a constant gap. With the assumptions that the transmitter only uses causal battery state information and that the receiver also has the energy information, [43] studies the general discrete case. Assuming the results of the references [4] and [5] therein can be generalized to finite state channels (defined in [2]) with input constraints, [43] suggests that the system may have a single-letter capacity formula under some extra assumptions.

In this work we study the capacity of a discrete energy harvesting channel with a finite battery in its full generality. First, in the rest of this chapter, we describe the channel models for two different energy information scenarios, as well as a related finite state channel based model. After transforming them to certain equivalent channels, we give the channel capacities in terms of a multi-letter capacity formula using the Verdú-Han general framework [1]. In Chapter 5 we impose some simplifying restrictions on the inputs of the equivalent channels, and derive the required stationarity and ergodicity conditions to use the Shannon-McMillan-Breiman theorem to compute some achievable rates, which are lower bounds for the channels. Then the generalized Blahut-Arimoto algorithm [44] is used to optimized these lower bounds. For the capacity upper bounds, in Chapter 6 we assume that channel side information is also known at the receiver, and use Gallager's methods [2] to obtain upper bounds in terms of block mutual information for each block size. These bounds have high computational complexity, and hence we relax them further to make the complexity linear. Finally, in Chapter 7 we study the pairwise error probabilities for the EHC under maximum likelihood (ML) decoding, which serves as a guideline for the code design.

4.1.1 Notation

In Part II of this dissertation we use the following notational conventions:

- For random variables:
  - capital letters denote the random variables, e.g., X_n, Y_n.
  - corresponding lowercase letters denote the realizations, e.g., x_n, y.

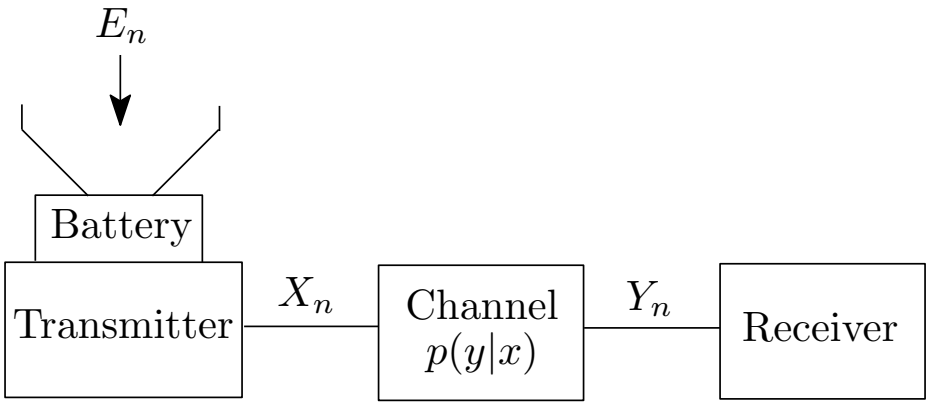


Figure 4.1: Energy harvesting system model

  - corresponding script letters denote the alphabets, e.g., $\mathcal{X}$, $\mathcal{Y}$.

- A vector of generic symbols $(z_m, z_{m+1}, \cdots, z_{m+n})$ is denoted by z_m^{m+n}, whereas $z^n \triangleq z_1^n$.
- Bold symbols denote vectors, e.g., $\boldsymbol{e}_i$, $\boldsymbol{v}_k$.
- $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function:

$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{o.w.} \end{cases}.$$

  When A is the solution set of an equation $f(x) = 0$, we write $\mathbf{1}_{\{f(x)=0\}}$ for succinctness.

- $\{\cdot\}_{n=n_1}^{n_2}$ denotes a sequence of symbols, indexed by n. For example, $\{E_n\}_{n=1}^{\infty}$ denotes the random process $E_1, E_2, \ldots, E_n, \ldots$ To be concise we sometimes drop the sub-/super- scripts and just write $\{E_n\}$ when the context is clear.

4.2 System Model, Two Scenarios and FSC-X

We consider a communication system powered by some energy harvesting mechanism with a battery, as depicted in Figure 4.1. At each transmission cycle n, the system first harvests some amount of energy, E_n, from the environment, and combines it with B_n, the energy stored in the battery after last transmission, to transmit

a symbol $X_n \in \mathcal{X}$. X_n consumes some amount of energy $\gamma(X_n)$, which cannot exceed the total available energy S_n for the current cycle. The remainder, not exceeding the battery capacity $\overline{B}$, is saved in the battery for future transmissions. The symbol X_n is sent over the channel $p(y|x)$ and at the receiver a symbol $Y_n \in \mathcal{Y}$ is received. The alphabets $\mathcal{X}$ and $\mathcal{Y}$ are assumed to be finite with $\mathcal{X} \subset \mathbb{R}$ or $\mathbb{C}$, and the channel is discrete memoryless.

To be precise, the energy constraint on the system can be written as

$$\begin{cases} S_n &= S(B_n, E_n) \\ \gamma(X_n) &\leq S_n \\ B_{n+1} &= \min\left\{ S_n - \gamma(X_n),\ \overline{B} \right\} \end{cases}, \tag{4.1}$$

where the total available energy S_n is expressed as a function S of the battery energy B_n and the harvested energy E_n. The form of $S(\cdot)$ depends on how the system combines and utilizes B_n and E_n. For example, if E_n is immediately available for transmission, then simply

$$S(B_n, E_n) = B_n + E_n. \tag{4.2}$$

However, if the system can only use E_n to charge the battery and draws energy solely from the battery for transmission, then

$$S(B_n, E_n) = \min\left\{ B_n + E_n,\ \overline{B} \right\}. \tag{4.3}$$

This energy model can also take account of more real world influences. For example, if the battery is inefficient at charging and has leakage, characterized by the ratios η and β, respectively, then the model (4.3) becomes

$$S(B_n, E_n) = \min\left\{ \beta B_n + \eta E_n,\ \overline{B} \right\}.$$

In view of the expression of B_{n+1} in (4.1), for $n \geq 1$ sometimes we also write

$$S_{n+1} = S(X_n, S_n, E_{n+1}) \tag{4.4}$$

Table 4.1: Energy harvesting channel notations

Symbol	Definition	Alphabet
E_n	Energy harvested between $(n-1)$-th and n-th transmission	$\mathcal{E}_H$
B_n	Energy stored in the battery after $(n-1)$-th transmission	$\mathcal{E}_B$
S_n	Energy available for n-th transmission	$\mathcal{S}$
X_n	Symbol transmitted at time n	$\mathcal{X}$
Y_n	Symbol received at time n	$\mathcal{Y}$
$\overline{B}$	Battery capacity limit	-
γ	Energy cost function	-

to emphasize the evolution of the $\{S_n\}$ process.

The energy cost function $\gamma(\cdot)$, in general, can be any non-negative function on the alphabet $\mathcal{X}$. However, in this work, we require $\mathcal{X}$ to always include a zero symbol 0 and that transmitting a zero does not consume any energy, i.e.,

$$\gamma(0) = 0. \tag{4.5}$$

This requirement is essential for the correct operation of the encoding/decoding process, otherwise when $S_n = 0$ the transmitter cannot send any symbol to the channel and the synchronization of the system breaks down. Moreover, γ is usually endowed with some physical meaning. For example, we often use the quadratic cost function to denote the instantaneous power:

$$\gamma(x) = |x|^2. \tag{4.6}$$

We summarize the notations for the energy harvesting system in Table 4.1.

Assume the initial energy B_1 stored in the battery is a random variable and the sequence of harvested energy $\{E_n\}_{n=1}^{\infty}$ is a random process independent of B_1. To simplify the problem, we only consider a *finite discrete* system. Specifically, we assume $\overline{B} < \infty$, and that all the energy quantities involved are quantized with the

same interval size, i.e., all E_n, B_n, S_n, $\gamma(X_n)$ and $\overline{B}$ are integral multiples of some common unit of energy Δ_E. Hence without loss of generality we can assume all these quantities are integers. Moreover, we further assume that the alphabet of E_n is a bounded set $\mathcal{E}_H$ of non-negative integers, so that B_n and S_n can also only take values in finite integral sets $\mathcal{E}_B$ and $\mathcal{S}$, respectively.

Remark 4.2.1. Because of the energy constraint (4.1), the energy harvesting channel is very different from an ordinary DMC and is much harder to analyze. During each transmission the transmitter is not free to choose any letter in $\mathcal{X}$; instead, at time n it can only send a symbol X_n that does not demand more than the current available energy S_n. Since it determines how much energy the system can spend for the current transmission, we also call S_n the *energy state* of the system. From the functional dependence of S_n on B_n and E_n, we see that $\{S_n\}$ is a random process. The input constraint (4.1) is unprecedented in traditional communication systems, as the constraint value S_n is *random*. In addition, this new constraint also differs from that for the usual channel with average cost constraint (e.g., average power constraint) in that it is *instantaneous*, and from peak power constraint in that it is *time-varying*. Furthermore, in the following, we will see that the major difficulty for the analysis of this system lies in the fact that the energy states $\{S_n\}$ has *memory* (as demonstrated below in Example 4.2.1).

Example 4.2.1. We use a simple example to demonstrate the interactions among the input symbols, the harvested energy, and the energy in the battery. Assume E_n is an i.i.d. Bernoulli(p) process, i.e.,

$$E_n = \begin{cases} 1 & \text{w.p.} \quad p \\ 0 & \text{w.p.} \quad 1-p \end{cases}$$

so $\mathcal{E}_H = \{0, 1\}$. The battery capacity $\overline{B} = 1$ and the alphabets $\mathcal{X} = \mathcal{Y} = \{0, 1\}$. We require E_n to be stored in the battery first and assume a quadratic energy cost

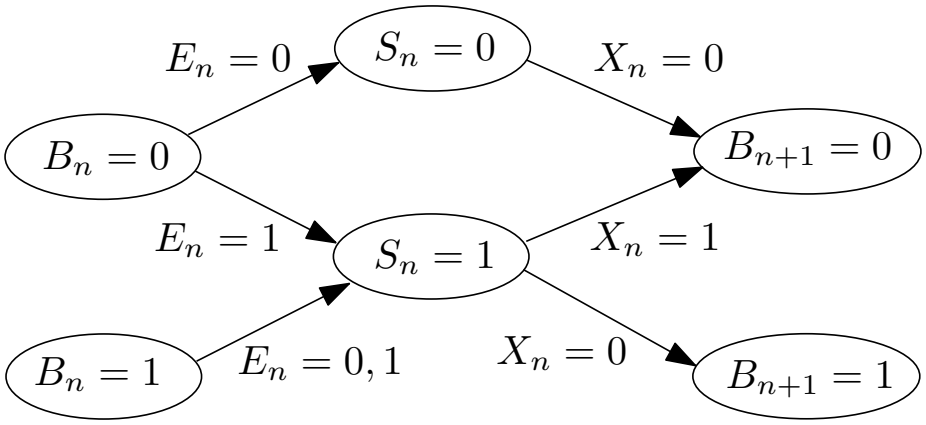


Figure 4.2: Evolution of battery energy

function, i.e., the energy model is (4.3) and (4.6), with

$$\gamma(0) = 0, \qquad \gamma(1) = 1.$$

Thus $\mathcal{S} = \{0, 1\}$ in this configuration. The DMC is binary symmetric (BSC) with crossover probability q, namely

$$p(y|x) = \begin{cases} 1-q & \text{if} \quad y = x \\ q & \text{if} \quad y \neq x \end{cases}.$$

With the constraint (4.1) and energy model (4.3), we show the evolution of the energy quantities for this system in Figure 4.2.

We study two scenarios with regard to the availability of energy information in the energy harvesting channel. In the first scenario (abbreviation: EH-SC1), before the n-th transmission only the energy state S_n is observed at the transmitter. In the second scenario (abbreviation: EH-SC2), however, the transmitter knows the initial battery level B_1 and observes E_n at time n. In both cases the receiver has no energy information. The second scenario is more general, since by (4.1), with X^n the transmitter can deduce S^n from E^n and B_1, but not vice versa. Observe that the energy information in the system is a certain form of channel side information causally known at the transmitter, which is reminiscent of the channels with causal

CSIT [38,45]. In conventional channels with states and CSI, the channel states usually affect the channel transition probabilities; in the energy harvesting channels, however, the energy states affect the input alphabets instead. As we will see, the first scenario for the energy harvesting channel is closely related to a certain finite state channel with causal CSIT and state-dependent input constraints (abbreviation: FSC-X). The second scenario also shares the same method of analysis with the FSC-X, only it requires an (often) more complicated theory.

In what follows we describe the encoding, decoding, and channel capacity for each of the above models, with a formal definition of the FSC-X. Note that all these channels are subject to random input constraints, so an ordinary channel encoding scheme cannot function properly. In fact, if a message is mapped to any fixed input vector x^N, then chances are that some symbol x_n does not satisfy the input constraint at the time of transmission, since the constraint itself at that time takes a random value, which might be incompatible with x_n. Nevertheless, since these random constraint values can be computed at the transmitter through the side information, we can define new encoding schemes analogous to those in [38,45]. These schemes not only take account of the causal side information, but also take the random input constraints into consideration when producing input symbols, thus resolving the above input incompatibility issue. For the definitions below let us denote the set of messages to be transmitted as

$$\mathcal{M} = \{1, 2, \cdots, M\}.$$

4.2.1 EH-SC1 and EH-SC2

Definition 4.2.1. A block code $f^{(N)}$ of length N for EH-SC1 is defined by a sequence of N encoding functions

$$f_n : \mathcal{M} \times \mathcal{S}^n \to \mathcal{X} \qquad 1 \le n \le N,$$

such that $\forall m \in \mathcal{M}$ and $\forall s^N \in \mathcal{S}^N$,

1) the output x^N of the encoder $f^{(N)}$ takes the form

$$x_n = f_n(m, s^n) \qquad 1 \leq n \leq N,$$

which means f_n is causal in $\{s_n\}$;

2) the energy constraint (4.1) is satisfied, in other words:

$$\gamma(x_n) \leq s_n \qquad 1 \leq n \leq N.$$

Definition 4.2.2. A block code $f^{(N)}$ of length N for EH-SC2 is defined by a sequence of N encoding functions

$$f_n : \mathcal{M} \times \mathcal{E}_B \times \mathcal{E}_H^n \to \mathcal{X} \qquad 1 \leq n \leq N,$$

such that $\forall m \in \mathcal{M}$, $\forall b_1 \in \mathcal{E}_B$ and $\forall e^N \in \mathcal{E}_H^N$,

1) the output x^N of the encoder $f^{(N)}$ takes the form

$$x_n = f_n(m, b_1, e^n) \qquad 1 \leq n \leq N,$$

which means f_n is causal in $\{e_n\}$;

2) the energy constraint (4.1) is satisfied.

The decoder for both scenarios are defined as usual:

$$g^{(N)} : \mathcal{Y}^N \to \mathcal{M},$$

which receives the output y^N of the channel and estimates the transmitted message m. With the block codes properly defined, the definitions of probability of error, code rate, achievable rate, and channel capacity follow standard texts (see, e.g., [46]).

Remark 4.2.2. The capacity for the first scenario is smaller than or at most equal to that of the second, since the latter has more energy information at the transmitter, as mentioned above. Hence any capacity lower bound/achievable rate for EH-SC1 is

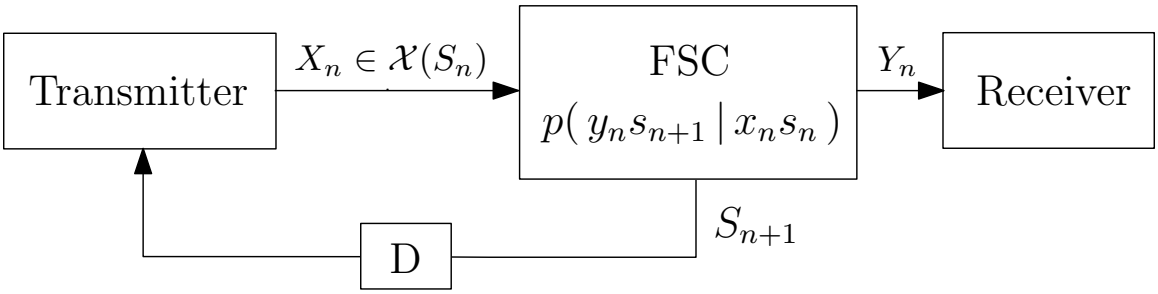


Figure 4.3: FSC-X: FSC with input constraint and Causal CSIT

also a capacity lower bound/achievable rate for EH-SC2, while any capacity upper bound of EH-SC2 is also a capacity upper bound of EH-SC1. That said, whether the first scenario has a strictly smaller capacity, or how much smaller the capacity is, is still an open question which is out of the scope of this dissertation.

4.2.2 FSC-X

We introduce a certain type of channels with states, which is derived from Gallager's finite state channel (FSC) model [2] and is illustrated in Figure 4.3. In this channel, the input, output, and state symbols at time n are denoted by X_n, Y_n, and S_n, respectively, whose respective alphabets $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{S}$ are all finite. The probability law governing the transitions of X_n, Y_n, and S_n follows that of the FSC, i.e., it is described by a conditional probability $p(y_n s_{n+1} \,|\, x_n s_n)$[2], which satisfies

$$p(y_n s_{n+1} \,|\, x^n s^n y^{n-1}) = p(y_n s_{n+1} \,|\, x_n s_n) \tag{4.7}$$

and which is *time-invariant* (i.e., independent of n). What distinguishes our new channel from an ordinary FSC is that the transmitter has causal CSI, and the input is constrained by the current state. Specifically, at time n, S_n is fed to the encoder, which limits the input X_n to a subset $\mathcal{X}(S_n) \subseteq \mathcal{X}$. For the ease of presentation we assume the receiver has no CSI, but by treating the CSIR as part of the output, the capacity can be analyzed in the same way.

[2]Compared to the original definition in [2], we increase the indices of the states by 1 to better accommodate our channel model (which gives a more natural physical meaning for the states).

We refer to this channel as FSC-X, which is short for "finite state channel with extra constraints and conditions". The encoding, decoding, and channel capacity are defined similarly to the subsection above.

Definition 4.2.3. A block code $f^{(N)}$ of length N for FSC-X is defined by a sequence of N encoding functions

$$f_n : \mathcal{M} \times \mathcal{S}^n \to \mathcal{X} \qquad 1 \le n \le N,$$

such that $\forall m \in \mathcal{M}$ and $\forall s^N \in \mathcal{S}^N$,

1) the output x^N of the encoder $f^{(N)}$ takes the form

$$x_n = f_n(m, s^n) \qquad 1 \le n \le N,$$

which means f_n is causal in $\{s_n\}$;

2) the input constraint is satisfied: $x_n \in \mathcal{X}(s_n)$, $1 \le n \le N$.

4.2.3 Relating EH-SC1 and FSC-X

In this section we show that if the energy harvesting process $\{E_n\}$ is *i.i.d.*, then EH-SC1 is a special case of FSC-X, whose states are exactly the energy states $\{S_n\}$ for EH-SC1.

First we consider the transition probabilities for the input, output, and state random variables. By (4.4) and the DMC property,

$$\begin{aligned} p(y_n s_{n+1} \,|\, x^n s^n y^{n-1}) &= p(y_n \,|\, x_n) p(s_{n+1} \,|\, x^n y^n s^n) \\ &= p(y_n \,|\, x_n) \Pr\left(S(x_n, s_n, E_{n+1}) = s_{n+1} \,|\, x^n y^n s^n\right) \\ &\overset{(*)}{=} p(y_n|x_n) \sum_{e_{n+1}} p(e_{n+1}) \cdot \mathbf{1}_{\{S(x_n,s_n,e_{n+1})=s_{n+1}\}} \\ &= p(y_n s_{n+1} \,|\, x_n s_n), \end{aligned} \tag{4.8}$$

where $(*)$ holds because by the i.i.d. property of $\{E_n\}$, E_{n+1} is independent of

all previous random variables (with index less than $n+1$). Furthermore, (4.8) is independent of n, hence defines an FSC.

Second, by (4.1) the input is constrained by $X_n \in \mathcal{X}(S_n)$, where $\forall s \in \mathcal{S}$,

$$\mathcal{X}(s) \triangleq \{x \in \mathcal{X} : \gamma(x) \leq s\}. \tag{4.9}$$

Since S_n is causally known at the transmitter, the channel model for EH-SC1 fits exactly into the regime of FSC-X.

4.3 Equivalent Channels without CSI and Constraints

As discussed before, the energy harvesting channel has an unprecedented random input constraint (4.1), which is instantaneous, time-varying, and has memory. In fact from the evolution of $\{S_n, B_n\}$, we see that the energy state S_n depends on the full history of the harvested energy E^n, all the past transmitted symbols X^{n-1}, and the initial battery level B_1. This ever-growing memory of the energy constraint poses a major difficulty for the analysis of the channel: since the energy information is causally know at the transmitter, the system can be treated using approaches for channels with causal CSIT [38,45], with encoding/decoding defined in Section 4.2.1. If $\{E_n\}$ is i.i.d. and the battery capacity $\overline{B} = 0$, then the system is actually memoryless, and it is easy to show that the channel for either scenario is simply equivalent to a DMC with an enlarged alphabet (using similar analysis as in [38]). When this is not the case, however, the system has infinite memory and the most general approach in [38, 45] has to be invoked to convert it to an equivalent channel without side information or constraints, which is much more complicated. The channel FSC-X, likewise, has infinite memory and causal CSIT, and hence is treated in the same way.

A general channel with memory (but without feedback, channel states, constraints, etc.) is defined by describing the input/output alphabets and the transition probabilities for each block length N. For each of our three models, the equivalent channel

can be expressed as

$$\boldsymbol{W} \triangleq \left\{ \mathcal{U}^{(N)},\ p(y^N \,|\, u^N),\ \mathcal{Y}^N \right\}_{N=1}^{\infty}, \tag{4.10}$$

where we use the random variables U_n and Y_n to denote the new input and output symbols, respectively. For each N this new channel corresponds to N operations of the original channel, starting from the beginning of transmission. The output alphabet is still the same, which is the Cartesian product $\mathcal{Y}^N$ for block length N. The input alphabet, however, is different: an input symbol U_n at each time n is now a function of the causal side information, which respects the input constraints. It turns out that the input alphabet for block length N is no longer the Cartesian product of N copies of the alphabet for a single symbol, and thus it is denoted by $\mathcal{U}^{(N)}$ instead of $\mathcal{U}^N$. The equivalent channel operates as follows: it looks at the function U_n and the (causal) side information to produce a symbol X_n, which is then sent to the original channel to output a symbol Y_n. Hence the new transition probabilities are now the ones averaged over the randomness of the environment or channel states. The precise definitions for the input alphabets and transition probabilities of the equivalent channels are given below for each model separately, starting from the simplest case, FSC-X.

4.3.1 FSC-X

The n-th input symbol for the equivalent channel is a function

$$u_n : \mathcal{S}^n \to \mathcal{X},$$

which can also be viewed as a vector in $\mathcal{X}^{|\mathcal{S}|^n}$. The function needs to satisfy the input constraint

$$u_n(s^n) \in \mathcal{X}(s_n), \qquad \forall s^n \in \mathcal{S}^n.$$

Thus the input alphabets for time n and for block length N are, respectively,

$$\mathcal{U}_n = \prod_{s \in \mathcal{S}} \mathcal{X}(s)^{|\mathcal{S}|^{n-1}}, \qquad \mathcal{U}^{(N)} = \prod_{n=1}^{N} \mathcal{U}_n.$$

The N-symbol channel transition probability is determined as follows. First with the FSC probability model (4.7) and the functional relation $x_n = u_n(s^n)$,

$$\begin{aligned} p\left(y^N s_2^{N+1} \,|\, u^N s_1\right) &= \prod_{n=1}^{N} p\left(y_n s_{n+1} \,|\, u^N s^n y^{n-1}\right) \\ &= \prod_{n=1}^{N} p\left(y_n s_{n+1} \,|\, u^N s^n y^{n-1} x_n\right) \\ &= \prod_{n=1}^{N} p\left(y_n s_{n+1} \,|\, x_n = u_n(s^n), s_n\right). \end{aligned} \tag{4.11}$$

Then, since in the new channel the initial state is not known at the transmitter, S_1 is independent of U^N:

$$\begin{aligned} p\left(y^N \,|\, u^N\right) &= \sum_{s^{N+1}} p\left(y^N s^{N+1} \,|\, u^N\right) \\ &= \sum_{s_1} \sum_{s_2^{N+1}} p\left(s_1 \,|\, u^N\right) p\left(y^N s_2^{N+1} \,|\, u^N s_1\right) \\ &= \sum_{s_1} p(s_1) \sum_{s_2^{N+1}} \prod_{n=1}^{N} p\left(y_n s_{n+1} \,|\, x_n = u_n(s^n), s_n\right). \end{aligned}$$

4.3.2 EH-SC1

The input symbols/alphabets take the same forms as the previous case, with $\mathcal{X}(s)$ defined by (4.9). Nevertheless, we restate it here for easier reference. The n-th input symbol for the equivalent channel is a function

$$u_n : \mathcal{S}^n \to \mathcal{X},$$

which satisfies the energy constraint (4.1)

$$\gamma(u_n(s^n)) \leq s_n, \qquad \forall s^n \in \mathcal{S}^n.$$

So the input alphabets for time n and for block length N are, respectively,

$$\mathcal{U}_n = \prod_{s\in\mathcal{S}} \mathcal{X}(s)^{|\mathcal{S}|^{n-1}}, \qquad \mathcal{U}^{(N)} = \prod_{n=1}^{N} \mathcal{U}_n$$

where $\mathcal{X}(s)$ is defined in (4.9).

The N-symbol channel transition probability is

$$\begin{aligned} p(y^N \,|\, u^N) &= \sum_{b_1 e^N} p(b_1 e^N y^N \,|\, u^N) \\ &\overset{(a)}{=} \sum_{b_1, e^N} p(b_1) p(e^N) p(y^N \,|\, b_1 e^N u^N) \\ &\overset{(b)}{=} \sum_{b_1, e^N} p(b_1) p(e^N) p(y^N \,|\, b_1 e^N u^N x^N) \\ &\overset{(c)}{=} \sum_{b_1, e^N} p(b_1) p(e^N) \prod_{n=1}^{N} p\left(y_n \,|\, x_n = u_n\left(s^n\left(b_1, e^n, u^{n-1}\right)\right)\right), \end{aligned} \tag{4.12}$$

where (a) holds because B_1 and E^N are independent, and in the equivalent channel they are unknown at the transmitter, hence are independent of U^N. (b) holds because x^N is determined by b_1, e^N, and u^N from the recursion

$$\begin{cases} s_n &= S(b_n, e_n) \\ x_n &= u_n(s^n) \\ b_{n+1} &= \min\left\{ s_n - \gamma(x_n),\ \overline{B}\right\} \end{cases},$$

which also specifies the functional dependence of s^n on b_1, e^n and u^{n-1}. (c) holds because y^N is produced by the DMC, whose input is x^N.

4.3.3 EH-SC2

The n-th input symbol for the equivalent channel is a function

$$u_n : \mathcal{E}_B \times \mathcal{E}_H^n \to \mathcal{X},$$

which can also be viewed as a vector in $\mathcal{X}^{|\mathcal{E}_B|\cdot|\mathcal{E}_H|^n}$. The function u_n needs to be compatible with the previous input symbols, u^{n-1}, in terms of the the energy constraint (4.1). In particular, for each block length N, a feasible input vector u^N needs to satisfy the following requirements: $\forall b_1 \in \mathcal{E}_B$ and $\forall e^N \in \mathcal{E}_H^N$,

$$\gamma(u_n(b_1, e^n)) \le s_n, \quad \forall 1 \le n \le N,$$

where $s_n = s_n(b_1, e^n, u^{n-1})$ is determined recursively by

$$\begin{cases} s_n &= S(b_n, e_n) \\ x_n &= u_n(b_1, e^n) \\ b_{n+1} &= \min\left\{ s_n - \gamma(x_n),\ \overline{B} \right\} \end{cases} . \tag{4.13}$$

Thus the permitted function values of u_n depends not only on the energy sequence (b_1, e^n), but also on all previous input symbols u^{n-1}. In other words, the input alphabet for time n is

$$\mathcal{U}_n(u^{n-1}) = \prod_{b_1, e^n} \mathcal{X}\left(s_n(b_1, e^n, u^{n-1})\right),$$

where $\mathcal{X}(\cdot)$ is defined in (4.9). Furthermore, the input alphabet for block length N is

$$\mathcal{U}^{(N)} = \left\{ (u_1, \cdots, u_N) : u_n \in \mathcal{U}_n(u^{n-1}),\ \forall 1 \le n \le N \right\}.$$

That is, $\mathcal{U}^{(N)}$ is the collection of all vectors of N causal functions on the energy sequence, that are consistent with the energy constraint.

The N-symbol transition probabilities $p(y^N \,|\, u^N)$ for the new channel is

$$\begin{aligned} p(y^N \,|\, u^N) &= \sum_{b_1, e^N} p(b_1 e^N y^N \,|\, u^N) \\ &\overset{(a)}{=} \sum_{b_1, e^N} p(b_1) p(e^N) p(y^N \,|\, b_1 e^N u^N) \end{aligned}$$

$$\stackrel{(b)}{=} \sum_{b_1, e^N} p(b_1)p(e^N)p(y^N \,|\, b_1 e^N u^N x^N)$$

$$\stackrel{(c)}{=} \sum_{b_1, e^N} p(b_1)p(e^N) \prod_{n=1}^{N} p\left(y_n \,|\, x_n = u_n(b_1, e^n) \right), \tag{4.14}$$

where as before, (a) holds because B_1 and E^N are independent, and are independent of U^N; (b) holds because x^N is determined by b_1, e^N, and u^N through (4.13); (c) holds because y^N is produced by the DMC, whose input is x^N.

Since there is no CSI or constraints for the equivalent channel, the encoding/decoding maps are defined as usual:

$$f^{(N)} : \mathcal{M} \to \mathcal{U}^{(N)},$$

$$g^{(N)} : \mathcal{Y}^N \to \mathcal{M}.$$

For each model this new channel is equivalent to the original one, in the sense that they have the same capacity. In fact, as stated in [38], block codes for the original and the equivalent channels can be translated into each other with the same probability of error. To be specific, fix N and let $f_o^{(N)}$ be a block code for the original channel. It is easy to check that for each $m \in \mathcal{M}$, $f_o^{(N)}(m, \cdot) \in \mathcal{U}^{(N)}$, which can be used to define an encoder for the equivalent channel:

$$f_e^{(N)}(m) = f_o^{(N)}(m, \cdot), \qquad \forall m \in \mathcal{M}.$$

By the definition of the transition probability for the equivalent channel, when $m \in \mathcal{M}$ is sent the output conditional probability $p(y^N \,|\, m)$ is the same as the original channel. Hence under the same decoder $g^{(N)}$, the error probabilities for both cases are equal. On the other hand, if $f_e^{(N)}$ is a block code for the new channel, for the original channel we can define an encoder

$$f_o^{(N)} : (m, CSIT) \mapsto f_e^{(N)}(m)(CSIT), \qquad \forall m \in \mathcal{M},$$

where $CSIT$ denotes the corresponding side information for each channel model. Again under the same decoder, the error probabilities for both cases are equal.

The equivalent channel avoids the difficulty of dealing with either the CSIT or the input constraints, at the cost of a more complicated input alphabet, whose size is larger and ever-growing. Roughly speaking, the cardinality for the input alphabet at time n grows double-exponentially with n: the alphabet size for CSIT grows exponentially, and hence without input constraints the number of functions on CSIT grows double-exponentially. With constraints this number is reduced, but the growth rate is still double-exponential. We illustrate such a computation for input alphabets in the following example.

Example 4.3.1. We continue with the setup of Example 4.2.1 and assume the second scenario (EH-SC2). For $N = 1, 2$, all possible values of B_1, E^2 and the corresponding energy states S^2 are listed in Table 4.2, together with all possible input function values. The cardinalities of the input alphabets can be computed from the table as

$$\left|\mathcal{U}^{(1)}\right| = 1 \cdot 2 \cdot 2 \cdot 2 = 8 = 2^3,$$

$$\left|\mathcal{U}^{(2)}\right| = (1 \cdot 2) \cdot (2 \cdot 2 + 1 \cdot 2) \cdot (2 \cdot 2 + 1 \cdot 2) \cdot (2 \cdot 2 + 1 \cdot 2) = 432 = 2^4 \cdot 3^3.$$

With the help of computer we can further obtain $\left|\mathcal{U}^{(3)}\right| = 2^{14} \cdot 3^4$. Roughly speaking, the cardinality of $\mathcal{U}^{(N)}$ is on the order of $2^{2^{N+1}}$.

4.4 Channel Capacities

To compute the capacity for a channel as general as (4.10), we need to invoke Verdu and Han's general capacity formula for arbitrary channels without feedback [1]. First we define an *input distribution process* $\boldsymbol{U}$ to be a sequence of probability distributions defined on $\mathcal{U}^{(N)}$ for each N, which need not have any relation among them. Equivalently, $\boldsymbol{U}$ can be represented by a collection of random vectors $\left\{U^{(N)}\right\}_{N=1}^{\infty}$, where each $U^{(N)}$ is a random vector in $\mathcal{U}^{(N)}$ that corresponds exactly to the N-th distribution of $\boldsymbol{U}$. The corresponding *output distribution process* $\boldsymbol{Y} = \left\{Y^{(N)}\right\}_{N=1}^{\infty}$ is

Table 4.2: Input symbols for the equivalent channel

B_1	E_1	S_1	$U_1(B_1, E_1)$	B_2	E_2	S_2	$U_2(B_1, E^2)$
0	0	0	0	0	0	0	0
					1	1	0
							1
	1	1	0	1	0	1	0
							1
					1	1	0
							1
			1	0	0	0	0
					1	1	0
							1
1	0	1	0	1	0	1	0
							1
					1	1	0
							1
			1	0	0	0	0
					1	1	0
							1
	1	1	0	1	0	1	0
							1
					1	1	0
							1
			1	0	0	0	0
					1	1	0
							1

the collection of random vectors $Y^{(N)}$ in $\mathcal{Y}^N$, where each $Y^{(N)}$ is induced by the input random vector $U^{(N)}$ and the N-symbol channel transition probability $p(y^N \mid u^N)$.

Remark 4.4.1. Note that in this context we use $U^{(N)}$, instead of the usual U^N, to denote a random input vector of length N. The reason is, the latter notation by default assumes that the first $N-1$ entries of U^N necessarily agrees with U^{N-1}, which is clearly not the case in the definition of $\boldsymbol{U}$. In contrast, $U^{(N)}$ need not have any relation to $U^{(N-1)}$ and so is a more appropriate notation. For the same reason we use

the notation $Y^{(N)}$ for the output vector.

Definition 4.4.1. Let $U^{(N)}$ be the input random vector for the channel (4.10) for block length N. Define the *information density* between $U^{(N)}$ and $Y^{(N)}$ as

$$i_N\left(u^N; y^N\right) = i_{U^{(N)};Y^{(N)}}\left(u^N; y^N\right) = \log \frac{p\left(y^N \mid u^N\right)}{p\left(y^N\right)}$$

for all $u^N \in \mathcal{U}^{(N)}$, $y^N \in \mathcal{Y}^N$, where the output distribution $p(y^N)$ is induced by the distribution of $U^{(N)}$ and the channel transition probability.

Definition 4.4.2. Let $\{A_N\}_{N=1}^{\infty}$ be a sequence of random variables, not necessarily having a joint distribution. Define the *liminf in probability* of $\{A_N\}$ to be the supremum of all the real numbers α for which $\Pr(A_N \leq \alpha)$ vanishes as $N \to \infty$. In other words, we write

$$\operatorname*{liminf}_{N\to\infty}^{(\mathrm{P})} A_N \triangleq \sup\left\{\alpha \in \mathbb{R} \;\middle|\; \lim_{N\to\infty} \Pr(A_N \leq \alpha) = 0\right\}.$$

Example 4.4.1. Let the probability density function of A_N be

$$p_{A_N}(a) = \frac{1}{2}\delta(a - N) + \frac{1}{2}\sqrt{\frac{N}{2\pi}} \exp\left(-\frac{Na^2}{2}\right),$$

where δ is the Dirac delta function. With probability $1/2$, A_N takes the value N, and with probability $1/2$, A_N assumes a Gaussian distribution $\mathcal{N}(0, 1/N)$. From Fig 4.4, we can easily see that $\operatorname{liminf}^{(\mathrm{P})}_{N\to\infty} A_N = 0$.

Definition 4.4.3. Let $\boldsymbol{U} = \left\{U^{(N)}\right\}_{N=1}^{\infty}$ be an input distribution process for the channel (4.10), which induces an output distribution process $\boldsymbol{Y} = \left\{Y^{(N)}\right\}_{N=1}^{\infty}$. Define the *inf-information rate* $\underline{\boldsymbol{I}}(\boldsymbol{U}; \boldsymbol{Y})$ between $\boldsymbol{U}$ and $\boldsymbol{Y}$ as the liminf in probability of the sequence of normalized information densities, which is a sequence of random variables:

$$\underline{\boldsymbol{I}}(\boldsymbol{U}; \boldsymbol{Y}) \triangleq \operatorname*{liminf}_{N\to\infty}^{(\mathrm{P})} \frac{1}{N} i_N\left(U^{(N)}; Y^{(N)}\right).$$

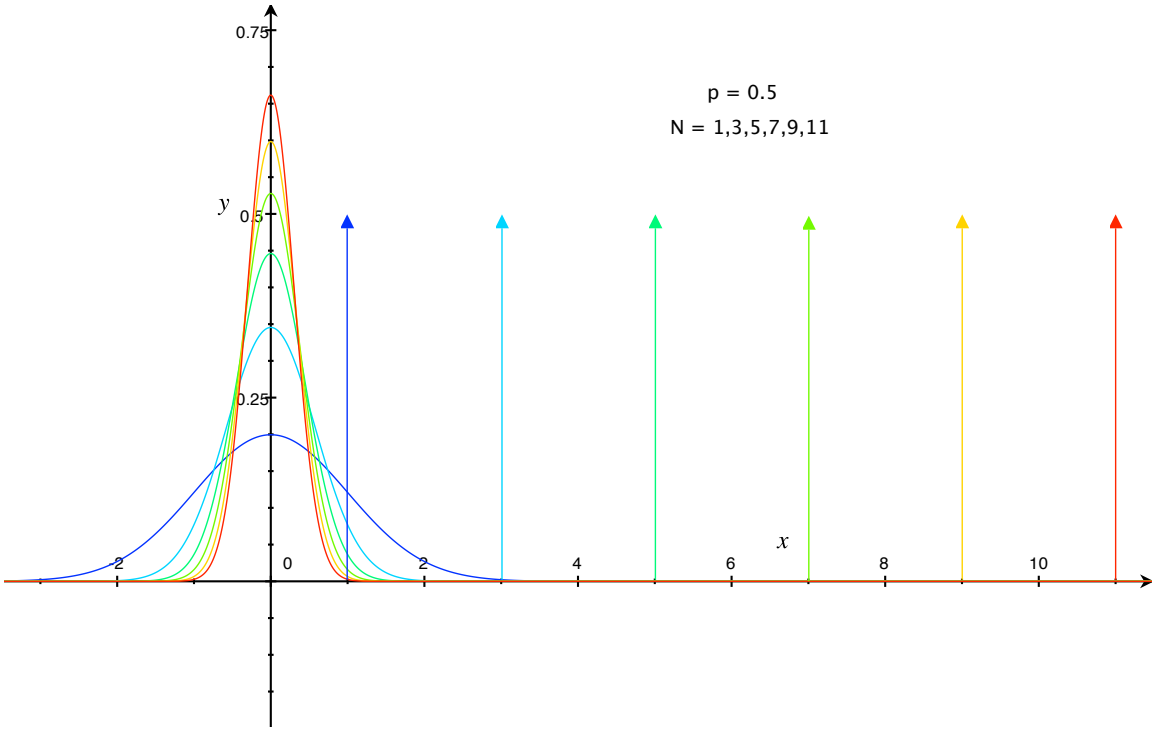


Figure 4.4: Probability density functions of A_N

Theorem 4.4.1 (Verdu-Han formula [1])**.** *The capacity of the channel* $\boldsymbol{W}$ *in* (4.10) *is given by*

$$C = \sup_{\boldsymbol{U}} \underline{\boldsymbol{I}}(\boldsymbol{U}; \boldsymbol{Y}), \tag{4.15}$$

where the supremum is taken over all input distribution processes $\boldsymbol{U}$*.*

The channel capacities of the three models we study can all be obtained from the definitions of their respective equivalent channels in Section 4.3 and Theorem 4.4.1. However, owing to the following issues, the capacity formula (4.15) is not easy to evaluate:

1) The supremum is taken over all possible input distribution processes, which is hard to enumerate/parameterize.
2) Given an arbitrary input distribution processes $\boldsymbol{U}$, the inf-information rate is not always readily computable, as the asymptotic behavior for the corresponding random sequence can be arbitrary.
3) As mentioned in the previous section, the input alphabet size $\left|\mathcal{U}^{(N)}\right|$ grows

double-exponentially, which results in a double-exponential complexity when computing either the information density or the mutual information for a single block length N.

Nonetheless, this formula gives us a means of analyzing the channel capacities. In the next chapter we will try to resolve these difficulties under some simplifying conditions and assumptions to make the computation tractable. Such simplifications give us achievable rates for these channels, which are lower bounds of their respective capacities.

Chapter 5

Achievable Rates

In this chapter we study the computation and optimization of some achievable rates for our channel models from the general capacity formula (4.15). First we discuss the general methodology, then for each channel model we give the detailed description of the conditions and specific considerations. After that we use some numerical examples to illustrate the computation. For the energy harvesting models EH-SC1 and EH-SC2, we also have a conjecture on the form of the optimal input functions.

5.1 Methodology

To address the issues in computing the channel capacities in the previous chapter, we restrict the input symbols of the channel model (4.10) to a constant-sized subset of its alphabet and obtain a surrogate channel $\boldsymbol{W}'$, whose capacity C' provides a lower bound for the capacity C of the channel $\boldsymbol{W}$. To be specific, instead of the full CSI history, the input functions now can only depend on a limited amount of the causal side information. In addition, from such limited side information the transmitter should still be able to compute the the input constraint values. (Otherwise, as they are random and unknown, we still have the input incompatibility issue discussed in Section 4.2.)

Let V_n denote the new input function at time n and $\mathcal{V}$ denote its (constant-sized)

alphabet. Similar to (4.10), the surrogate channel can be expressed as

$$\boldsymbol{W}' \triangleq \{\, \mathcal{V}^N,\ p(y^N \,|\, v^N),\ \mathcal{Y}^N \,\}_{N=1}^{\infty}. \tag{5.1}$$

It turns out that in many cases we are interested in, $\boldsymbol{W}'$ becomes a finite state channel (FSC). The capacity of a general FSC is studied in [2,47], both of which give a series of capacity upper and lower bounds, in terms of the mutual information between input and output vectors for each block size N. When the FSC is *indecomposable*[1], the upper and lower bounds both converge to the capacity. However, these bounds are not very desirable in our case, since i) the upper bounds are not useful since the capacity of $\boldsymbol{W}'$ is only a lower bound for C; ii) the computational complexity of such bounds is exponential in N; iii) the bounds in [2] are too loose for small N and their convergence is slow (see [47]); iv) the bounds in [47] are supposed to be tighter, but their computation is not easy for a general N.

Another way of describing the capacity C' is through the Verdu-Han formula (cf. Theorem 4.4.1):

$$C' = \sup_{\boldsymbol{V}} \underline{\boldsymbol{I}}(\boldsymbol{V}; \boldsymbol{Y}), \tag{5.2}$$

for which we define the same concepts and similar notations, as in Section 4.4, with respect to the surrogate channel $\boldsymbol{W}'$. The supremum in (5.2) is taken over all input distribution processes $\boldsymbol{V}$. Although in general this formula is not computable, for any given input distribution process that yields a computable inf-information rate we still obtain an achievable rate for $\boldsymbol{W}'$ (and hence also for the channel $\boldsymbol{W}$), which is a lower bound of the capacity C' (and C). In particular, assume the input distribution process $\boldsymbol{V}$ is induced by a source random process $\{V_n\}$, so that the N-th distribution of $\boldsymbol{V}$ corresponds exactly to the random vector V^N for each N. Assume further that the induced joint input-output process $\{V_n, Y_n\}$ satisfies the Shannon-McMillan-Breiman (SMB) theorem (see Section B.4 in Appendices), then the sample entropies for $\{V_n, Y_n\}$ converge almost surely to their respective entropy rates. Accordingly,

[1] See Definition B.2.2.

the normalized information density, which can be written as

$$\frac{1}{N} i_N\left(V^N; Y^N\right) = \frac{1}{N}\log p\left(V^N, Y^N\right) - \frac{1}{N}\log p\left(V^N\right) - \frac{1}{N}\log p\left(Y^N\right), \tag{5.3}$$

converges almost surely to the mutual information rate[2]

$$I(\mathcal{V}, \mathcal{Y}) \triangleq \lim_{N\to\infty} \frac{1}{N} I(V^N; Y^N) = H(\mathcal{V}) + H(\mathcal{Y}) - H(\mathcal{V}, \mathcal{Y}),$$

where $H(\mathcal{V})$, $H(\mathcal{Y})$, and $H(\mathcal{V}, \mathcal{Y})$ denote the (joint) entropy rates of $\{V_n\}$, $\{Y_n\}$, and $\{V_n, Y_n\}$, respectively. As a result, the liminf in probability of $\left\{\frac{1}{N} i_N\left(V^N; Y^N\right)\right\}_{N=1}^{\infty}$ evaluates to the same value $I(\mathcal{V}, \mathcal{Y})$, and so the inf-information rate corresponding to $\boldsymbol{V}$ becomes the mutual information rate

$$\underline{\boldsymbol{I}}(\boldsymbol{V}; \boldsymbol{Y}) = I(\mathcal{V}, \mathcal{Y}),$$

which yields a (at least theoretically) computable achievable rate. Alternatively, since AEP holds in this case (see Section B.4 in Appendices), we can use the idea of typical set decoding as in [46] to directly prove the achievability of the rate $I(\mathcal{V}, \mathcal{Y})$.

The Shannon-McMillan-Breiman theorem demands certain stationarity and ergodicity properties of the joint input-output process, which in turn require the source and channel to satisfy some conditions in that aspect. Specifically, the version of SMB theorem (Theorem B.4.1) suitable for our models requires the joint process $\{V_n, Y_n\}$ to be *asymptotically mean stationary*[3] (AMS) and ergodic. When the surrogate channel $\boldsymbol{W}'$ is an FSC, it belongs to the category of *Markov channels* and always produces an AMS joint input-state-output process for any AMS or stationary source. For such a channel $\boldsymbol{W}'$, if (i) the source $\{V_n\}$ is stationary and ergodic while $\boldsymbol{W}'$ satisfies some further ergodicity conditions with respect to the source, or (ii) the source $\{V_n\}$ is finite-order Markov and induces a joint source-channel Markov chain with some irreducibility condition, then the joint input-state-output process is AMS

[2] Also called the *information rate*.

[3] See Appendix B for this and other concepts in the stationarity and ergodicity theory.

and ergodic, and so is the process $\{V_n, Y_n\}$[4]. Due to its technical nature, we defer the exposition of a more detailed stationarity and ergodicity theory to Appendix B, which is largely based on the theory of Markov channels developed in [48, 49].

In practice, the computation of the mutual information rate $I(\mathcal{V}, \mathcal{Y})$ for general source processes $\{V_n\}$ is a challenging problem. One can use the sequence of finite block length mutual information to approximate $I(\mathcal{V}, \mathcal{Y})$, but since the alphabet sizes grow exponentially with the block length, so does the computational complexity. Moreover, the convergence of such a sequence is often rather slow. With the above stationarity and ergodicity conditions for the source and channel, however, we have the SMB theorem and so can estimate the information rate using the sample entropies (through (5.3)) of a very long sample sequence, which can be computed using the transition probabilities in (5.1) and the input distribution. In addition to that, when the source is a finite-order Markov process and the channel is an FSC, the computation of the sample entropies in (5.3) has a complexity linear in N; in fact one can use the well-known BCJR algorithm [50] (a.k.a. the sum-product algorithm [51]) to compute them. This stochastic method for information rate computation was proposed independently in [52–54], and is summarized in [55].

So far by restricting the input alphabet and imposing extra stationarity and ergodicity conditions on the source and channel, we are able to resolve the issues 2) and 3) in Section 4.4, and efficiently compute some achievable rates for the channel $\boldsymbol{W}$. If we further fix the order of a Markov input process, under some conditions (described below) we can maximize the achievable rate over a given set of transition probabilities for the Markov chain, thus also resolving the issue 1) in Section 4.4 to some extent. Specifically, we use the generalized Blahut-Arimoto algorithm (GBAA) for the achievable rate optimization, which is proposed by Vontobel et al. in [44]. In their work, the traditional Blahut-Arimoto algorithm [56], originally used for computing the capacity of a DMC, is generalized in the setting of an indecomposable FSC with a finite-order Markov input process, whose underlying chain is stationary, ergodic, and aperiodic, to optimize the information rate over all possible transition probabilities

[4]See Section B.5.1 in Appendices.

of the input Markov chain.[5] The core part of the GBAA is to estimate the so-called "T-values" defined in [44, Definition 41] through the algorithms in [44, Lemma 70] in each iteration, which are then used both to calculate the information rate and to update the optimization parameters (i.e., the transition probabilities). As we examine the derivations and proofs in [44], we find that, to the best of our knowledge, the sole purpose of both the indecomposable assumption of the FSC and the ergodicity and aperiodicity of the input Markov chain is to guarantee the almost sure convergence of the estimated T-values in [44, Lemma 70]. However, the details of that step are not included in the proof. Nevertheless, we speculate that the required convergence still holds as long as the joint input-state-output process satisfies the SMB theorem; that is, the joint process is AMS and ergodic. Such a requirement is fulfilled when the source is a stationary finite-order Markov process whose underlying chain is irreducible,[6] while the channel is an FSC with the ergodicity conditions mentioned earlier (which are weaker than indecomposability). Therefore we conjecture that the GBAA still works under these relaxed conditions. Also, in this setting we can use the GBAA primarily as a means to find a good set of input process parameters (i.e., the Markov transition probabilities); the resulting information rates can always be cross-checked with those obtained using the stochastic methods described above, since the SMB theorem applies. Therefore, when these conditions hold, we apply the GBAA to our surrogate channel $\boldsymbol{W}'$ for each fixed Markov order of the input process[7] to find an optimized achievable rate. Apart from the GBAA, Han [57] also gives a stochastic method for the information rate optimization of a finite state channel. However, the assumptions on the channel are more stringent in [57], which limits the type of channels this algorithm can be applied to; hence it is not used in our work.

[5] In fact, we found that the algorithm as it is in [44] is not applicable to all indecomposable FSC's, as the calculation of the critical T-values is erroneous for some channel models. However, surprisingly, this issue does not affect the correct calculation of the information rate at each iteration, but only affects the selection of the new optimization parameters for the next iteration. Furthermore, after we communicated with them, the authors fixd this issue by adding certain correction terms to the original T-values.

[6] A finite alphabet stationary Markov process is ergodic iff the chain is irreducible; see Theorem B.3.1.

[7] Recall that when the order of the Markov process is k, the states of the underlying Markov chain are the tuples of k successive input symbols.

Lastly, we want to comment on a subtle technical issue of the surrogate channels. In our models we will always restrict the input function to depend on a finite duration of historical side information, which gives a constant-sized alphabet. However, this might cause some problem for the first few input functions, as there is not enough "history" to be used as their arguments. Nevertheless, we can fix this issue by providing some deterministic "dummy" pre-historical state variables, which also contribute to the determination of the surrogate channel transition probability. The detailed discussion of these dummy variables is presented in Section B.5.2 in Appendices.

5.2 FSC-X

We restrict the input function u_n to depend only on the m most recent states, where $m > 0$ is a fixed integer. To be specific, let $\mathcal{V}$ be the collection of all functions $v : \mathcal{S}^m \to \mathcal{X}$ such that

$$v(s^m) \in \mathcal{X}(s_m), \qquad \forall s^m \in \mathcal{S}^m.$$

Therefore $\mathcal{V} = \prod_{s\in\mathcal{S}} \mathcal{X}(s)^{|\mathcal{S}|^{m-1}}$ has a constant alphabet size. We restrict u_n in such a way that each u_n is associated with a symbol $v_n \in \mathcal{V}$, and

$$u_n(s^n) = v_n(s_{n-m+1}^n), \tag{5.4}$$

where we provide the dummy variables $s_{-m+2}, \cdots, s_0 \in \mathcal{S}$ as the *pre-historical states*, when $m > 1$. Note that these states are artificial and are only used as arguments of v_n for $n < m$, but do not affect the distribution of S_1 (which is determined by the environment/nature). With such a configuration we define a surrogate channel $\boldsymbol{W}'$ with the input alphabet $\mathcal{V}$, whose transition probability is defined through the corresponding u^N for each N. In other words, according to (4.11),

$$p\left(y^N s_2^{N+1} \,|\, v^N s_1\right) = \prod_{n=1}^{N} p\left(y_n s_{n+1} \,|\, x_n = v_n(s_{n-m+1}^n), s_n\right)$$

$$p\left(y^N \,|\, v^N\right) = \sum_{s_1} p(s_1) \sum_{s_2^{N+1}} \prod_{n=1}^{N} p\left(y_n s_{n+1} \,|\, x_n = v_n(s_{n-m+1}^n), s_n\right). \tag{5.5}$$

We claim that channel $\boldsymbol{W}'$ is an FSC for $n \geq m$,[8] whose state is defined as

$$Z_n = S_{n-m+1}^n,$$

with alphabet $\mathcal{Z} = \mathcal{S}^m$. In fact, for $n \geq m$, the transition probability satisfies the following: if z_n is compatible with z_{n+1}, i.e., for some $s^{n+1} \in \mathcal{S}^{n+1}$, $z_n = s_{n-m+1}^n$ while $z_{n+1} = s_{n-m+2}^{n+1}$, then by the FSC transition probability (4.7),

$$\begin{aligned} p(y_n z_{n+1} \,|\, v^n z^n y^{n-1}) &= p(y_n s_{n-m+2}^{n+1} \,|\, v^n s^n y^{n-1}) \\ &= p(y_n s_{n+1} \,|\, v^n s^n y^{n-1}, x_n = v_n(s_{n-m+1}^n)) \\ &= p(y_n s_{n+1} \,|\, x_n = v_n(s_{n-m+1}^n), s_n) \\ &= p(y_n z_{n+1} \,|\, v_n z_n). \end{aligned} \tag{5.6}$$

If z_n is not compatible with z_{n+1}, then both the first and the last term of the above equality chain is 0.

For the required stationarity and ergodicity properties for the SMB theorem, we provide the following two set of simple conditions. We also have some stronger but more complicated conditions, see Corollary B.2.2 and Lemma B.3.2 in Appendix B.

Lemma 5.2.1. *Assume the input process $\{V_n\}$ of the surrogate channel $\boldsymbol{W}'$ for FSC-X is stationary and ergodic. Then the joint process $\{V_n, Y_n\}$ is AMS and ergodic, if any of the following holds.*

i) *$\boldsymbol{W}'$ is indecomposable.*

ii) *There is a finite vector v_m^N with $\Pr(V_m^N = v_m^N) > 0$ satisfying the following property: given $V_m^N = v_m^N$, for any $z_m, z_m' \in \mathcal{Z}$, there exists $y_N \in \mathcal{Y}$ and $z_{N+1} \in \mathcal{Z}$ such that when $Z_m = z_m$ or z_m', we both have $Y_N Z_{N+1} = y_N z_{N+1}$ with positive probability.*

[8] This restriction does not affect the information rate computation, see Section B.5.3 in Appendices.

Proof. The first condition follows from Lemma B.2.5 and Theorem B.2.4 (or Corollary B.2.2), and the second follows from Corollary B.2.3. □

Lemma 5.2.2. *Assume the input process $\{V_n\}$ of the surrogate channel $\boldsymbol{W}'$ for FSC-X is finite-order Markov, then so is the joint process $\{(V_n, Y_n, Z_{n+1})\}$. If the underlying Markov chain for the latter is irreducible, then $\{V_n, Y_n\}$ is AMS and ergodic.*

Proof. This lemma is a simple case of Lemma B.3.2. □

5.3 EH-SC1

Again we restrict the input function u_n to depend only on the $m > 0$ most recent (energy) states, and supply the dummy pre-historical states $s_{-m+2}, \cdots, s_0 \in \mathcal{S}$ when $m > 1$. Then the surrogate channel $\boldsymbol{W}'$ has the same input alphabet as in the previous section, with $\mathcal{X}(s)$ defined in (4.9). According to (4.12) the transition probabilities are (note that $s^0_{-m+2}(\cdot)$ are provided by the dummy variables)

$$p\left(y^N \,|\, v^N\right) = \sum_{b_1, e^N} p(b_1) p(e^N) \prod_{n=1}^{N} p\left(y_n \,|\, x_n = v_n\left(s^n_{n-m+1}\left(b_1, e^n, u^{n-1}\right)\right)\right).$$

If the energy harvesting process $\{E_n\}$ is i.i.d., then, as shown in Section 4.2.3, the channel EH-SC1 is an instance of FSC-X, and by the previous section $\boldsymbol{W}'$ is an FSC with state variable $Z_n = S^n_{n-m+1}$. Note that the argument s^n_{n-m+1} for v_n is contained in z_n, by (4.8) and (5.6) we have

$$p(y_n z_{n+1} \,|\, v_n z_n) = p(y_n \,|\, v_n z_n) p(z_{n+1} \,|\, v_n z_n). \tag{5.7}$$

More generally, if $\{E_n\}$ is Markov of order $r > 0$, the surrogate channel is still an FSC for $n \geq \max\{m, r\}$, with the states

$$Z_n = E^n_{n-r+1} S^n_{n-m+1},$$

whose alphabet is $\mathcal{Z} = \mathcal{E}_H^r \times \mathcal{S}^m$. In fact, for $n \geq \max\{m, r\}$, the transition probability

satisfies the following: if z_n is compatible with z_{n+1}, i.e., for some $e^{n+1} \in \mathcal{E}_H^{n+1}$ and $s^{n+1} \in \mathcal{S}^{n+1}$, $z_n = e_{n-r+1}^n s_{n-m+1}^n$ while $z_{n+1} = e_{n-r+2}^{n+1} s_{n-m+2}^{n+1}$, then

$$\begin{aligned} p(y_n z_{n+1} \,|\, v^n z^n y^{n-1}) &= p(y_n e_{n-r+2}^{n+1} s_{n-m+2}^{n+1} \,|\, v^n e^n s^n y^{n-1}) \\ &= p(y_n e_{n+1} s_{n+1} \,|\, v^n e^n s^n y^{n-1} x_n) \\ &= p(y_n \,|\, x_n) \cdot p(e_{n+1} \,|\, e_{n-r+1}^n) \cdot \mathbf{1}_{\{S(x_n, s_n, e_{n+1}) = s_{n+1}\}} \cdot \mathbf{1}_{\left\{x_n = v_n(s_{n-m+1}^n)\right\}} \\ &= p(y_n z_{n+1} \,|\, v_n z_n), \end{aligned}$$

by the structure of the channel. If z_n is not compatible with z_{n+1}, then both the first and the last term of the above equality chain is 0. Again as the argument s_{n-m+1}^n for v_n is contained in z_n, (5.7) holds.

Remark 5.3.1. We have a conjecture regarding the form of the optimal input functions for both of the energy harvesting channels, EH-SC1 and EH-SC2. Observe that the energy state S_n contains all the information about the energy constraint on the current immediate input symbol X_n, which is the only influence the full history of energy information has on the transmission. Thus we conjecture that for the channel $\boldsymbol{W}$ it is enough to only consider input functions u_n that depends only on the current energy state s_n, i.e., setting $m = 1$ in the surrogate channel $\boldsymbol{W}'$ does not lose optimality—the capacity $C' = C$. But we are not able to prove it yet.

Now for the stationarity and ergodicity conditions, since (5.7) is true, by Section B.5.1 in Appendices we can just consider a smaller FSC $p(z_{n+1} \,|\, v_n z_n)$. Hence similar to the previous section, we have the following two set of simple conditions as well as some stronger but more complicated conditions, Corollary B.2.2 and B.3.1.

Lemma 5.3.1. *Assume the input process $\{V_n\}$ of the surrogate channel $\boldsymbol{W}'$ for EH-SC1 is stationary and ergodic. Then the joint process $\{V_n, Y_n\}$ is AMS and ergodic, if any of the following holds.*

i) *$\boldsymbol{W}'$ is indecomposable.*

ii) *There is a finite vector v_m^N with $\Pr(V_m^N = v_m^N) > 0$ satisfying the following property: given $V_m^N = v_m^N$, for any $z_m, z_m' \in \mathcal{Z}$, there exists $z_{N+1} \in \mathcal{Z}$ such that*

when $Z_m = z_m$ or z'_m, we both have $Z_{N+1} = z_{N+1}$ with positive probability.

Lemma 5.3.2. *Assume the input process $\{V_n\}$ of the surrogate channel $\boldsymbol{W}'$ for EH-SC1 is finite-order Markov, then so is the joint process $\{(V_n, Z_{n+1})\}$. If the underlying Markov chain for the latter is irreducible, then $\{V_n, Y_n\}$ is AMS and ergodic.*

Proof. This lemma is a simple case of Corollary B.3.1. □

In the following simpler setting, we have more concrete conditions.

Theorem 5.3.1. *For the FSC $\boldsymbol{W}'$ with $m = 1$, assume $\{E_n\}$ is i.i.d., the energy model is (4.2) or (4.3), and the distribution of E_n is supported on the full $\mathcal{E}_H$.*

i) *If there exists N such that for each input sequence $v = \{v_n\}$ and any $S_1 = s_1$, $B_N = \overline{B}$ with a positive probability, then $\boldsymbol{W}'$ is indecomposable.*

ii) *If $\mathcal{E}_H$ is a continuous interval of non-negative integers and $\max \mathcal{E}_H - \min \mathcal{E}_H \geq \overline{B}$ for the energy model (4.2), or $\max \mathcal{E}_H \geq \overline{B}$ for the model (4.3), then $\boldsymbol{W}'$ is indecomposable.*

iii) *If $\{V_n\}$ is stationary and ergodic, and there is v^N with $\Pr(V^N = v^N) > 0$ such that for any $S_1 = s_1$, either $B_N = 0$ or $B_N = \overline{B}$ with positive probability, then $\{V_n, Y_n\}$ is AMS and ergodic.*

iv) *Both i) and iii) hold if $\max \mathcal{E}_H > \max\{\gamma(x) : x \in \mathcal{X}\}$.*

Proof. Note that in this case $Z_n = S_n$.

i): Whenever such N exists, the strong positive column condition holds and so $\boldsymbol{W}'$ is indecomposable. (See comments below Definition B.2.2.)

ii): With a positive probability S_2 can always be boosted up to $s_2 = \overline{B} + \min \mathcal{E}_H$ for the model (4.2), or $\overline{B}$ for the model (4.3), hence the strong positive column condition holds.

iii): This is a straightforward application of Lemma 5.3.1, condition ii).

iv): If $\max \mathcal{E}_H > \max\{\gamma(x) : x \in \mathcal{X}\}$, then for any v and s_1, at most after $n = \overline{B}$ transmissions, $S_n - \gamma(X_n) \geq \overline{B}$ with a probability no smaller than $[\Pr(E_n = \max \mathcal{E}_H)]^n > 0$, in which case $B_{n+1} = \overline{B}$. □

Remark 5.3.2. The conditions in ii) and iv) are satisfied if E_n can reach a relatively high energy level (compared to $\mathcal{X}$ or $\overline{B}$) with even a very small positive probability, which is not a harsh requirement for many natural energy sources. Alternatively, if the input process $\{V_n\}$ is stationary ergodic, and put a positive probability on a moderately long sequence of "all zero" functions (that is, $v_n(s_n) = 0$ for all s_n), or "all-consume" functions (that is, $\gamma(v_n(s_n)) = s_n$ for all s_n), then condition iii) is satisfied.

5.4 EH-SC2

Note that as commented in Remark 4.2.2, any achievable rates from EH-SC1 is also achievable for EH-SC2. In fact, for this scenario we can also restrict the input function to depend only on the m most recent energy states, but we need to be more careful about the definition. Consider an input symbol $u^N = (u_1, \cdots, u_N)$ whose n-th coordinate function u_n is only a function of S_{n-m+1}^n, then its alphabet size is a constant that does not depend on n. To be precise, each u_n is associated with an auxiliary function $v_n \in \mathcal{V}$, which is defined as above. The input function u_n is defined through v_n in the following way: for each (b_1, e^n), it first computes $s_n = s_n(b_1, e^n, u^{n-1})$ through the recursion (4.13), then together with the previously computed s_{n-m+1}^{n-1}, u_n assigns the function value

$$u_n(b_1, e^n) = v_n(s_{n-m+1}^n).$$

Hence the vector $v^N = (v_1, \cdots, v_N)$ uniquely determines the input symbol u^N, and for each N there is a one-to-one correspondence between the collection $\mathcal{U}'^{(N)}$ of all such special input symbols u^N and $\mathcal{V}^N$.

With such a restriction for the side information, we see indeed the way that the energy information is used falls in the regime of scenario 1, hence we can use the results from the previous section for this scenario. Also for EH-SC2 we have the same optimal input conjecture as stated in Remark 5.3.1.

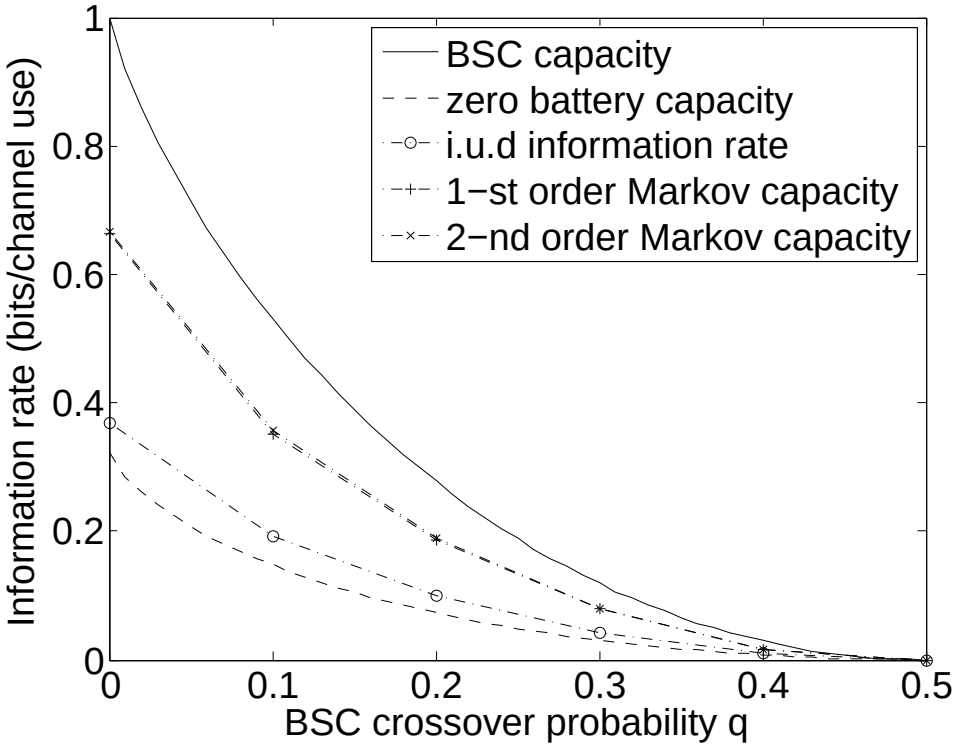


Figure 5.1: The information rates

5.5 Numerical Computation

We use the following example to demonstrate the computation of the achievable rates of the energy harvesting channels.

Example 5.5.1. Assume $\{E_n\}$ is an i.i.d. Bernoulli(0.5) process with $\mathcal{E}_H = \mathcal{X} = \{0, 1\}$. The energy model is (4.2) and (4.6). Assume $\overline{B} = 1$, then $\mathcal{S} = \{0, 1, 2\}$. Let $m = 1$, then $\mathcal{V} = \{v_a, v_b, v_c, v_d\}$ with

$$v_a = (0, 0, 0),\ v_b = (0, 0, 1),\ v_c = (0, 1, 0),\ v_d = (0, 1, 1).$$

Let the DMC be a binary symmetric (BSC) with crossover probability q. Then the condition in case ii) of Theorem 5.3.1 is satisfied and $\boldsymbol{W}'$ is indecomposable. The

channel state transition probability matrices are

$$P(\cdot|v_a,\cdot) = P(\cdot|v_b,\cdot) = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \\ 0 & 0.5 & 0.5 \end{bmatrix},$$

$$P(\cdot|v_c,\cdot) = P(\cdot|v_d,\cdot) = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \end{bmatrix},$$

where, e.g., the (1,3) entry of $P(\cdot|v_a,\cdot)$ represents $P(0|v_a,2)$.

The capacity for the same BSC without energy constraint is $1 - H(q)$, which is an upper bound for the case of infinite battery. When energy arrival process is i.i.d. Bernoulli(p) with $p \geq 0.5$, using the same analysis in [37], we can prove this bound is tight.

On the other extreme, when there is no battery, as commented in Section 4.3, Shannon's method [38] can be modified to obtain an equivalent DMC with input alphabet $\mathcal{U} = \{u_a, u_b\}$, where $u_a = (0,0)$ and $u_b = (0,1)$, both of which are functions of E_n. The transition probability of this DMC is

$$p(y|u) = \sum_{e \in \mathcal{E}_H} p(e)p(y|u(e)), \quad \forall y \in \mathcal{Y},\ u \in \mathcal{U}.$$

in particular, $p(y|u_a) = p(y|0)$ and $p(y|u_b) = 0.5$. The capacity of this DMC is

$$H\left(\frac{1}{1+\alpha}\right) - 1 + r(1 - H(q)), \quad \text{where}$$

$$\alpha = 2^{-\frac{1-H(q)}{0.5-q}} \quad \text{and} \quad r = \frac{(1+\alpha)^{-1} - 0.5}{0.5 - q}.$$

Its capacity can be easily calculated analytically.

For the channel $\boldsymbol{W}'$ we compute the i.u.d. rate, which is the information rate for the i.i.d. uniform input process, and optimize the information rate over Markov input processes of order 1 and 2. The numerical results are shown in Figure 5.1. For

comparison, the capacities for the same BSC without energy constraint and with zero battery are also shown.

We have the following remarks on the numerical results:

1) The minimal non-zero battery storage can give us a great boost on the capacity; it even achieves a significant fraction (around 70%) of the capacity without energy constraints.
2) The Markov input processes achieve higher rate than the i.u.d. input, and with higher order the information rate is higher. So memory in the input helps, but increasing the Markov order by 1 only slightly increases the information rate.

5.6 Discussions

The results in this chapter can be extended to continuous energy harvesting channels, especially when the input alphabet is finite, e.g., AWGN channel with binary input. Although the FSC and Markov channel results are both for finite alphabets, we can consider only the state process itself as the output of an Markov channel, and then a continuous memoryless channel is connected to the output, as discussed in Section B.5.1. For such a case we can still derive ergodicity results and apply the SMB theorem.

On the other hand, we can deal with finite energy harvesting channels with channel memory. For example, if the channel is not DMC, but an FSC, we can still use the approach in this chapter to obtain some achievable rates.

Chapter 6

Capacity Bounds

Compared to the achievable rates (lower bounds), nontrivial upper bounds for the energy harvesting channels are much more difficult to obtain. For the binary noiseless case (with scenario 1) [41] derives an upper bound, assuming full CSI at the receiver (CSIR). [58] tries to tighten this bound, though the approach seems not fully mathematically rigorous.

In Gallager's study of finite state channels [2], two convergent sequences in the form of finite block length mutual information are proved to bound the channel capacity from above and below, respectively. In particular, each term of the sequences gives an upper/lower bound. In this chapter we study our channel models and obtain capacity bounds for them, most of which are in Gallager's flavor. Especially, for EH-SC2 we derive upper bounds when the energy harvesting process is finite-order Markov[1]. For EH-SC1 we only consider the case when the energy harvesting process is i.i.d., which is a special case of FSC-X, as proved in Section 4.2.3. Thus in the following we only present the bounds for FSC-X and EH-SC2. These bounds, although computable in theory for each block length N, are not practical to compute for large N as the complexity is double exponential. To address this issue for the upper bounds, we relax them further to allow for a dynamic programming recursion, which has linear complexity.

We begin with an approach we use for our major upper bound results, which is

[1]Note that in this scenario when the process is i.i.d., a proof sketch for the same bounds also appears in [59].

based on techniques of Verdú and Han and Gallager.

6.1 A General Gallager-type Upper Bound

A general channel $\{\mathcal{X}^{(N)}, p(y^N \mid x^N), \mathcal{Y}^{(N)}\}_{N=1}^{\infty}$ without feedback is defined by describing the input/output alphabets $\mathcal{X}^{(N)}$, $\mathcal{Y}^{(N)}$ and the transition probabilities $p(y^N \mid x^N)$ for each block length N. Using Fano inequality Verdú and Han [1] showed that its capacity is upper bounded by

$$\liminf_{N\to\infty} C_N, \tag{6.1}$$

$$C_N \triangleq \sup_{P_{X^N}} \frac{1}{N} I(X^N; Y^N).$$

In general, the upper bound is not easy to compute, since the limiting behavior of C_N is unknown. On the other hand, Gallager [2] uses the following lemma to derive a series of computable upper bounds for finite state channels:

Lemma 6.1.1 (Fekete's subadditive lemma)**.** *If the sequence* $\{a_n\}_{n=1}^{\infty}$ *is subadditive, i.e.,* $a_{m+n} \le a_m + a_n$ *for all* m, n*, then the limit* $\lim_{n\to\infty} \frac{a_n}{n}$ *exists and is equal to* $\inf \frac{a_n}{n}$.

If we can show that for each N, there is a $\overline{C}_N$ such that

(R1) $C_N \le \overline{C}_N$,

(R2) $\{N\overline{C}_N\}_{N=1}^{\infty}$ is subadditive,

then by Fekete's lemma, $\lim_{N\to\infty} \overline{C}_N$ exists and is equal to $\inf \overline{C}_N$. Hence (6.1) is upper bounded by $\liminf_{N\to\infty} \overline{C}_N = \inf \overline{C}_N$, and so $\overline{C}_N$ is an upper bound for the general channel capacity for each finite N. In other words, the limiting process in (6.1) is not needed anymore, which greatly simplifies the computation of upper bounds, especially when such computable $\overline{C}_N$'s can be easily found.

6.2 FSC-X Upper Bounds

We use the technique above on the equivalent channel (see Section 4.3.1) to derive a series of Gallager-type upper bounds. First of all, the capacity can be upper bounded by a system with full receiver side channel state information (CSIR). Then in (6.1) we consider the mutual information $I(U^N; Y^N S^{N+1})$. As S_1 is independent of U^N,

$$\begin{aligned} I\left(U^N; Y^N S^{N+1}\right) &= I\left(U^N; Y^N S_2^{N+1} \,|\, S_1\right) \\ &\leq \max_{s_1} I\left(U^N; Y^N S_2^{N+1} \,|\, s_1\right), \end{aligned}$$

where $I(\cdot\,;\,\cdot \,|\, s_1) := I(\cdot\,;\,\cdot \,|\, S_1 = s_1)$. Define

$$\overline{C}_N = \max_{P_{U^N}} \max_{s_1} \frac{1}{N} I\left(U^N; Y^N S_2^{N+1} \,|\, s_1\right), \tag{6.2}$$

then $\overline{C}_N$ satisfies (R1) in the previous section. Furthermore,

Theorem 6.2.1. *For each N, $\overline{C}_N$ defined in* (6.2) *is an upper bound for the capacity of the channel FSC-X.*

Proof. As described above, we can use (6.1) for the full CSIR case as an upper bound. Since $\overline{C}_N$ satisfies (R1) for this upper bound, if we can show it satisfies (R2) as well, then $\overline{C}_N$ is an upper bound for each N by Section 6.1.

Let N be arbitrary and let m, n be positive integers that sum to N. In the following we will show that

$$N\overline{C}_N \leq n\overline{C}_n + m\overline{C}_m, \tag{6.3}$$

i.e., $\{N\overline{C}_N\}_{N=1}^{\infty}$ is subadditive. For any P_{U^N} and s_1 consider the decomposition

$$\begin{aligned} I\left(U^N; Y^N S_2^{N+1} \mid s_1\right) &= I\left(U^N; Y^n S_2^{n+1} \mid s_1\right) + I\left(U^N; Y_{n+1}^N S_{n+2}^{N+1} \mid Y^n S_2^{n+1} s_1\right) \\ &= I\left(U^n; Y^n S_2^{n+1} \mid s_1\right) + I\left(U_{n+1}^N; Y^n S_2^{n+1} \mid U^n s_1\right) \\ &\quad + I\left(U_{n+1}^N; Y_{n+1}^N S_{n+2}^{N+1} \mid Y^n S_2^{n+1} s_1\right) \end{aligned}$$

$$
\begin{aligned}
&+ I\left(U^n; Y_{n+1}^N S_{n+2}^{N+1} \mid U_{n+1}^N Y^n S_2^{n+1} s_1\right) \\
=\;& I_1 + I_2 + I_3 + I_4, \qquad (6.4)
\end{aligned}
$$

where I_1–I_4 are respectively defined as the first to fourth terms in the line above them. By the definition (6.2), $I_1 \le n\overline{C}_n$. Next, using the property of FSC conditional probabilities as in (4.11), for I_2 and I_4 we respectively have

$$
\begin{aligned}
p\left(y^n s_2^{n+1} \mid u_{n+1}^N u^n s_1\right) &= \prod_{i=1}^{n} p\left(y_i s_{i+1} \mid x_i = u_i(s^i), s_i\right) \\
&= p\left(y^n s_2^{n+1} \mid u^n s_1\right),
\end{aligned}
$$

$$
\begin{aligned}
p\big(y_{n+1}^N s_{n+2}^{N+1} \mid u^n u_{n+1}^N y^n s_2^{n+1} s_1\big) &= \prod_{i=n+1}^{N} p\left(y_i s_{i+1} \mid x_i = u_i(s^i), s_i\right) \\
&= p\left(y_{n+1}^N s_{n+2}^{N+1} \mid u_{n+1}^N y^n s_2^{n+1} s_1\right).
\end{aligned}
$$

Therefore $I_2 = I_4 = 0$. Furthermore,

$$
I_3 = \sum_{y^n s_2^{n+1}} p\left(y^n s_2^{n+1} \mid s_1\right) I\left(U_{n+1}^N; Y_{n+1}^N S_{n+2}^{N+1} \mid y^n s^{n+1}\right). \qquad (6.5)
$$

Fix $y^n s^{n+1}$. For each u_{n+1}^N and $k = 1, \ldots, m$, let

$$
\tilde{u}_k : \mathcal{S}^k \to \mathcal{X}
$$

be the projection $\tilde{u}_k(\cdot) = u_{n+k}(s^n, \cdot)$, i.e.,

$$
\tilde{u}_k\left(t^k\right) = u_{n+k}\left(s^n, t^k\right), \qquad \forall t^k \in \mathcal{S}^k.
$$

Then $\forall t^k \in \mathcal{S}^k$, $\tilde{u}_k\left(t^k\right) \in \mathcal{X}(t_k)$ and so $\tilde{u}_k \in \mathcal{U}_k$. By (4.11) again

$$
p\big(y_{n+1}^N s_{n+2}^{N+1} \mid u_{n+1}^N y^n s^{n+1}\big) = \prod_{i=n+1}^{N} p\left(y_i s_{i+1} \mid x_i = u_i(s^i), s_i\right)
$$

$$
\begin{aligned}
&= \prod_{k=1}^{m} p\left(y_{n+k}s_{n+k+1} \,\middle|\, x_{n+k} = u_{n+k}\left(s^n, s_{n+1}^{n+k}\right), s_{n+k}\right) \\
&= \prod_{k=1}^{m} p\left(y_{n+k}s_{n+k+1} \,\middle|\, x_{n+k} = \tilde{u}_k\left(s_{n+1}^{n+k}\right), s_{n+k}\right) \\
&= Q\left(y_{n+1}^N s_{n+2}^{N+1} \,|\, \tilde{u}^m s_{n+1}\right),
\end{aligned}
$$

where $Q \triangleq P_{Y^m S_2^{m+1} \,|\, U^m S_1}$ is the m-block channel transition probability given S_1. Denote the projection map

$$T : u_{n+1}^N \mapsto \tilde{u}^m,$$

which depends on s^n. Then T and $P_{U_{n+1}^N \,|\, y^n s^{n+1}}$ induce a probability distribution $\tilde{P}$ on $\mathcal{U}^{(m)}$: for all $\tilde{u}^m \in \mathcal{U}^{(m)}$,

$$
\begin{aligned}
\tilde{P}\left(\tilde{u}^m\right) &= \Pr\left(T\left(U_{n+1}^N\right) = \tilde{u}^m \,\middle|\, Y^n S^{n+1} = y^n s^{n+1}\right). \\
&= \sum_{u_{n+1}^N : T\left(u_{n+1}^N\right) = \tilde{u}^m} p\left(u_{n+1}^N \,|\, y^n s^{n+1}\right).
\end{aligned}
$$

Now it is easy to verify that

$$
\begin{aligned}
p\left(y_{n+1}^N s_{n+2}^{N+1} \,|\, y^n s^{n+1}\right) &= \sum_{u_{n+1}^N} p\left(u_{n+1}^N \,|\, y^n s^{n+1}\right) p\left(y_{n+1}^N s_{n+2}^{N+1} \,|\, u_{n+1}^N y^n s^{n+1}\right) \\
&= \sum_{u_{n+1}^N} p\left(u_{n+1}^N \,|\, y^n s^{n+1}\right) Q\left(y_{n+1}^N s_{n+2}^{N+1} \,|\, \tilde{u}^m s_{n+1}\right) \\
&= \sum_{\tilde{u}^m} \tilde{P}\left(\tilde{u}^m\right) Q\left(y_{n+1}^N s_{n+2}^{N+1} \,|\, \tilde{u}^m s_{n+1}\right) \\
&= \tilde{R}\left(y_{n+1}^N s_{n+2}^{N+1} \,|\, s_{n+1}\right),
\end{aligned}
$$

where $\tilde{R}(\cdot \,|\, s_{n+1})$ is the m-block channel output distribution given $S_1 = s_{n+1}$, induced by $\tilde{P}$ and the channel Q. Thus if we denote the relative entropy

$$D_{u_{n+1}^N \,|\, y^n s^{n+1}} \triangleq D\left(P_{Y_{n+1}^N S_{n+2}^{N+1} \,|\, u_{n+1}^N y^n s^{n+1}} \,\middle\|\, P_{Y_{n+1}^N S_{n+2}^{N+1} \,|\, y^n s^{n+1}}\right),$$

then

$$D_{u_{n+1}^N \,|\, y^n s^{n+1}} = D\big(Q\left(\,\cdot \,|\, \tilde{u}^m s_{n+1}\right) \,\|\, \tilde{R}(\cdot \,|\, s_{n+1})\big).$$

Therefore we can write

$$\begin{aligned}
I\big(U_{n+1}^N; Y_{n+1}^N S_{n+2}^{N+1} \,|\, y^n s^{n+1}\big) &= \sum_{u_{n+1}^N} p\left(u_{n+1}^N \,|\, y^n s^{n+1}\right) D_{u_{n+1}^N \,|\, y^n s^{n+1}} \\
&= \sum_{u_{n+1}^N} p\left(u_{n+1}^N \,|\, y^n s^{n+1}\right) D\big(Q(\,\cdot \,|\, \tilde{u}^m s_{n+1}) \,\|\, \tilde{R}(\cdot \,|\, s_{n+1})\big) \\
&= \sum_{\tilde{u}^m} \tilde{P}(\tilde{u}^m) D\big(Q(\,\cdot \,|\, \tilde{u}^m s_{n+1}) \,\|\, \tilde{R}(\cdot \,|\, s_{n+1})\big) \\
&= I_{\tilde{P}}\left(U^m; Y^m S_2^{m+1} \,|\, S_1 = s_{n+1}\right) \\
&\le \max_{s_1} I_{\tilde{P}}\left(U^m; Y^m S_2^{m+1} \,|\, s_1\right) \\
&\le m\overline{C}_m,
\end{aligned}$$

where $I_{\tilde{P}}$ denotes the mutual information induced by the input distribution $\tilde{P}$. Since this inequality holds for all $y^n s^{n+1}$, by (6.5) we have

$$I_3 \le m\overline{C}_m.$$

Combining the results for I_1–I_4 with (6.4), we have

$$I\left(U^N; Y^N S_2^{N+1} \,|\, s_1\right) \le n\overline{C}_n + m\overline{C}_m.$$

This inequality is true for all P_{U^N} and s_1, so it must be true for the maximization over them, and thus (6.3) holds. □

Remark 6.2.1. Note that since the order of maximization in (6.2) can be exchanged, $\overline{C}_N$ can be calculated by finding the capacities of $|\mathcal{S}|$ discrete memoryless channels (DMC), which can be efficiently computed using the Blahut-Arimoto algorithms (see, e.g., [56]).

6.3 EH-SC2

For this model we will obtain some new capacity bounds in the form of finite block length mutual information, especially the Gallager-type upper bounds. We again used the equivalent channel model (see Section 4.3.3). Note that the energy harvesting process and the initial battery level are independent, and in the equivalent channel neither of them is known at the transmitter, so they are also independent of the input. For $r \geq 0$, define $e^{-r} := e^0_{-r+1}$. We have

$$p\left(y^N e^N | u^N b_1 e^{-r}\right) = p(e^N|e^{-r}) \prod_{n=1}^{N} p\left(y_n | x_n = u_n(b_1, e^n)\right). \tag{6.6}$$

Next we develop some preliminary results on the input alphabet and block conditional mutual information. Recall that $\mathcal{U}^{(N)}$ is the collection of all causal mappings $\mathcal{E}_B \times \mathcal{E}_H^N \to \mathcal{X}^N$ that are consistent with the energy constraint (4.1). Define $\mathcal{U}_{b_1}^{(N)}$ as the "b_1-th section of $\mathcal{U}^{(N)}$", which consists of all causal mappings $\mathcal{E}_H^N \to \mathcal{X}^N$ (which are denoted by $\mathcal{V}^{(N)}$) that together with $B_1 = b_1$ satisfy the energy constraint:

$$\mathcal{U}_{b_1}^{(N)} \triangleq \left\{ v^N = u^N(b_1, \cdot) \mid u^N \in \mathcal{U}^{(N)} \right\}.$$

Let $b_1 \leq b_1'$. For each e^N, x^N satisfies (4.1) with b_1' whenever it does with b_1, so $\mathcal{U}_{b_1}^{(N)} \subseteq \mathcal{U}_{b_1'}^{(N)}$. In particular,

$$\mathcal{U}_{b_1}^{(N)} \subseteq \mathcal{U}_{\overline{B}}^{(N)}, \qquad \forall b_1 \in \mathcal{E}_B. \tag{6.7}$$

Now fix b_1. Define the projection map

$$\begin{aligned} T : \mathcal{U}^{(N)} &\to \mathcal{V}^{(N)} \\ u^N &\mapsto u^N(b_1, \cdot) \end{aligned}$$

and denote

$$\hat{u}^N = T\left(u^N\right),$$

whose image is in $\mathcal{U}_{b_1}^{(N)}$. Furthermore, for $v^N \in \mathcal{V}^{(N)}$ and $r \geq 0$ we define

$$p\left(y^N e^N \,|\, v^N e^{-r}\right) = p\left(e^N \,|\, e^{-r}\right) \prod_{n=1}^{N} p\left(y_n \,|\, x_n = v_n(e^n)\right),$$

$$p\left(y^N \,|\, v^N\right) = \sum_{e^N} p\left(y^N e^N \,|\, v^N\right)$$

through $p(e^N)$ and $p(y|x)$. Then by (6.6) we have

$$p\left(y^N e^N \,|\, u^N b_1 e^{-r}\right) = p\left(y^N e^N \,|\, \hat{u}^N e^{-r}\right),$$
$$p\left(y^N \,|\, u^N b_1\right) = p\left(y^N \,|\, \hat{u}^N\right).$$

By the same argument as in the proof of Theorem 6.2.1,

$$I(U^N; Y^N E^N \,|\, b_1 e^{-r}) = I(\hat{U}^N; Y^N E^N \,|\, e^{-r}), \tag{6.8}$$
$$I(U^N; Y^N \,|\, b_1) = I(\hat{U}^N; Y^N), \tag{6.9}$$

where the distribution of $\hat{U}^N = T(U^N)$ is supported on $\mathcal{U}_{b_1}^{(N)}$.

Lemma 6.3.1. *Let $\mathcal{P}_{b_1}^{(N)}$ denote the family of all probability distributions on $\mathcal{U}_{b_1}^{(N)}$. We have*

$$\max_{P_{U^N}} I(U^N; Y^N E^N | b_1 e^{-r}) = \max_{P_{V^N} \in \mathcal{P}_{b_1}^{(N)}} I(V^N; Y^N E^N | e^{-r}),$$
$$\max_{P_{U^N}} I(U^N; Y^N \,|\, b_1) = \max_{P_{V^N} \in \mathcal{P}_{b_1}^{(N)}} I(V^N; Y^N).$$

Proof. We only prove the second equation since the proof of the first is essentially the same. Denote the LHS and RHS of the second equation by C_U and C_V, respectively. For any P_{U^N}, we have $P_{\hat{U}^N} \in \mathcal{P}_{b_1}^{(N)}$ and so

$$I(U^N; Y^N \,|\, b_1) \leq C_V$$

by (6.9), and hence $C_U \leq C_V$. On the other hand, $T^{-1}\left(v^N\right) \neq \emptyset$ for every $v^N \in \mathcal{U}_{b_1}^{(N)}$.

Thus for any $P_{V^N} \in \mathcal{P}_{b_1}^{(N)}$, define a $P_{U^{(N)}}$ such that

$$P_{U^{(N)}}\left(T^{-1}\left(v^N\right)\right) = P_{V^N}\left(v^N\right)$$

for all $v^N \in \mathcal{U}_{b_1}^{(N)}$, then P_{V^N} is induced by P_{U^N} and T. Then by (6.9) again,

$$C_U \geq I(V^N; Y^N)$$

and so $C_U \geq C_V$. □

Now we are ready to present the new capacity bounds.

Theorem 6.3.1. *If* $\{E_n\}_{n=1}^{\infty}$ *is* i.i.d.*, then for each* N

$$\underline{C}_N := \max_{P_{U^N}} \frac{1}{N} I(U^N; Y^N \mid B_1 = 0)$$

is a lower bound of the channel capacity for EH-SC2.

Proof. Consider using the channel in blocks of length N and restrict the input functions to those that i) ignore the initially stored energy in the battery, and ii) essentially comprise concatenations of functions in $\mathcal{U}_0^{(N)}$. That is, for $k > 0$ the input u^{kN} is only a function of e^{kN} and can be identified with the collection

$$\left\{\boldsymbol{v}_i \in \mathcal{U}_0^{(N)}, 1 \leq i \leq k\right\},$$

where for any b_1 and e^{kN},

$$u^{kN}\left(b_1, e^{kN}\right) = \left(\boldsymbol{v}_1\left(e^N\right), \ldots, \boldsymbol{v}_k\left(e_{(k-1)N+1}^{kN}\right)\right).$$

It is a legitimate input symbol since between the transition of blocks the function ignores the remaining battery energy, thus is always compatible with the energy constraint (4.1).

Let $\boldsymbol{y}_i$ and $\overline{e}_i$ denote $y_{(i-1)N+1}^{iN}$ and $e_{(i-1)N+1}^{iN}$, respectively. By (6.6) and the *i.i.d.*

assumption for E_n,

$$p\left(y^{kN}e^{kN} \mid u^{kN}b_1\right) = \prod_{i=1}^{k} p(\overline{e}_i)p(\boldsymbol{y}_i \mid \boldsymbol{v}_i(\overline{e}_i)),$$

$$p\left(y^{kN} \mid u^{kN}\right) = \prod_{i=1}^{k} p\left(\boldsymbol{y}_i \mid \boldsymbol{v}_i\right).$$

Note that since E_n is *i.i.d.*, $p\left(\boldsymbol{y}_i \mid \boldsymbol{v}_i\right)$ is independent of i. Thus kN times of using the original channel in the specified manner is equivalent to k times of using a discrete memoryless channel $p\left(\boldsymbol{y} \mid \boldsymbol{v}\right)$ with input alphabet $\mathcal{U}_0^{(N)}$, whose capacity is

$$\max_{P_{V^N} \in \mathcal{P}_0^{(N)}} I(V^N; Y^N).$$

By Lemma 6.3.1 and considering the block length N, $\underline{C}_N$ is achievable. □

Theorem 6.3.2. *If $\{E_n\}_{n=1}^{\infty}$ is a homogeneous Markov chain of order $r \geq 0$, then for each N*

$$\overline{C}_N := \max_{P_{U^N}} \max_{e^{-r}} \frac{1}{N} I(U^N; Y^N \mid E^N, B_1 = \overline{B}, e^{-r})$$

is an upper bound of the channel capacity for EH-SC2.

Proof. We use the upper bounding technique in Section 6.2 and the proof parallels that of Theorem 6.2.1. By providing full CSIR to the receiver, in (6.1) we consider

$$\begin{aligned} I\left(U^N; Y^N E_{-r+1}^N B_1\right) &= I\left(U^N; Y^N E^N \mid B_1 E^{-r}\right) \\ &\leq \max_{b_1, e^{-r}} I\left(U^N; Y^N E^N \mid b_1 e^{-r}\right), \end{aligned}$$

due to the independence between B_1E^{-r} and U^N. Now define

$$\overline{C}_N = \max_{P_{U^N}} \max_{b_1, e^{-r}} \frac{1}{N} I\left(U^N; Y^N E^N \mid b_1 e^{-r}\right). \tag{6.10}$$

We will show that it is equivalent to the definition in the theorem. For each b_1e^{-r}, by

Lemma 6.3.1, (6.7) and the independence between $\{E_n\}_{n=1}^{\infty}$ and the input symbols,

$$\begin{aligned}\max_{P_{U^N}} I(U^N; Y^N E^N \,|\, b_1 e^{-r}) &\le \max_{P_{V^N} \in \mathcal{P}_{\overline{B}}^{(N)}} I(V^N; Y^N E^N \,|\, e^{-r}) \\ &= \max_{P_{U^N}} I(U^N; Y^N E^N \,|\, B_1 = \overline{B}, e^{-r}) \\ &= \max_{P_{U^N}} I(U^N; Y^N \,|\, E^N, B_1 = \overline{B}, e^{-r})\end{aligned}$$

with the equality attained when $b_1 = \overline{B}$. Now, taking the maximum of both sides over e^{-r} and exchanging the order of maximization, we see the equivalence of both definitions.

From the analysis above $\overline{C}_N$ satisfies (R1) in Section 6.1. Next we will show the subadditivity (6.3) and then the theorem is proved. Let N be arbitrary and let m, n be positive integers that sum to N. We have the decomposition

$$\begin{aligned}I\left(U^N; Y^N E^N \,|\, b_1 e^{-r}\right) &= I\left(U^n; Y^n E^n \,|\, b_1 e^{-r}\right) + I\left(U_{n+1}^N; Y^n E^n \,|\, U^n b_1 e^{-r}\right) \\ &\quad + I\left(U_{n+1}^N; Y_{n+1}^N E_{n+1}^N \,|\, Y^n E^n b_1 e^{-r}\right) \\ &\quad + I\left(U^n; Y_{n+1}^N E_{n+1}^N \,|\, U_{n+1}^N Y^n E^n b_1 e^{-r}\right) \\ &= I_1 + I_2 + I_3 + I_4, \end{aligned} \tag{6.11}$$

where I_1–I_4 are respectively defined as the first to fourth terms above. By the definition (6.10), $I_1 \le n\overline{C}_n$. Next using (6.6) we can show that $I_2 = I_4 = 0$. Furthermore,

$$I_3 = \sum_{y^n e^n} p\left(y^n e^n \,|\, b_1 e^{-r}\right) I\left(U_{n+1}^N; Y_{n+1}^N E_{n+1}^N \,|\, y^n b_1 e_{-r+1}^n\right). \tag{6.12}$$

Fix $y^n b_1 e_{-r+1}^n$. For each u_{n+1}^N define the projection map

$$u_{n+1}^N \mapsto \tilde{u}^m := u_{n+1}^N(b_1 e^n, \cdot).$$

Since u_{n+1}^N is extracted from a legal input function $u^N \in \mathcal{U}^{(N)}$, for any e_{n+1}^N the output $\tilde{u}^m(e_{n+1}^N) = u_{n+1}^N(b_1 e^n, e_{n+1}^N)$ needs to satisfy (4.1) with the intermediate battery level

b_{n+1}, which is determined by u^n and $b_1 e^n$. Hence

$$\tilde{u}^m \in \mathcal{U}_{b_{n+1}}^{(m)} \subseteq \mathcal{U}_{\overline{B}}^{(m)}$$

by (6.7). Now by (6.6)

$$\begin{aligned} p\big(y_{n+1}^N e_{n+1}^N \,|\, u_{n+1}^N y^n b_1 e_{-r+1}^n\big) &= p(e_{n+1}^N \,|\, e_{n-r+1}^n) \cdot p\big(y_{n+1}^N \,|\, u_{n+1}^N (b_1 e^n, e_{n+1}^N)\big) \\ &= p(e_{n+1}^N \,|\, e_{n-r+1}^n) \cdot p\big(y_{n+1}^N \,|\, \tilde{u}^m (e_{n+1}^N)\big) \\ &= P_{Y^m E^m \,|\, V^m E^{-r}} \left(y_{n+1}^N e_{n+1}^N \,|\, \tilde{u}^m e_{n-r+1}^n\right), \end{aligned}$$

where we used the Markov property of E_n. Again similar to Theorem 6.2.1, for an induced distribution $\tilde{P}$ on $\mathcal{U}_{\overline{B}}^{(m)}$

$$\begin{aligned} I\big(U_{n+1}^N; Y_{n+1}^N E_{n+1}^N \,|\, y^n b_1 e_{-r+1}^n\big) &= I_{\tilde{P}}\big(V^m; Y^m E^m \,|\, e_{n-r+1}^n\big) \\ &\le \max_{P_{V^m} \in \mathcal{P}_{\overline{B}}^{(m)}} I(V^m; Y^m E^m \,|\, e_{n-r+1}^n) \\ &= \max_{P_{U^m}} I(U^m; Y^m E^m | B_1 = \overline{B}, e_{n-r+1}^n) \\ &\le m\overline{C}_m, \end{aligned}$$

where we used Lemma 6.3.1. Since this inequality holds for all $y^n b_1 e_{-r+1}^n$, by (6.12) we have

$$I_3 \le m\overline{C}_m.$$

Combining the results for I_1–I_4 with (6.11), we have

$$I\left(U^N; Y^N E^N \,|\, b_1 e^{-r}\right) \le n\overline{C}_n + m\overline{C}_m$$

for arbitrary P_{U^N} and $b_1 e^{-r}$, thus (6.3) holds. □

Remark 6.3.1. As stated in Remark 6.2.1, $\overline{C}_N$ can be computed by finding the capacities of a finite number of DMC's.

6.4 Linear Complexity Upper Bounds

As remarked at the end of Section 4.3, for the equivalent channel models the alphabet size for each channel use grows double exponentially, and hence so does the computational complexity. This poses a problem in practice for the computation of the block mutual information bounds above. In the following we will loosen the upper bounds to keep the alphabet size fixed, which also leads to a nice dynamic programming recursion that gives a linear complexity algorithm. If we relax the bounds even further, the recursions can be solved analytically. Although these relaxed bounds are looser than the original ones for each block length N, as we can compute them for very large N, the resulting bounds are often tighter in practice.

These relaxation methods are inspired by the study of FSC with feedback [60] [61].

6.4.1 FSC-X

For each $s \in \mathcal{S}$ define $\mathcal{P}_s^*$ to be the set of all probability distributions on $\mathcal{X}(s)$ and $\mathcal{P}^* = \prod_{s\in\mathcal{S}} \mathcal{P}_s^*$. We say the conditional distribution $P_{X|S} \in \mathcal{P}^*$ iff $P_{X|S}(\cdot \,|\, s) \in \mathcal{P}_s^*$ for all $s \in \mathcal{S}$. Let $p_n \triangleq P_{X_n|S_n}$, we write $\{p_n\}_{n=1}^N \subset \mathcal{P}^*$ if $p_n \in \mathcal{P}^*$ for all $1 \le n \le N$. Moreover, define $Q = P_{S_{n+1}|X_nS_n}$, then

$$Q(s_{n+1} \,|\, x_n s_n) = \sum_{y_n} p(y_n s_{n+1} \,|\, x_n s_n).$$

Also define

$$I(p, s) = I(Y_nS_{n+1}; X_n \,|\, S_n = s)\big|_{p_n=p} \,.$$

Now we begin relaxing $\overline{C}_N$ in Section 6.2. For a fixed $S_1 = s_1$,

$$I\big(U^N; Y^N S_2^{N+1} \,|\, s_1\big) = \sum_{n=1}^{N} I\left(U^N; Y_nS_{n+1} \,|\, Y^{n-1}S^n\right).$$

Observe that by (4.7), (4.11), and $X_n = U_n(S^n)$, we have

$$I\big(U^N; Y_nS_{n+1} \,|\, Y^{n-1}S^n\big) = H(Y_nS_{n+1} \,|\, Y^{n-1}S^n) - H(Y_nS_{n+1} \,|\, Y^{n-1}S^nU^N)$$

$$
\begin{aligned}
&= H(Y_n S_{n+1} \,|\, Y^{n-1} S^n) - H(Y_n S_{n+1} \,|\, Y^{n-1} S^n U^N X_n) \\
&\le H(Y_n S_{n+1} \,|\, S_n) - H(Y_n S_{n+1} \,|\, S_n X_n) \\
&= I(Y_n S_{n+1}; X_n \,|\, S_n),
\end{aligned}
$$

$$
I\big(U^N; Y^N S_2^{N+1} \,|\, s_1\big) \le \sum_{n=1}^{N} I(Y_n S_{n+1}; X_n \,|\, S_n). \tag{6.13}
$$

For any P_{U^N} and P_{S_1}, (U^N, S^{N+1}, Y^N) induces a random tuple (X^N, S^{N+1}, Y^N) through the functional dependence of U^N and X^N. It further induces a set of conditional probabilities $\{p_n\}_{n=1}^N \subset \mathcal{P}^*$, which together with P_{S_1} and $p(y_n s_{n+1} \,|\, x_n s_n)$ uniquely determines $\sum_{n=1}^N I(Y_n S_{n+1}; X_n \,|\, S_n)$. (Cf. [61, App. VIII]). On the other hand, given P_{S_1} and any set of conditional probabilities $\{p_n\}_{n=1}^N \subset \mathcal{P}^*$, we can always construct a compatible random tuple (X^N, S^{N+1}, Y^N). Therefore for any P_{U^N},

$$
I\big(U^N; Y^N S_2^{N+1} \,|\, s_1\big) \le \tilde{c}_{N,s},
$$

$$
\tilde{c}_{N,s} \triangleq \max_{\{p_n\}_{n=1}^N \subset \mathcal{P}^*} \sum_{n=1}^{N} I(Y_n S_{n+1}; X_n \,|\, S_n) \Big|_{S_1 = s}.
$$

Hence we have a new upper bound:

Theorem 6.4.1. *For each N,*

$$
\overline{C}_N \le \tilde{C}_N \triangleq \frac{1}{N} \max_{s \in \mathcal{S}} \tilde{c}_{N,s} \ , \tag{6.14}
$$

which is a capacity upper bound for the channel FSC-X.

Observe that the optimization for $\tilde{c}_{N,s}$ is over the distributions $\{p_n\}_{n=1}^N$, whose alphabet sizes are fixed. Furthermore, the following theorem gives a recursive algorithm to compute $\tilde{c}_{N,s}$, which has a complexity linear in N.

Theorem 6.4.2. *Let S_1 have an arbitrary distribution π,*

$$
\tilde{c}_N(\pi) \triangleq \max_{\{p_n\}_{n=1}^N \subset \mathcal{P}^*} \sum_{n=1}^{N} I(Y_n S_{n+1}; X_n \,|\, S_n).
$$

Then $\tilde{c}_N(\pi) = \sum_{s\in\mathcal{S}} \pi(s)\cdot \tilde{c}_{N,s}$ *where for each* s

$$\tilde{c}_{N,s} = \max_{p(\cdot|s)\in\mathcal{P}_s^*} \left[I(p,s) + \sum_x p(x|s) \sum_t Q(t|xs)\cdot \tilde{c}_{N-1,t} \right],$$

with the initial condition $\tilde{c}_{0,s} = 0$, $\forall s \in \mathcal{S}$.

Proof. From the definitions, $\tilde{c}_{N,s} = \tilde{c}_N(\delta_s)$ where δ_s puts probability 1 on s. For $N = 1$, the theorem is true (note that the optimization for $\tilde{c}_1(\pi)$ is over $\{p_1(\cdot|s) : s \in \mathcal{S}\}$, which can be separated). Assume it is true for $N = k$, then for $N = k+1$,

$$\tilde{c}_{k+1}(\pi) = \max_{p_1\in\mathcal{P}^*} \left[\sum_s \pi(s) I(p_1, s) + \max_{\{p_n\}_{n=2}^{k+1}\subset\mathcal{P}^*} \sum_{n=2}^{k+1} I(Y_n S_{n+1}; X_n \,|\, S_n) \right].$$

Given π and p_1,

$$P_{S_2}(t) = \sum_{x,s} \pi(s) p_1(x|s) Q(t|xs).$$

Now define

$$(\hat{X}^k, \hat{S}^{k+1}, \hat{Y}^k) = (X_2^{k+1}, S_2^{k+2}, Y_2^{k+1})$$

and so

$$\hat{p}_n \triangleq P_{\hat{X}_n|\hat{S}_n} = p_{n+1}, \quad 1 \le n \le k.$$

Using the theorem for $N = k$,

$$\begin{aligned}
\max_{\{p_n\}_{n=2}^{k+1}\subset\mathcal{P}^*} \sum_{n=2}^{k+1} I(Y_n S_{n+1}; X_n \,|\, S_n) &= \max_{\{\hat{p}_n\}_{n=1}^{k}\subset\mathcal{P}^*} \sum_{n=1}^{k} I(\hat{Y}_n \hat{S}_{n+1}; \hat{X}_n \,|\, \hat{S}_n) \\
&= \tilde{c}_k(P_{\hat{S}_1}) \\
&= \tilde{c}_k(P_{S_2}) \\
&= \sum_{t\in\mathcal{S}} P_{S_2}(t)\cdot \tilde{c}_{k,t},
\end{aligned}$$

since the value of $\sum_{n=1}^{k} I(\hat{Y}_n \hat{S}_{n+1}; \hat{X}_n \,|\, \hat{S}_n)$ is uniquely determined by $P_{\hat{S}_1}$, $\{\hat{p}_n\}_{n=1}^{k}$

and the time-invariant transition probabilities $p(y_n s_{n+1} \,|\, x_n s_n)$. Thus

$$\begin{aligned}\tilde{c}_{k+1}(\pi) &= \max_{p_1 \in \mathcal{P}^*} \left[\sum_s \pi(s) I(p_1, s) + \sum_t P_{S_2}(t) \cdot \tilde{c}_{k,t} \right] \\ &= \max_{p_1 \in \mathcal{P}^*} \left[\sum_s \pi(s) I(p_1, s) + \sum_s \pi(s) \sum_x p_1(x|s) \sum_t Q(t|xs) \cdot \tilde{c}_{k,t} \right] \\ &= \sum_s \pi(s) \max_{p_1(\cdot|s) \in \mathcal{P}_s^*} \left[I(p_1, s) + \sum_x p_1(x|s) \sum_t Q(t|xs) \cdot \tilde{c}_{k,t} \right].\end{aligned}$$

Letting $\pi = \delta_s$ we obtain the statement for $\tilde{c}_{k+1,s}$, which can be plugged back for every s into the expression above to obtain the result for $\tilde{c}_{k+1}(\pi)$. So the theorem is true for $N = k + 1$ and hence true for all N. □

Note that for every recursion we only need to maximize the sum of a concave function $I(\cdot, s)$ and a linear term over the same space $\mathcal{P}_s^*$, which is simple to implement using convex optimization. The recursion can even be solved analytically if we relax $\tilde{C}_N$ further. From (6.13) and

$$I(Y_n S_{n+1}; X_n \,|\, S_n) \leq H(X_n \,|\, S_n) \tag{6.15}$$

we can replace the mutual information in the definitions of $\tilde{c}_{N,s}, \tilde{C}_N$ and $\tilde{c}_N(\pi)$ by the corresponding conditional entropies to define $\tilde{c}'_{N,s}, \tilde{C}'_N$ and $\tilde{c}'_N(\pi)$ and obtain a corresponding new theorem:

Theorem 6.4.3. *Assume the base of* log *is e. We have*

$$\tilde{C}_N \leq \tilde{C}'_N,$$

$$\tilde{c}'_N(\pi) = \sum_{s \in \mathcal{S}} \pi(s) \cdot \tilde{c}'_{N,s},$$

$$\tilde{c}'_{N,s} = \log \sum_{x \in \mathcal{X}(s)} \exp \left[\sum_t Q(t|xs) \cdot \tilde{c}'_{N-1,t} \right],$$

with the initial condition $\tilde{c}'_{0,s} = 0$, $\forall s \in \mathcal{S}$.

Proof. By (6.15), $\tilde{C}_N \le \tilde{C}'_N$. Using arguments similar to Theorem 6.4.2 and defining

$$H(p,s) = H(X_n \,|\, S_n = s)\big|_{p_n=p},$$

we have

$$\tilde{c}'_N(\pi) = \sum_{s\in\mathcal{S}} \pi(s)\cdot \tilde{c}'_{N,s},$$

$$\tilde{c}'_{N,s} = \max_{p(\cdot|s)\in\mathcal{P}^*_s} \left[H(p,s) + \sum_x p(x|s) \sum_t Q(t|xs)\cdot \tilde{c}'_{N-1,t} \right].$$

Denote

$$\alpha_x = \sum_t Q(t|xs)\cdot \tilde{c}'_{N-1,t}, \qquad r_x = p(x|s).$$

The optimization problem above can be written as

$$\begin{array}{rc} \text{maximize} & -\sum_x r_x \log r_x + \sum_x r_x \alpha_x \\ \text{s.t.} & \sum_x r_x = 1 \\ & r_x \ge 0,\ \forall x \in \mathcal{X}(s) \end{array},$$

whose solution r^* can be easily found using KKT conditions:

$$r^*_x = \frac{e^{\alpha_x}}{\sum_{x'\in\mathcal{X}(s)} e^{\alpha_{x'}}}.$$

Plugging into the objective function, we obtain the desired formula for $\tilde{c}'_{N,s}$. □

Remark 6.4.1. When X_n is uniquely determined by Y_n (e.g., $Y_n = X_n$), (6.15) holds with equality and $\tilde{C}'_N = \tilde{C}_N$.

6.4.2 EH-SC2

We want to use the techniques above to obtain a linear complexity relaxation for the upper bound in Theorem 6.3.2. For that purpose we introduce the "overall" state

$$Z_n \triangleq E^n_{n-r+1} S_n,$$

with alphabet $\mathcal{Z} \triangleq \mathcal{E}_H^r \times \mathcal{S}$. When $r = 0$, E_n is i.i.d. and $Z_n = S_n$, whose transition probability is described by (4.8). Now assume $r > 0$. For each $z_n = e_{n-r+1}^n s_n \in \mathcal{Z}$, if $z_{n+1} = e_{n-r+2}^{n+1} s_{n+1}$ for some $e_{n+1}s_{n+1}$, then the transition probability

$$p(z_{n+1} \,|\, x_n z_n) = p(e_{n+1} \,|\, e_{n-r+1}^n) \cdot \mathbf{1}_{\{S(x_n, s_n, e_{n+1}) = s_{n+1}\}},$$

otherwise

$$p(z_{n+1} \,|\, x_n z_n) = 0.$$

Note that the transition probabilities are time-invariant (independent of n).

Similar to Section 6.4.1, we define $\mathcal{P}_z^*, \mathcal{P}^*, p_n$, and Q w.r.t. the state Z_n and with

$$\mathcal{X}(z) \triangleq \{x \in \mathcal{X} : |x|^2 \le s(z)\}, \quad \forall z \in \mathcal{Z},$$

where $s(z)$ is the $\mathcal{S}$-component of z. Next define

$$I(p, z) = I(Y_n E_{n+1} Z_{n+1}; X_n \mid Z_n = z)\big|_{p_n = p}.$$

Moreover, let $\pi_{b_1, e^{-r}}$ denote the distribution of Z_1 when $B_1 = b_1$ and $E^{-r} = e^{-r}$, which is determined by

$$p(e_1 s_1 | b_1 e^{-r}) = p(e_1 \,|\, e^{-r}) \cdot \mathbf{1}_{\{S(b_1, e_1) = s_1\}}.$$

We start relaxing $\overline{C}_N$ in Section 6.3 by providing more energy information to the receiver. First note that $\overline{C}_N$ can also be written as

$$\max_{e^{-r}} \max_{P_{U^N}} \frac{1}{N} I(U^N; Y^N E^N \,|\, B_1 = \overline{B}, e^{-r}).$$

For fixed $B_1 = \overline{B}$ and $E^{-r} = e^{-r}$,

$$\begin{aligned} I(U^N; Y^N E^N \,|\, B_1 E^{-r}) &\le I(U^N; Y^N E^{N+1} S^{N+1} \,|\, B_1 E^{-r}) \\ &= I(U^N; E_1 S_1 \,|\, B_1 E^{-r}) \end{aligned}$$

$$+ \sum_{n=1}^{N} I\big(U^N; Y_n E_{n+1} S_{n+1} \mid Y^{n-1} E_{-r+1}^n S^n B_1\big).$$

The first term is 0, because given $B_1 E^{-r}$, $E_1 S_1$ is independent of the input. Note that $X_n = U_n(B_1, E^n)$ and given $X_n Z_n$, $Y_n E_{n+1} S_{n+1}$ is independent of all previous random variables as well as U^N. Thus by decomposing the mutual information

$$\begin{aligned} H\big(Y_n E_{n+1} S_{n+1} \mid Y^{n-1} E_{-r+1}^n S^n B_1 U^N\big) &= H\big(Y_n E_{n+1} S_{n+1} \mid Y^{n-1} E_{-r+1}^n S^n B_1 U^N X_n\big) \\ &= H\big(Y_n E_{n+1} S_{n+1} \mid X_n Z_n\big) \\ &= H\big(Y_n E_{n+1} Z_{n+1} \mid X_n Z_n\big), \end{aligned}$$

$$H\big(Y_n E_{n+1} S_{n+1} \mid Y^{n-1} E_{-r+1}^n S^n B_1\big) \le H(Y_n E_{n+1} Z_{n+1} \mid Z_n),$$

$$I\big(U^N; Y_n E_{n+1} S_{n+1} \mid Y^{n-1} E_{-r+1}^n S^n B_1\big) \le I\big(Y_n E_{n+1} Z_{n+1}; X_n \mid Z_n\big),$$

$$I\big(U^N; Y^N E^N \mid B_1 E^{-r}\big) \le \sum_{n=1}^{N} I(Y_n E_{n+1} Z_{n+1}; X_n \mid Z_n).$$

For any P_{U^N} and $P_{B_1 E^{-r}}$, $\big(U^N, B_1, E_{-r+1}^{N+1}, S^{N+1}, Y^N\big)$ determines a random tuple $(X^N, E_2^{N+1}, Z^{N+1}, Y^N)$. Similar to the analysis in Section 6.4.1, for any P_{U^N}

$$I(U^N; Y^N E^N \mid B_1 = \overline{B}, e^{-r}) \le \tilde{c}_N\big(\pi_{\overline{B}, e^{-r}}\big),$$

where for any distribution π of Z_1 define

$$\tilde{c}_N(\pi) = \max_{\{p_n\}_{n=1}^N \subset \mathcal{P}^*} \sum_{n=1}^{N} I(Y_n E_{n+1} Z_{n+1}; X_n \mid Z_n).$$

Now we can establish the following theorems.

Theorem 6.4.4. *For each N,*

$$\tilde{C}_N \triangleq \frac{1}{N} \max_{e^{-r}} \tilde{c}_N\big(\pi_{\overline{B}, e^{-r}}\big) \ge \overline{C}_N$$

is an upper bound for the channel capacity of EH-SC2.

Theorem 6.4.5. *Define* $\tilde{c}_{N,z} = \tilde{c}_N(\delta_z)$ *for all* $z \in \mathcal{Z}$. *Then*

$$\tilde{c}_N(\pi) = \sum_{z \in \mathcal{Z}} \pi(z) \cdot \tilde{c}_{N,z},$$

$$\tilde{c}_{N,z} = \max_{p(\cdot|z) \in \mathcal{P}_z^*} \left[I(p,z) + \sum_x p(x|z) \sum_w Q(w|xz) \cdot \tilde{c}_{N-1,w} \right],$$

with the initial condition $\tilde{c}_{0,z} = 0$, $\forall z \in \mathcal{Z}$.

Note that since E_{n+1} is independent of X_n given Z_n,

$$I(p,z) = I(Y_n S_{n+1}; X_n \,|\, E_{n+1}, Z_n = z)$$

with $P_{X_n|E_{n+1}Z_n} = p_n$, which simplifies the computation. Also since

$$I(Y_n E_{n+1} Z_{n+1}; X_n \mid Z_n) \le H(X_n \,|\, Z_n),$$

we can replace the mutual information in the definitions of $\tilde{c}_{N,z}, \tilde{C}_N$ and $\tilde{c}_N(\pi)$ by the corresponding conditional entropies to define $\tilde{c}'_{N,z}, \tilde{C}'_N$ and $\tilde{c}'_N(\pi)$ and obtain:

Theorem 6.4.6. *Assume the base of* log *is e. We have*

$$\tilde{C}_N \le \tilde{C}'_N,$$

$$\tilde{c}'_N(\pi) = \sum_{z \in \mathcal{Z}} \pi(z) \cdot \tilde{c}'_{N,z},$$

$$\tilde{c}'_{N,z} = \log \sum_{x \in \mathcal{X}(z)} \exp \left[\sum_w Q(w|xz) \cdot \tilde{c}'_{N-1,w} \right],$$

with the initial condition $\tilde{c}'_{0,z} = 0$, $\forall z \in \mathcal{Z}$.

Remark 6.4.2. The proofs for Theorems 6.4.5 and 6.4.6 are similar to Theorems 6.4.2 and 6.4.3, respectively, and are omitted. Again when X_n is uniquely determined by Y_n (e.g., $Y_n = X_n$), $\tilde{C}'_N = \tilde{C}_N$.

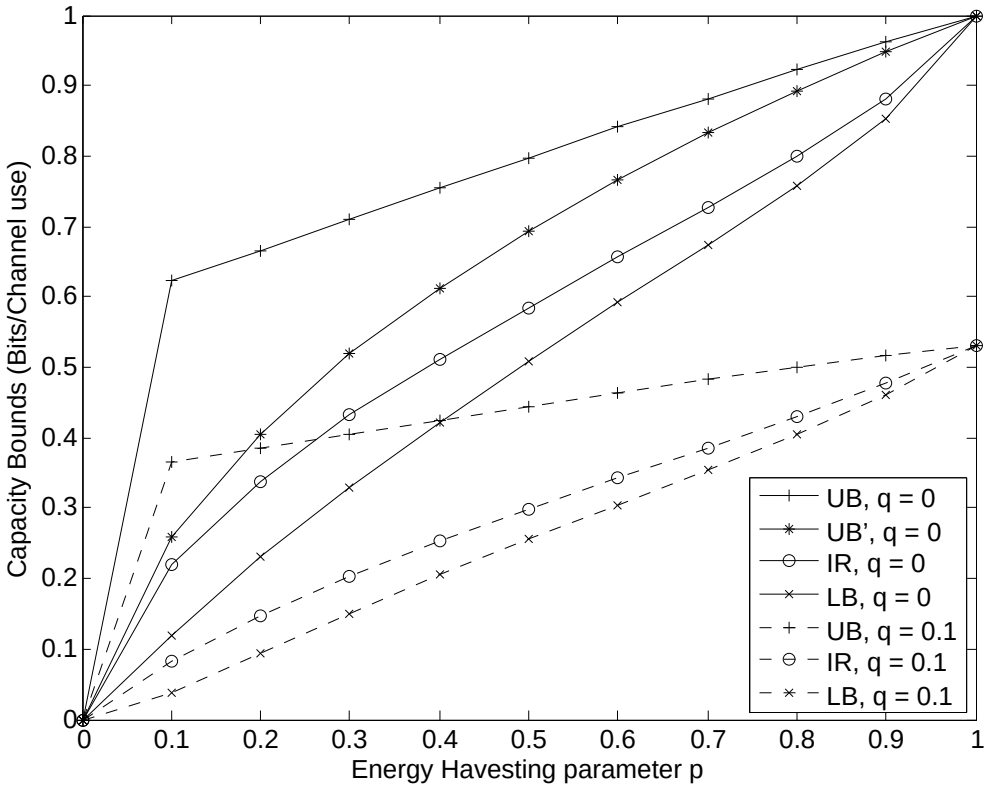


Figure 6.1: Capacity bounds for an energy harvesting channel

6.5 Numerical Results

We use an energy harvesting example to demonstrate the computation of our bounds. In this model E_n is *i.i.d.* Bernoulli(p), $\overline{B} = 1$, $\mathcal{X} = \mathcal{Y} = \{0, 1\}$. The DMC is binary symmetric with crossover probability q and we require E_n to be stored in the battery first, i.e., we use energy models (4.3) and (4.6).

Figure 6.1 shows different bounds and achievable rates from Chapter 5, [41] and Theorems 6.3.1 and 6.3.2, where UB and UB' denote upper bounds from Theorem 6.3.2 ($N = 4$) and [41] respectively, IR denotes the achievable rate for optimal *i.i.d.* input from Chapter 5 (m=1), and LB denotes the lower bound from Theorem 6.3.1 ($N = 4$). Note that UB and IR work for both scenarios, whereas LB only works for EH-SC2. For $q = 0$ the upper bound in [41] only works for EH-SC1, whereas for $q > 0$ the result is not applicable.

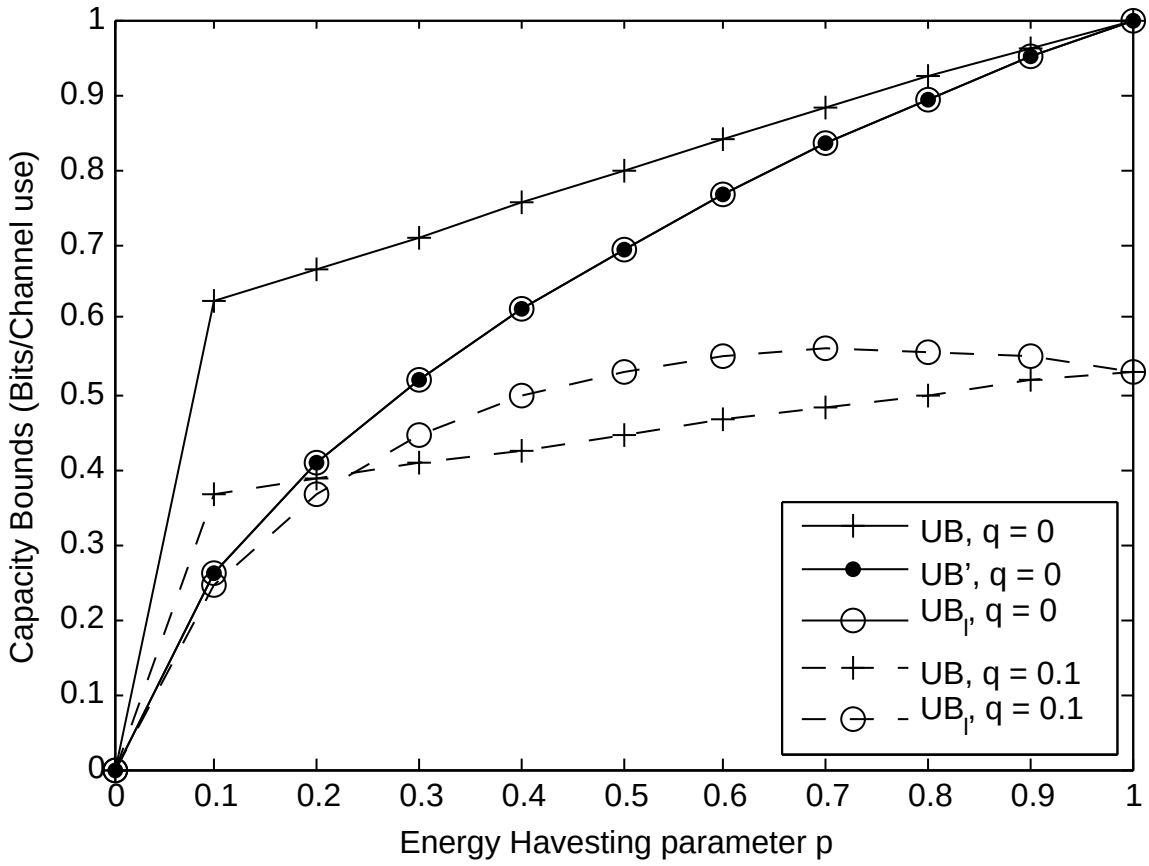


Figure 6.2: Capacity upper bounds comparison

Figure 6.2 shows the bounds $\overline{C}_N$ and $\tilde{C}_N$ for EH-SC2, as well as the upper bound in [41] for EH-SC1. Here UB denotes $\overline{C}_4$, UB$'$ denotes the upper bound from [41], and UB$_\text{l}$ denotes $\tilde{C}_{10^4}$. Note that $\tilde{C}'_N$ for all q are equal to the $\tilde{C}_N$ for $q = 0$. For $\overline{C}_N$ we are only able to compute its value up to $N = 4$, whereas for $\tilde{C}_N$ we can compute it up to $N = 10^4$ easily and the values appear to have converged to their limit. Although $\tilde{C}_N$ is looser than $\overline{C}_N$ for any fixed N, since we can make N much larger for the former, the bound can be (much) tighter, especially when q or p is small. For $q = 0$, $\tilde{C}_{10^4}$ for EH-SC2 coincides numerically with both the $\tilde{C}_{10^4}$ for EH-SC1 and the upper bound in [41], suggesting that the best full CSIR upper bounds for both scenarios are the same for a perfect channel.

Chapter 7

Pairwise Error Probability

In this chapter we study the pairwise error probabilities for Maximum-Likelyhood (ML) decoding of the energy harvesting channel, with the setting of the more general scenario 2. Suppose there are two codewords, u^N and v^N. Assume u^N is sent over the channel and Y^N is received. The decoder makes a mistake if

$$p\big(Y^N \,|\, v^N\big) > p\big(Y^N \,|\, u^N\big),$$

or makes a mistake with probability $1/2$ if

$$p\big(Y^N \,|\, v^N\big) = p\big(Y^N \,|\, u^N\big).$$

Denote this error probability by $P_e\big(v^N \,|\, u^N\big)$, then

$$\begin{aligned} P_e\big(v^N \,|\, u^N\big) \;&=\; \Pr\big(\, p\big(Y^N \,|\, v^N\big) > p\big(Y^N \,|\, u^N\big) \mid u^N \,\big) \\ &\quad + \frac{1}{2}\Pr\big(\, p\big(Y^N \,|\, v^N\big) = p\big(Y^N \,|\, u^N\big) \mid u^N \,\big), \end{aligned}$$

where $p\big(Y^N|\,\cdot\,\big)$ is defined by (6.6) and (4.14). Define the pairwise error probability between u^N and v^N as

$$P_e\big(u^N, v^N\big) = \big[P_e\big(v^N \,|\, u^N\big) + P_e\big(u^N \,|\, v^N\big)\big]\,/2,$$

which we want to minimize when designing codes. The formula above with a general DMC is complicated, but for the two cases below we have some simplification. Since the two terms in $P_e\big(v^N \,|\, u^N\big)$ are similar, for simplicity we only analyze the first one, which is denoted by $P'_e\big(v^N \,|\, u^N\big)$.

7.1 Noiseless Channel

Assume the DMC in Figure 4.1 is noiseless, i.e., $Y_n = X_n$. Then by (6.6) and (4.14)

$$p\big(y^N \,|\, v^N\big) = \sum_{b_1, e^N} p(b_1) p\big(e^N\big) \cdot \mathbf{1}_{\{y^N = v^N(b_1, e^N)\}}.$$

Define $g_{u^N, v^N}\big(\overline{b}_1, \overline{e}^N\big)$ as the probability of v^N yielding $x^N = u^N\big(\overline{b}_1, \overline{e}^N\big)$, i.e.,

$$g_{u^N, v^N}\big(\overline{b}_1, \overline{e}^N\big) \triangleq \sum_{b_1, e^N} p(b_1) p\big(e^N\big) \cdot \mathbf{1}_{\left\{u^N\left(\overline{b}_1, \overline{e}^N\right) = v^N(b_1, e^N)\right\}}.$$

When u^N is sent, $Y^N = u^N\big(B_1, E^N\big)$ and so

$$\begin{aligned} P'_e\big(v^N \,|\, u^N\big) &= \Pr\left(g_{u^N, v^N}\big(B_1, E^N\big) > g_{u^N, u^N}\big(B_1, E^N\big)\right) \\ &= \sum_{\overline{b}_1, \overline{e}^N} p\big(\overline{b}_1\big) p\big(\overline{e}^N\big) \cdot \mathbf{1}_{\left\{g_{u^N, v^N}\left(\overline{b}_1, \overline{e}^N\right) > g_{u^N, u^N}\left(\overline{b}_1, \overline{e}^N\right)\right\}}. \end{aligned}$$

This gives us an easier formula to compute the pairwise error probability. When B_1, E^N have uniform distributions, the expression simplifies further. Let us define $K_{u^N, v^N}\big(\overline{b}_1, \overline{e}^N\big)$ to be the cardinality of

$$\left\{\big(b_1, e^N\big) \mid v^N\big(b_1, e^N\big) = u^N\big(\overline{b}_1, \overline{e}^N\big)\right\},$$

which is the number of pairs $\big(b_1, e^N\big)$ that under v^N yields $x^N = u^N\big(\overline{b}_1, \overline{e}^N\big)$. Then

$$g_{u^N, v^N}\big(\overline{b}_1, \overline{e}^N\big) = \frac{K_{u^N, v^N}\left(\overline{b}_1, \overline{e}^N\right)}{|\mathcal{E}_B| \cdot |\mathcal{E}_H|^N},$$

$$P'_e\big(v^N \,|\, u^N\big) = \frac{1}{|\mathcal{E}_B||\mathcal{E}_H|^N} \cdot \Big|\big\{\big(\overline{b}_1, \overline{e}^N\big) : K_{u^N,v^N}\big(\overline{b}_1, \overline{e}^N\big) > K_{u^N,u^N}\big(\overline{b}_1, \overline{e}^N\big)\big\}\Big|,$$

which is proportional to the number of pairs $\big(\overline{b}_1, \overline{e}^N\big)$ having the following property: the number of pairs $\big(b_1, e^N\big)$ yielding the same vector $x^N = u^N\big(\overline{b}_1, \overline{e}^N\big)$ under the function v^N is more than that of u^N. Suppose each function u^N is represented by its matrix of values, whose rows and columns are indexed by the time n and different pairs $\big(b_1, e^N\big)$, respectively—that is, the $\big(n, \big(b_1, e^N\big)\big)$-th entry denote $x_n = u_n\big(b_1, e^N\big)$. Then $K_{u^N,v^N}\big(\overline{b}_1, \overline{e}^N\big)$ is the number of columns in v^N that is identical to the $\big(\overline{b}_1, \overline{e}^N\big)$-th column in u^N.

Example 7.1.1. Assume the setting of Example 4.3.1 with $B_1 \sim$ Bernoulli(1/2), $p = 1/2$, $q = 0$. Then the conditions above are all satisfied. Let $N = 1$, $u_1 = (0, 1, 0, 1)$ and $v_1 = (0, 1, 1, 1)$, with the four entries denoting the output x_1 when

$$(b_1, e_1) = (0, 0), (0, 1), (1, 0), (1, 1),$$

respectively. Let $\big(\overline{b}_1, \overline{e}_1\big) = (1, 1)$, with $u_1\big(\overline{b}_1, \overline{e}_1\big) = 1$. Then $K_{u_1,v_1}\big(\overline{b}_1, \overline{e}_1\big) = 3$, which is larger than $K_{u_1,u_1}\big(\overline{b}_1, \overline{e}_1\big) = 2$. The number of such pairs $\big(\overline{b}_1, \overline{e}_1\big)$ is 2, so $P'_e\big(v_1 \,|\, u_1\big) = 1/2$. Similarly $P'_e\big(u_1 \,|\, v_1\big) = 1/4$. Indeed $P_e(u^N, v^N) = 3/8$. Presumably as N grows we can find u^N and v^N such that the pairwise error probability is small.

7.2 Binary Symmetric Channel

Assume the DMC in Figure 4.1 is BSC(q), i.e., $\mathcal{X} = \mathcal{Y} = \{0, 1\}$,

$$Y_n = X_n \oplus Z_n$$

with $Z_n \sim$ i.i.d Bernoulli(q). Then

$$p\left(y^N \,|\, x^N\right) = q^{w_H\left(y^N \oplus x^N\right)} \cdot (1 - q)^{N - w_H\left(y^N \oplus x^N\right)},$$

where w_H denotes the Hamming weight. Denote this probability by $f(y^N \,|\, x^N)$ and define

$$\begin{aligned} g_{u^N,v^N}(\bar{b}_1, \bar{e}^N, z^N) &= p\left(y^N = u^N\left(\bar{b}_1, \bar{e}^N\right) \oplus z^N \mid v^N\right) \\ &= \sum_{b_1, e^N} p(b_1) p\left(e^N\right) f\left(u^N\left(\bar{b}_1, \bar{e}^N\right) \oplus z^N \mid v^N\left(b_1, e^N\right)\right). \end{aligned}$$

When u^N is sent,

$$Y^N = u^N(B_1, E^N) \oplus Z^N$$

and so

$$\begin{aligned} P_e'\left(v^N \,|\, u^N\right) &= \Pr\left(g_{u^N,v^N}\left(B_1, E^N, Z^N\right) > g_{u^N,u^N}\left(B_1, E^N, Z^N\right)\right) \\ &= \sum_{\bar{b}_1, \bar{e}^N, z^N} p\left(\bar{b}_1\right) p\left(\bar{e}^N\right) p\left(z^N\right) \cdot \mathbf{1}_{\left\{g_{u^N,v^N}\left(\bar{b}_1, \bar{e}^N, z^N\right) > g_{u^N,u^N}\left(\bar{b}_1, \bar{e}^N, z^N\right)\right\}}, \end{aligned}$$

which also gives us an easier formula.

Appendix A

Appendices for Part I

A.1 Proofs and Calculations in Section 2.6

A.1.1 Structures of M, K, K', J, J'

When the characteristic p of $\mathbb{F}_q$ equals 2, $K = K'$ and $J = J'$. So for the analysis of K' and J' we only consider the case $p \neq 2$.

Observe that $|A_\alpha| = p$ for each $\alpha \in \mathbb{F}_q^\times$, and

$$|C| = 3, \quad |B_1| = 2, \quad |B| = |B'| = |P| = |P'| = q - 1.$$

As $(CB_1)^2 = I$, we have $M \cong D_6 \cong S_3$. It is easy to check that $\forall \alpha \in \mathbb{F}_q$,

$$A_\alpha^B = A_{t^{-1}\alpha}, \quad A_\alpha^{B'} = A_{-t^{-1}\alpha}, \quad A_\alpha^P = A_{t\alpha}, \quad A_\alpha^{P'} = A_{-t\alpha}.$$

Therefore, N is a normal subgroup of all K, K', J, J' and

$$K = N \cdot \langle B \rangle, \quad K' = N \cdot \langle B' \rangle, \quad J = N \cdot \langle P \rangle, \quad J' = N \cdot \langle P' \rangle.$$

Also N trivially intersects each of $\langle B \rangle, \langle B' \rangle, \langle P \rangle$, and $\langle P' \rangle$, thus

$$K \cong N \rtimes \langle B \rangle, \quad K' \cong N \rtimes \langle B' \rangle, \quad J \cong N \rtimes \langle P \rangle, \quad J' \cong N \rtimes \langle P' \rangle,$$

all of which are semidirect products $\mathbb{Z}_p^m \rtimes \mathbb{Z}_{q-1}$. We claim that $K \cong J$ and $K' \cong J'$.

Moreover, in the case $p \neq 2$, all the four groups are isomorphic if and only if $\frac{q-1}{2}$ is even.

To see this, first consider the bijections $\sigma : K \to J$ and $\sigma' : K' \to J'$, where $\forall \alpha \in \mathbb{F}_q$, $\forall k \in \mathcal{K}_q$,

$$\sigma\left(A_\alpha B^k\right) = A_\alpha P^{-k}, \quad \sigma'\left(A_\alpha (B')^k\right) = A_\alpha (P')^{-k}.$$

Observe that $\forall \alpha, \beta \in \mathbb{F}_q$, $\forall k, l \in \mathcal{K}_q$,

$$\begin{aligned}\sigma\left(A_\alpha B^k \cdot A_\beta B^l\right) &= \sigma\left(A_{\alpha + t^k \beta} B^{k+l}\right) = A_{\alpha + t^k \beta} P^{-k-l} \\ &= A_\alpha P^{-k} \cdot A_\beta P^{-l} = \sigma\left(A_\alpha B^k\right) \cdot \sigma\left(A_\beta B^l\right),\end{aligned}$$

so σ is indeed an isomorphism. Similarly σ' is also an isomorphism.

Next observe that in the case $p \neq 2$, when $\frac{q-1}{2}$ is even, $\frac{q-1}{4}$ is an integer and so

$$\left(\frac{q+1}{2}\right)^2 = \left(\frac{q-1}{2} + 1\right)^2 = \frac{(q-1)^2}{4} + (q-1) + 1 \equiv 1 \pmod{q-1},$$

$$\left((B')^{\frac{q+1}{2}}\right)^{\frac{q+1}{2}} = B', \qquad \left\langle (B')^{\frac{q+1}{2}} \right\rangle = \langle B' \rangle.$$

In addition, since $\mathbb{F}_q^\times$ is cyclic of an even order $q-1$, we have $-1 = t^{\frac{q-1}{2}}$, and thus

$$(-t)^{\frac{q+1}{2}} = \left(t^{\frac{q+1}{2}}\right)^{\frac{q+1}{2}} = t.$$

Consider $\tau : K \to K'$, where

$$\tau\left(A_\alpha B^k\right) = A_\alpha (B')^{\frac{q+1}{2}k}, \quad \forall \alpha \in \mathbb{F}_q, \ \forall k \in \mathcal{K}_q.$$

Apparently τ is a bijection. Also we can show that it is a homomorphism by calculating $\tau\left(A_\alpha B^k \cdot A_\beta B^l\right)$ with the following fact:

$$A_\alpha (B')^{\frac{q+1}{2}k} \cdot A_\beta (B')^{\frac{q+1}{2}l} = A_{\alpha + (-t)^{\frac{q+1}{2}k} \beta} (B')^{\frac{q+1}{2}(k+l)} = A_{\alpha + t^k \beta} (B')^{\frac{q+1}{2}(k+l)}.$$

Thus when $\frac{q-1}{2}$ is even, $K \cong K'$ and the four groups are all isomorphic.

When $\frac{q-1}{2}$ is odd, however, τ is not a bijection anymore, because this time $B' \notin \left\langle (B')^{\frac{q+1}{2}} \right\rangle$ and $\tau(K) \neq K'$. Furthermore, we can prove that in this case K and K' are not isomorphic, by showing that K and J have generalized flower structures whenever $q > 2$, whereas if $p \neq 2$, K' and J' only have flower structures when $\frac{q-1}{2}$ is even. Since $K \cong J$ and $K' \cong J'$, it is enough to only show the analysis of K and K'. Pick $\alpha \in \mathbb{F}_q^\times$ and assume $k, l \in \mathcal{K}_q$. Similar to the G_2 in Section 2.5.2, we have the relation

$$(B^k)^{A_\alpha} = B^l \iff k = l = 0,$$

and thus K has a generalized flower structure whenever $q > 2$. On the other hand, for K' we have

$$(B'^k)^{A_\alpha} = B'^l \iff \begin{bmatrix} (-1)^k & 0 \\ t^k\alpha & t^k \end{bmatrix} = \begin{bmatrix} (-1)^l & 0 \\ (-1)^l\alpha & t^l \end{bmatrix},$$

which requires $k = l$ and $t^l = (-1)^l$. Thus for $p \neq 2$, l can only be 0 or $\frac{q-1}{2}$. If $\frac{q-1}{2}$ is even, we have $(-1)^{\frac{q-1}{2}} = 1$ and so $k = l = 0$, then K' also has a generalized flower structure (as expected since here $K \cong K'$). If $\frac{q-1}{2}$ is odd, however, this is not true: in this case $(-1)^{\frac{q-1}{2}} = -1$, so $k = l = 0$ or $\frac{q-1}{2}$ in the above relation. Thus $\forall \alpha \in \mathbb{F}_q^\times$,

$$\langle B' \rangle \bigcap \langle B' \rangle^{A_\alpha} = \langle -I \rangle \cong \mathbb{Z}_2.$$

When $q = 3$, $B' = -I$ and

$$K' = \langle A \rangle \times \langle -I \rangle \cong \mathbb{Z}_3 \times \mathbb{Z}_2 \cong \mathbb{Z}_6,$$

when $q > 3$, $\langle B' \rangle$ and $\langle B' \rangle^{A_\alpha}$ are distinct groups but have nontrivial intersection. Therefore, in neither case does K' have a generalized flower structure.

A.1.2 Intersections in Instances 8 and 9

Let $p \neq 2$. Observe that K' and J' are both subgroups of the G_2 in Instance 1, so all the intersections in both instances are subgroups of their respective counterparts in Instance 1. In Instance 8, since $G_{12} \leq \langle tI, B_1 \rangle$ and the (1,1)-entry for every matrix in $G_2 = K'$ is always ± 1, we have $G_{12} \leq \langle -I, B_1 \rangle$. This further limits the (2,2)-entry to be ± 1 for each matrix in G_{12}. As the (2,2)-entry in K' takes the form t^k for some k, this k can only be 0 or $\frac{q-1}{2}$. By examining the parity of $\frac{q-1}{2}$, we have

$$G_{12} = \begin{cases} \langle B_1 \rangle \cong \mathbb{Z}_2 & \text{if } \frac{q-1}{2} \text{ is even} \\ \langle -I \rangle \cong \mathbb{Z}_2 & \text{otherwise} \end{cases},$$

$$G_{123} = G_{124} = \begin{cases} 1 & \text{if } \frac{q-1}{2} \text{ is even} \\ \langle -I \rangle \cong \mathbb{Z}_2 & \text{otherwise} \end{cases}.$$

Similarly we can calculate G_{12}, G_{123}, and G_{124} for Instance 9.

In both instances, G_{24} is simply the subgroup of all diagonal matrices in G_2, and $G_{23} \leq T$. As matrices in K' and J' can be respectively written as

$$(-1)^k \begin{bmatrix} 1 & 0 \\ \alpha' & (-t)^k \end{bmatrix} = (-1)^k \begin{bmatrix} 1 & 0 \\ \alpha' & (t^{\frac{q+1}{2}})^k \end{bmatrix},$$

$$t^k \begin{bmatrix} 1 & 0 \\ \alpha'' & (-t^{-1})^k \end{bmatrix} = t^k \begin{bmatrix} 1 & 0 \\ \alpha'' & (t^{\frac{q-3}{2}})^k \end{bmatrix}$$

for some $\alpha', \alpha'' \in \mathbb{F}_q$ and $k \in \mathcal{K}_q$, we see that $G_{23} = \langle -B_3^{\frac{q+1}{2}} \rangle$ and $\langle tB_3^{\frac{q-3}{2}} \rangle$, respectively, where

$$(-B_3^{\frac{q+1}{2}})^k = \begin{bmatrix} (-1)^k & 0 \\ t^k - (-1)^k & t^k \end{bmatrix}, \qquad (tB_3^{\frac{q-3}{2}})^k = \begin{bmatrix} t^k & 0 \\ (-1)^k - t^k & (-1)^k \end{bmatrix}.$$

Thus $G_{23} \cong \mathbb{Z}_{q-1}$ in both cases.

A.1.3 The case $p = 3$ for Instance 15

In Instance 15, $G_1 = M = \langle C, B_1 \rangle$ and $G_2 = (J')^E$. We can show that $G_1G_2 = G_2G_1$ when $p = 3$, thus Condition 2.3.3 is satisfied. Observe that

$$G_2 = \{X_{\alpha,j} \,|\, \alpha \in \mathbb{F}_q, j \in \mathcal{K}_q\},$$

$$X_{\alpha,j} \triangleq \begin{bmatrix} (-1)^j - \alpha & \alpha \\ (-1)^j - t^j - \alpha & t^j + \alpha \end{bmatrix}.$$

When $p = 3$, we have $2 = -1$, and thus $C = X_{1,0} \in G_2$. It is easy to check that for each α and j

$$X_{\alpha,j}^{B_1} = \begin{bmatrix} (-1)^j + \alpha & -\alpha \\ (-1)^j - t^j + \alpha & t^j - \alpha \end{bmatrix} = X_{-\alpha,j} \in G_2.$$

Thus G_1 normalizes G_2. In particular, $\forall X \in G_2$ and $\forall Y \in G_1$, we have $X^Y \in G_2$ and $X^{Y^{-1}} \in G_2$, which imply $XY \in G_1G_2$ and $YX \in G_2G_1$, respectively. Therefore $G_1G_2 = G_2G_1$.

A.1.4 Intersections in Instances 12–15

Most intersections are easily obtained by comparing the formulae of the matrices in the subgroups involved. For the intersection of M with any of $J^E, (J')^E, J^Q$, or $(J')^Q$, we can utilize the properties below to facilitate calculation. Let $\vec{c}_i(X)$ denote the i-th column of a matrix X, and we have

$$\vec{c}_1(X) + \vec{c}_2(X) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \ \forall X \in J^E;$$

$$\vec{c}_1(X) + \vec{c}_2(X) = \pm \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \ \forall X \in (J')^E;$$

$$\vec{c}_1(X) - 2\vec{c}_2(X) = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \ \forall X \in J^Q;$$

$$\vec{c}_1(X) - 2\vec{c}_2(X) = \pm \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \ \forall X \in (J')^Q.$$

Thus, we need only seek elements of M which share these properties.

We also want to mention the calculation of G_{34} for Instances 13 and 15 when $p > 3$. In Instance 13, finding G_{34} is equivalent to solving the following set of equations:

$$\begin{cases} (-1)^j - \alpha = (-1)^i + 2\beta \\ \alpha = \beta \\ (-1)^j - t^j - \alpha = 2\left(t^i - 2\beta - (-1)^i\right) \\ t^j + \alpha = t^i - 2\beta \end{cases} \iff \begin{cases} \alpha = \beta \\ 3\beta = (-1)^j - (-1)^i \\ t^i = (-1)^j \\ t^j = (-1)^i \end{cases}.$$

From the last two equations, we can see that i and j can only be 0 or $\frac{q-1}{2}$. If $\frac{q-1}{2}$ is even, then $(-1)^{\frac{q-1}{2}} = 1$, so i and j must both be 0, which yields that $G_{34} = 1$. If $\frac{q-1}{2}$ is odd, then $i = 0$ implies that $j = 0$, and $i = \frac{q-1}{2}$ implies that $j = \frac{q-1}{2}$. In both cases $\alpha = \beta = 0$, therefore $G_{34} = \langle -I \rangle$. For G_{34} in Instance 15, we have similar equations and the same discussion also applies.

A.2 Group Network Codes: Details

A.2.1 Code Construction

To establish the encoding and decoding process, we need an auxiliary lemma.

Lemma A.2.1. *Let K_1, K_2 be two subgroups of G with $K_1 \leq K_2$. Then the coset mapping*

$$\begin{aligned} \pi : G/K_1 &\to G/K_2 \\ xK_1 &\mapsto xK_2 \end{aligned} \tag{A.1}$$

is a well defined onto function, where xK_1 is mapped to the unique coset in G/K_2 that contains it. Furthermore, if Λ_1 is a uniform random variable on G/K_1, then $\pi(\Lambda_1)$ is uniform on G/K_2.

Proof. π is well defined since $xK_2 = x'K_2$ whenever $xK_1 = x'K_1$. Note that K_2 is

partitioned by the m distinct cosets $\{\, y_i K_1 : 1 \le i \le m \,\}$, where $m = |K_2/K_1|$ and $y_i \in K_2$ for $i = 1, 2, \ldots, m$. Therefore, each $xK_2 \in G/K_2$ is also partitioned by the m cosets $\{\, (xy_i)K_1 : 1 \le i \le m \,\}$, which are precisely the m preimages of xK_2 under π. Thus $\pi(\Lambda_1)$ is uniform on G/K_2. □

For any collection α of subgroups of G, the intersection mapping (2.1) is a bijection. Consider the collection of all source subgroups. Let

$$\mathcal{X}_{\mathcal{S}} = \{\, (xG_s : s \in \mathcal{S}) \mid x \in G \,\} \subseteq \prod_{s \in \mathcal{S}} \mathcal{Y}_s,$$

then we have the bijective intersection mapping $\Theta_{\mathcal{S}} : \mathcal{X}_{\mathcal{S}} \to G/G_{\mathcal{S}}$. But with (R1), $\left|\prod_{s\in\mathcal{S}} \mathcal{Y}_s\right| = |G/G_{\mathcal{S}}| = |\mathcal{X}_{\mathcal{S}}|$ and so

$$\mathcal{X}_{\mathcal{S}} = \prod_{s \in \mathcal{S}} \mathcal{Y}_s.$$

This means that any coset tuple $(x_s G_s : s \in \mathcal{S})$ in $\prod_{s\in\mathcal{S}} \mathcal{Y}_s$ can be represented in the form $(xG_s : s \in \mathcal{S})$ for a common $x \in G$, and the intersection of $\{\, x_s G_s : s \in \mathcal{S} \,\}$ is equal to $xG_{\mathcal{S}}$. Therefore, we can rewrite the bijection $\Theta_{\mathcal{S}}$ as

$$\Theta_{\mathcal{S}} : \prod_{s \in \mathcal{S}} \mathcal{Y}_s \to G/G_{\mathcal{S}},$$

which maps a tuple to the intersection of all its cosets.

Moreover, let t be an edge or a sink node, and define

$$\mathcal{X}_{\mathcal{I}(t)} = \{\, (xG_f : f \in \mathcal{I}(t)) \mid x \in G \,\}, \qquad G_{\mathcal{I}(t)} = \bigcap_{f \in \mathcal{I}(t)} G_f.$$

Then the intersection mapping

$$\Theta_{\mathcal{I}(t)} : \mathcal{X}_{\mathcal{I}(t)} \to G/G_{\mathcal{I}(t)}$$

is a bijection. With (R2) and (R3), we can also define coset mappings for edges and

source/sink pairs as follows. For each edge e, since $G_{\mathcal{I}(e)} \leq G_e$ by (R2), define the coset mapping π_e as (A.1) with $K_1 = G_{\mathcal{I}(e)}$ and $K_2 = G_e$. Similarly for each source s with $u \in \mathcal{D}(s)$, since $G_{\mathcal{I}(u)} \leq G_s$ by (R3), define $\pi_{u,s}$ with $K_1 = G_{\mathcal{I}(u)}$ and $K_2 = G_s$.

Now we can define the encoding and decoding functions. At each edge e, let the encoding function be $\phi_e = \pi_e \circ \Theta_{\mathcal{I}(e)}$. For each source s with $u \in \mathcal{D}(s)$, let the decoding function be $\phi_{u,s} = \pi_{u,s} \circ \Theta_{\mathcal{I}(u)}$. In other words, at an edge or a sink node t, the encoding/decoding function takes an input coset tuple $(Y_f : f \in \mathcal{I}(t))$ and first forms the intersection of them, which is a coset of $G_{\mathcal{I}(t)}$, then maps this coset to the unique coset of G_e (or G_s, whichever is appropriate) that contains it. Such network operations define a proper network code, since by the proposition below the decoding functions always yield correct source symbols at each sink node.

Proposition A.2.1. *Assume (R1) holds, and let the encoding and decoding functions be defined as above. Then for some common $x \in G$, $\forall s \in \mathcal{S}$, $Y_s = xG_s$ and $\forall e \in \mathcal{E}$, $Y_e = xG_e$. Also for each source s with $u \in \mathcal{D}(s)$, Y_s is recovered by the decoding function $\phi_{u,s}$.*

Proof. Let the source symbols $(Y_s : s \in \mathcal{S})$ be an arbitrary tuple from $\prod_{s\in\mathcal{S}} \mathcal{Y}_s$. Since (R1) is true, as discussed above, for all $s \in \mathcal{S}$, $Y_s = xG_s$ with a common $x \in G$. As $\mathcal{G}$ is directed and acyclic, we can define the "depth" of each node v as the length of the longest path from a source node to v, and define the depth of an edge to be the depth of its tail node. Note that e is always "deeper" than f if $f \in \mathcal{I}(e)$. Also if $Y_f = xG_f$ for all $f \in \mathcal{I}(e)$, then

$$Y_e = \phi_e(Y_f : f \in \mathcal{I}(e)) = xG_e.$$

So by induction on the depths of the edges, $Y_e = xG_e$ for all $e \in \mathcal{E}$.

Furthermore, for each $s \in \mathcal{S}$ with $u \in \mathcal{D}(s)$, since $Y_f = xG_f$ for all $f \in \mathcal{I}(u)$,

$$\phi_{u,s}(Y_f : f \in \mathcal{I}(u)) = xG_s = Y_s.$$

Thus the source symbol Y_s is successfully recovered at u. □

Remark A.2.1. Note that the encoding/decoding function for an edge or a sink node t is only defined on $\mathcal{X}_{\mathcal{I}(t)}$, but not on the entire Cartesian product $\prod_{f\in\mathcal{I}(t)}\mathcal{Y}_f$. This is because for an arbitrary tuple in $\prod_{f\in\mathcal{I}(t)}\mathcal{Y}_f$, it is possible that the intersection of all cosets is the empty set, which is not a coset of $G_{\mathcal{I}(t)}$. However, with (R1) this is not a problem, as Proposition A.2.1 guarantees that $(Y_f : f \in \mathcal{I}(t))$ is always a tuple in $\mathcal{X}_{\mathcal{I}(t)}$.

Remark A.2.2. From the proof above, even without (R1) these encoding and decoding functions still constitute a valid network code, if the sources cooperate in such a way that the transmit tuples are always from $\mathcal{X}_{\mathcal{S}}$. But in this case the source random variables are dependent.

A.2.2 The Entropy Vector

Here we analyze the global mappings of this group network code, and show that the entropy vector is characterizable by the group G and its subgroups $\{G_t : t \in \mathcal{S}\cup\mathcal{E}\}$ when the sources are independent and uniform. First we give another auxiliary lemma.

Lemma A.2.2. *Let $K \leq G$ and let G_i, $i = 1, \ldots, n$, be subgroups of G containing K. For each i let π_i be the coset mapping defined as* (A.1) *with $K_1 = K$ and $K_2 = G_i$. Let Λ_K be a uniform random variable on G/K, and define $X_i = \pi_i(\Lambda_K)$ for each i. Then the entropy vector of $\{X_1, X_2, \ldots, X_n\}$ is exactly the group characterizable vector induced by G and $\{G_1, G_2, \ldots, G_n\}$.*

Proof. For each nonempty subset $\alpha \subseteq \mathcal{N}$, since $K \leq G_\alpha$, we can define the coset mapping π_α with K and G_α. As in Section 2.1.1, the alphabet of X_α is still

$$\mathcal{X}_\alpha = \{\, (xG_i : i \in \alpha) \mid x \in G \,\},$$

and the intersection mapping Θ_α is a bijection. Also $\Theta_\alpha(X_\alpha) = \pi_\alpha(\Lambda_K)$, which is

uniform on G/G_α by Lemma A.2.1. So the joint entropy

$$H(X_\alpha) = H(\Theta_\alpha(X_\alpha)) = \log \frac{|G|}{|G_\alpha|}$$

and the lemma follows. □

For each $s \in \mathcal{S}$ define the coset mapping π'_s as (A.1) with $K_1 = G_\mathcal{S}$ and $K_2 = G_s$. For every edge e we can similarly define a new coset mapping π'_e with $K_1 = G_\mathcal{S}$ and $K_2 = G_e$, since according to the following proposition, $G_\mathcal{S} \leq G_e$.

Proposition A.2.2. *If (R2) is satisfied, then* $\forall e \in \mathcal{E}$, $G_\mathcal{S} \leq G_e$.

Proof. The proposition is trivially true if e is emitted from a source node. Also if $G_\mathcal{S} \leq G_f$ for all $f \in \mathcal{I}(e)$, then by (R2) we have $G_\mathcal{S} \leq G_e$. Similar to Proposition A.2.1, by induction on the depths of the edges the proof follows. □

Proposition A.2.3. $\forall e \in \mathcal{E}$, *the global mapping at* e *for the above group network code is* $\varphi_e = \pi'_e \circ \Theta_\mathcal{S}$. *In other words,* φ_e *first forms the intersection of all the source cosets to obtain a coset of* $G_\mathcal{S}$, *and then maps this coset to the unique coset of* G_e *containing it.*

Proof. Assume the source symbols $(Y_s : s \in \mathcal{S})$ are transmitted and let

$$\Lambda_\mathcal{S} = \Theta_\mathcal{S}(Y_s : s \in \mathcal{S}).$$

Then $\Lambda_\mathcal{S} = xG_\mathcal{S}$ for some $x \in G$, and $Y_s = xG_s = \pi'_s(\Lambda_\mathcal{S})$ for all $s \in \mathcal{S}$. By Proposition A.2.1, $Y_e = xG_e = \pi'_e(\Lambda_\mathcal{S})$, so $\varphi_e = \pi'_e \circ \Theta_\mathcal{S}$. □

Let the source random variables $\{Y_s : s \in \mathcal{S}\}$ be independent and uniformly distributed, so the joint distribution is uniform on $\prod_{s\in\mathcal{S}} \mathcal{Y}_s$. Let $\Lambda_\mathcal{S} = \Theta_\mathcal{S}(Y_s : s \in \mathcal{S})$, then $\Lambda_\mathcal{S}$ is uniform on $G/G_\mathcal{S}$ as $\Theta_\mathcal{S}$ is bijective. From Proposition A.2.3, $\forall t \in \mathcal{S} \cup \mathcal{E}$, $Y_t = \pi'_t(\Lambda_\mathcal{S})$, and so by Lemma A.2.2, the entropy vector for $\{Y_t : t \in \mathcal{S} \cup \mathcal{E}\}$ is characterizable by the group G and its subgroups $\{G_t : t \in \mathcal{S} \cup \mathcal{E}\}$.

A.2.3 Inclusion of Linear Network Codes

In this section we carry over the group theory notations in Section 2.2 to vector spaces, but with additive notation. For example, the left coset is now written as $v + W$ for a vector v and a subspace W. Further, we use $\oplus$ to denote the direct sum of vector spaces. In the following we show that for each linear network code, there exists an equivalent group network code, with essentially the same network operations and hence the same encoding/decoding results.

Consider a linear network code $\mathcal{C}$ over a finite field F. For each $t \in \mathcal{S} \cup \mathcal{E}$, the alphabet $\mathcal{Y}_t$ is a finite dimensional vector space over F. Let v denote the concatenation of all the source vectors $(Y_s : s \in \mathcal{S})$, then v is a vector in

$$V \triangleq \oplus_{s \in \mathcal{S}} U_s,$$

where $U_s \triangleq \mathcal{Y}_s$. Then for each edge e, the global mapping φ_e is a linear transformation from V to $\mathcal{Y}_e$, whose range is denoted by U_e. Also for each source s, let

$$\varphi_s : V \to U_s$$

be the linear projection that maps $v \in V$ to its s-th section. Thus $\forall t \in \mathcal{S} \cup \mathcal{E}$, we can write $Y_t = \varphi_t(v)$. Let W_t be the null space of φ_t, then by the First Isomorphism Theorem [25],

$$\psi_t : v + W_t \mapsto \varphi_t(v)$$

is a vector space isomorphism between the quotient space V/W_t and U_t.

Let t be an edge or a sink node. If $Y_f = 0$ for all $f \in \mathcal{I}(t)$, then $Y_t = 0$ as the encoding/decoding functions are linear. Thus $\bigcap_{f \in \mathcal{I}(t)} W_f \leq W_t$. Further, for each source s

$$W_s = \{\, v \in V \mid s\text{-th section of v is } 0 \,\} \cong \oplus_{r \in \mathcal{S} \setminus \{s\}} U_r,$$

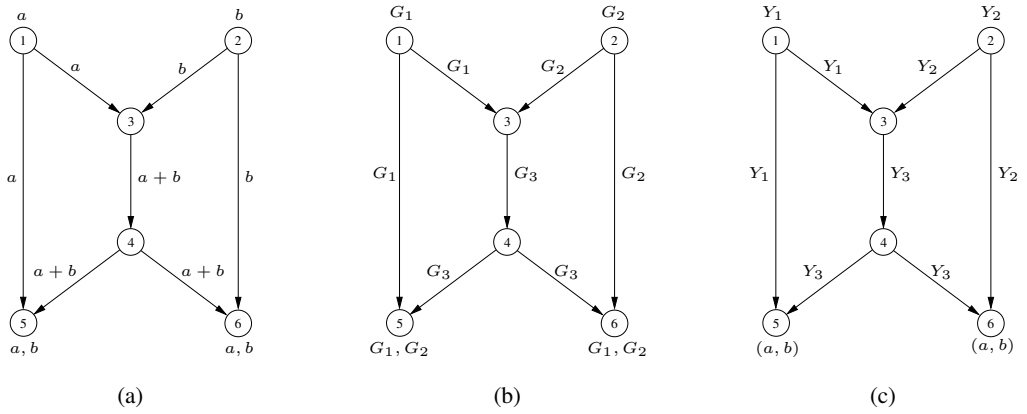


Figure A.1: Two network codes on the the butterfly network. (a) A linear network code; (b) the subgroup assignment for the corresponding group network code; (c) the transmitted symbols in the group network code.

so $\bigcap_{s\in\mathcal{S}} W_s = 0$. Since $V/W_s \cong U_s$, we have

$$\prod_{s\in\mathcal{S}} |V/W_s| = |V|.$$

Let $G = V$, $G_t = W_t$ for all $t \in \mathcal{S}\cup\mathcal{E}$. As V is a finite dimensional vector space over a finite field, G is a finite group. It is straightforward to check that the requirements (R1)–(R3) are all satisfied, so we can define a group network code $\mathcal{C}'$ with these groups.

This network code is equivalent to $\mathcal{C}$, since $\{\psi_t : t \in \mathcal{S}\cup\mathcal{E}\}$ provides a set of bijections between their codewords at each source/edge, and these bijections respect the encoding/decoding operations. In particular, assume in $\mathcal{C}$ that the source vectors yield some $v \in V$, and so $Y_t = \varphi_t(v)$ is transmitted at each source/edge t. Then with ψ_t the corresponding symbol for $\mathcal{C}'$ is $v + W_t$, which is consistent with the encoding/decoding result of $\mathcal{C}'$ at each edge/sink node by Proposition A.2.1.

For example, Figure A.1 demonstrates a linear network code over $\mathbb{F}_q$ for the well-known butterfly network (Figure A.1-(a)), and the corresponding group network code (Figure A.1-(b),(c)). Here for the linear network code, we have

$$V = \mathbb{F}_q^2$$

$$U_1 = U_2 = U_{e_{34}} = \mathbb{F}_q$$

$$W_1 = \{\, (0, x) : x \in \mathbb{F}_q \,\}$$

$$W_2 = \{\, (y, 0) : y \in \mathbb{F}_q \,\}$$

$$W_{e_{34}} = \{\, (z, -z) : z \in \mathbb{F}_q \,\}.$$

If we set $G = V$, $G_1 = W_1$, $G_2 = W_2$, and $G_3 = W_{e_{34}}$, then the resulting group network code is equivalent to the original linear one. In particular, for the group network code, the transmitted symbols are

$$Y_1 = \{\, (a, x) : x \in \mathbb{F}_q \,\}$$

$$Y_2 = \{\, (y, b) : y \in \mathbb{F}_q \,\}$$

$$Y_3 = \{\, (a + z, b - z) : z \in \mathbb{F}_q \,\}.$$

Appendix B

Appendices for Part II: A Theory of Stationarity and Ergodicity

To apply the Shannon-McMillan-Breiman theorem in our channel models, we need to derive the required stationarity and ergodicity conditions. For that purpose we introduce the theory of stationarity and ergodicity for Markov channels, mostly established by [48,49] (also see Gray's books [62,63]). It turns out, however, that some results in Gray et al. [49] are inaccurate and/or not properly proved; thus before making use of them we must fix these issues first. In addition, from the existing theory we want to develop some extended results tailored for our own purposes, especially for the application in the finite state channels derived from energy harvesting systems. In this appendix we first present the necessary background and preliminary results, then state the relevant theory of Markov channels from the literature, correct or supplement it if necessary, and in the meantime derive some extended or additional stationarity/ergodicity results of our own. Moreover, for our application we study the special case of a finite state channel with a finite-order Markov input, and also obtain some stationarity and ergodicity results. Following that we present the Shannon-McMillan-Breiman theorem in the setting of an AMS ergodic process, and then develop some specific results for the models used in this work.

Throughout this section we follow the notations in [48,49], which uses a convention different from the main text. At first glance, it might cause confusion; however, we wish to make the notational conventions consistent with the related literature, to

make it more convenient for the readers to have a coherent understanding of the related material.

B.1 Preliminaries

We gather the most frequently used notations, concepts and preliminary results here, and refer the interested readers to the original papers [48, 49] for the rest. Most of the terminology and basic results can also be found in Gray's books [62, 63].

B.1.1 General Properties

Let $(\Omega, \boldsymbol{F})$ be a measurable space and $T : \Omega \to \Omega$ be a measurable mapping on it. Define a probability measure μ on $(\Omega, \boldsymbol{F})$ to be *stationary*[1] if

$$\mu(T^{-1}F) = \mu(F), \qquad \forall F \in \boldsymbol{F}.$$

For a probability measure μ on $(\Omega, \boldsymbol{F})$, if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}F)$$

exists for all $F \in \boldsymbol{F}$, we say μ is *asymptotically mean stationary* (AMS). The above equation also defines a stationary probability measure on $(\Omega, \boldsymbol{F})$, which is called the *stationary mean* of μ and is usually denoted by $\bar{\mu}$. Define an event F to be *invariant* if $T^{-1}F = F$. μ is *ergodic* if $\mu(F)$ is either 0 or 1 for all invariant events F. Note that an AMS measure is ergodic iff its stationary mean is [62, Lemma 6.7.1].

We say a dynamical system $(\Omega, \boldsymbol{F}, \mu, T)$ is stationary, AMS, or ergodic if the measure μ is. The following lemmas provide some useful results regarding the AMS property of a dynamical system (see [62, Sec. 6.2–6.3]):

[1] This and many subsequent notions are defined *with respect to* T. But for conciseness we usually omit this modifying phrase.

Lemma B.1.1. *$(\Omega, \boldsymbol{F}, \mu, T)$ is AMS iff there exists a probability measure $\bar{\mu}$ on $(\Omega, \boldsymbol{F})$ which is stationary and which agrees with μ on each invariant event.*

Lemma B.1.2. *$(\Omega, \boldsymbol{F}, \mu, T)$ is AMS if there exists a stationary probability measure $\bar{\mu}$ on $(\Omega, \boldsymbol{F})$ such that for any invariant $F \in \boldsymbol{F}$, $\mu(F) = 0$ whenever $\bar{\mu}(F) = 0$.*

B.1.2 Sources, Channels, and Hookups

The dynamical systems we are interested in are sources and source-channel hookups, both of which can be either two-sided or one-sided. Let $(A, \boldsymbol{A})$ be a measurable space, on which we want to define the one- and two-sided sequence spaces and sources. Let $(A_1^\infty, \boldsymbol{A}_1^\infty)$ denote the measurable space of one-side sequences from alphabet A, whose sample space is composed of all sequences $(x_1, x_2, \cdots)$ from A and whose σ-field $\boldsymbol{A}_1^\infty$ is the usual product σ-field of A_1^∞. Let T be the left shift on A_1^∞, i.e.,

$$T : (x_1, x_2, \cdots) \mapsto (x_2, x_3, \cdots),$$

which is a measurable map. A dynamical system $(A_1^\infty, \boldsymbol{A}_1^\infty, \mu, T)$ of this form is called a one-sided *source*, or *process*, and is abbreviated to $[A, \mu]$. A two-sided source $(A^\infty, \boldsymbol{A}^\infty, \mu, T)$ is defined analogously: the sample space A^∞ consists of all two-sided sequences $(\cdots, x_{-1}, x_0, x_1, \cdots)$ from A and the σ-field $\boldsymbol{A}_1^\infty$ is the corresponding product σ-field. Again, T is the left shift, which maps a sequence $x = \{x_i\}_{i=-\infty}^\infty \in A^\infty$ to $Tx \in A^\infty$, where

$$(Tx)_i = x_{i+1}, \qquad \forall i \in \mathbb{Z}.$$

Note that in this case T has an inverse (the right shift), and both T and T^{-1} are measurable.

The same notation T is used for the left shifts on both spaces, but context should make clear what the underlying space is. Furthermore, for unified treatment of both cases, let $(\Sigma_A, \boldsymbol{\Sigma}_A)$ denote the one- or two-sided sequence space of $(A, \boldsymbol{A})$, and let $\boldsymbol{I}$ denote the time index set, which equals $\mathbb{Z}^+ \triangleq \{1, 2, \cdots\}$ or $\mathbb{Z}$ for the one- or two-sided cases, respectively. Recall that the basic events of the sequence spaces are the (finite

dimensional) *rectangles*, also called the *cylinder sets*, which are subsets F of the form

$$F = \{\, x \in \Sigma_A : x_i \in F_i,\ \forall i \in \boldsymbol{J} \,\},$$

where $\boldsymbol{J}$ is a finite subset of $\boldsymbol{I}$ and $F_i \in \boldsymbol{A}$ for all $i \in \boldsymbol{J}$. The sets F_i, $i \in \boldsymbol{J}$ are called the *coordinate events*. When F_i is a singleton for each $i \in \boldsymbol{J}$, F is called a *thin cylinder*.

A *channel* $[A, \nu, B]$ with input alphabet A and output alphabet B is defined by a family of probability measures $\{\nu_x : x \in \Sigma_A\}$ on $(\Sigma_B, \boldsymbol{\Sigma}_B)$ such that for each event $F \in \boldsymbol{\Sigma}_B$, the map

$$x \mapsto \nu_x(F)$$

from $(\Sigma_A, \boldsymbol{\Sigma}_A)$ into $[0, 1]$ with its Borel σ-field is measurable. A channel is called one- or two-sided if the underlying sequence space is. Given a source $[A, \mu]$ and a channel $[A, \nu, B]$, the *source-channel hookup*, or the *input-output process*, is the process $[A \times B, \mu\nu]$, where the measure $\mu\nu$ is defined by

$$\mu\nu(F) = \int_{\Sigma_A} \nu_x(F_x)\, \mathrm{d}\mu(x), \qquad \forall F \in \boldsymbol{\Sigma}_{A\times B},$$

with F_x being the section of F at x:

$$F_x \triangleq \{y \in \Sigma_B \mid (x, y) \in F\}.$$

The corresponding left shift for this process is still denoted by T, with

$$T(x, y) = (Tx, Ty), \qquad \forall (x, y) \in \Sigma_A \times \Sigma_B.$$

Sometimes when the alphabets are understood we simply denote the above source, channel, and their hookup by the corresponding measures μ, ν, and $\mu\nu$, respectively. As usual, the random processes corresponding to the source and hookup can be denoted by their respective sequences of coordinate random variables $\{X_n\}_{n\in\boldsymbol{I}}$ and

$\{(X_n, Y_n)\}_{n\in\boldsymbol{I}}$, where for any n we define

$$X_n : \Sigma_A \to A, \quad x \mapsto x_n$$
$$Y_n : \Sigma_B \to B, \quad y \mapsto y_n.$$

Sometimes we also drop the subscript $n \in \boldsymbol{I}$ when there is no confusion. We say these random processes are stationary, AMS, or ergodic if the underlying dynamic systems are. Furthermore, for convenience we define the projection map π between the one- and two-sided spaces on A as

$$\begin{array}{rl} \pi : A^\infty \to A_1^\infty \\ x \ \mapsto x_1^\infty \end{array},$$

where $x = \{x_i\}_{i=-\infty}^{\infty}$ and

$$x_1^\infty \triangleq (x_1, x_2, \cdots).$$

Similarly define the projection maps for the alphabets B and $A \times B$, which are still denoted by π. It is easy to verify that π is always measurable and *stationary*, namely, $\pi T = T\pi$.

A channel $[A, \nu, B]$ is said to be *stationary* if $\forall x \in \Sigma_A$, $\forall F \in \boldsymbol{\Sigma}_B$

$$\nu_{Tx}(F) = \nu_x(T^{-1}F).$$

The term "stationary" is justified by [63, Lemma 9.3.1], which shows that connecting a stationary source to a stationary channel yields a stationary input-output process. The channel is said to be *AMS* if, for every AMS source, the source-channel hookup is AMS. An AMS channel ν is *ergodic* if the hookup $\mu\nu$ is ergodic whenever μ is AMS and ergodic.

A simple example of stationary channels is the family of *stationary memoryless channels*[2]. Every channel $[A, \nu, B]$ in this family is associated with a collection of probability measures $\{q_a : a \in A\}$ on $(B, \boldsymbol{B})$, such that for each output rectangle

[2] In [63] such channels are simply called memoryless channels.

$F \in \mathbf{\Sigma}_B$,

$$\nu_x(F) = \prod_{i \in \boldsymbol{J}} q_{x_i}(F_i),$$

where $\boldsymbol{J}$ is the index set and F_i, $i \in \boldsymbol{J}$ are the coordinate events of F. When A and B are finite sets, ν is called a *discrete memoryless channel* (DMC).

B.1.3 Markov Channels and Finite State Channels

Fix the input and output measurable spaces $(A, \boldsymbol{A})$ and $(B, \boldsymbol{B})$, where $(A, \boldsymbol{A})$ is arbitrary, but B is a finite set with cardinality K and $\boldsymbol{B}$ consists of all subsets of B. Let $\boldsymbol{P}$ denote the space of all $K \times K$ stochastic matrices P, whose (i, j)-th entry is denoted by $P(i, j)$ for $1 \leq i, j \leq K$. Using the Euclidean metric on $\boldsymbol{P}$ we can construct its Borel σ-field to form a measurable space, which in turn induces a one- or two-sided sequence space $(\Sigma_P, \mathbf{\Sigma}_P)$. Given a sequence $P \in \Sigma_P$, let $\boldsymbol{M}(P)$ denote the set of all probability measures on $(\Sigma_B, \mathbf{\Sigma}_B)$ with respect to which $Y_m, Y_{m+1}, \cdots$ forms a (non-homogeneous) Markov chain with transition matrices $P_m, P_{m+1}, \cdots$ for any integer $m \in \boldsymbol{I}$. That is, $\lambda \in \boldsymbol{M}(P)$ iff $\forall m \in \boldsymbol{I}$, $\forall n > m$, and $\forall y_m, \cdots, y_n \in B$,

$$\lambda(Y_m = y_m, \cdots, Y_n = y_n) = \lambda(Y_m = y_m) \prod_{i=m}^{n-1} P_i(y_i, y_{i+1}).$$

In the one-sided case only $m = 1$ need be verified.

As before we say a map $\phi : \Sigma_A \to \Sigma_P$ is *stationary* if $\phi T = T\phi$. A channel $[A, \nu, B]$ is called *Markov* if there exists a stationary measurable map $\phi : \Sigma_A \to \Sigma_P$ such that

$$\nu_x \in \boldsymbol{M}(\phi(x)), \qquad \forall x \in \Sigma_A.$$

The major results proved in [48] by Kieffer and Rahe for Markov channels is summarized in the following theorem:

Theorem B.1.1. *Every one- and two-sided Markov channel is AMS.*

Now let A also be finite and let $\{P_a : a \in A\} \subset \boldsymbol{P}$. If a one-sided Markov channel

$[A, \nu, B]$ satisfies

$$\phi(x)_n \triangleq [\phi(x)]_n = P_{x_n}, \qquad \forall n > 0,$$

then ν is called a *finite state channel*. In this case, the matrix produced by ϕ at time n depends only on the input at that time, x_n. This definition is equivalent to Gallager's finite state channel (FSC) defined in [2], in terms of channel transitions. In fact, for the latter definition we have finite input, output, and state alphabets with respective symbols X_n, Y_n, and S_n that fulfill the conditional probability requirement[3]

$$\Pr\left(\begin{array}{c|l} \begin{array}{c} Y_n = y_n, \\ S_{n+1} = s_{n+1} \end{array} & \begin{array}{l} Y_i = y_i,\ 0 < i < n; \\ S_j = s_j,\ 0 < j \le n; \\ X_k = x_k,\ k > 0 \end{array} \end{array}\right) = p(y_n s_{n+1} \,|\, x_n s_n). \tag{B.1}$$

In other words, conditioned on (X_n, S_n), the pair (Y_n, S_{n+1}) is independent of all prior inputs, outputs, and states[4]. If we define the new output Y'_n of the channel as the output-state pair (Y_{n-1}, S_n) with

$$P_{x_n}\left(y'_n, y'_{n+1}\right) = P_{x_n}\left(\,(y_{n-1}, s_n), (y_n, s_{n+1})\,\right) \triangleq p(y_n s_{n+1} \,|\, x_n s_n),$$

then Gallager's model fits in the definition here. The other direction is obvious if we define $S_n = Y_n$. In light of their equivalence, we do not explicitly distinguish the two definitions in this appendix. Most of the time we will find out that it is more convenient to work with the first one when studying the general theory, while the second one provides more flexibility when dealing with specific channel models.

[3] As in the main text, the state index is increased by 1 compared to the original definition in [2].

[4] Actually from (B.1), (Y_n, S_{n+1}) is also conditionally independent of the future inputs. In the definition of FSC in [2] this requirement is not explicitly stated, however, it is indeed implicitly assumed when computing the block conditional probability (equation (4.6.1) in [2]). This subtle requirement is reasonable: since in FSC the input symbol does not depend on any past state or output information, (Y_n, S_{n+1}) should have no dependence on future inputs either (see Section B.3.3 for more discussion). Also, it is indeed satisfied by the FSC models we study in Chapter 5.

B.1.4 Constructions by Kieffer and Rahe

To prove Theorem B.1.1, Keiffer and Rahe establish some intermediate source and channel constructions in [48], which we will need for the relevant ergodicity results and are summarized below.

Let $[A, \mu]$ be an AMS source and $[A, \nu, B]$ be a Markov channel, with ϕ being the corresponding stationary map. Since μ is AMS, by Lemma B.1.1 there is a stationary measure $\bar{\mu}$ on $(\Sigma_A, \mathbf{\Sigma}_A)$ that agrees with μ on each invariant event in $\mathbf{\Sigma}_A$. ($\bar{\mu}$ can be simply taken to be the stationary mean of μ.) Define a two-sided stationary source $[A, \bar{\mu}^*]$ as follows: if the original source is two-sided, then $\bar{\mu}^* = \bar{\mu}$; otherwise let $\bar{\mu}^*$ be the two-sided stationary extension[5] of the one-sided measure $\bar{\mu}$, which is specified by

$$\bar{\mu}^*(\,(X_m, X_{m+1}, \cdots) \in F\,) = \bar{\mu}(F), \quad \forall m \in \mathbb{Z},\ \forall F \in \boldsymbol{A}_1^\infty.$$

In particular, considering $m = 1$ we have

$$\bar{\mu}^*(\pi^{-1}F) = \bar{\mu}(F).$$

Also, define a two-sided stationary map ϕ' by setting $\phi' = \phi$ if the original system is two-sided, and defining

$$\phi'(x)_i = \phi(x_i^\infty)_1 \qquad \forall i \in \mathbb{Z},\ \forall x \in A^\infty$$

otherwise, where $x_i^\infty \triangleq (x_i, x_{i+1}, \cdots)$. In particular, for the latter case

$$\phi'(x)_1^\infty = \phi(x_1^\infty) = \phi(\pi(x)).$$

Furthermore, [48] constructs a measurable subset $R \subset \boldsymbol{P}^\infty$ and proves that the measurable set

$$R' = (\phi')^{-1}(R) = \{x \in A^\infty : \phi'(x) \in R\}$$

[5]Such an extension is always possible and unique by the Kolmogorov extension theorem if the measurable space $(A, \boldsymbol{A})$ is *standard*, which is true for countable or Euclidean spaces. Interested readers can consult [62, Ch. 2,3] for details.

is invariant and has probability 1 under any stationary probability measure on $(A^\infty, \boldsymbol{A}^\infty)$, in particular

$$\bar{\mu}^*(R') = 1.$$

With these constructions Kieffer and Rahe define a two-sided channel $[A, \hat{\nu}, B]$ which has the following properties:

1) $\hat{\nu}$ is stationary and hence so is the input-output process $\bar{\mu}^*\hat{\nu}$.

2) $\hat{\nu}_x \in \boldsymbol{M}(\phi'(x))$ for $x \in R'$, so $\hat{\nu}$ has the same transition structure as ν, $\bar{\mu}^*$-a.e.

Besides, if the original system is two-sided, then $\mu\nu$ is absolutely continuous w.r.t. $\bar{\mu}^*\hat{\nu}$. In particular, for any invariant event $F \in \boldsymbol{A}^\infty \times \boldsymbol{B}^\infty$, $\mu\nu(F) = 0$ whenever $\bar{\mu}^*\hat{\nu}(F) = 0$, whereas if ν is one-sided, [48] defines the "one-sided restriction" of the two-sided measure $\bar{\mu}^*\hat{\nu}$ as

$$(\bar{\mu}^*\hat{\nu})' \triangleq (\bar{\mu}^*\hat{\nu})\pi^{-1},$$

which is also stationary since π is. Moreover, if $F \in \boldsymbol{A}_1^\infty \times \boldsymbol{B}_1^\infty$ is invariant and $(\bar{\mu}^*\hat{\nu})'(F) = 0$, then also $\mu\nu(F) = 0$. Therefore in both cases $\mu\nu$ is AMS by Lemma B.1.2, and so is ν.

Remark B.1.1. In [49] property 2) of $\hat{\nu}$ is assumed to be true for all $x \in A^\infty$, which is not the case in the original construction of [48]. This misrepresentation is one source of inaccuracy for Lemma 2 and the proof of Theorem 2 in [49], which we will fix in later sections.

From these facts we can also obtain the following two results regarding the ergodicity of certain related processes, which are indispensable in current approaches for proving ergodicity of Markov channels. Although their proofs are not difficult and [48] uses these results without explicitly proving them, we provide the proofs below for the sake of clarity and completeness.

Lemma B.1.3. *If μ is ergodic, then so is the auxiliary measure $\bar{\mu}^*$ for both one- and two-sided systems.*

Proof. By construction $\bar{\mu}$ is ergodic iff μ is, so for the two-sided case we are done. For the one-sided case, by the generating field structure of $\boldsymbol{A}^\infty$ and [62, Lemma 6.7.4] it

is enough to prove that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\bar{\mu}^*(T^{-i}F\cap G)=\bar{\mu}^*(F)\bar{\mu}^*(G) \tag{B.2}$$

for all rectangles $F, G \in \boldsymbol{A}^\infty$ when μ is ergodic. But by the stationarity of $\bar{\mu}^*$, without loss of generality we can assume the relevant coordinates for the rectangles F and G are positive. Thus there exists rectangles $F', G' \in \boldsymbol{A}_1^\infty$ such that $F = \pi^{-1}F'$ and $G = \pi^{-1}G'$. Now by the relation of $\bar{\mu}$ and $\bar{\mu}^*$ and the stationarity of π, (B.2) becomes

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\bar{\mu}(T^{-i}F'\cap G')=\bar{\mu}(F')\bar{\mu}(G'),$$

which is true by [62, Lemma 6.7.3] when $\bar{\mu}$ is ergodic. □

Lemma B.1.4. *If the auxiliary measure $\bar{\mu}^*\hat{\nu}$ is ergodic, then so is $\mu\nu$ for both one- and two-sided systems.*

Proof. Observe that the complement of an invariant event is also invariant. In the two-sided case, if $\bar{\mu}^*\hat{\nu}(F) = 1$ for an invariant F, then $\bar{\mu}^*\hat{\nu}(F^c) = 0$ and so $\mu\nu(F^c) = 0$, and thus $\mu\nu(F) = 1$. Hence ergodicity of $\bar{\mu}^*\hat{\nu}$ implies ergodicity of $\mu\nu$. For the one-sided case, let $F \in \boldsymbol{A}_1^\infty \times \boldsymbol{B}_1^\infty$ be invariant, then $\pi^{-1}F \in \boldsymbol{A}^\infty \times \boldsymbol{B}^\infty$ is also invariant as π is stationary. Assume $\bar{\mu}^*\hat{\nu}$ is ergodic, then

$$(\bar{\mu}^*\hat{\nu})'(F) = [(\bar{\mu}^*\hat{\nu})\pi^{-1}](F) = \bar{\mu}^*\hat{\nu}(\pi^{-1}F),$$

which is either 1 or 0. Again by the same argument, $\mu\nu(F) = 1$ or 0 and hence $\mu\nu$ is also ergodic. □

B.2 Ergodicity Results for Markov Channels

We are now ready to present the relevant results in [49], together with our comments, amendments, and corrections. In the meantime, we will develop some supplementary or extended results to apply in our own work.

B.2.1 Weak Ergodicity of Markov Channels

Assume the same setting as the previous section, where we have an AMS source $[A, \mu]$ and a Markov channel $[A, \nu, B]$ with the corresponding auxiliary constructions. For any $m \in \boldsymbol{I}$, $\forall n > m$ and $\forall x \in \Sigma_A$, we denote the output transition probability matrix for ν from time m to n by $H_{mn}(x) = H_{m,n}(x)$. In other words, for $1 \le j, k \le K$,

$$[H_{mn}(x)]_{jk} \triangleq \nu_x(Y_n = b_k \,|\, Y_m = b_j),$$

where we fix an ordered enumeration $\{b_1, b_2, \cdots, b_K\}$ of B. Since $\nu_x \in \boldsymbol{M}(\phi(x))$,

$$H_{mn}(x) = \prod_{i=m}^{n-1} \phi(x)_i, \qquad \forall x \in \Sigma_A. \tag{B.3}$$

Similarly, for the auxiliary two-sided channel $\hat{\nu}$, for any $m < n \in \mathbb{Z}$ and $\forall x \in A^\infty$ we define the matrix $H^*_{mn}(x) = H^*_{m,n}(x)$ by

$$[H^*_{mn}(x)]_{jk} \triangleq \hat{\nu}_x(Y_n = b_k \,|\, Y_m = b_j)$$

for $1 \le j, k \le K$. Since $\hat{\nu}_x \in \boldsymbol{M}(\phi'(x))$ on R', we have

$$H^*_{mn}(x) = \prod_{i=m}^{n-1} \phi'(x)_i, \qquad \forall x \in R'. \tag{B.4}$$

Thus if ν is two-sided, then $\phi' = \phi$ and so

$$H^*_{mn}(x) = H_{mn}(x), \qquad \forall x \in R',\ \forall n \ge m \in \mathbb{Z}, \tag{B.5}$$

whereas if ν is one-sided, as $\phi'(x)_1^\infty = \phi(\pi(x))$ for all $x \in A^\infty$,

$$H^*_{1n}(x) = H_{1n}(\pi(x)), \qquad \forall x \in R',\ \forall n \ge 1. \tag{B.6}$$

Definition B.2.1. Let H_{mn} denote the transition matrix from time m to n for a non-homogeneous Markov chain with K states, for $0 < m < n$. The Markov chain is

called *weakly ergodic* if

$$\lim_{n\to\infty} |(H_{mn})_{ij} - (H_{mn})_{kj}| = 0, \qquad \forall m > 0,\ \forall 1 \le i, j, k \le K. \tag{B.7}$$

A Markov channel ν is *weakly ergodic* if for all $x \in \Sigma_A$,

$$\lim_{n\to\infty} |[H_{mn}(x)]_{ij} - [H_{mn}(x)]_{kj}| = 0, \qquad \forall m \in \boldsymbol{I},\ \forall 1 \le i, j, k \le K. \tag{B.8}$$

We also say it is *weakly ergodic on a set* F if (B.8) holds for all $x \in F$. Furthermore, ν is called *weakly ergodic* μ*-a.e.* for a probability measure μ if it is weakly ergodic on a set with μ-measure 1.

Since ϕ is stationary, by (B.3)

$$H_{mn}(x) = H_{1,(n-m+1)}(T^{m-1}x).$$

This relation is true for both one- and two-sided channels for all $m \in \boldsymbol{I}$, noting that in the latter case T is invertible and so $T^{m-1}x$ is always a single point. Hence we only need to verify (B.8) for the special case $m = 1$ to prove the weak ergodicity of a Markov channel. Similarly, for the almost everywhere definition we have

Lemma B.2.1 (Lemma 1 in [49])**.** *Suppose μ is a stationary source. Then a Markov channel ν is weakly ergodic μ-a.e. iff for $m = 1$,* (B.8) *holds with μ-probability 1.*

Given a $K \times K$ stochastic matrix P, define

$$\delta(P) = \max_{s,t} \sum_{1\le k\le K} (P_{tk} - P_{sk})^+,$$

where $(a)^+ \triangleq \max\{0, a\}$. It is the maximum total variation distances between the rows of P, with $0 \le \delta(P) \le 1$. P is called *scrambling* if $\delta(P) < 1$, which holds iff for any two rows i and k there is at least one column j for which both $P_{ij} > 0$ and $P_{kj} > 0$; or equivalently, no two rows of P are orthogonal. Moreover, for any

stochastic matrices P and Q,

$$\delta(PQ) \le \delta(P)\delta(Q). \tag{B.9}$$

Observe that for any fixed m, (B.7) is true iff

$$\lim_{n\to\infty} \delta(H_{mn}) = 0.$$

This gives an equivalent definition for the weak ergodicity of a non-homogeneous Markov chain. By the same token we have the following lemma. Its first part comes from [49, Lemma 2] with the issue of $\hat{\nu}$ (metioned in Remark B.1.1) fixed, while the second part comprises two statements supplemented by ourselves.

Lemma B.2.2 (Lemma 2 in [49], amended for R' and extended)**.** *A Markov channel ν is weakly ergodic iff*

$$\lim_{n\to\infty} \delta(H_{1n}(x)) = 0, \qquad \forall x \in \Sigma_A.$$

In this case, the induced channel $\hat{\nu}$ is weakly ergodic on R'. Given a source $[A, \mu]$, a Markov channel ν is weakly ergodic μ-a.e. iff the event

$$F \triangleq \left\{ x \in \Sigma_A : \lim_{n\to\infty} \delta(H_{mn}(x)) = 0,\ \forall m \in \boldsymbol{I} \right\}$$

has μ-probability 1. If the μ is stationary, then only $m = 1$ need be considered. Furthermore, if μ is AMS, then ν is weakly ergodic μ-a.e. iff $\bar{\mu}$-a.e., in which case $\hat{\nu}$ is also weakly ergodic on a subset of R' with $\bar{\mu}^*$-probability 1.

Proof. The first statement follows from the observation above. The second will be considered together with the last one at the end of this proof. The μ-a.e. condition follows again from the previous observation, and the statement for stationary μ is a consequence of Lemma B.2.1.

Next for the original channel define

$$F_m = \left\{ x \in \Sigma_A : \lim_{n\to\infty} \delta(H_{mn}(x)) = 0 \right\}, \qquad \forall m \in \boldsymbol{I}.$$

Note that $F = \bigcap_{m\in\boldsymbol{I}} F_m$, which we will prove to be invariant. As shown before, $H_{mn}(x) = H_{1,(n-m+1)}(T^{m-1}x)$ for all m, hence $F_m = T^{-m+1}F_1$ and

$$F = \bigcap_{m\in\boldsymbol{I}} T^{-m+1}F_1.$$

For the two-sided case $\boldsymbol{I} = \mathbb{Z}$ and so $T^{-1}F = F$. For the one-sided case, however,

$$T^{-1}F = \bigcap_{m\geq 1} T^{-m}F_1 = \bigcap_{m\geq 1} F_{m+1}.$$

Note that for a Markov chain with transition matrices $\{H_{mn}\}_{m<n}$,

$$H_{mn} = H_{m,(m+1)}H_{(m+1),n}. \tag{B.10}$$

By (B.9) and the fact that $\delta(\cdot) \leq 1$,

$$\delta(H_{mn}) \leq \delta(H_{m,(m+1)}) \cdot \delta(H_{(m+1),n}) \leq \delta(H_{(m+1),n}). \tag{B.11}$$

Therefore, $\lim_{n\to\infty} \delta(H_{(m+1),n}(x)) = 0$ implies that $\lim_{n\to\infty} \delta(H_{mn}(x)) = 0$, hence $F_{m+1} \subset F_m$, and so for the one-sided channel

$$F = \bigcap_{m\geq 1} F_m = \bigcap_{m\geq 2} F_m = T^{-1}F.$$

Now assume μ is AMS. Since F is always invariant, $\mu(F) = \bar{\mu}(F)$ by the definition of $\bar{\mu}$. So ν is weakly ergodic μ-a.e. iff $\bar{\mu}$-a.e..

Finally let us consider the two-sided channel $\hat{\nu}$. Define

$$F^*_m = \left\{ x \in A^\infty : \lim_{n\to\infty} \delta(H^*_{mn}(x)) = 0 \right\}, \qquad \forall m \in \mathbb{Z}$$

and set $F^* = \bigcap_{m\in\mathbb{Z}} F_m^*$. Then by definition $\hat{\nu}$ is weakly ergodic on $F^* \cap R'$.

Assume ν is two-sided. By virtue of (B.5), $F_m^* \cap R' = F_m \cap R'$ and so

$$F^* \cap R' = F \cap R'.$$

If ν is weakly ergodic, then $F = A^\infty$ and $\hat{\nu}$ is weakly ergodic on R'. If ν is weakly ergodic μ-a.e. for an AMS source μ, then as $\bar{\mu}^* = \bar{\mu}$,

$$\bar{\mu}^*(F) = \bar{\mu}(F) = \mu(F) = 1.$$

As we already have $\bar{\mu}^*(R') = 1$, $F \cap R'$ also has $\bar{\mu}^*$-probability 1, on which $\hat{\nu}$ is weakly ergodic.

When ν is one-sided, by the stationarity of ϕ' and (B.4),

$$H_{mn}^*(x) = H_{1,(n-m+1)}^*(T^{m-1}x), \qquad \forall x \in R'.$$

Then considering the invariance of R',

$$\begin{aligned} F_m^* \cap R' &= \Big\{ x \in A^\infty : \lim_{n\to\infty} \delta(H_{1n}^*(T^{m-1}x)) = 0 \Big\} \cap R' \\ &= T^{-m+1} F_1^* \cap R' \\ &= T^{-m+1}(F_1^* \cap R'). \end{aligned}$$

Furthermore, according to (B.6),

$$\begin{aligned} F_1^* \cap R' &= \Big\{ x \in A^\infty : \lim_{n\to\infty} \delta[\, H_{1n}(\pi(x))\,] = 0 \Big\} \cap R' \\ &= \pi^{-1} F_1 \cap R'. \end{aligned}$$

If ν is weakly ergodic, then $F_1 = A_1^\infty$ and $\pi^{-1}F_1 = A^\infty$. Hence $F_m^* \cap R' = R'$ and so

$$F^* \cap R' = R',$$

on which $\hat{\nu}$ is weakly ergodic. If ν is weakly ergodic μ-a.e. for an AMS source μ, then as $\bar{\mu}(F) = \mu(F) = 1$ and $F \subseteq F_1$, we also have $\bar{\mu}(F_1) = 1$. Now since $\bar{\mu}^*(\pi^{-1}F_1) = \bar{\mu}(F_1)$ by construction, also $\bar{\mu}^*(R') = 1$, we have $\bar{\mu}^*(F_1^* \cap R') = 1$ and so

$$\bar{\mu}^*(F_m^* \cap R') = 1$$

by the stationarity of $\bar{\mu}^*$. Hence the countable intersection $F^* \cap R'$ also has $\bar{\mu}^*$-probability 1, on which $\hat{\nu}$ is weakly ergodic. □

The first main result in [49] provides an alternative characterization of a.e. weakly Markov channels. Let $\mathbb{E}[\,\cdot\,]$ denote expectation, i.e., the integration w.r.t. the corresponding measure.

Theorem B.2.1 (Theorem 1 in [49]). *A necessary condition for a Markov channel ν to be weakly ergodic μ-a.e. for a stationary measure μ is that there exists an N such that*

$$\mathbb{E}[\,\ln \delta(H_{1N}(X))\,] < 0. \tag{B.12}$$

A sufficient condition for ν to be weakly ergodic μ-a.e. for a stationary and ergodic measure μ is that there exists an N such that (B.12) *holds.*

Gray et al. further derive three corollaries of this theorem in [49]. However, all of them are inaccurate in that they all require an additional condition to hold: the source μ need be ergodic, apart from being stationary. That is because essentially the proofs all need to use the sufficient condition of the theorem. Below we state these corollaries as lemmas, together with the corrections and some extended results.

Lemma B.2.3 (Corollary 1 in [49], corrected and amended). *Given a Markov channel ν and a stationary ergodic source μ the following conditions are equivalent.*

a) *The channel is weakly ergodic μ-a.e..*

b) *For μ-a.e. each x, $\exists n$ such that no two rows of $H_{1n}(x)$ are orthogonal; or equivalently, $H_{1n}(x)$ is scrambling, i.e., $\delta(H_{1n}(x)) < 1$.*

c) *The channel has the "positive column property" μ-a.e.; that is, for μ-a.e. each x there is an n for which $H_{1n}(x)$ has a positive column.*

Proof. The proof provided in [49] is mostly correct, except that the result that b) implies a) does require the sufficient condition of Theorem B.2.1. To prove that result, assume b) is true but a) is false. Then $\mathbb{E}[\ln \delta(H_{1n}(X))] = 0$ for all n, otherwise by the sufficient condition ν is indeed weakly ergodic μ-a.e.. As $\ln \delta(\cdot) \leq 0$, for each n we must have $\ln \delta(H_{1n}(x)) = 0$ on a set F_n with μ-probability 1. Thus the intersection $\bigcap_{n>1} F_n$ also has μ-probability 1, on which $\delta(H_{1n}(x)) = 1$ for all n. As a result, the set

$$E \triangleq \Big(\bigcap_{n>1} F_n \Big)^c = \{ x \in \Sigma_A : \exists n > 1 \text{ s.t. } \delta(H_{1n}(x)) < 1 \}$$

is null, i.e., $\mu(E) = 0$. This is a contradiction, since $\mu(E) = 1$ by b). □

From the proof above, the contradiction still exists as long as E—the set on which the requirement for b) holds—has a positive μ-probability. Also, for each point x the requirement for b) is implied by that of c). Hence we can relax the conditions b) and c), to only requiring them to hold on a set with positive μ-probability, and the lemma is still correct. However, actually this is not a true relaxation, in view of our next lemma.

Lemma B.2.4. *Let μ be stationary and ergodic. The corresponding requirement for each condition of Lemma B.2.3 holds μ-a.e. iff it holds on a set of positive μ-probability.*

Proof. First consider condition a). The event F defined in Lemma B.2.2 is exactly the set on which the channel is weakly ergodic. Since the proof of that lemma shows that F is invariant, $\mu(F)$ is either 0 or 1 as μ is ergodic. Therefore $\mu(F) > 0$ iff $\mu(F) = 1$.

Next for condition b), let F denote the set on which the corresponding requirement holds, i.e.,

$$F \triangleq \{ x \in \Sigma_A : \exists n > 1 \text{ s.t. } \delta(H_{1n}(x)) < 1 \}.$$

Assume $x \in T^{-1}F$, then there is an n such that $\delta(H_{1n}(Tx)) < 1$. By (B.3) and the

stationarity of ϕ, this is equivalent to $\delta(H_{2,(n+1)}(x)) < 1$. But by (B.11)

$$\delta(H_{1,(n+1)}(x)) \leq \delta(H_{2,(n+1)}(x)) < 1,$$

we have $x \in F$ and so $T^{-1}F \subset F$.

As a further consequence,

$$F \supseteq T^{-1}F \supseteq T^{-2}F \supseteq \cdots$$

while $\mu(T^{-m}F) = \mu(F)$ for all $m > 0$, by the stationarity of μ. Define

$$E = \bigcap_{m \geq 0} T^{-m}F,$$

then $T^{-m}F \downarrow E$ and $\mu(E) = \lim_{m\to\infty} \mu(T^{-m}F) = \mu(F)$ by continuity of measures. Moreover, it is easy to check that $T^{-1}E = E$, so E is an invariant set. Since μ is ergodic, $\mu(E)$ is either 0 or 1 and so is $\mu(F)$. Therefore if $\mu(F) > 0$, then $\mu(F) = 1$.

Finally for condition c), similarly define F as the set on which the requirement of c) holds, i.e., $x \in F$ iff $\exists n$ such that $H_{1n}(x)$ has a positive column. Assume $x \in T^{-1}F$, then there is an n such that $H_{1n}(Tx) = H_{2,(n+1)}(x)$ has a positive column. But $H_{1,(n+1)}(x)$ also has a positive column as by (B.10)

$$H_{1,(n+1)}(x) = H_{1,2}(x)H_{2,(n+1)}(x),$$

all of which are stochastic matrices. Therefore $x \in F$ and $T^{-1}F \subset F$. By the previous paragraph again, $\mu(F) > 0$ iff $\mu(F) = 1$. □

Furthermore, note that for both conditions b) and c), the corresponding properties only need to hold on a finite segment of a sequence. Combining this observation with the definition of finite state channel, we have the following corollary.

Corollary B.2.1. *Let μ be a stationary ergodic source and ν be a Markov channel. For either condition b) or c) of Lemma B.2.3, if there exists a finite dimensional*

rectangle F possessing positive μ-probability such that the corresponding requirement holds for all $x \in F$, then ν is weakly ergodic μ-a.e. In particular, when ν is a finite state channel and F is a thin cylinder, we have a specific result: let $(a_1, \cdots, a_n) \in A^n$, if

1) $\mu(X_1 = a_1, \cdots, X_n = a_n) > 0$,

2) $\prod_{i=1}^{n} P_{a_i}$ *is scrambling, or has a positive column,*

then ν is weakly ergodic μ-a.e.

Proof. The first statement follows from the two lemmas above. For a finite state channel ν, let F be the thin cylinder with coordinate events $F_i = \{a_i\}$ for $1 \le i \le n$. Then by (B.3)

$$H_{1,(n+1)}(x) = \prod_{i=1}^{n} \phi(x)_i = \prod_{i=1}^{n} P_{a_i}, \qquad \forall x \in F.$$

Hence the second statement holds as a special case of the first one. □

The second corollary of Theorem B.2.1 deals with Gallager's concept of indecomposable finite state channels [2], which is generalized to all Markov channels in [49] as follows.

Definition B.2.2. A Markov channel ν is *indecomposable in the Gallager sense*[6] if for every $\epsilon > 0$ there is an N such that for all $n \ge N$

$$|[H_{1n}(x)]_{ij} - [H_{1n}(x)]_{kj}| < \epsilon, \qquad \forall x \in \Sigma_A,\ \forall 1 \le i, j, k \le K.$$

Remark B.2.1. For a Markov channel both the indecomposability in the Gallager sense and the weak ergodicity require that asymptotically the rows of the transition matrix become more and more alike. However, the former requires uniform convergence for all input sequences x while the latter does not.

If a Markov channel ν is indecomposable in the Gallager sense, then ν has the *strong positive column property*, that is, there is an n such that $H_{1n}(x)$ has a positive

[6] In the main text we only use the term *indecomposability* in the context of an FSC and it refers exclusively to this definition.

column for every x. If ν is a finite state channel, then [2] shows that the relation is indeed *if and only if*. Since obviously strong positive column property implies positive column property, by Lemma B.2.3 we have the following lemma.

Lemma B.2.5 (Corollary 2 in [49], corrected)**.** *A sufficient condition for a Markov channel to be weakly ergodic μ-a.e. for a stationary and ergodic source μ is that it is indecomposable in the Gallager sense μ-a.e.*

The third corollary of Theorem B.2.1 is not used in our work and requires some extra definitions, hence we only correct it below and refer the interested readers to the original paper of Gray et al. for the concept of indecomposability for a Markov channel (which is different from Definition B.2.2).

Lemma B.2.6 (Corollary 3 in [49], corrected)**.** *A sufficient condition for a Markov channel to be weakly ergodic μ-a.e. for a stationary and ergodic source μ is that it is indecomposable μ-a.e.*

Remark B.2.2. Since Lemma B.2.5 and B.2.6 essentially use Lemma B.2.3, by Lemma B.2.4 we only need their corresponding conditions to hold on a set of positive probability.

B.2.2 Mixing and Ergodic Markov Channels

Before presenting the main ergodicity results for Markov channels, we require yet another definition of a class of channels, which was first introduced by Adler in [64].

Definition B.2.3. A channel ν is called *strongly mixing*, or *output mixing* [63], or *asymptotically independent of the remote past* [64] if for all output rectangles F and G and all input sequences x

$$\lim_{n\to\infty} \left|\nu_x(F \cap T^{-n}G) - \nu_x(F)\nu_x(T^{-n}G)\right| = 0. \tag{B.13}$$

It is called *strongly mixing μ-a.e.* for a probability measure μ if the above condition holds for all x in a set of μ-measure 1.

Remark B.2.3. Immediately from the definition we can see that stationary memoryless channels are strongly mixing. In fact, the strongly mixing channels are proposed in [64] to generalize the idea of channels with finite memory (which obviously include the memoryless channels).

The importance of strongly mixing channels lies in the following theorem, which is adapted from [64] and [63, Lemma 9.4.3].

Theorem B.2.2 (Adler's Theorem)**.** *Let ν be a stationary channel. If μ is a stationary ergodic source and ν is strongly mixing μ-a.e., then $\mu\nu$ is also stationary and ergodic. Similarly, if μ is AMS ergodic and ν is strongly mixing μ-a.e., then $\mu\nu$ is also AMS and ergodic.*

Proof. For the statement with stationary μ, see [64] or [63, Lemma 9.4.3] for a proof. For the AMS case the proof can be easily adapted from the stationary case with [63, Lemma 9.3.2]. □

The following lemma connects the a.e. weak ergodicity and a.e. strongly mixing property of Markov channels.

Lemma B.2.7 (Lemma 3 in [49], corrected)**.** *Given a stationary source μ, if a Markov channel is weakly ergodic μ-a.e., then it is also strongly mixing μ-a.e.*

Remark B.2.4. The original statement of Lemma 3 in [49] claims that the reverse direction is also true. However, the proof for this direction has a missing link: equation (12) in [49] is not necessarily true when $\nu_x(F) = 0$, thus one cannot deduce weak ergodicity from strongly mixing property by (12). Nevertheless, since the reverse direction is not used in our work, we will not discuss the possible fixes of that proof.

The proof of the above lemma in [49] indeed gives the following specific pointwise result, which we will use later.

Lemma B.2.8. *Let $[A, \nu, B]$ be a channel (not necessarily Markov) and $x \in \Sigma_A$. If ν_x corresponds to a weakly ergodic Markov chain, namely,* (B.8) *is true for x, then* (B.13) *holds for x for all output rectangles F and G.*

Next we state the second main result in [49].

Theorem B.2.3 (Theorem 2 in [49])**.** *If a stationary Markov channel ν is weakly ergodic μ-a.e. for a stationary and ergodic source μ, then $\mu\nu$ is stationary and ergodic. A Markov channel is ergodic if it is weakly ergodic μ-a.e. with respect to all stationary measures μ (e.g., if it is weakly ergodic everywhere).*

Remark B.2.5. In fact the condition for the second statement can be weakened to just requiring ν to be weakly ergodic μ-a.e. with respect to all stationary and ergodic measures μ.

The proof of this theorem in [49] is mostly correct, except that the proof for the second statement has the issue of $\hat{\nu}$ mentioned in Remark B.1.1. Also it is too sketchy. In the following we use the same proof idea to extend this theorem to a more specific one tailored for our own purposes. Its proof not only rigorously assembles various results built up in this appendix, but also demonstrates the proper treatment of the corresponding measurable sets on which the desired properties hold. In particular, the above issue of $\hat{\nu}$ is fixed in this proof.

Theorem B.2.4. *Let ν be a Markov channel and μ be an AMS ergodic source. If ν is weakly ergodic μ-a.e., then the input-output process $\mu\nu$ is also AMS and ergodic.*

Proof. Construct the auxiliary measures/processes $\bar{\mu}$ and $\bar{\mu}^*$ and the auxiliary two-sided channel $\hat{\nu}$ as in Section B.1.4. First from Theorem B.1.1 we know $\mu\nu$ is AMS and by Lemma B.1.3 the stationary measure $\bar{\mu}^*$ is also ergodic. Next, as ν is weakly ergodic μ-a.e. and μ is AMS, $\hat{\nu}$ is weakly ergodic on a subset $R^* \subseteq R'$ with $\bar{\mu}^*$-probability 1 by Lemma B.2.2. Hence by Lemma B.2.8 the condition in Definition B.2.3 for the channel $\hat{\nu}$ holds for all $x \in R^*$, so $\hat{\nu}$ is strongly mixing $\bar{\mu}^*$-a.e. Now as $\hat{\nu}$ is also stationary while $\bar{\mu}^*$ is stationary and ergodic, $\bar{\mu}^*\hat{\nu}$ is also stationary and ergodic by Theorem B.2.2. Finally, $\mu\nu$ is also ergodic by Lemma B.1.4. □

Corollary B.2.2. *Let ν be a Markov channel and μ be a stationary ergodic source. If any one of the conditions in Lemmas B.2.3, B.2.5, and B.2.6 holds on a set of positive μ-probability, then $\mu\nu$ is AMS and ergodic.*

Proof. The result is obtained by combining Lemmas B.2.3–B.2.6, and Remark B.2.2 together with Theorem B.2.4. □

Corollary B.2.3. *Let ν be a finite state channel and μ be a stationary ergodic source. Let $(a_1, \cdots, a_n) \in A^n$, if*

1) $\mu(X_1 = a_1, \cdots, X_n = a_n) > 0$,

2) $\prod_{i=1}^{n} P_{a_i}$ *is scrambling, or has a positive column,*

then $\mu\nu$ is AMS and ergodic.

Proof. The result is obtained by combining Corollary B.2.1 and Theorem B.2.4. □

B.3 Results for Finite State Channels with Markov sources

In this section we specialize to the case of connecting a finite-order Markov input process to a finite state channel, and obtain some stationarity and ergodicity results. These results provide an alternative set of sufficient conditions for the Shannon-McMillan-Breiman theorem.

B.3.1 Extension Functions and Projection Mappings

As a preliminary step we introduce the k-step extension function g of a one-sided process and its left inverse, the sequence projection mapping f. Consider a random process $\{X_n\}_{n>0}$ with a finite alphabet A and process measure μ. For $k > 0$ define the k-step extension function

$$g : A_1^\infty \to (A^k)_k^\infty,$$

which maps a sequence $\{x_n\}_{n>0}$ to $\{w_n\}_{n\geq k}$, where w_n denotes the k-tuple

$$(x_{n-k+1}, \cdots, x_n)$$

for all $n \geq k$. As A is finite, g is measurable (which follows easily from [62, Lemma 1.4.1]). Let $\{W_n\}_{n\geq k}$ denote the resulting random process, then the n-th coordinate variable satisfies

$$W_n = (X_{n-k+1}, \cdots, X_n), \qquad \forall n \geq k.$$

According to the functional relationship, the process measure is μg^{-1}.

Conversely, let $\{W_n\}_{n\geq k}$ be a random process with alphabet A^k and process measure η. Let $\hat{f} : A^k \to A$ denote the projection function that maps $(x_1, \cdots, x_k)$ to its first coordinate x_1. Further define the sequence projection mapping

$$f : (A^k)_k^\infty \to A_1^\infty,$$

such that a sequence $\{w_n\}_{n\geq k}$ is mapped to $\{x_n\}_{n>0}$, with

$$x_{n-k+1} = \hat{f}(w_n)$$

for all $n \geq k$. Again, as A is finite, f is measurable. Let $\{X_n\}_{n>0}$ denote the random process produced by f, then its process measure is ηf^{-1}.

Observe that fg is the identity mapping and so f is a left inverse of g. However, gf is not identity and g is not invertible—it is one-to-one but not onto. It is straightforward to verify that both g and f are stationary, i.e.,

$$gT = Tg, \qquad fT = Tf.$$

These stationary mappings respect the stationarity, AMS property, and ergodicity of the processes. For example, given $\{X_n\}$ with a stationary μ, it is easy to check that the measure μg^{-1} is also stationary, and hence the corresponding process $\{W_n\}$ as well. If, instead, μ is AMS, then for any event $F \in \left(\boldsymbol{A}^k\right)_k^\infty$ we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu g^{-1}(T^{-i}F) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(g^{-1}T^{-i}F)$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu\left(T^{-i}(g^{-1}F)\right),$$

which always exists by the AMS property of μ, and thus $\{W_n\}$ is also AMS. As another example, assume $\{W_n\}$ is given with a (stationary or AMS) ergodic η. Let F be an invariant event for the space $(A_1^\infty, \boldsymbol{A}_1^\infty)$, then $f^{-1}F$ is also invariant:

$$T^{-1}(f^{-1}F) = f^{-1}(T^{-1}F) = f^{-1}T.$$

Hence $(\eta f^{-1})(F) = \eta(f^{-1}F) = 0$ or 1, and so ηf^{-1} and the corresponding process $\{X_n\}$ are also ergodic.

B.3.2 Finite-order Markov Processes

We start with the ergodicity of finite-order Markov processes in this subsection, and then extend to finite state channels with finite-order Markov sources in the next one. The main theoretical tool is the following theorem for the ergodicity of stationary Markov chains from [65].

Theorem B.3.1 (Theorem 1.19 in [65])**.** *Consider a Markov chain on a finite state space $\{1, 2, \cdots, K\}$ with transition matrix P. Assume the initial distribution π is a positive stationary distribution for this chain, namely, $\pi P = P$ and $\pi_i > 0$ for all $1 \le i \le K$. Then the corresponding stationary random process is ergodic iff P is irreducible, in which case π is the unique stationary distribution for P.*

Assume $\{X_n\}_{n>0}$ is a Markov process of order k, with a finite alphabet A. Let W_n denote the state $(X_{n-k+1}, \cdots, X_n)$ of the underlying Markov chain for $n \ge k$, then the state process $\{W_n\}_{n\ge k}$ is exactly given by the k-step extension function g applied to $\{X_n\}_{n>0}$. Conversely, $\{X_n\}_{n>0}$ is also given by the sequence projection mapping f and $\{W_n\}_{n\ge k}$. Hence by the previous subsection, the stationarity, AMS property, or ergodicity of one process implies the same property for the other. Let P denote the transition matrix of the Markov chain. The process measure η of $\{W_n\}$ is determined by P and the initial distribution, and is AMS by [48, Theorem 9].

Let $\bar{\eta}$ be the stationary mean of η and π be the initial distribution for $\bar{\eta}$, then π is a stationary distribution of P.[7] Denote the support of π by Γ, which is called the *contingent stationary support* of the Markov process $\{W_n\}$ (since it depends on the initial distribution). It is easy to see that Γ is a *closed* subset of A^k, that is, $P_{ij} = 0$ for all $i \in \Gamma, j \notin \Gamma$.

Now assume that the Markov chain P is irreducible on Γ. As the conditions for Theorem B.3.1 are satisfied on Γ with the initial distribution π, the stationary measure $\bar{\eta}$ is ergodic, and so is η (see [62, Lemma 6.7.1]). Hence $\{W_n\}$ and $\{X_n\}$ are AMS ergodic processes. Conversely, if $\{X_n\}$ or $\{W_n\}$ is ergodic, then η, and so $\bar{\eta}$ are ergodic, and by Theorem B.3.1, P is irreducible on Γ.

Moreover, when either of the above conditions holds, Theorem B.3.1 states that π is the unique stationary distribution for the chain on Γ. Thus if another initial distribution on the Markov chain induces a process measure $\tilde{\eta}$, whose stationary mean has a (stationary) initial distribution $\tilde{\pi}$ that is also supported on Γ, then necessarily $\tilde{\pi} = \pi$ and the stationary mean is $\bar{\eta}$. In particular, if $\Gamma = A^k$, or equivalently, (the full matrix) P is irreducible, then the stationary process measures for $\{W_n\}$ and $\{X_n\}$ are unique.

Summarizing the discussions above we have the following lemma.

Lemma B.3.1. *Let $\{X_n\}$ be a finite-alphabet finite-order Markov process, with an underlying state process $\{W_n\}$, whose Markov transition matrix is P. Then both $\{X_n\}$ and $\{W_n\}$ are AMS. Let Γ denote the contingent stationary support of $\{W_n\}$, then $\{W_n\}$ (and $\{X_n\}$) are ergodic iff P is irreducible on Γ. Furthermore, when this is the case, any other initial distribution of the Markov chain that leads to the same contingent stationary support induces the same stationary mean for $\{W_n\}$ (and hence also the same stationary mean for $\{X_n\}$), which are ergodic measures. In particular, if Γ is the full state space, or equivalently, P is irreducible, then these stationary process measures are unique.*

[7] A stationary distribution always exists for any finite-state Markov chain [66].

B.3.3 Finite State Channels with Markov sources

Consider a finite state channel defined in Gallager's form (B.1). Assume the source process $\{X_n\}_{n>0}$ is Markov of order $k > 0$ and is independent of the initial state S_1 of the FSC, then the joint process $\{(X_n, Y_n, S_{n+1})\}_{n>0}$ is also Markov of order k. To see that, let p denote the joint process measure. Observe that for all $n \geq 1$ and when the joint probability is nonzero, by (B.1),

$$\begin{aligned} p(x^n, y^n, s^{n+1}) &= p(s_1) \cdot p(x^n) \cdot p(y^n s_2^{n+1} \,|\, x^n s_1) \\ &= p(s_1) \cdot \prod_{i=1}^{n} p\left(x_i \,|\, x^{i-1}\right) \cdot \prod_{i=1}^{n} p(y_i s_{i+1} \,|\, x_i s_i), \end{aligned}$$

whereas by the expansion of joint probability,

$$\begin{aligned} p(x^n, y^n, s^{n+1}) &= p(s_1) \cdot \prod_{i=1}^{n} p\left(x_i, y_i, s_{i+1} \,|\, x^{i-1} y^{i-1} s^i\right) \\ &= p(s_1) \cdot \prod_{i=1}^{n} p\left(x_i \,|\, x^{i-1} y^{i-1} s^i\right) \cdot \prod_{i=1}^{n} p(y_i s_{i+1} \,|\, x_i s_i). \end{aligned}$$

Comparing the two expressions and induct on n, we have

$$p(x_n \,|\, x^{n-1} y^{n-1} s^n) = p(x_n \,|\, x^{n-1}) \tag{B.14}$$

for all n. Namely, the input does not depend on the past state or output (which is natural, since in the FSC model there is no output feedback or state information available at the transmitter).[8] Therefore, for $n \geq k$, by (B.1) and (B.14)

$$\begin{aligned} p(x_n y_n s_{n+1} \,|\, x^{n-1} y^{n-1} s^n) &= p(x_n \,|\, x^{n-1} y^{n-1} s^n) \cdot p(y_n s_{n+1} \,|\, x_n s_n) \\ &= p(x_n \,|\, x^{n-1}) \cdot p(y_n s_{n+1} \,|\, x_n s_n) \\ &\overset{(a)}{=} p(x_n \,|\, x_{n-k}^{n-1}) \cdot p(y_n s_{n+1} \,|\, x_n s_n) \\ &\overset{(b)}{=} p(x_n y_n s_{n+1} \,|\, x_{n-k}^{n-1} y_{n-k}^{n-1} s_{n-k+1}^{n}), \end{aligned} \tag{B.15}$$

[8] This is also in accordance with the comment after (B.1) (in the footnote). Moreover, in fact, from (B.14) and (4.7) we can also deduce (B.1), by reversing the direction of the above derivation and induct on i for the expansion of $p(y^n s_2^{n+1} \,|\, x^n s_1)$.

where (a) follows since $\{X_n\}$ is Markov of order k, (b) follows from expanding the joint probability

$$p(x_n y_n s_{n+1}, x^{n-k-1} y^{n-k-1} s^{n-k} \,|\, x_{n-k}^{n-1} y_{n-k}^{n-1} s_{n-k+1}^{n})$$

and summing over $(x^{n-k-1}, y^{n-k-1}, s^{n-k})$. Now since (B.15) is independent of n, $\{(X_n, Y_n, S_{n+1})\}_{n>0}$ is a (homogeneous) Markov process of order k.

In addition, when $\{X_n\}$ is i.i.d. (i.e., $k = 0$), from the derivation above we see that $\{(X_n, Y_n, S_{n+1})\}_{n>0}$ is simply Markov (i.e., of order-1). Hence combining with Lemma B.3.1, we have:

Lemma B.3.2. *If the source $\{X_n\}$ of an FSC is an order-k Markov process with $k \geq 0$, then $\{(X_n, Y_n, S_{n+1})\}$ is a Markov process of order $\max\{k, 1\}$. If the underlying Markov chain for the latter is irreducible on the contingent stationary support, then $\{(X_n, Y_n, S_{n+1})\}$ is AMS and ergodic.*

In our energy harvesting channels we often encounter FSC's that satisfy

$$p(y_n s_{n+1} \,|\, x_n s_n) = p(y_n \,|\, x_n s_n) p(s_{n+1} \,|\, x_n s_n), \tag{B.16}$$

for which we will show that if the input-state process is AMS ergodic, then so is the full joint process (see Lemma B.5.1 in Section B.5.1). Thus for such channels we have:

Corollary B.3.1. *If the source $\{X_n\}$ of an FSC satisfying* (B.16) *is an order-k Markov process with $k \geq 0$, then $\{(X_n, S_{n+1})\}$ is a Markov process of order $\max\{k, 1\}$. If the underlying Markov chain for the latter is irreducible on the contingent stationary support, then $\{(X_n, Y_n, S_{n+1})\}$ is AMS and ergodic.*

B.4 The Shannon-McMillan-Breiman Theorem

For a finite alphabet random process $\{X_n\}$ whose probability measure is denoted by p, we are interested in the convergence of the sample entropy $-\frac{1}{n} \log p(X^n)$ to the

entropy rate

$$H(\mathcal{X}) \triangleq \lim_{n\to\infty} \frac{1}{n} H(X^n) \tag{B.17}$$

whenever the limit exists. In information theory, this property is called the *asymptotic equipartition property* (AEP) [46]. When the process is i.i.d., AEP is easily proved using law of large numbers. When $\{X_n\}$ is stationary and ergodic, the Shannon-McMillan-Breiman (SMB) theorem for stationary processes [46] also gives the AEP; in particular, the sample entropy converges to the entropy rate with probability 1. Yet this result is still not general enough for our application in the energy harvesting systems, since the joint input-output process produced by the surrogate channel is often not stationary, but AMS instead. Hence we require an SMB theorem for AMS processes, which is also called the entropy ergodic theorem in [63].

Theorem B.4.1. *(Shannon-McMillan-Breiman / Entropy Ergodic Theorem [63]) Let* $\{X_n\}$ *be a finite alphabet random process with an AMS ergodic process distribution* p*, whose stationary mean is denoted by* $\bar{p}$*. Then the entropy rate* (B.17) *exists and*

$$\lim_{n\to\infty} -\frac{1}{n} \log p(X^n) = H(\mathcal{X}),$$

where the convergence is both p*-a.e. and in* L^1*-norm. Furthermore, the value of* $H(\mathcal{X})$ *is the same as* $H_{\bar{p}}(\mathcal{X})$*, the entropy rate defined under the stationary measure* $\bar{p}$*.*

B.5 Some Specific Results

In this section we derive some specific results regarding stationarity and ergodicity, which are to be used for the FSC models in Chapter 5.

B.5.1 Joint and Marginal Processes

In our application we always have a joint process, say $\{V_n, S_n, Y_n\}$, and want to apply the SMB theorem on its various marginal processes, e.g., $\{V_n, Y_n\}$ or $\{Y_n\}$. It is enough to show the required AMS and ergodic properties for the joint process

$\{V_n, S_n, Y_n\}$, since from their respective definitions we can easily see that these properties are inherited by the marginal processes from the joint one.

We also have some remarks for the other direction. Consider a general channel $[A, \nu, B']$ whose input and output symbols are X_n and Y'_n, respectively. Let $[A \times B', \eta, B]$ be another channel, whose input symbols are the pairs (X_n, Y'_n) and output symbols are Y_n. Assume η is a stationary memoryless channel, then it is stationary and strongly mixing and so Adler's theorem applies. In particular, if a source $[A, \mu]$ gives an AMS ergodic hookup $\mu\nu$, then by Theorem B.2.2, connecting $\mu\nu$ to η gives an AMS ergodic hookup $(\mu\nu)\eta$. In other words, the joint process $\{(X_n, Y'_n, Y_n)\}_{n\in \boldsymbol{I}}$ is also AMS and ergodic.

For the application in our energy harvesting channels, consider a special class of FSC models whose transition probability satisfies

$$p(y_n s_{n+1} \,|\, x_n s_n) = p(y_n \,|\, x_n s_n) p(s_{n+1} \,|\, x_n s_n). \tag{B.18}$$

We can view $p(s_{n+1} \,|\, x_n s_n)$ as the transition probability of a smaller finite state channel ν, with input symbols X_n and output symbols $Y'_n = S_n$. Furthermore, Y_n can be viewed as the output of another DMC η, whose input symbols are the pairs (X_n, Y'_n) with transition probability

$$p(y_n \,|\, x_n y'_n) = p(y_n \,|\, x_n s_n).$$

Applying the argument from the previous paragraph to the channels ν and η, we have the lemma below. As a result, to show the full joint process $\{(X_n, S_n, Y_n)\}_{n>0}$ is AMS and ergodic we only need to consider the smaller finite state channel $p(s_{n+1} \,|\, x_n s_n)$.

Lemma B.5.1. *For the FSC model* (B.18) *let* $\{X_n\}_{n>0}$ *be an input process that yields an AMS ergodic joint input-state process* $\{(X_n, S_n)\}_{n>0}$*, then the joint input-state-output process* $\{(X_n, S_n, Y_n)\}_{n>0}$ *is also AMS ergodic.*

B.5.2 Pre-historical State Variables

To construct the surrogate channel models in Chapter 5, we restrict the input function to depend on a finite duration of historical side information and provide some dummy pre-historical state variables. In this subsection the influences of such dummy variables are discussed. We use the FSC-X model as an example, and the analysis and conclusions for the other two channel models follow similarly.

In Section 5.2, for the FSC-X we define the pre-historical states $s^0_{-m+2} \in \mathcal{S}^{m-1}$, which is deterministic. With these states, for each $0 < n < m$, each input function v_n for the surrogate channel $\boldsymbol{W}'$ uniquely determines an input function u_n for the channel $\boldsymbol{W}$:

$$u_n = v_n(s^0_{n-m+1}, \cdot).$$

Hence together with (5.4) for $n \geq m$, we see that once s^0_{-m+2} is given, each input sequence $\{v_n\}_{n>0}$ for $\boldsymbol{W}'$ uniquely determines an input sequence $\{u_n\}_{n>0}$ for $\boldsymbol{W}$. A different choice of pre-historical states leads to a potentially different input sequence for $\boldsymbol{W}$, thus defining a potentially different surrogate channel: since for each v^N the channel transition probability $p(y^N \,|\, v^N)$ of $\boldsymbol{W}'$ is determined by the corresponding u^N through $p(y^N \,|\, u^N)$ of $\boldsymbol{W}$. We can also directly see the influence of s^0_{-m+2} on the transition probability of $\boldsymbol{W}'$ from (5.5).

Therefore, given a random process $\{V_n\}$, different choices of pre-historical states lead to potentially different surrogate channels and hence potentially different achievable rates. (Sometimes the rates are indeed the same, see the next subsection.) Yet these rates are all achievable for the channel $\boldsymbol{W}$, since each of them corresponds to a particular input distribution process $\boldsymbol{U}$.

As commented in Section 5.2, these dummy pre-historical states are only used to help determining the function values of v_n for $n < m$, but do not affect the real initial channel state S_1, which is determined by the environment/nature/physical mechanism. However, they do affect the transitions of the channel states $S_2, \cdots, S_m$ indirectly through the function values x_n of v_n, and hence also the initial state Z_m of the FSC $\boldsymbol{W}'$ (which starts from $n = m$). The impact on the ergodicity of the

joint input-state-output process for a given source $\{V_n\}$, is thus different in different situations: if the FSC $\boldsymbol{W}'$ together with the source satisfies one of the ergodicity conditions in Section B.2, all of which deal solely with channel transition probabilities, then the pre-historical states do not influence the ergodicity of the system; if, instead, $\boldsymbol{W}'$ and the source satisfy an ergodicity condition in Section B.3, then the choice of pre-historical states might influence the ergodicity of the joint process, since the conditions therein are contingent on the initial state of the system.

B.5.3 Starting Time and Initial State for Surrogate Channels

In Chapter 5, the surrogate channels are only finite state channels for $n \geq M$, where M is a time index strictly larger than 1 in many cases. Our stationarity and ergodicity theory, as well as the simulation/optimization algorithms, are mostly prepared for FSC's, which starts from time index M; however, the computation of information rates via sample entropies (5.3) requires a starting time index 1. Nevertheless, we can resolve this conflict by setting the starting point of sample entropy computation to M, by virtue of the following lemma:

Lemma B.5.2 (Lemma 3.4.1 in [63])**.** *Let $\{X_n\}$ be a finite alphabet random process. For $M > 1$, almost surely we have*

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{p(X^n)}{p(X_M^n)} = 0.$$

As stated in the previous subsection, the initial state Z_M of the FSC $\boldsymbol{W}'$ is influenced by the choice of pre-historical states. It necessarily affects the information rate computation, but in the following case two different initial state distributions of $\boldsymbol{W}'$ give rise to the same rate. Assume the source is a finite-order Markov process and the source-channel process is ergodic under some initial state distribution. Then by Lemma B.3.2 and B.3.1, the joint input-output-state process is also finite-order Markov, whose underlying chain is irreducible on the contingent stationary support. Also, if two such initial state distributions of $\boldsymbol{W}'$ lead to the same contingent stationary support, then the corresponding stationary process measures are the same. Since

by the SMB theorem (Theorem B.4.1), the information rates are determined by the stationary measures, these two initial state distributions induce the same information rate.

Bibliography

[1] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.

[2] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, 1968.

[3] Z. Zhang and R. W. Yeung, "A non-Shannon-type conditional inequality of information quantities," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1982–1986, Nov. 1997.

[4] X. Yan, R. Yeung, and Z. Zhang, "The capacity for multi-source multi-sink network coding," in *Proc. of the 2007 IEEE International Symposium on Information Theory*, Nice, France, Jun. 2007, pp. 116–120.

[5] B. Hassibi and S. Shadbakht, "Normalized entropy vectors, network information theory and convex optimization," in *Proc. of the 2007 IEEE Information Theory Workshop*, Bergen, Norway, Jul. 2007, pp. 1–6.

[6] R. Dougherty, C. Freiling, and K. Zeger, "Networks, matroids, and non-shannon information inequalities," in *IEEE Transactions on Information Theory*, June 2007, pp. 1949–1969.

[7] S. Shadbakht and B. Hassibi, "Cayley's hyperdeterminant, the principal minors of a symmetric matrix and the antropy region of 4 Gaussian random variables," in *Proc. of the 46th Annual Allerton Conference on Communication, Control and Computing*, Monticello, IL, Sep. 2008.

[8] T. Chan, "A combinatorial approach to information inequalities," in *Communications in Information and Systems*, vol. 1, no. 3, 2001, pp. 241–253.

[9] T. H. Chan and R. W. Yeung, "On a relation between information inequalities and group theory," *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1992–1995, Jul. 2002.

[10] A. Ingleton, "Representation of matroids," in *Combinatorial Mathematics and its Applications*, 1971, pp. 149–167.

[11] D. Hammer, A. Romashchenko, A. Shen, and N. Vereshchagin, "Inequalities for shannon entropy and kolmogorov complexity," *Journal of Computer and System Sciences*, vol. 60, no. 2, pp. 442–464, Apr. 2000.

[12] F. Matúš, "Conditional independences among four random variables III: Final conclusion," *Combinatorics, Probability and Computing*, vol. 8, pp. 269–276, 1999.

[13] R. Dougherty, C. Freiling, and K. Zeger, "Insufficiency of linear network coding in network information flow," in *IEEE Transactions on Information Theory*, 2005, pp. 2745–2759.

[14] R. Kinser, "New inequalities for subspace arrangements," *Journal of Combinatorial Theory*, vol. 118, no. 1, pp. 152–161, Jan. 2011.

[15] R. Dougherty, C. Freiling, and K. Zeger, "Linear rank inequalities on five or more variables," 2010, preprint. [Online]. Available: arXiv:0910.0284v3

[16] R. Dougherty, "Computations of linear rank inequalities on six variables," in *Proc. of the 2014 IEEE International Symposium on Information Theory*, Honolulu, HI, Jun. 2014, pp. 2819–2823.

[17] R. Dougherty, C. Freiling, and K. Zeger, "Characteristic-dependent linear rank inequalities and network coding applications," in *Proc. of the 2014 IEEE In-*

ternational Symposium on Information Theory, Honolulu, HI, Jun. 2014, pp. 101–105.

[18] T. Chan, A. Grant, and D. Pflüger, "Truncation technique for characterizing linear polymatroids," *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6364–6378, Oct. 2011.

[19] T. H. Chan, "Group characterizable entropy functions," in *Proc. of the 2007 IEEE International Symposium on Information Theory*, Nice, France, Jun. 2007, pp. 506–510.

[20] Z. Zhang and R. W. Yeung, "On characterization of entropy function via information inequalities," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 1440–1452, Jul. 1998.

[21] W. Mao and B. Hassibi, "Violating the Ingleton inequality with finite groups," in *Proc. of the 47th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sep./Oct. 2009.

[22] N. Boston and T.-T. Nan, "Large violations of the Ingleton inequality," in *Proc. of the 50th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2012.

[23] P. Paajanen, "Finite p-groups, entropy vectors, and the Ingleton inequality for nilpotent groups," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3821–3824, Jul. 2014.

[24] N. Markin, E. Thomas, and F. Oggier, "Groups and information inequalities in 5 variables," in *Proc. of the 51th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2013.

[25] D. S. Dummit and R. M. Foote, *Abstract algebra*, 3rd ed. Hoboken, NJ: Wiley, 2004.

[26] *GAP – Groups, Algorithms, and Programming, Version 4.4.12*, The GAP Group, 2008. [Online]. Available: http://www.gap-system.org

[27] H. Li and E. K. P. Chong, "On connections between group homomorphisms and the ingleton inequality," in *Proc. of the 2007 IEEE International Symposium on Information Theory*, Nice, France, Jun. 2007, pp. 1996–2000.

[28] D. L. Johnson, *Presentations of Groups*, ser. London Mathematical Society Student Texts. Cambridge: Cambridge University Press, 1990, vol. 15.

[29] R. Dougherty, C. Freiling, and K. Zeger, "Non-shannon information inequalities in four random variables," 2011, preprint. [Online]. Available: arXiv:1104.3602v1

[30] F. Matúš and L. Csirmaz, "Entropy region and convolution," *Combinatorics, Probability and Computing*, submitted. [Online]. Available: arXiv:1310.5957v1

[31] S. Shadbakht and B. Hassibi, "MCMC methods for entropy optimization and nonlinear network coding," in *Proc. of the 2010 IEEE International Symposium on Information Theory*, Austin, TX, Jun. 2010.

[32] S. Shadbakht, "Entropy region and network information theory," Ph.D. dissertation, California Institute of Technology, 2011.

[33] T. H. Chan, "On the optimality of group network codes," in *Proc. of the 2005 IEEE International Symposium on Information Theory*, Adelaide, Australia, Sep. 2005, pp. 1992–1996.

[34] ——, "Capacity regions for linear and abelian network codes," in *Proc. of the 2007 Information Theory and Applications Workshop*, La Jolla, CA, Jan./Feb. 2007, pp. 73–78.

[35] O. Ozel, J. Yang, and S. Ulukus, "Optimal broadcast scheduling for an energy harvesting rechargeable transmitter with a finite capacity battery," *IEEE Transactions on Wireless Communications*, vol. 11, no. 6, pp. 2193–2203, Jun 2012.

[36] J. Yang and S. Ulukus, "Optimal packet scheduling in an energy harvesting communication system," *IEEE Transactions on Communications*, vol. 60, no. 1, pp. 220–230, January 2012.

[37] O. Ozel and S. Ulukus, "Achieving awgn capacity under stochastic energy harvesting," *IEEE Transactions on Information Theory*, vol. 58, no. 10, pp. 6471–6483, October 2012.

[38] C. Shannon, "Channels with side information at the transmitter," *IBM Journal of Research and Development*, vol. 2, no. 4, pp. 289–293, Oct. 1958.

[39] O. Ozel and S. Ulukus, "AWGN channel under time-varying amplitude constraints with causal information at the transmitter," in *Proc. of the 45th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2011.

[40] K. Tutuncuoglu and A. Yener, "Optimum transmission policies for battery limited energy harvesting nodes," *IEEE Transactions on Wireless Communications*, vol. 11, no. 3, pp. 1180–1189, Mar. 2012.

[41] K. Tutuncuoglu, O. Ozel, A. Yener, and S. Ulukus, "Binary energy harvesting channel with finite energy storage," in *Proc. of 2013 IEEE International Symposium on Information Theory*, Istanbul, Turkey, Jul. 2013.

[42] Y. Dong and A. Özgür, "Approximate capacity of energy harvesting communication with finite battery," in *Proc. of 2014 IEEE International Symposium on Information Theory*, Honolulu, HI, Jun. 2014.

[43] O. Ozel, K. Tutuncuoglu, S. Ulukus, and A. Yener, "Capacity of the discrete memoryless energy harvesting channel with side information," in *Proc. of 2014 IEEE International Symposium on Information Theory*, Honolulu, HI, Jun. 2014.

[44] P. O. Vontobel, A. Kavčić, D. M. Arnold, and H.-A. Loeliger, "A generalization of the blahut–arimoto algorithm to finite-state channels," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 1887–1918, May 2008.

[45] G. Caire and S. Shamai (Shitz), "On the capacity of some channels with channel state information," *IEEE Transactions on Information Theory*, vol. 45, no. 6, pp. 2007–2019, Sep. 1999.

[46] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, N.J: Wiley-Interscience, 2006.

[47] J. Chen, H. Permuter, and T. Weissman, "Tighter bounds on the capacity of finite-state channels via Markov set-chains," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3660–3691, Aug. 2010.

[48] J. C. Kieffer and M. Rahe, "Markov channels are asymptotically mean stationary," *Siam Journal of Mathematical Analysis*, vol. 12, no. 3, pp. 293–305, 1981.

[49] R. M. Gray, M. O. Dunham, and R. L. Gobbi, "Ergodicity of markov channels," *IEEE Transactions on Information Theory*, vol. 33, no. 5, pp. 656–664, Sep. 1987.

[50] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 284–287, Mar. 1974.

[51] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[52] D. Arnold and H.-A. Loeliger, "On the information rate of binary-input channels with memory," in *Proc. of 2001 IEEE International Conference on Communications*, Helsinki, Finland, Jun. 2001, pp. 2692–2695.

[53] V. Sharma and S. K. Singh, "Entropy and channel capacity in the regenerative setup with applications to Markov channels," in *Proc. of 2001 IEEE International Symposium on Information Theory*, Washington, DC, Jun. 2001, p. 283.

[54] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite-state ISI channels," in *Proc of 2001 IEEE Global Telecommunications Conference (GLOBECOM '01)*, San Antonio, TX, Nov. 2001, pp. 2992–2996.

[55] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3498–3508, August 2006.

[56] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Transactions on Information Theory*, vol. 18, no. 4, pp. 460–473, Jul. 1972.

[57] G. Han, "A randomized approach to the capacity of finite-state channels," in *Proc. of 2013 IEEE International Symposium on Information Theory*, Istanbul, Turkey, Jul. 2013.

[58] K. Tutuncuoglu, O. Ozel, A. Yener, and S. Ulukus, "Improved capacity bounds for the binary energy harvesting channel," in *Proc. of 2014 IEEE International Symposium on Information Theory*, Honolulu, HI, Jul. 2014.

[59] O. Ozel, K. Tutuncuoglu, S. Ulukus, and A. Yener, "Capacity of the energy harvesting channel with energy arrival information at the receiver," in *Proc. of the 2014 Information Theory Workshop*, Hobart, Australia, Nov. 2014.

[60] J. Chen and T. Berger, "The capacity of finite-state Markov channels with feedback," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 780–798, Mar. 2005.

[61] H. H. Permuter, T. Weissman, and A. J. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 644–662, Feb. 2009.

[62] R. M. Gray, *Probability, Random Processes, and Ergodic Properties*, 1st ed. Springer, 1987.

[63] ——, *Entropy and Information Theory.* New York: Springer-Verlag, 1990.

[64] R. L. Adler, "Ergodic and mixing properties of infinite memory channels," *Proceedings of the American Mathematical Society*, vol. 12, no. 6, pp. 924–930, 1961.

[65] P. Walters, *An Introduction to Ergodic Theory*, ser. Graduate texts in mathematics. New York: Springer-Verlag, 1982, vol. 79.

[66] Y. Ephraim and N. Merhav, "Hidden markov processes," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1518–1569, Jun. 2002.