On the Capacity of Noisy Computations

François Simon[1]

Telecom SudParis, 9 rue Charles Fourier, 91011 Evry, FRANCE

Abstract

This paper presents an analysis of the concept of *capacity* for *noisy computations*, i.e. functions implemented by unreliable or random devices. An information theoretic model of noisy computation of a perfect function f (measurable function between sequence spaces) thanks to an unreliable device (random channel) F is given: a noisy computation is a product $f \times F$ of channels. A model of reliable computation based on input encoding and output decoding is also proposed. These models extend those of noisy communication channel and of reliable communication through a noisy channel. The capacity of a noisy computation is defined and justified by a coding theorem and a converse. Under some constraints on the encoding process, capacity is the upper bound of *input rates* allowing reliable computation, i.e. decodability of noisy outputs into expected outputs.

These results hold when the one-sided random processes under concern are asymptotic mean stationary (AMS) and ergodic. Conditions on f and F for the noisy computation $f \times F$ be AMS and ergodic are also given.

In addition, some characterizations of AMS and ergodic noisy computations are given based on stability properties of the perfect function f and of the random channel F. These results are derived from the more general framework of channel products. Finally, a way to apply the noisy and reliable computation models to cases where the perfect function f is a defined according to a formal computational model is proposed.

Keywords:
Noisy Computation, Capacity of Noisy Computation, Coding Theorem, Channel Product.

Email address: Francois.Simon@telecom-sudparis.eu (François Simon)

[1] Institut Mines-Telecom ; Telecom SudParis ; CITI

1. Introduction

Reliable computation with unreliable devices, or in the presence of noise to use a terminology close to the Information Theory field, has been the subject of numerous works within the vast field of fault-tolerant computing but much less in Information Theory.

A major difference between reliable communication and reliable computation in the presence of noise is that unreliable computing devices can be made reliable not only using information redundancy but also component redundancy. For example, self-checking circuits may rely on information redundancy but also on gate redundancy, see [17]. Among theoretical analyses, some aim at identifying theoretical boundaries on the amount of necessary and/or sufficient redundancy to achieve reliability.

Recent references continue to extend the stream opened by Von Neumann's seminal paper [16], for example [24], [7], [12], [21] and [19]. These works identify theoretical limits or bounds (e.g., depth and size of circuits) but also propose frameworks to design reliable computations mainly thanks to gate redundancy. Nevertheless these papers do not address the question of a capacity for noisy computations.

When it exists, the capacity of a channel sets a boundary between codes which allow reliable communication and codes which cannot. Channel capacity can be viewed as the upper limit of the ratio k/n of encoding k-length input blocks in n-length blocks of binary symbols to allow almost perfect error correction for arbitrarily large k and n.

The question of whether a noisy computation possesses a capacity like communication channels do (or equivalently whether some coding theorems for noisy computation hold and in which cases) has been raised by [4]. There is a practical consequence in answering positively this theoretical question. This would mean that, given a noisy implementation of an expected function, it is possible to find families of efficient input codes which asymptotically allow an almost perfect computation. Efficiency means having an input encoding rate which could remain strictly positive or arbitrarily close to a capacity when the length of the code tends to infinity.

While the concept of capacity has been thoroughly studied, in Information Theory, for data communication, it has not been the case for computation and very few positive results have been obtained.

In a strongly constrained context (independent encoding of operands for bit-by-bit boolean operations), [4] obtained negative first results on the existence of a noisy computation capacity . This work was deepened by [18] and by [25]. One of the major conclusions of these studies was that reliable computation with asymptotic positive rate (the ratio k/n of encoding k-length input block in n-length blocks of binary symbols) in the presence of noise is not possible for some boolean operations (e.g., AND) under some assumptions (independent coding of operands, bijective decoding and bit-by-bit operation). This led to the conclusion that, under these assumptions, there is no capacity for such noisy operations. It is worth noting that the assumptions were made to forbid the reliable encoder and decoder to "participate" to the computation of the expected operation.

[1] went into the subject in greater depth and made an important contribution in characterizing contexts in which a capacity for noisy computations cannot exist. It appears that the characteristics of the decoding function play a fundamental role. If the inverse of the decoding function is injective and monotonic then weak converse theorems hold for the average and maximal error probabilities. If, in addition, the inverse of the decoding function preserves the logical AND (this implies monotonicity), then strong converses hold[2]. The hypotheses made by [4], [18] and [25], i.e., independent encoding of operands and bijective decoding, imply monotonicity of the inverse of the decoding function. On this aspect, [1] supersedes [4], [18] and [25].

Nevertheless, these negative results do not imply the absolute impossibility to identify a capacity for noisy computation. They characterize codes and the companion encoding and decoding processes which cannot open this ability. To relevantly define a capacity for noisy computations, the assumptions must be relaxed.

To the author's best knowledge, the first positive answer given through a definition of a capacity of a noisy computation (in fact similar to the one for a noisy channel) and a companion coding theorem came from [26]. [26] considers the special case of noisy functions called decomposable modules. Decomposable modules are noisy functions which can be modeled by a perfect function followed by a noisy transmission channel: the error probability depends on the desired output value not on the input value. The quantity $max_X(I(f(X);Z))$ is shown to be a capacity, where X is the input pro-

[2] Weak converse states that the rate of encoding tends to 0 when the block code length tends to infinity and for an error probability which tends to zero. Strong converse states the same result for an error probability arbitrarily close to 1.

cess, f is the expected perfect function, Z the decomposable module output and $I(f(X); Z)$ is the mutual information between the expected output process and the noisy one. These peculiar noisy functions make the context equivalent to that where the reliable encoder computes and encodes the expected function result before transmission through a noisy channel. Due to the restriction of considering decomposable modules, [26] did not completely succeed in proposing a noisy computation capacity in a general scope (see theorem 6.3, pages 47-48 in [26]). In fact the key property of decomposable modules is that $X \to f(X) \to Z$ is a Markov chain (see 6.2 below).

Noisy computation capacity is also considered in reliable reconstruction of a function of sources over a multiple access channel. Much more recently, a definition of noisy computation capacity is established by [15] and is totally consistent with the one proposed here. Nazer and Gastpar demonstrate the possible advantages of joint source-channel coding of multiple sources over a separation-based scheme, allowing a decoder to retrieve a value which is a function of input sources. This context makes relevant the proposed distributed encoding process which perfectly performs a computation equivalent to the desired function. The encoder outputs are then transmitted through a noisy multiple access channel (MAC) to a decoder, see proofs of Theorems 1 and 2 of [15]. This models a noisy computation as a perfect computation followed by a noisy transmission of the result like [26] and, thus, does not cover in full generality noisy computation of functions. It can be noticed that [26] and [15] relax the assumptions of [4], [18], [25] and [1] in a similar way: all goes as if the operands are jointly coded into an encoded form of the expected function result before being handled by the noisy device which transmits with possible noise. Joint coding means that, for example, a pair of operands (x_1, x_2) is encoded in a pair (u_1, u_2) where u_1 is obtained from (x_1, x_2) and not only from x_1 (the same for u_2).

The question close to the fundamental one raised by Elias is under what conditions, for an algorithm and a noisy implementation of it, does there exist a family of codes which permits almost perfectly reliable computation and for what cost ? To the author's best knowledge, no general answer to this question has been given.

The present paper extends and makes more rigorous ideas sketched in [22] and [23]. Noisy computations are addressed following an information theoretic approach and not a computability theoretic one.

Part I is devoted to preliminaries on sources and channels. Section 2 and Section 3 review the definitions and key properties of information sources, which are random processes, and random channels, which are probability kernels. Some technical lemmas and some properties of one-sided channels used latter are proved. Channel products are defined.

Part II deals with the main subject of this paper. Section 4 proposes a model for noisy computations which is in fact an extension of the classical approach for noisy channels: a noisy computation is a channel $X \to (f(X), Z)$ where f is the perfect or expected function and Z the (possibly noisy) output of a random channel. An important remark is that functions under consideration are not, at this stage, assumed to be computable according to a formal computational model (e.g., Turing computable functions). They are measurable functions between sequence spaces. Section 5 motivates and defines the notions of typical input rate and typical input capacity of a noisy computation. The adjective "typical" comes from the fact that the encoding process assumed in section 7 will be constrained in using only typical sequences. Section 6 extends Feinstein theorem and the so-called Feinstein channel codes to noisy computations. Section 7 proposes a complete model of reliable computation of a function thanks to a sequence "encoder-noisy computation-decoder" and then states and proves a "joint source-computation coding theorem" and its converse. These results assume that the functions under concern are unary. m-ary functions can be modeled as unary ones by concatenating m input values in one "meta"-input and thus modeling a *joint coding* of operands. This relaxes the assumption of independent coding of operands made by [4], [18], [25] and [1]. To some extent, computation induces a generalization of the problem (and so some results) set by communication. The concept of code applies well: a code for computation is a couple of an input code (linked to a function f) and an "associated" output code. The capacity of a noisy computation generalizes well the existing concept for communication as the latter appears to be a special case of the former. Finally, as an example, Section 8 establishes that correcting results of a computation on noisy input offers a better input rate that correcting noisy inputs before computation.

The results established in part II hold for AMS and ergodic noisy computations. Then the characterization of classes of AMS and ergodic noisy computations is an important question in this context. Part III is a first analysis of AMS and ergodic noisy computations. Some results are obtained for channel products and straightforwardly applied to noisy computations. In Section 9, it is proved that a channel product is resp. stationary, AMS (w.r.t. a stationary source), recurrent or output weakly mixing if both channels are resp. stationary, AMS (w.r.t. a stationary source), recurrent or output weakly mixing. Based on these results, Section 10 gives sufficient condition for ergodicity of channel products and thus noisy computations. Section 11 proposes a way to apply the noisy computation model (and thus the reliable computation model, coding theorem . . .) to cases where the perfect function is defined in a formal computational model and its noisy realization in the corresponding formal noisy computation model. In fact, this is obtained through the more general case where the perfect function is a function acting on finite length sequences and producing finite length sequences. Two very simple examples are given.

Part I

Preliminaries: information sources and channels

2. Information Sources

This section reviews the definition of information sources and some properties of sources.

Information sources can be described following two equivalent models: random processes and dynamical systems (see [10]). Both models will be used below. Notations follow partly [10] and partly [14].

Definition 1. *Let $\mathcal{I}$ be a countable index set. Let $(A, \mathcal{B}_A)$ be a measurable space, called the alphabet. Let $(A^{\mathcal{I}}, \mathcal{B}_{A^{\mathcal{I}}})$ be the measurable sequence space where $\mathcal{B}_{A^{\mathcal{I}}}$ is the σ-field generated by the sets of rectangles $\{x = (x_i)_{i \in \mathcal{I}} \in A^{\mathcal{I}} / x_i \in B_i, B_i \in \mathcal{B}_A$ for any $i \in \mathcal{J}\}$, $\mathcal{J}$ finite subsets of $\mathcal{I}$. A source $[A, X]$ is a random process $X = \{X_i; i \in \mathcal{I}\}$ with values in $(A^{\mathcal{I}}, \mathcal{B}_{A^{\mathcal{I}}})$. The distribution of the source $[A, X]$ is denoted by P_X.*

If the index set $\mathcal{I}$ is $\mathbb{N}$ then the process X is said to be one-sided. If the index set $\mathcal{I}$ is $\mathbb{Z}$ then the process X is said to be two-sided. Let $T_A : A^{\mathcal{I}} \to A^{\mathcal{I}}$ be the shift transform on $A^{\mathcal{I}}$. For a one-sided process:

$$T_A(x_0 x_1 \ldots x_i \ldots) = x_1 x_2 \ldots x_{i+1} \ldots$$

For a two-sided process:

$$T_A(\ldots x_{-j} \ldots x_0 x_1 \ldots x_i \ldots) = \ldots x_{-j+1} \ldots x_1 x_2 \ldots x_{i+1} \ldots$$

In the latter case, the shift T_A is invertible.

T_A is $(\mathcal{B}_{A^{\mathcal{I}}}, \mathcal{B}_{A^{\mathcal{I}}})$-measurable. If μ is a probability on $(A^{\mathcal{I}}, \mathcal{B}_{A^{\mathcal{I}}})$ then $(A^{\mathcal{I}}, \mathcal{B}_{A^{\mathcal{I}}}, T_A, \mu)$ is a dynamical system.

Definition 2. *A source on the alphabet $(A, \mathcal{B}_A)$ is a dynamical system $(A^{\mathcal{I}}, \mathcal{B}_{A^{\mathcal{I}}}, T_A, \mu)$.*

Let Π_0 denote the "zero-time sampling function": $\Pi_0(x) = x_0$ for any $x = x_0 x_1 \ldots x_i \ldots$ (or $x = \ldots x_{-j} \ldots x_0 x_1 \ldots x_i \ldots$ in the two-sided case). The two definitions of a source are equivalent. A dynamical system $(A^{\mathcal{I}}, \mathcal{B}_{A^{\mathcal{I}}}, T_A, \mu)$ determines a random process $X = \{X_i; i \in \mathcal{I}\}$ where $X_i(\omega) = \Pi_0(T_A^i(x)) = x_i$ and the distribution of X is μ: $P_X = \mu$.

In the sequel, each model will be used when the most relevant. Moreover, the word source will be used to name either the random process, the dynamical system, the distribution P_X or the probability μ. If $x = x_0 \ldots x_n \ldots \in A^{\mathcal{I}}$, x^n stands for $x_0 \cdots x_{n-1}$.

Definition 3. *Let $[A, X]$ be a source with distribution P_X. The n^{th} extension of the source $[A, X]$ is the random value $X^n = (X_0, \ldots, X_{n-1})$.*

The distribution P_{X^n} of the n^{th} extension of the source $[A, X]$ is given by:

$$\forall E \in \mathcal{B}_{A^n}, P_{X^n}(E) = P_X(c(E))$$

where $c(E) = \{x \in A^{\mathcal{I}} / x^n \in E\}$.

Definition 4. *A source $[A, X]$ with distribution μ is* stationary *if $\forall E \in \mathcal{B}_{A^{\mathcal{I}}}$, $\mu(T_A^{-1}E) = \mu(E)$.*

A source is asymptotically mean stationary *(AMS) if $\forall E \in \mathcal{B}_{A^{\mathcal{I}}}$, the limit $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(T^{-k}E)$ exists. If $\overline{\mu}(E)$ is this limit, then $\overline{\mu}$ is a stationary probability called the stationary mean of μ.*

A source is recurrent *if, $\forall E \in \mathcal{B}_{A^{\mathcal{I}}}$, $\mu(E \setminus \cup_{k\geq 1} T_A^{-k}E) = 0$.*

A recurrent and AMS source is called R-AMS.

A source with distribution μ is ergodic *if for any T_A-invariant $E \in \mathcal{B}_{A^{\mathcal{I}}}$ (i.e., $T_A^{-1}E = E$), $\mu(E) = 0$ or $\mu(E) = 1$.*

If μ is AMS, then its stationary mean $\overline{\mu}$ asymptotically dominates μ ($\mu \ll^a \overline{\mu}$): $\overline{\mu}(E) = 0 \Rightarrow \lim_{n\to\infty} \mu(T^{-n}E) = 0$ ([9]). μ is AMS if and only if there exists a stationary distribution η such that $\mu \ll^a \eta$.

A stationary source is recurrent ([9]). An AMS source distribution μ is dominated by its stationary mean ($\mu \ll \overline{\mu}$: $\overline{\mu}(E) = 0 \Rightarrow \mu(E) = 0$) if and only if the source is recurrent, see [9], theorem 7.4.

The next lemma is a rephrasing of Theorem 3 of [11].

Lemma 1. *Let $[A, X]$ be a source with distribution μ, let $[A, Y]$ be a stationary source with distribution η. Let $(\mathcal{B}_{A^{\mathcal{I}}})_\infty$ denote the tail σ-field $\cap_{n\geq 0} T_A^{-n}\mathcal{B}_{A^{\mathcal{I}}}$, μ_∞ and η_∞ denote the restrictions of μ and η to $(\mathcal{B}_{A^{\mathcal{I}}})_\infty$. Then η asymptotically dominates μ if and only if η_∞ dominates μ_∞:*

$$\mu \ll^a \eta \Leftrightarrow \mu_\infty \ll \eta_\infty$$

The following lemma asserts that asymptotic mean stationarity and ergodicity jointly imply recurrence.

Lemma 2. *An ergodic and AMS process is recurrent.*

Proof. Let μ be an ergodic AMS process distribution on the space $(\Omega, \mathcal{B})$. μ is AMS then $\mu \ll^a \overline{\mu}$ and $\overline{\mu}$ is ergodic.

From the Lebesgue decomposition theorem there exist two probabilities μ_1, μ_2 on $(\Omega, \mathcal{B})$ and $\lambda \in [0, 1]$ such that $\mu = \lambda\mu_1 + (1 - \lambda)\mu_2$, $\mu_1 \ll \overline{\mu}$, μ_2 and $\overline{\mu}$ are mutually singular.

$\overline{\mu}$ is stationary (thus recurrent) ergodic and dominates μ_1 then μ_1 is AMS ergodic and recurrent. This implies (by theorem 7.4, page 220, of [9]) that $\mu_1 \ll \overline{\mu}_1$. μ_1 being ergodic so is $\overline{\mu}_1$. $\overline{\mu}$ and $\overline{\mu}_1$ are stationary ergodic probabilities on the same space, they are thus equal or mutually singular ([14], Lemma 1 p. 75). Since they both dominate the same probability (μ_1), they are equal: $\overline{\mu} = \overline{\mu}_1$

$\mu = \lambda\mu_1 + (1 - \lambda)\mu_2$ is AMS and μ_1 is AMS then, necessarily, μ_2 is AMS. This also implies that: $\overline{\mu} = \lambda\overline{\mu_1} + (1 - \lambda)\overline{\mu_2}$

Since $\overline{\mu} = \overline{\mu_1}$, necessarily $\lambda = 1$, hence $\mu = \mu_1$ which is recurrent. □

Definition 5. *Let $[A, X]$ be a source. The entropy rate of the source is defined by*

$$\overline{H}(X) = \limsup_{n\to\infty} \frac{H(X_0, \dots, X_{n-1})}{n}$$

where $H(X_0, \dots, X_{n-1}) = -\sum_{x^n \in A^n} P_{X^n}(\{x^n\}) ln(P_{X^n}(\{x^n\}))$ is the entropy of the random value $(X_0, \dots, X_{n-1})$.

If a source $[A, X]$ is AMS and ergodic then

$$\overline{H}(X) = \lim_{n\to\infty} \frac{H(X_0, \dots, X_{n-1})}{n}$$

A consequence of the Shannon-Mc Millan-Breimann theorem is the Asymptotic Equipartition Property ([10]):

Theorem 1 (Asymptotic Equipartition Property)**.** *Let $[A, X]$ be an ergodic AMS source. For $\epsilon > 0$, let $A^n(\epsilon)$ be the set of ϵ-typical n-sequences*

$$A^n(\epsilon) = \left\{ x^n \in A^n / \left| \frac{-log(P_{X^n}(x^n))}{n} - \overline{H}(X) \right| < \epsilon \right\}$$

Then:

1. $\lim_{n\to\infty} P_{X^n}(A^n(\epsilon)) = 1$
2. *for any $\delta > 0$ and any $\epsilon > 0$, there exists N such that $\forall n \geq N$:*

$$(1 - \delta)e^{n(\overline{H}(X) - \epsilon)} \leq |A^n(\epsilon)| \leq e^{n(\overline{H}(X) + \epsilon)}$$

where $|A^n(\epsilon)|$ is the cardinality of $A^n(\epsilon)$.

3. Channels

This section reviews the definition of noisy communication channels and some properties of channels, used latter, are recalled (references are [6] and [14]). Some technical lemmas are given for further use. In what follows the shifts under consideration are not assumed to be invertible, channels under consideration may be one-sided.

3.1. Noisy or random channels

Let $(A, \mathcal{B}_A)$ and $(B, \mathcal{B}_B)$ be alphabets, $(A^{\mathcal{I}}, \mathcal{B}_{A^{\mathcal{I}}})$ and $(B^{\mathcal{I}}, \mathcal{B}_{B^{\mathcal{I}}})$ be the two corresponding sequence spaces. The shifts (assumed non-invertible) on $(A^{\mathcal{I}}, \mathcal{B}_{A^{\mathcal{I}}})$ and $(B^{\mathcal{I}}, \mathcal{B}_{B^{\mathcal{I}}})$ are respectively denoted T_A and T_B.

Let $\mathcal{B}_{A^{\mathcal{I}} \times B^{\mathcal{I}}}$ be the σ-field generated by the rectangles $\{E \times G / E \in A^{\mathcal{I}}, G \in B^{\mathcal{I}}\}$. T_A and T_B define a measurable shift T_{AB} on the space $(A^{\mathcal{I}} \times B^{\mathcal{I}}, \mathcal{B}_{A^{\mathcal{I}} \times B^{\mathcal{I}}})$ where $T_{AB}(x, y) = (T_A x, T_B y)$.

Definition 6. *A noisy communication channel $[A, \nu, B]$ is a function $\nu : A^{\mathcal{I}} \times \mathcal{B}_{B^{\mathcal{I}}} \to [0, 1]$ such that:*

- *for any $x \in A^{\mathcal{I}}$, the set function $G \mapsto \nu(x, G)$ is a probability on the space $(B^{\mathcal{I}}, \mathcal{B}_{B^{\mathcal{I}}})$*
- *for any $G \in \mathcal{B}_{B^{\mathcal{I}}}$, the function $x \mapsto \nu(x, G)$ is measurable.*

If the noisy channel $[A, \nu, B]$ takes a source $[A, X]$ as input, it produces as an output an information source $[B, Y]$. Let P_{XY} be the distribution of the joint process (X, Y) induced by the source X fed into the noisy channel. P_{XY} is defined on rectangles by:

$$\forall E \in \mathcal{B}_{A^{\mathcal{I}}}, \forall G \in \mathcal{B}_{B^{\mathcal{I}}}, P_{XY}(E \times G) = \int_E \nu(x, G) dP_X$$

P_{XY} will also be denoted by $\mu\nu$ (the hookup of μ and ν where $\mu = P_X$).

In fact a noisy communication channel is a probability kernel. In the sequel, the alphabets $(A, \mathcal{B}_A)$ and $(B, \mathcal{B}_B)$ will be assumed standard. Then the sequence spaces $(A^{\mathcal{I}}, \mathcal{B}_{A^{\mathcal{I}}})$ and $(B^{\mathcal{I}}, \mathcal{B}_{B^{\mathcal{I}}})$ are also standard. This assumption is made to ensure that conditional probabilities defined are regular (see [9] or [5]). Thus, on such spaces, given a joint random process (X, Y), it will always be possible to define a channel ν, unique P_X-a.s., taking X as an input and inducing the joint process (X, Y) ([10]).

Given a channel $[A, \nu, B]$ and a source $[A, X]$, and denoting $[B, Y]$ the corresponding output source, the probabilities $\nu(x, .)$ will also be denoted by $P_{Y|X}(.|x)$.

Definition 7. *Let $[A, X]$ be a source and $[A, \nu, B]$ be a channel. Let (X, Y) be the generated joint process and $[B, Y]$ be the output of the channel when $[A, X]$ is the input. The set of probabilities $\{P_{Y^n|X^n}(.|x^n), x^n \in A^n\}$ such that, $\forall a^n \in A^n$ such that $P_{X^n}(\{a^n\}) \neq 0$, $\forall G \in \mathcal{B}_{B^n}$:*

$$P_{Y^n|X^n}(G|a^n) = \frac{1}{P_X(c(a^n))} \int_{c(a^n)} P_{Y|X}(c(G)|x) dP_X$$

is called the induced channel (of order n).

Obviously, since $P_{XY}(c(\{a^n\}) \times c(G)) = \int_{c(a^n)} P_{Y|X}(c(G)|x) dP_X$

$$P_{XY}(c(\{a^n\}) \times c(G)) = P_X(c(\{a^n\})).P_{Y^n|X^n}(G|a^n)$$

thus

$$P_{X^nY^n}(\{a^n\} \times G) = P_{X^n}(\{a^n\}).P_{Y^n|X^n}(G|a^n)$$

and, $\forall E \in \mathcal{B}_{A^n}$:

$$P_{X^nY^n}(E \times G) = P_{X^n}(E).P_{Y^n|X^n}(G|E)$$

Given an input source and a channel, the induced channel is the (finite) channel linking the n^{th} extensions of the input and output sources.

Definition 8. *A channel $[A, \nu, B]$ is* stationary *with respect to a stationary source distribution μ if the hookup $\mu\nu$ is stationary.*

A channel $[A, \nu, B]$ is asymptotically mean stationary (AMS) *with respect to an AMS source distribution μ if the hookup $\mu\nu$ is AMS.*

A stationary (resp. AMS) channel $[A, \nu, B]$ is ergodic *with respect to a stationary (resp. AMS) ergodic source distribution μ if the hookup $\mu\nu$ is ergodic.*

A channel $[A, \nu, B]$ is output weakly mixing *w.r.t. μ if $\forall G, G' \in \mathcal{B}_{B^{\mathcal{I}}}$ $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |\nu(x, T_B^{-i} G \cap G') - \nu(x, T_B^{-i} G).\nu(x, G')| = 0$ μ-a.e.*

A channel $[A, \nu, B]$ is recurrent *with respect to a recurrent source distribution μ if the hookup $\mu\nu$ is recurrent.*

A channel $[A, \nu, B]$ is R-AMS *with respect to an R-AMS source distribution μ if the hookup $\mu\nu$ is R-AMS.*

The following lemma is immediate (c.f. [14]).

Lemma 3. *A channel $[A,\nu,B]$ is* stationary *with respect to a stationary source μ if and only if $\forall G \in \mathcal{B}_{B^{\mathcal{I}}}$, $\nu(x,T_B^{-1}G)=\nu(T_Ax,G)$ μ.a.e.*

Definition 9. *Let A and B be two finite alphabets. Let $[A,\nu,B]$ be a channel, $[A,X]$ an input source and $[B,Y]$ the corresponding output source. The information rate of (X,Y) is defined by*

$$\overline{I}(X,Y)=\limsup_{n\to\infty}\frac{I(X^n,Y^n)}{n}$$

where

$$I(X^n,Y^n)=\sum_{x^n\in A^n,y^n\in B^n}P_{X^nY^n}(\{x^n\}\times\{y^n\})ln\left(\frac{P_{X^nY^n}(\{x^n\}\times\{y^n\})}{P_{X^n}(\{x^n\}).P_{Y^n}(\{y^n\})}\right)$$

is the mutual information of (X^n,Y^n).

If the source $[A,X]$ and the channel $[A,\nu,B]$ are AMS and ergodic then (Lemma 8.1, [10]):

$$\overline{I}(X,Y)=\lim_{n\to\infty}\frac{I(X^n,Y^n)}{n}$$

In this case, the conditional entropy $\overline{H}(X|Y)=\overline{H}(X)-\overline{I}(X,Y)$ is well defined and a "conditional" Asymptotic Equipartition Property holds.

Theorem 2 (Conditional AEP)**.** *Let A and B be two finite alphabets. Let $[A,\nu,B]$ be an ergodic AMS channel, $[A,X]$ an ergodic AMS input source and $[B,Y]$ the corresponding output source. For $\epsilon>0$ and $y\in B^{\mathcal{I}}$, let $A^n(\epsilon,y^n)$ be the set of ϵ-typical sequences given y^n:*

$$A^n(\epsilon,y^n)=\left\{x^n\in A^n/\left|\frac{-log(P_{X^n|Y^n}(x^n|y^n))}{n}-\overline{H}(X|Y)\right|<\epsilon\right\}$$

Then:

1. $\lim_{n\to\infty}P_{X^n|Y^n}(A^n(\epsilon,y^n)|y^n)=1$ $P_Y-a.e.$
2. *for any $\delta>0$ and any $\epsilon>0$, for n large enough:*

$$(1-\delta)e^{n(\overline{H}(X|Y)-\epsilon)}\leq|A^n(\epsilon,y^n)|\leq e^{n(\overline{H}(X|Y)+\epsilon)}\ P_Y^n-a.e.$$

3.2. Restriction of a channel to the tail σ-field

The following lemma asserts that it is possible to consistently define the restriction $[A, \nu_\infty, B]$ of the channel $[A, \nu, B]$ meaning a probability kernel such that $\mu_\infty \nu_\infty = (\mu\nu)_\infty$.

Lemma 4. *Let μ be the distribution of a source $[A, X]$ and let μ_∞ denote its restriction to the tail σ-field $(\mathcal{B}_{A^\mathcal{I}})_\infty = \cap_{n \geq 0} T_A^{-n} \mathcal{B}_{A^\mathcal{I}}$. Let $[A, \nu, B]$ be a channel. Then*

- $\left(\sigma\left(\mathcal{B}_{A^\mathcal{I}} \times \mathcal{B}_{B^\mathcal{I}}\right)\right)_\infty = \sigma\left((\mathcal{B}_{A^\mathcal{I}})_\infty \times (\mathcal{B}_{B^\mathcal{I}})_\infty\right)$
- $(\mu\nu)_\infty = \mu_\infty \nu_\infty$ *where* $\nu_\infty(x, .)$ *is the restriction of* $\nu(x, .)$ *to the tail* σ*-field* $(\mathcal{B}_{B^\mathcal{I}})_\infty$ μ_∞*-a.e.*

Proof. Let $\mathcal{G} = \mathcal{B}_{A^\mathcal{I}} \times \mathcal{B}_{B^\mathcal{I}}$. Since T_A and T_B are respectively $(\mathcal{B}_{A^\mathcal{I}}, \mathcal{B}_{A^\mathcal{I}})$-measurable and $(\mathcal{B}_{B^\mathcal{I}}, \mathcal{B}_{B^\mathcal{I}})$-measurable , $\forall i \in \mathbb{N}$, $T_{AB}^{-(i+1)}\mathcal{G} \subset T_{AB}^{-(i)}\mathcal{G}$ and, thus, $\sigma(T_{AB}^{-(i+1)}\mathcal{G}) \subset \sigma(T_{AB}^{-(i)}\mathcal{G})$. Then $\forall n \in \mathbb{N}$

$$\sigma(T_{AB}^{-n}\mathcal{G}) = \cap_{i=0}^{n} \sigma(T_{AB}^{-i}\mathcal{G}) = \sigma(\cap_{i=0}^{n} T_{AB}^{-i}\mathcal{G})$$

This implies that $\cap_{i \geq 0} \sigma(T_{AB}^{-i}\mathcal{G}) = \sigma(\cap_{i \geq 0} T_{AB}^{-i}\mathcal{G})$, thus

$$\left(\sigma\left(\mathcal{B}_{A^\mathcal{I}} \times \mathcal{B}_{B^\mathcal{I}}\right)\right)_\infty = \sigma\left((\mathcal{B}_{A^\mathcal{I}})_\infty \times (\mathcal{B}_{B^\mathcal{I}})_\infty\right)$$

Defining $\nu_\infty(x, .)$ as the restriction of $\nu(x, .)$ to the tail σ-field $(\mathcal{B}_{B^\mathcal{I}})_\infty$ for μ_∞-almost $x \in A^\mathcal{I}$, $(\mu\nu)_\infty$ and $\mu_\infty\nu_\infty$ are probabilities defined on the same σ-field $\left(\sigma\left(\mathcal{B}_{A^\mathcal{I}} \times \mathcal{B}_{B^\mathcal{I}}\right)\right)_\infty$ which coincide on the generating semi-algebra $(\mathcal{B}_{A^\mathcal{I}})_\infty \times (\mathcal{B}_{B^\mathcal{I}})_\infty$. They are thus equal. □

3.3. Recurrent channels

The following lemma gives necessary and sufficient conditions for a channel to be recurrent.

Lemma 5. *Let $[A, X]$ be a source with distribution μ and $[A, \nu, B]$ a channel. Let $\mathcal{G} = \{F \times G / F \in \mathcal{B}_{A^\mathcal{I}}, G \in \mathcal{B}_{B^\mathcal{I}}\}$. Then*

- *the dynamical system $(A^\mathcal{I} \times B^\mathcal{I}, \mathcal{B}_{A^\mathcal{I} \times B^\mathcal{I}}, T_{AB}, \mu\nu)$ is recurrent if and only if for any $F \times G \in \mathcal{G}$, $\mu\nu(F \times G \setminus \cup_{k \geq 1} T_{AB}^{-k} F \times G) = 0$.*
- $[A, \nu, B]$ *is recurrent with respect to* μ *if and only if* μ *is recurrent and* $\nu(x, .)$ *is recurrent* μ*-a.e.*

Proof. Assume that for any $F \times G \in \mathcal{G}$, $\mu\nu(F \times G \setminus \cup_{k\geq 1} T_{AB}^{-k}(F \times G)) = 0$. Let $\mathcal{B}'$ be the set $\{O \in \mathcal{B}_{A^{\mathcal{I}} \times B^{\mathcal{I}}} / \mu\nu(O \setminus \cup_{k\geq 1} T_{AB}^{-k} O) = 0\}$. Let $(O_i)_i$ be a countable family of elements of $\mathcal{B}'$. Then

$$\mu\nu((\cup_{i\geq 0} O_i) \setminus (\cup_{k\geq 1} T_{AB}^{-k}(\cup_{i\geq 0} O_i)) = \mu\nu((\cup_{i\geq 0} O_i) \setminus (\cup_{i\geq 0} \cup_{k\geq 1} T_{AB}^{-k} O_i))$$

Given four sets Θ, Λ, Γ and Δ, $(\Theta \cup \Lambda) \setminus (\Gamma \cup \Delta) \subset (\Theta \setminus \Gamma) \cup (\Lambda \setminus \Delta)$, then:

$$\mu\nu((\cup_{i\geq 0} O_i) \setminus (\cup_{k\geq 1} T_{AB}^{-k}(\cup_{i\geq 0} O_i)) \leq \mu\nu(\cup_{i\geq 0}(O_i \setminus \cup_{k\geq 1} T_{AB}^{-k} O_i))$$

$$\mu\nu((\cup_{i\geq 0} O_i) \setminus (\cup_{k\geq 1} T_{AB}^{-k}(\cup_{i\geq 0} O_i)) \leq \sum_{i\geq 0} \mu\nu(O_i \setminus \cup_{k\geq 1} T_{AB}^{-k} O_i) = 0$$

Thus $\cup_{i\geq 0} O_i \in \mathcal{B}'$. Then $\mathcal{B}'$ is stable by countable union.

Let $\mathcal{F}$ be the field generated by $\mathcal{G}$. Any element of the field $\mathcal{F}$ is a finite union of rectangles. Then any countable union of field elements is a countable union of rectangles. Then, from above, for any element R of $\mathcal{F}$, $\mu\nu(R \setminus \cup_{k\geq 1} T_{AB}^{-k} R) = 0$.

Let $O \in \mathcal{B}_{\mathcal{A}^{\mathcal{I}} \times \mathcal{B}^{\mathcal{I}}}$. The probability $\mu\nu$ on $(A^{\mathcal{I}} \times B^{\mathcal{I}}, \mathcal{B}_{\mathcal{A}^{\mathcal{I}} \times \mathcal{B}^{\mathcal{I}}})$ is the extension of the set function $\mu\nu$ on the field generated by the rectangles and verifies ([9]):

$$\mu\nu(O) = \inf_{(R_i)_{i\geq 0} / O \subset \cup_{i\geq 0} R_i} \mu\nu(\cup_{i\geq 0} R_i)$$

where the families $(R_i)_{i\geq 0}$ are countable covers of O made of elements of the field generated by the rectangles (see [9]). Let $\epsilon > 0$, then there exist countable families of field elements $(R_i)_{i\geq 0}$ and $(R'_i)_{i\geq 0}$ respectively covering O and O^c such that:

$$\mu\nu(\cup_{i\geq 0} R_i) - \frac{\epsilon}{2} < \mu\nu(O) \leq \mu\nu(\cup_{i\geq 0} R_i)$$

$$\mu\nu(\cup_{i\geq 0} R'_i) - \frac{\epsilon}{2} < \mu\nu(O^c) \leq \mu\nu(\cup_{i\geq 0} R'_i)$$

Let $\alpha = \cup_{i\geq 0} R_i$ and $\beta = \cup_{i\geq 0} R'_i$. Obviously $T_{AB}^{-k}\beta^c \subset T_{AB}^{-k} O \subset T_{AB}^{-k}\alpha$ for any k. Then:

$$O \setminus \cup_{k\geq 1} T_{AB}^{-k} O \subset \alpha \setminus \cup_{k\geq 1} T_{AB}^{-k}\beta^c \subset (\beta^c \setminus \cup_{k\geq 1} T_{AB}^{-k}\beta^c) \cup (\alpha \setminus \beta^c)$$

β^c is a countable union of elements of the field generated by rectangles then $\mu\nu(\beta^c \setminus \cup_{k\geq 1} T_{AB}^{-k}\beta^c) = 0$. Moreover, $\mu\nu(\alpha \setminus \beta^c) < \epsilon$. Then $\mu\nu(O \setminus \cup_{k\geq 1} T_{AB}^{-k} O) < \epsilon$. This holds for any $\epsilon > 0$, then

$$\mu\nu(O \setminus \cup_{k\geq 1} T_{AB}^{-k} O) = 0$$

Then the dynamical system $(A^{\mathcal{I}} \times B^{\mathcal{I}}, \mathcal{B}_{A^{\mathcal{I}} \times B^{\mathcal{I}}}, T_{AB}, \mu\nu)$ is recurrent. The reciprocal is obvious.

Assume that μ is recurrent and $\nu(x,.)$ is recurrent μ-a.e. Let $F \times G \in \mathcal{G}$.

$$\begin{aligned}
\mu\nu\left(F \times G \setminus \cup_{k\geq 1} T_{AB}^{-k} F \times G\right) &= \mu\nu\left(F \times G \cap \bigcap_{k\geq 1} (T_{AB}^{-k}(F \times G))^c\right) \\
&= \mu\nu\left(F \times G \cap \bigcap_{k\geq 1} (T_A^{-k} F^c \times T_B^{-k} G)\right) \\
&\quad +\mu\nu\left(F \times G \cap \bigcap_{k\geq 1} (T_A^{-k} F \times T_B^{-k} G^c\right) \\
&\quad +\mu\nu\left(F \times G \cap \bigcap_{k\geq 1} (T_A^{-k} F^c \times T_B^{-k} G^c)\right)
\end{aligned}$$

Since μ is recurrent:

$$\mu\nu\left(F \times G \cap \bigcap_{k\geq 1} (T_A^{-k} F^c \times T_B^{-k} G)\right) \leq \mu(F \cap \bigcap_{k\geq 1} T_A^{-k} F^c) = 0$$

Since $\nu(x,.)$ is recurrent μ-a.e. :

$$\begin{aligned}
&\mu\nu\left(F \times G \cap \bigcap_{k\geq 1} (T_A^{-k} F \times T_B^{-k} G^c\right) \leq \\
&\qquad \mu\nu(A^{\mathcal{I}} \times (G \cap_{k\geq 1} T_B^{-k} G^c)) = \int \nu(x, G \cap_{k\geq 1} T_B^{-k} G^c) d\mu = 0
\end{aligned}$$

and

$$\begin{aligned}
&\mu\nu\left(F \times G \cap \bigcap_{k\geq 1} (T_A^{-k} F^c \times T_B^{-k} G^c)\right) = \\
&\qquad \mu\nu((F \cap_{k\geq 1} T_A^{-k} F^c) \times (G \cap_{k\geq 1} T_B^{-k} G^c)) = \\
&\qquad\qquad \int_{F \cap_{k\geq 1} T_A^{-k} F^c} \nu(x, G \cap_{k\geq 1} T_B^{-k} G^c) d\mu = 0
\end{aligned}$$

then $\mu\nu\left(F \times G \setminus \cup_{k\geq 1} T_{AB}^{-k} F \times G\right) = 0$. From above, $[A, \nu, B]$ is recurrent w.r.t. μ. The reciprocal is obvious. □

3.4. AMS Channels

Let μ be the distribution of an AMS source with stationary mean $\overline{\mu}$: $\mu \ll^a \overline{\mu}$. Let ν be an AMS channel. Then $\mu\nu$ is AMS: $\mu\nu \ll^a \overline{\mu\nu}$. The "input marginal" of $\overline{\mu\nu}$ is $\overline{\mu}$, hence there exists a (unique modulo $\overline{\mu}$) channel $\overline{\nu}_\mu$ stationary w.r.t $\overline{\mu}$ such that $\overline{\mu\nu} = \overline{\mu}\ \overline{\nu}_\mu$. The channel $[A, \overline{\nu}_\mu, B]$ such that $\overline{\mu\nu} = \overline{\mu}\ \overline{\nu}_\mu$ is the *stationary mean* of the AMS channel ν with respect to the AMS source distribution μ.

The following lemma shows that the stationary mean of a channel is made of AMS probabilities and that the channel made by the stationary means of these AMS probabilities is stationary. The lemma gives also relationships, based on asymptotic dominance, between the corresponding hookups.

Lemma 6. *Let μ be the distribution of an AMS source and $[A, \nu, B]$ be a channel AMS w.r.t. μ. Let $\overline{\nu}_\mu$ be the stationary mean of ν w.r.t. μ.*

1. *The probability $\overline{\nu}_\mu(x, .)$ is AMS $\overline{\mu}$-a.e.*
2. *Let $S\overline{\nu}_\mu$ be the channel defined by*

$$S\overline{\nu}_\mu(x, G) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \overline{\nu}_\mu(x, T_B^{-i}G)\ \overline{\mu}\text{-}a.e.$$

   $S\overline{\nu}_\mu$ is stationary w.r.t. $\overline{\mu}$ and
   (a) $\mu\nu \ll^a \overline{\mu}\ \overline{\nu}_\mu$
   (b) $\mu\nu \ll^a \overline{\mu} S\overline{\nu}_\mu$
   (c) $\overline{\mu}\ \overline{\nu}_\mu \ll \overline{\mu} S\overline{\nu}_\mu$

Proof of Lemma 6 relies on the following lemma.

Lemma 7. *Let μ be the distribution of a stationary source and $[A, \nu, B]$ a channel stationary w.r.t. μ. Then*

1. *$\nu(x, .)$ is an AMS probability μ-a.e.*
2. *let $S\nu$ be the channel defined by $S\nu(x, G) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \nu(x, T_B^{-i}G)$ when $\nu(x, .)$ is AMS and $S\nu(x, G) = \nu(x, G)$ else. Then $[A, S\nu, B]$ is stationary w.r.t. μ.*
3. *$\mu\nu \ll \mu S\nu$*

Proof. 1. Since ν is stationary w.r.t. μ, by Lemma 3 (p. 13):

$$\forall n \in \mathbb{N}, \frac{1}{n} \sum_{i=0}^{n-1} \nu(x, T_B^{-i}G) = \frac{1}{n} \sum_{i=0}^{n-1} \nu(T_A^i x, G)\ \mu\text{-a.e.}$$

μ is stationary thus, thanks to the pointwise ergodic theorem, $S\nu(x,G) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \nu(T_A^i x, G)$ exists and $S\nu(x,G) = S\nu(T_A x, G)$ μ-a.e. Then $S\nu(x,G) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \nu(x, T_B^{-i} G)$ exists and, by the Vitali-Hahn-Saks theorem, is a (stationary) probability.

2. From 1.):

$$S\nu(T_A x, G) = S\nu(x,G) = S\nu(x, T_B^{-1} G) \ \mu\text{-a.e.}$$

   Then, thanks to Lemma 3 (p. 13), $S\nu$ is stationary w.r.t. μ

3. Let $E \in \left(\sigma\left(\mathcal{B}_{A^{\mathcal{I}}} \times \mathcal{B}_{B^{\mathcal{I}}}\right)\right)_\infty$ such that $(\mu S\nu)_\infty(E) = 0$.
   Then $S\nu(x, E_x)_\infty = 0$ μ-a.e. where $E_x \in (\mathcal{B}_{B^{\mathcal{I}}})_\infty$ is the section of E at x. Since $\nu(x,.) \ll^a S\nu(x,.)$ and since $S\nu(x,.)$ is stationary, by Lemma 1 (p. 9), $\nu(x,.)_\infty \ll S\nu(x,.)_\infty$. Then

$$\nu(x, E_x)_\infty = 0 \ \mu\text{-a.e.}$$

   This implies that $(\mu\nu)_\infty(E) = 0$. Then $(\mu\nu)_\infty \ll (\mu S\nu)_\infty$. From 2.), $\mu S\nu$ is stationary then, by Lemma 1 (p. 9), $\mu\nu \ll^a \mu S\nu$. $\mu\nu$ is stationary then $\mu\nu \ll \mu S\nu$.

□

Proof of Lemma 6. By definition $\overline{\nu}_\mu$ is stationary w.r.t. $\overline{\mu}$. Then, thanks to Lemma 7 (p. 17):

- $\overline{\nu}_\mu(x,.)$ is AMS $\overline{\mu}$-a.e.
- $S\overline{\nu}_\mu$ is stationary w.r.t. $\overline{\mu}$
- $\overline{\mu}\,\overline{\nu}_\mu \ll \overline{\mu} S\overline{\nu}_\mu$

ν is AMS w.r.t. the AMS probability μ thus $\mu\nu \ll^a \overline{\mu}\,\overline{\nu}_\mu$.

Since $\mu\nu \ll^a \overline{\mu}\,\overline{\nu}_\mu$ and $\overline{\mu}\,\overline{\nu}_\mu \ll \overline{\mu} S\overline{\nu}_\mu$:

$$\mu\nu \ll^a \overline{\mu} S\overline{\nu}_\mu$$

□

The following lemma is Lemma 2 of [6]. It also holds for restrictions of sources and channels to the tail σ-fields.

Lemma 8. *Let μ and η be the distributions of two sources on the same alphabet A. Then, for any channel $[A, \nu, B]$*

$$\mu \ll \eta \Rightarrow \mu\nu \ll \eta\nu$$

The following lemma gathers Lemma 3 and Lemma 4 of [6]. It states that "hookup dominance" is equivalent to "channel dominance". It also holds for restrictions of sources and channels to the tail σ-fields.

Lemma 9. *Let $[A, X]$ be a source with distribution μ. Let $[A, \nu, B]$ and $[A, \nu', B]$ be two arbitrary channels. Then*

$$\nu(x, .) \ll \nu'(x, .) \ \mu\text{-a.e.} \ \Leftrightarrow \mu\nu \ll \mu\nu'$$

The following lemma gives a necessary and sufficient condition for a channel to be AMS w.r.t a stationary source.

Lemma 10. *Let $[A, X]$ be a stationary source with distribution μ. A channel $[A, \nu, B]$ is AMS w.r.t. to μ if and only if there exists a channel $\overline{\nu}_\mu$ stationary w.r.t. μ such that $\nu(x, .) \ll^a S\overline{\nu}_\mu(x, .)$ μ-a.e. where $S\overline{\nu}_\mu(x, .)$ is the stationary mean of $\overline{\nu}_\mu(x, .)$.*

Proof. Assume that ν is AMS w.r.t. μ. Then there exists a stationary channel $\overline{\nu}_\mu$ such that $\mu\nu \ll^a \mu\overline{\nu}_\mu$ and (thanks to Lemma 6, p. 17) such that $\mu\nu \ll^a \mu S\overline{\nu}_\mu$. Then, since $\mu S\overline{\nu}_\mu$ is stationary, thanks to Lemma 1 (p. 9), $\mu_\infty\nu_\infty \ll \mu_\infty(S\overline{\nu}_\mu)_\infty$. By Lemma 9 (p. 19), $(\nu(x, .))_\infty \ll (S\overline{\nu}_\mu(x, .))_\infty$ μ-a.e. Since $S\overline{\nu}_\mu(x, .)$ is stationary, by Lemma 1 (p. 9), $\nu(x, .) \ll^a S\overline{\nu}_\mu(x, .)$ μ-a.e.

Assume that $\nu(x, .) \ll^a S\overline{\nu}_\mu(x, .)$ μ-a.e.. Then, by Lemma 1 (p. 9), $(\nu(x, .))_\infty \ll (S\overline{\nu}_\mu(x, .))_\infty$ μ-a.e. By Lemma 9 (p. 19), $\mu_\infty\nu_\infty \ll \mu_\infty(S\overline{\nu}_\mu)_\infty$. Since $\mu S\overline{\nu}_\mu$ is stationary (thanks to Lemma 7, p. 17), by Lemma 1 (p. 9), $\mu\nu \ll^a \mu S\overline{\nu}_\mu$. Then $\mu\nu$ is AMS. □

Lemma 10 can also be stated as:

Lemma 11. *Let $[A, X]$ be a stationary source with distribution μ. A channel $[A, \nu, B]$ is AMS w.r.t. to μ if and only if $\nu(x, .)$ is AMS μ-a.e.*

3.5. Channel product

Given two channels with the same input alphabet, it is possible to define a product of channels as follows.

Definition 10. *Let $[A, \nu_1, B]$ and $[A, \nu_2, C]$ be two channels. The* product *of ν_1 and ν_2 is the channel $[A, \nu_1 \times \nu_2, B \times C]$ defined by, $\forall x \in A^{\mathcal{I}}$, $\forall G \in \mathcal{B}_{B^{\mathcal{I}}}$ and $\forall H \in \mathcal{B}_{C^{\mathcal{I}}}$:*

$$\nu_1 \times \nu_2(x, G \times H) = \nu_1(x, G).\nu_2(x, H)$$

Remark: if $[A, X]$ is an input source to the channel product $[A, \nu_1 \times \nu_2, B \times C]$ and $[B \times C, (Y, Z)]$ the corresponding output source, then $Y \to X \to Z$ is a Markov chain.

The following lemma states that the product of channel restrictions to the tail σ-fields is the restriction of the channel product.

Lemma 12. *Let $[A, \nu_1, B]$ and $[A, \nu_2, C]$ be two channels. Then, for any $x \in A^{\mathcal{I}}$, $(\nu_1 \times \nu_2(x, .))_\infty = (\nu_1(x, .))_\infty \times (\nu_2(x, .))_\infty$ i.e. the restriction of $[A, \nu_1 \times \nu_2, B \times C]$ to the tail σ-field $(\sigma(\mathcal{B}_{B^{\mathcal{I}}} \times \mathcal{B}_{C^{\mathcal{I}}}))_\infty$ is the probability kernel $[A, (\nu_1)_\infty \times (\nu_2)_\infty, B \times C]$:*

$$(\nu_1 \times \nu_2)_\infty = (\nu_1)_\infty \times (\nu_2)_\infty$$

Proof. For any $x \in A^{\mathcal{I}}$, $(\nu_1 \times \nu_2(x, .))_\infty$ and $(\nu_1(x, .))_\infty \times (\nu_2(x, .))_\infty$ coincide on the semi-algebra of rectangles $G \times H \in (\mathcal{B}_{B^{\mathcal{I}}})_\infty \times (\mathcal{B}_{C^{\mathcal{I}}})_\infty$. Then they coincide on $\sigma((\mathcal{B}_{B^{\mathcal{I}}})_\infty \times (\mathcal{B}_{C^{\mathcal{I}}})_\infty)$.

Thanks to Lemma 4 (p. 14), $\sigma((\mathcal{B}_{B^{\mathcal{I}}})_\infty \times (\mathcal{B}_{C^{\mathcal{I}}})_\infty) = (\sigma(\mathcal{B}_{B^{\mathcal{I}}} \times \mathcal{B}_{C^{\mathcal{I}}}))_\infty$ □

Part II
Capacity of a noisy computation

4. Model for noisy computations

The model of noisy computation proposed in this section is not linked to a noisy version of a peculiar formal computational model (e.g., boolean circuits made of noisy boolean gates, [16]). Functions considered here are "general" measurable functions from a sequence space to a sequence space and their noisy realizations are viewed as random channels. Section 11 is devoted to the question of the instantiation of the noisy computation model to formal computational models and to their noisy versions.

The input and output sequences of a computation will be modeled as random processes with values in sequence spaces built from measurable alphabets. Since the purpose is to handle computations which could be processes evolving from an initial state, the time shifts will not be assumed invertible. The processes considered are thus one-sided. The alphabets $(A, \mathcal{B}_A)$, $(B, \mathcal{B}_B)$ and $(C, \mathcal{B}_C)$, considered below, are assumed to be standard measurable spaces. Let $(A^{\mathcal{I}}, \mathcal{B}_{A^{\mathcal{I}}})$, $(B^{\mathcal{I}}, \mathcal{B}_{B^{\mathcal{I}}})$ and $(C^{\mathcal{I}}, \mathcal{B}_{C^{\mathcal{I}}})$ be the corresponding measurable sequence spaces. Let $[A, X]$ be a source.

A noisy computing device is modeled as a noisy channel $[A, \nu, C]$: given an input sequence x, the output sequence z is not deterministically determined but, due to possible errors, randomly produced. The noisy computing device takes X as input and produces as an output a random process Z on $(C^{\mathcal{I}}, \mathcal{B}_{C^{\mathcal{I}}})$. The hookup $P_{XZ} = P_X\nu$ represents the actual noisy calculation of the device operating on a given input flow represented by X.

In the sequel, the noisy channel/computing device will be denoted by F and, with an abuse of notation, the output Z by $F(X)$. Thus $\nu(x, .)$ will be denoted by $P_{F(X)|X}(.|x)$.

The perfect computation is represented by a $(\mathcal{B}_{A^{\mathcal{I}}}, \mathcal{B}_{B^{\mathcal{I}}})$-measurable function $f : A^{\mathcal{I}} \to B^{\mathcal{I}}$ (the expected function). f defines a deterministic channel $\left\{P_{f(X)|X}(.|x), x \in A^{\mathcal{I}}\right\}$ such that, $\forall G \in \mathcal{B}_{B^{\mathcal{I}}}$, $P_{f(X)|X}(G|x) = 1_{f^{-1}(G)}(x)$ P_X-a.e. ([10]) where 1_E denotes the characteristic function of a set E .

Since f is a function, $F(X) \to X \to f(X)$ is a Markov chain. Then the measurable functions $x \mapsto P_{f(X)F(X)}(E|x)$, $E \in \mathcal{B}_{B^{\mathcal{I}} \times C^{\mathcal{I}}}$, are such that, $\forall G \in \mathcal{B}_{B^{\mathcal{I}}}$ and $\forall H \in \mathcal{B}_{C^{\mathcal{I}}}$:

$$P_{f(X)F(X)|X}(G \times H|x) = P_{f(X)|X}(G|x).P_{F(X)|X}(H|x)$$

The collection of probabilities $\left\{P_{f(X)F(X)}(.|x), x \in A^{\mathcal{I}}\right\}$ defines a channel product which will be denoted $[A, f \times F, B \times C]$ and called the noisy computation of f by F.

$$\begin{tikzcd}
\begin{array}{c}\text{Input source}\\ \left(A^{\mathcal{I}},\mathcal{B}_{A^{\mathcal{I}}},P_X\right)\end{array} \arrow[r, "F"] \arrow[dr, dotted, "f"] & \begin{array}{c}\text{Noisy result}\\ \left(C^{\mathcal{I}},\mathcal{B}_{C^{\mathcal{I}}},P_{F(X)}\right)\end{array} \\
 & \begin{array}{c}\text{Expected result}\\ \left(B^{\mathcal{I}},\mathcal{B}_{B^{\mathcal{I}}},P_{f(X)}\right)\end{array}
\end{tikzcd}$$

Figure 1: Model for Noisy Computation

Definition 11. *Let $\mathcal{I}$ be a countable index set. Let $(A^{\mathcal{I}},\mathcal{B}_{A^{\mathcal{I}}})$, $(B^{\mathcal{I}},\mathcal{B}_{B^{\mathcal{I}}})$ and $(C^{\mathcal{I}},\mathcal{B}_{C^{\mathcal{I}}})$ be standard sequence spaces. Let $f : A^{\mathcal{I}} \to B^{\mathcal{I}}$ be a $(\mathcal{B}_{A^{\mathcal{I}}},\mathcal{B}_{B^{\mathcal{I}}})$-measurable function and $[A, F, C]$ a channel. The channel product $[A, f \times F, B \times C]$ is called the* noisy computation *of f by F. $[[A, X]; [A, f \times F, B \times C]]$ is called the* hookup *of the source $[A, X]$ and of the noisy computation $[A, f \times F, B \times C]$. It determines the process $(X, f(X), F(X))$.*

Asymptotic mean stationarity (or stationarity) and ergodicity of processes are key properties for the Asymptotic Equipartition Property (AEP) to hold ([8]). Thus coding theorems generally rely on these properties. Asymptotic mean stationarity and ergodicity will be also fundamental properties for noisy computations.

Definition 12. *A noisy computation $[A, f \times F, B \times C]$ is stationary (respectively AMS) if, for any stationary (respectively AMS) source $[A, X]$, the processes $(X, f(X))$, $(X, F(X))$ and $(f(X), F(X))$ are stationary (respectively AMS).*

Definition 13. *A stationary (resp. AMS) noisy computation $[A, f \times F, B \times C]$ is ergodic if, for any stationary (resp. AMS) and ergodic source $[A, X]$, the processes $(X, f(X))$, $(X, F(X))$ and $(f(X), F(X))$ are ergodic.*

The following lemma is a straightforward application of the conditional AEP to the special case where $Y = f(X)$.

Lemma 13. *Let A and B be two finite alphabets. Let $f : A^{\mathcal{I}} \to B^{\mathcal{I}}$ be a $(\mathcal{B}_{A^{\mathcal{I}}},\mathcal{B}_{B^{\mathcal{I}}})$-measurable function such that $[A, f, B]$ is an ergodic AMS channel. Let $[A, X]$ be an ergodic AMS input source and $[B, Y]$ the corresponding* *output source. For $\epsilon > 0$ and $y \in B^{\mathcal{I}}$, let $A^n(\epsilon, y^n)$ be the set of ϵ-typical sequences given y^n such that:*

$$A^n(\epsilon, y^n) = \left\{ x^n \in (f^{-1}(y))^n / \left| \frac{-log(P_{X^n|Y^n}(x^n|y^n))}{n} - \overline{H}(X^n|Y^n) \right| < \epsilon \right\}$$

where $(f^{-1}(y))^n$ is the set of prefixes (of length n) of inverse images of y by f.

Then:

1. $\lim_{n\to\infty} P_{X^n|Y^n}(A^n(\epsilon, y^n)|y^n) = 1 \; P_Y - a.e.$
2. *for any $\delta > 0$ and any $\epsilon > 0$, for n large enough:*

$$(1-\delta)e^{n(\overline{H}(X|Y)-\epsilon)} \leq |A^n(\epsilon, y^n)| \leq e^{n(\overline{H}(X|Y)+\epsilon)} \; P_Y^n - a.e.$$

Proof. The lemma follows immediately from the conditional AEP and from the fact that

$$\lim_{n\to\infty} P_{X^n|Y^n}(\{x^n \in (f^{-1}(y))^n\}|y^n) = 1$$

□

It should be noticed that $\overline{H}(f(X)|X) = 0$ and then

$$\overline{H}(X) = \overline{H}(X|f(X)) + \overline{H}(f(X))$$

and

$$\overline{I}(X, f(X)) = \overline{H}(f(X))$$

Let f be a measurable function such that $f^n : x^n \mapsto (f(x))^n$ is a $(\mathcal{B}_{A^n}, \mathcal{B}_{B^n})$-measurable function for any large n. This means that $\forall x, x' \in A^{\mathcal{I}}$ such that $x^n = x'^n$, $(f(x))^n = (f(x'))^n$. In other words $(f^{-1}(y))^n = (f^n)^{-1}(y^n)$, for large n. Such functions will be called weakly causal functions.

Definition 14. *A $(\mathcal{B}_{A^{\mathcal{I}}}, \mathcal{B}_{B^{\mathcal{I}}})$-measurable function $f : A^{\mathcal{I}} \to B^{\mathcal{I}}$ is weakly causal if there exists $N_f \in \mathbb{N}$ such that $\forall n \geq N_f$, $\forall x, x' \in A^{\mathcal{I}}$ such that $x^n = x'^n$, then $(f(x))^n = (f(x'))^n$ and $f^n : x^n \mapsto (f(x))^n$ is a $(\mathcal{B}_{A^n}, \mathcal{B}_{B^n})$-measurable function.*

5. Typical input Rate of a Hookup and Typical Input Capacity of a Noisy Computation

In this section, the definition of the typical input rate of a hookup and the definition of the typical input capacity of a noisy computation are given. In a next section, this rate will be shown to characterize the rate at which a source should produce typical inputs for a random channel in order to allow decodability of the output. Decodability means ability to recover the desired function results by decoding. Assumptions and notations used throughout this section are:

- the alphabets A, B and C are assumed finite, the σ-fields $\mathcal{B}_A$, $\mathcal{B}_B$ and $\mathcal{B}_C$ are the sets of subsets of the alphabets A, B and C
- the noisy computation $[A, f \times F, B \times C]$ and the source $[A, X]$ are assumed AMS and ergodic
- the function f is assumed weakly causal.

Defining a notion of capacity for noisy computations, following the classical approach for communication channels, boils down to characterizing the maximum number of *input* n-sequences that can be selected in order to allow an asymptotically perfect correction/decoding process. The definition of capacity proposed below can be informally justified as follows.

Assume that only sequences x^n conditionally typical given y^n, typical sequence of $f^n(X^n)$, are allowed as inputs (n large). Assume also that only typical sequences z^n of $F^n(X^n)$ which are conditionally typical given any allowed input x^n (x^n conditionally typical given a typical y^n) are retained for decoding, the others are simply ignored. Then such a z^n is conditionally typical given y^n.

Hence, the set of possible output values given $x^n \in (f^n)^{-1}(y^n)$ submitted to decoding has "almost" $e^{n\overline{H}(F(X)|f(X))}$ elements (thanks to the conditional AEP). Thus the number of non-overlapping such sets built from the set of $e^{n\overline{H}(F(X))}$ typical sequences z^n is at most $e^{n\overline{H}(F(X))}/e^{n(\overline{H}(F(X)|f(X)))}$.

Each inverse image $(f^n)^{-1}(y^n)$ offers $e^{n\overline{H}(X|f(X))}$ different possible input n-sequences. So among the $e^{n\overline{H}(X)}$ possible input n sequences, at most $e^{n(\overline{H}(X|f(X))+\overline{H}(F(X))-\overline{H}(F(X)|f(X)))}$ of them can be used in order to reach unambiguous decodability.

Since $\overline{H}(F(X)) - \overline{H}(F(X)|f(X))) = \overline{I}(F(X), f(X))$, this leads to define the capacity of the noisy computation $[A, f \times F, B \times C]$ as:

$$C_f(F) = \sup_X [\overline{H}(X|f(X)) + \overline{I}(F(X), f(X))]$$

$\overline{I}(F(X), f(X)) = \overline{H}(f(X)) - \overline{H}(f(X)|F(X)))$ and since f is a function, $\overline{H}(X|f(X)) + \overline{H}(f(X)) = \overline{H}(X)$. Then:

$$C_f(F) = \sup_X \left(\overline{H}(X) - \overline{H}(f(X)|F(X))\right)$$

Definition 15 (Typical input rate)**.** *Let $[A, f \times F, B \times C]$ be an AMS and ergodic noisy computation on finite alphabets and let $[A, X]$ be an AMS and ergodic source. Assume that f is a $(\mathcal{B}_{A^\mathcal{I}}, \mathcal{B}_{B^\mathcal{I}})$-measurable weakly causal function. $\forall n \geq N_f$, the n^{th}* order typical input rate *of the hookup $[[A, X]; [A, f \times F, B \times C]]$ is:*

$$B^n(X^n, f^n, F^n) = \frac{1}{n}\left[H(X^n) - H(f^n(X^n)|F^n(X^n))\right]$$

The typical input rate of the hookup $[[A, X]; [A, f \times F, B \times C]]$ is:

$$\overline{B}(X, f, F) = \overline{H}(X) - \overline{H}(f(X)|F(X))$$

Let $\mathfrak{P}_{ae}(A^\mathcal{I})$ denote the set of ergodic and AMS probability measures on $(A^\mathcal{I}, \mathcal{B}_{A^\mathcal{I}})$.

Definition 16 (Typical input capacity)**.** *Let $[A, f \times F, B \times C]$ be an AMS and ergodic noisy computation on finite alphabets. Assume that f is a $(\mathcal{B}_{A^\mathcal{I}}, \mathcal{B}_{B^\mathcal{I}})$-measurable weakly causal function. The typical input capacity (for AMS and ergodic sources) of the noisy computation $[A, f \times F, B \times C]$ is*

$$C_f(F) = \sup_{P_X \in \mathfrak{P}_{ae}(A^\mathcal{I})} \overline{B}(X, f, F)$$

The equivalent expression

$$C_f(F) = \sup_{P_X \in \mathfrak{P}_{ae}(A^\mathcal{I})} \left(\overline{H}(X) - \overline{H}(f(X)|F(X))\right)$$

immediately shows that this capacity boils down to the usual channel capacity when f is a bijection (in which case $\overline{H}(f(X)|F(X)) = \overline{H}(X|F(X))$.

The adjective typical is motivated by the forthcoming joint source-channel coding theorem: the encoding process will be constrained to encode typical sequences of a source by typical sequences of a source "achieving" capacity. This constraint will be discussed in section 7.

f being a function, $\overline{H}(f(X)|F(X)) \leq \overline{H}(X|F(X))$ thus $C_f(F)$ is greater than the *channel capacity* of the noisy function F (considered as a *communication channel*). This is totally consistent with the intuitive interpretation of equivocation: as computation "burns" information (i.e. $\overline{H}(f(X)) \leq \overline{H}(X)$), the amount of information that must be added to a noisy result to retrieve the input of the function is likely to be larger than the amount of information that must added to a noisy result to obtain the correct one.

From the definition of $C_f(F)$, the limit cases can be briefly discussed:

- if the computation by F is perfect almost surely (i.e. there is a deterministic way of retrieving $f(x)$ from $F(x)$ almost everywhere) then $\overline{H}(f(X)|F(X)) = 0$ and thus $C_f(F) = \sup_{P_X \in \mathfrak{P}_{ae}(A^{\mathcal{I}})} \overline{H}(X)$. In other words, almost any possible input sequence allows decodability

- if the computation is totally noisy meaning that the noisy output $F(X)$ and the expected output $f(X)$ are independent for any X, then $\overline{H}(f(X)|F(X)) = \overline{H}(f(X))$ and $C_f(F) = \sup_{P_X \in \mathfrak{P}_{ae}(A^{\mathcal{I}})} \overline{H}(X|f(X))$; the joint source-noisy computation coding theorem will show that this means that to allow decodability it is necessary to restrict inputs to sequences belonging to a unique $f^{-1}(y_0)$; only one output sequence y_0 can be reliably computed i.e. only a constant function can be reliably computed with such a totally noisy apparatus

- if $C_f(F) = 0$ then, for any X, $\overline{H}(X|f(X)) + \overline{I}(f(X); F(X)) = 0$; this implies that both $\overline{H}(X|f(X)) = 0$ and $\overline{I}(f(X); F(X)) = 0$; the latter condition is equivalent to the preceding case (totally noisy computation) and the former condition means that f is injective almost everywhere.

6. Extension of Feinstein Theorem to Noisy Computations

In this section, extensions of Feinstein lemma and theorem covering noisy computations are given. Based on the model given above, the existence of Feinstein codes for noisy computations is derived from classical results dealing with communication channels. While a Feinstein code for a channel is a set of pairs (x_i, Γ_i) where x_i is an input sequence and Γ_i a set of output sequences (the decodability region), a Feinstein code for a noisy computation will be defined as a set of pairs (A_i, Γ_i) where A_i is a set of input sequences and Γ_i the associated decodability region.

6.1. Feinstein lemma for noisy computations

If B is countable, for any $y \in B^{\mathcal{I}}$, the set $\{y\}$ is measurable (countable intersection of cylinders $c(y_i), i = 1, \cdots$ if $y = y_1 y_2 \ldots$) and $f^{-1}(y)$ is measurable as f is measurable. This justifies the assumptions on the alphabets made by the following lemmas and corollaries.

Feinstein lemma for noisy computation will be stated in a form consistent with the one of Feinstein lemma for communication channels given in [10] (Lemma 14.1) and assumptions close to [10] : alphabets A and C are standard, B is countable and standard.

In the framework of the noisy computation model, from now on, the joint probabilities $P_{XF(X)}$ and $P_{f(X)F(X)}$ are assumed to be dominated by (i.e., absolutely continuous w.r.t) the corresponding product probabilities: $P_{XF(X)} \ll P_X \times P_{F(X)}$ and $P_{f(X)F(X)} \ll P_{f(X)} \times P_{F(X)}$.

Lemma 14 (Feinstein lemma for noisy computation)**.** *Let A and C be standard alphabets and B be a countable standard alphabet. Let $[A, f \times F, B \times C]$ be a noisy computation. Let $[A, X]$ be a source. Assume that $P_{f(X)F(X)} \ll P_{f(X)} \times P_{F(X)}$. Let $\psi_{f(X)F(X)}$ be the Radon-Nikodym derivative $\frac{dP_{f(X)F(X)}}{d(P_{f(X)} \times P_{F(X)})}$ and $i_{f(X)F(X)} = ln(\psi_{f(X)F(X)})$. Then $\forall M \in \mathbb{N}$, $\forall a > 0$ such that*

$$Me^{-a} + P_{f(X)F(X)}(i_{f(X)F(X)} \leq a) < \frac{1}{4}$$

there exist a set $\left\{y_k \in B^{\mathcal{I}}, k = 1 \cdots M\right\}$, a collection of measurable disjoint sets $\Gamma_k, k = 1, ..., M$ members of $\mathcal{B}_{C^{\mathcal{I}}}$ and a collection of measurable sets $A_k \subset f^{-1}(y_k), k = 1, \cdots, M$ members of $\mathcal{B}_{A^{\mathcal{I}}}$ such that:

$$\forall k = 1, \cdots, M \;\; P_{X|f(X)}(A_k|y_k) > 1 - \lambda$$

and

$$\forall x \in A_k \ P_{F(X)|X}\left(\Gamma_k^c|x\right) \leq \epsilon$$

for any $\epsilon \in]0, \frac{1}{2}[$ *and* $\lambda \in]0, \frac{1}{2}[$ *such that* $\epsilon\lambda = Me^{-a} + P_{f(X)F(X)}(i_{f(X)F(X)} \leq a)$

Before proving the extension of Feinstein lemma for noisy computation, the following lemma is needed. Its proof is given in Appendix A:

Lemma 15. *Let A and C be standard alphabets and B be a countable standard alphabet. Let $[A, f \times F, B \times C]$ be a noisy computation. Let $[A, X]$ be a source. Let $\epsilon > 0$ and $\lambda > 0$. Let $\tilde{B}$ be a measurable subset of $B^{\mathcal{I}}$. Assume there exist a set $\left\{y_k \in \tilde{B}, k = 1 \cdots M\right\}$, a collection of measurable disjoint sets $\Gamma_k, k = 1, ..., M$ members of $\mathcal{B}_{C^{\mathcal{I}}}$ and a collection of measurable sets $A_k \subset f^{-1}(y_k), k = 1, \cdots, M$ members of $\mathcal{B}_{A^{\mathcal{I}}}$ such that:*

$$\forall k = 1, \cdots, M \ P_{X|f(X)}(A_k|y_k) > 1 - \lambda \text{ and } \forall x \in A_k \ P_{F(X)|X}\left(\Gamma_k^c|x\right) \leq \epsilon$$

If the set $\left\{y_k \in \tilde{B}, k = 1 \cdots M\right\}$ is maximal, meaning there do not exist any other $y_{M+1} \in \tilde{B}$, Γ_{M+1} and A_{M+1} complying with the given properties, then:

1. $\forall y_0 \in \tilde{B} \setminus \{y_1, \cdots, y_M\}$, $P_{F(X)|f(X)}\left(\bigcup_{k=1}^{M} \Gamma_k | y_0\right) > \lambda\epsilon$
2. $P_{F(X)}\left(\bigcup_{k=1}^{M} \Gamma_k\right) > min((1-\lambda)(1-\epsilon), \lambda\epsilon).P_{f(X)}(\tilde{B})$
3. $\tilde{B} \neq \emptyset \Rightarrow M \geq 1$

Proof of Lemma 14. This proof is derived from the proof of Feinstein lemma given in [10] (Lemma 14.1).

Let $a > 0$, M a positive integer, $\epsilon \in]0, \frac{1}{2}[$ and $\lambda = \in]0, \frac{1}{2}[$ be such that

$$\epsilon\lambda = Me^{-a} + P_{f(X)F(X)}(i_{f(X)F(X)} \leq a)$$

Let $G = \{(y, z) \in B^{\mathcal{I}} \times C^{\mathcal{I}} / i_{f(X)F(X)}(y, z) > a\}$. $P_{f(X)F(X)}(G^c) \leq \epsilon\lambda < 1$ then $P_{f(X)F(X)}(G) \geq 1 - \epsilon\lambda > 0$.

Pick up a $y_1 \in B^{\mathcal{I}}$ and determine a $\mathcal{B}_{\mathcal{A}^{\mathcal{I}}}$-measurable $A_1 \subset f^{-1}(y_1)$ and a $\mathcal{B}_{\mathcal{C}^{\mathcal{I}}}$-measurable Γ_1 such that:

$$P_{X|f(X)}(A_1|y_1) > 1 - \lambda \text{ and } \forall x \in A_1 \ P_{F(X)|X}\left(\Gamma_1^c|x\right) \leq \epsilon$$

Such a (y_1, A_1, Γ_1) exists, for instance $(y_1, f^{-1}(y_1), C^{\mathcal{I}})$. Recursively, pick up (if any) a y_i and determine a $\mathcal{B}_{\mathcal{A}^{\mathcal{I}}}$-measurable $A_i \subset f^{-1}(y_i)$ and a $\mathcal{B}_{\mathcal{C}^{\mathcal{I}}}$-measurable Γ_i such that:

$$\begin{aligned} P_{X|f(X)}(A_i|y_i) &> 1-\lambda \\ \text{and } \forall x \in A_i \; P_{F(X)|X}\left(\Gamma_i^c|x\right) &\leq \epsilon \\ \text{and } \forall j < i, \Gamma_i \bigcap \Gamma_j &= \emptyset \end{aligned}$$

until $B^{\mathcal{I}}$ is exhausted. Let n be the largest possible number of selected y_i and assume that $n < M$.

The selected set $\{(y_k, A_k, \Gamma_k), k = 1, \cdots, n\}$ is maximal. By Lemma 15:

$$\begin{aligned} \forall y \notin \{y_1, \cdots, y_n\}, P_{F(X)|f(X)}\left(\bigcup_{k=1}^{M} \Gamma_k|y\right) &> \lambda\epsilon \\ \Rightarrow P_{F(X)|f(X)}\left((\bigcup_{k=1}^{M} \Gamma_k)^c|y\right) < 1-\lambda\epsilon \end{aligned} \quad (1)$$

$f(X) \to X \to F(X)$ is a Markov Chain, then $\forall k = 1, \cdots, n$:

$$\begin{aligned} P_{F(X)|f(X)}(\Gamma_k|y_k) &= \int P_{F(X)|X}(\Gamma_k|x) dP_{X|f(X)}(x|y_k) \\ &\geq \int_{A_k} P_{F(X)|X}(\Gamma_k|x) dP_{X|f(X)}(x|y_k) > (1-\lambda)(1-\epsilon) \end{aligned}$$

which implies $P_{F(X)|f(X)}\left((\bigcup_{k=1}^{M} \Gamma_k)|y_k\right) > (1-\lambda)(1-\epsilon)$ and thus

$$P_{F(X)|f(X)}\left((\bigcup_{k=1}^{M} \Gamma_k)^c|y_k\right) \leq 1-(1-\lambda)(1-\epsilon) \quad (2)$$

$P_{f(X)F(X)}(G) = \Delta + \Theta$ where $\Delta = P_{f(X)F(X)}\left(G \bigcap (B^{\mathcal{I}} \times \bigcup_{k=1}^{M} \Gamma_k)\right)$ and $\Theta = P_{f(X)F(X)}\left(G \bigcap (B^{\mathcal{I}} \times \bigcup_{k=1}^{M} \Gamma_k)^c\right)$.

It holds that:

$$\Theta \leq P_{f(X)F(X)}\left(B^{\mathcal{I}} \times \bigcup_{k=1}^{M} \Gamma_k)^c\right) = \int P_{F(X)|f(X)}\left((\bigcup_{k=1}^{M} \Gamma_k)^c|y\right) dP_{f(X)}$$

Thus

$$\Theta \leq \int_{\{y_1,\cdots,y_n\}} P_{F(X)|f(X)}\left((\bigcup_{k=1}^{M}\Gamma_k)^c|y\right)dP_{f(X)}+$$
$$\int_{\{y_1,\cdots,y_n\}^c} P_{F(X)|f(X)}\left((\bigcup_{k=1}^{M}\Gamma_k)^c|y\right)dP_{f(X)}$$

By (1), (2) and the fact that $\epsilon \in]0,\frac{1}{2}[$ and $\lambda \in]0,\frac{1}{2}[$:

$$\Theta < \int_{\{y_1,\cdots,y_n\}}(1-(1-\epsilon)(1-\lambda))dP_{f(X)} + \int_{\{y_1,\cdots,y_n\}^c}(1-\epsilon\lambda)dP_{f(X)}$$

$$\Rightarrow \Theta < 1-\epsilon\lambda \tag{3}$$

$$\Delta = P_{f(X)F(X)}\left(G\bigcap(B^{\mathcal{I}}\times\bigcup_{k=1}^{M}\Gamma_k)\right) \leq P_{f(X)F(X)}(G) = \int_G dP_{f(X)F(X)}$$

Since, on G, $i_{f(X)F(X)}(y,z) > a$ or equivalently $\frac{\psi_{f(X)F(X)}}{e^a} > 1$:

$$\Delta \leq \int_G \frac{\psi_{f(X)F(X)}(y,z)}{e^a}dP_{f(X)F(X)} \leq e^{-a}\int \psi_{f(X)F(X)}dP_{f(X)F(X)}$$

But $\Delta = \int \Delta d(P_{f(X)}\times P_{F(X)})$ thus, since $\psi_{f(X)F(X)} = \frac{dP_{f(X)F(X)}}{d(P_{f(X)}\times P_{F(X)})}$:

$$\Delta \leq e^{-a}\int\int \psi_{f(X)F(X)}dP_{f(X)F(X)}d(P_{f(X)}\times P_{F(X)})$$
$$\leq e^{-a}\int\int dP_{f(X)F(X)}dP_{f(X)F(X)} = e^{-a}$$

$$\Rightarrow \Delta \leq n.e^{-a} \tag{4}$$

From (3) and (4), $P_{f(X)F(X)}(G) = \Delta + \Theta < 1-\epsilon\lambda + ne^{-a}$. But $\epsilon\lambda = Me^{-a} + P_{f(X)F(X)}(G^c) = Me^{-a} + 1 - P_{f(X)F(X)}(G)$ thus $P_{f(X)F(X)}(G) = 1-\epsilon\lambda+Me^{-a}$. This leads to $n > M$ contradicting the assumption $n < M$. Then $n = M$. □

Remark: since $Me^{-a} < 1$, M is necessarily strictly lower than e^a.

6.2. Decomposable modules

In the peculiar case where $X \to f(X) \to F(X)$ is a Markov chain (decomposable modules defined in [26] fulfill this assumption), a straightforward lemma covering noisy computations can be directly derived from the generalized Feinstein lemma, [10] (Lemma 14.1, page 362) or [13]. This assumption allows to decompose the channel $[A, F, C]$ as a cascade of the channels $[A, f, B]$ and $[B, f_X^{-1}F, C]$ where $[B, f_X^{-1}, A]$ is the reverse channel linking $f(X)$ to X for a fixed source $[A, X]$.

$$\begin{tikzcd}
\begin{matrix}\text{Input source}\\ \left(A^{\mathcal{I}}, \mathcal{B}_{A^{\mathcal{I}}}, P_X\right)\end{matrix} \arrow[r, "F"] \arrow[rd, dotted, shift left, "f"] & \begin{matrix}\text{Noisy result}\\ \left(C^{\mathcal{I}}, \mathcal{B}_{C^{\mathcal{I}}}, P_{F(X)}\right)\end{matrix} \\
& \begin{matrix}\text{Expected result}\\ \left(B^{\mathcal{I}}, \mathcal{B}_{B^{\mathcal{I}}}, P_{f(X)}\right)\end{matrix} \arrow[lu, dotted, shift left, "f_X^{-1}"] \arrow[u, dotted, "f_X^{-1}F"]
\end{tikzcd}$$

Figure 2: Model for Winograd and Cowan 's Decomposable Module

Lemma 16. *Let A and C be standard alphabets and B be a countable standard alphabet. Let $[A, f \times F, B \times C]$ be a noisy computation and $[A, X]$ a source such that $X \to f(X) \to F(X)$ is a Markov Chain. Let $\psi_{f(X)F(X)}$ be the Radon-Nikodym derivative $\frac{dP_{f(X)F(X)}}{d(P_{f(X)} \times P_{F(X)})}$ and $i_{f(X)F(X)} = ln(\psi_{f(X)F(X)})$. Then $\forall M \in \mathbb{N}$, $\forall a > 0$, there exist a set $\left\{y_k \in B^{\mathcal{I}}, k = 1 \cdots M\right\}$ and a collection of measurable disjoint sets $\Gamma_k, k = 1, ..., M$ members of $\mathcal{B}_{C^{\mathcal{I}}}$ such that $\forall k = 1, \cdots, M$, $\forall x \in f^{-1}(y_k)$:*

$$P_{F(X)|X}\left(\Gamma_k^c | x\right) \leq Me^{-a} + P_{f(X)F(X)}(i_{f(X)F(X)} \leq a)$$

Proof. f is a function then $f(X) \to X \to F(X)$ is a Markov chain so the channel $f(X) \to F(X)$ is a cascade of the channels $f(X) \to X$, denoted $[B, f_X^{-1}, A]$, and $X \to F(X)$ i.e. $[A, F, C]$. This cascade is the channel $[B, f_X^{-1}F, C]$.

Let $M \in \mathbb{N}$ and $a > 0$. By Feinstein lemma ([13]) applied to the channel $[B, f_X^{-1}F, C]$, there exist a set $\left\{y_k \in B^{\mathcal{I}}, k = 1 \cdots M\right\}$ and a collection of measurable disjoint sets $\Gamma_k, k = 1, ..., M$ members of $\mathcal{B}_{C^{\mathcal{I}}}$ such that, $\forall k = 1, \cdots, M$:

$$P_{F(X)|f(X)}\left(\Gamma_k^c | y_k\right) \leq Me^{-a} + P_{f(X)F(X)}(i_{f(X)F(X)} \leq a)$$

By assumption $X \to f(X) \to F(X)$ is also a Markov chain so the channel F is also a cascade of the channels ($[A, f, B]$ and $[B, f_X^{-1}F, C]$). Then:

$$\forall H \in \mathcal{B}_{C^{\mathcal{I}}} \; P_{F(X)|X}(H|x) = \int_{B^{\mathcal{I}}} P_{F(X)|f(X)}(H|y) dP_{f(X)|X}(y|x)$$

Thus, for any $k = 1, \cdots, M$:

$$P_{F(X)|X}(\Gamma_k^c|x) = \int_{B^{\mathcal{I}}} P_{F(X)|f(X)}(\Gamma_k^c|y) dP_{f(X)|X}(y|x)$$

$\{y_k\}$ is measurable then:

$$\begin{aligned} P_{F(X)|X}(\Gamma_k^c|x) = & \int_{\{y_k\}} P_{F(X)|f(X)}(\Gamma_k^c|y) dP_{f(X)|X}(y|x) \\ & + \int_{B^{\mathcal{I}} - \{y_k\}} P_{F(X)|f(X)}(\Gamma_k^c|y) dP_{f(X)|X}(y|x) \end{aligned}$$

$$\begin{aligned} & P_{F(X)|X}(\Gamma_k^c|x) \leq \\ & \quad P_{f(X)|X}\left(\{y_k\}|x\right) . \left(Me^{-a} + P_{f(X)F(X)}(i_{f(X)F(X)} \leq a)\right) \\ & \quad\quad + P_{f(X)|X}\left(B^{\mathcal{I}} - \{y_k\}|x\right) \end{aligned}$$

If $x \in f^{-1}(y_k)$ then $P_{f(X)|X}\left(\{y_k\}|x\right) = 1$ and $P_{f(X)|X}\left(B^{\mathcal{I}} - \{y_k\}|x\right) = 0$, hence $\forall x \in f^{-1}(y_k)$:

$$P_{F(X)|X}\left(\Gamma_k^c|x_k\right) \leq Me^{-a} + P_{f(X)F(X)}(i_{f(X)F(X)} \leq a)$$

□

6.3. Feinstein codes and Extension of Feinstein theorem to noisy computations

The alphabets A, B and C are now assumed *finite* and the noisy computation $[A, f \times F, B \times C]$ AMS and ergodic. This implies that the input rate $\overline{B}(X, f, F)$ is well defined for any AMS and ergodic source X. The assumption about alphabets is consistent, for example, with the case of digital computation i.e. if F represents a unreliable digital computer and f a computable function.

Definition 17 (Feinstein code for noisy computation)**.** *Let A, B and C be finite alphabets. Let $[A, f \times F, B \times C]$ be a noisy computation such that f is a measurable weakly causal function. Let $n \geq N_f$.*

A $[M, n, \epsilon, \lambda]$-Feinstein code for the hookup $[[A, X^n]; [A, f^n \times F^n, B \times C]]$ is a set $\{(A_i^n, \Gamma_i^n) \in \mathcal{B}_{A^n} \times \mathcal{B}_{C^n}, i = 1, \cdots, M\}$ such that:

1. *$\Gamma_i \cap \Gamma_j = \emptyset$, $i \neq j$*
2. *$\forall i = 1, \cdots, M$, $P_{F^n(X^n)|X^n}(\Gamma_i^c|x^n) \leq \epsilon$ for any $x^n \in A_i^n$*
3. *$\forall i = 1, \cdots, M$, there exists $y_i \in B^n$ such that $A_i^n \subset (f^n)^{-1}(y_i)$ and $P_{X^n|f^n(X^n)}(A_i^n|y_i) > 1 - \lambda$*

F^n (ϵ, λ)-reliably computes f^n on the code $\{(A_i^n, \Gamma_i^n), i = 1, \cdots, M\}$ for the n^{th} extension of the source $[A, X]$. $\{A_i^n, i = 1, \cdots, M\}$ is the input code and $\{\Gamma_i^n, i = 1, \cdots, M\}$ the output code.

Proposition 1 (Extension of Feinstein theorem to Noisy Computation)**.** *Let A, B and C be finite alphabets. Let $[A, f \times F, B \times C]$ be an AMS and ergodic noisy computation such that f is a measurable weakly causal function. Let $[A, X]$ be an AMS and ergodic source.*

Then, for any $R < \overline{B}(X, f, F)$, for any $\epsilon > 0$ and any $\lambda > 0$, there exists $n(\epsilon, \lambda) \geq N_f$ such that $\forall n > n(\epsilon, \lambda)$, there exists a $[\lfloor e^{n(R - \overline{H}(X|f(X)))} \rfloor, n, \epsilon, \lambda]$-Feinstein code for the hookup $[[A, X^n]; [A, f^n \times F^n, B \times C]]$.

Proof. The proof is derived from that of Feinstein theorem given in [10], chapter 14.

$$R < \overline{B}(X, f, F) \Rightarrow R' = R - \overline{H}(X|f(X)) < \overline{I}(f(X), F(X))$$

Let $\delta = \frac{\overline{I}(f(X), F(X)) - R'}{2} > 0$, $M = \lfloor e^{nR'} \rfloor$ and $a = n(R' + \delta)$. Let E_n be the measurable set $\{(y, z) \in B^n \times C^n / i_{f^n(X^n)F^n(X^n)}(y, z) \leq a\}$. Then E_n is the set:

$$\{(y, z) \in B^n \times C^n / \frac{i_{f^n(X^n)F^n(X^n)}(y, z)}{n} \leq (\overline{I}(f(X), F(X)) - \delta)\}$$

Since X and $[A, f \times F, B \times C]$ are AMS and ergodic then $(f(X), F(X))$ is AMS and ergodic. Thus, from the mutual information ergodic theorem ([10], Theorem 8.1), $\frac{i_{f^n(X^n)F^n(X^n)}}{n}$ converges in L^1 to $\overline{I}(f(X), F(X))$. This implies that

$$\lim_{n \to \infty} P_{f^n(X^n)F^n(X^n)}(E_n) = 0$$

because $\delta > 0$. As a consequence, for any $\epsilon > 0$ and $\lambda > 0$, a $n \geq N_f$ large enough ensures that:

$$Me^{-a} + P_{f^n(X^n)F^n(X^n)}(i_{f^n(X^n)F^n(X^n)} \leq a) < min(\frac{1}{4}, \epsilon\lambda)$$

$e^{n\delta} \leq Me^{-a}$, thus for any $\epsilon' \in]0, \frac{1}{2}[$ and $\lambda' \in]0, \frac{1}{2}[$ such that $\epsilon' \leq \epsilon$ and $\lambda' \leq \lambda$ and

$$\lambda'\epsilon' = e^{-n\delta} + P_{f^n(X^n)F^n(X^n)}(E_n) < min(\frac{1}{4}, \epsilon\lambda)$$

Lemma 14 applied to $[[A, X^n]; [A, f^n \times F^n, B \times C]]$ implies that there exists a $[M, n, \epsilon', \lambda']$-Feinstein code $\{(A_i^n, \Gamma_i^n) \in \mathcal{B}_{A^n} \times \mathcal{B}_{C^n}, i = 1, \cdots, M\}$ such that:

1. $\Gamma_i \cap \Gamma_j = \emptyset$, $i \neq j$
2. $P_{F^n(X^n)|X^n}(\Gamma_i^c|x^n) \leq \epsilon' \leq \epsilon$ for any $x^n \in A_i^n$, $i = 1, \cdots, M$
3. $\forall i = 1, \cdots, M$, there exists $y_i \in B^n$ such that $A_i^n \subset (f^n)^{-1}(y_i)$ and $P_{X^n|f^n(X^n)}(A_i^n|y_i) > 1 - \lambda' > 1 - \lambda$

□

This proposition is a first justification of the definition of the input capacity $C_f(F)$ of a noisy computation $[A, f \times F, B \times C]$.

7. Reliable computation in the presence of noise

In this section, an abstract model of reliable computation is proposed. A consequence of Proposition 1 is that, given a noisy computation $[A, f \times F, B]$, $\epsilon > 0$ and $\delta > 0$, F^n (ϵ, δ)-reliably computes f^n (n large enough) only on a strict subset of the domain of f^n (i.e., the input code). Thus, to achieve reliable computation of g^k where g is a weakly causal function on the entire domain of g^k requires to rely on an ancillary weakly causal function f and on encoding and decoding:

- an input k-sequence of g^k is encoded into an input n-sequence of $f^n \times F^n$ where F is the noisy implementation of f
- this input n-sequence is fed into F^n (the induced channel) and this produces an output n-sequence
- the output n-sequence is decoded to give a k-sequence which is an estimate of the expected result from g^k.

It is shown that this approach allows to achieve almost perfect reliability in computing $g^k(x^k)$ for large k provided that the encoding rate is strictly lower than the capacity of the noisy computation[3]. This is very similar to the case of reliable communication through noisy channels.

A converse is also proved: if the rate of encoding is strictly greater than the capacity then no code exists to ensure reliable computation.

Moreover the reliable computation model formally captures practical approaches: e.g., a regular arithmetic addition g obtained from the noisy actual circuit F implementing f which is an addition acting on residue encoded operands, see [20].

The weakly causal perfect functions f and g are assumed unary. m-ary functions can be modeled as unary ones by concatenating m input values in one "meta"-input and thus modeling a *joint coding* of operands. This relaxes the assumption of independent coding of operands made by [4], [18], [25] and [1].

[3] As in the classical reliable communication model, the encoder and decoder will be assumed perfectly reliable. This assumption will be discussed below.

7.1. Model of reliable computation

There is a need to constrain the encoding and decoding stages to forbid the model to reduce to (c.f. [4] and [25]):

- either an (assumed perfect) encoder which is equivalent to the expected function followed by an encoding of the result ($g^k(x'^k)$) suitable to transmission through the noisy device (then considered as a noisy transmission channel)
- or an encoder which encodes input values for reliable transmission through the noisy device (considered here also as a noisy transmission channel) and a decoder (assumed also reliable) which is equivalent to an (almost perfect) decoding followed by the perfect expected function.

In both cases, the channel coding theorem sets limits on rates within which reliable transmission (and thus computation in these peculiar cases) is possible. Assume that g is non-injective. Then, if X' is a source, $H(g(X')) < H(X')$ (equivalently $H(X'|g(X')) > 0$). The first case cannot be under the constraint that the encoding is an injection: a injective encoding cannot compress X' into $g(X')$. The second case can also be avoided if the decoding is also based on an injection: the decoding will injectively associate each decodability region to a given output sequence $g^k(x'^k)$. With this constraint, the decoding cannot be equivalent to f followed by an injective transcoding: the non-injective "part of the job" has to be done by the noisy device associating an input sequence to a decodability region[4].

The proposed model is thus mainly adapted to *non-injective* functions. Obviously, it comes down to the classical reliable communication model when g is an injection.

The model of the complete process to reliably compute prefixes of a weakly causal function $g : A'^{\mathcal{I}} \to B'^{\mathcal{I}}$ acting on a source X', thanks to a noisy implementation F of a weakly causal function $f : A^{\mathcal{I}} \to B^{\mathcal{I}}$ is the following:

- **encoding**: let X^n be the n^{th} extension of a source for which a Feinstein code $(A_i^n, \Gamma_i)_{i=1,\cdots,M}$ allowing to (ϵ, λ)-reliably compute $f^n(X^n)$

[4] [26] argues that, from an Information Theory point of view, a frontier between computation and transmission can be drawn considering that computation is a "true" computation if the entropy is reduced ($H(g(X')) < H(X')$ i.e. g is non injective) and that a bijection is equivalent to perfect transmission (preserving entropy).

by $F^n(X^n)$ (cf Proposition 1 and Definition 17) is given ; a typical k-sequence x' of X'^k is encoded as a *typical*[5] n-sequence of X^n by a function, say $\mathcal{U}$, such that $\mathcal{U}(x') \in A_i^n$ for some $i = 1, \cdots, M$.

If $\mathcal{U}$ is an injection then the encoding is said *injective.*

- **computation of the noisy function**: F^n is applied to $\mathcal{U}(x')$ producing a typical n-sequence $F^n(\mathcal{U}(x'))$ of $F^n(X^n)$ where $F^n(\mathcal{U}(x'))$ belongs to a given Γ_i (with probability greater than $1-\epsilon$)

- **decoding**: the first step is to associate to $F^n(\mathcal{U}(x'))$ the typical n-sequence y_i of $f^n(X^n)$ corresponding to Γ_i, the second step is to apply to y_i a function $\mathcal{V} : \{\mathbf{y_1}, \ldots, \mathbf{y_M}\} \rightarrow \{\text{typical k-sequences of } g^k(X'^k)\}$ such that $\mathcal{V}(y_i) = g^k(x')$. The first step of decoding considered as a function from $\{\Gamma_i, i = 1, \cdots, M\}$ to $\{y_i, i = 1, \cdots, M\}$ is injective. The characteristics of the second step are discussed below.

A decoding error occurs when one obtains a n-sequence y_j (or equivalently a Γ_j) such that $\widehat{g^k(x')} = \mathcal{V}(y_j) \neq g^k(x')$.

To be able to define a decoding function $\mathcal{V}$ (i.e, a *deterministic* decoding), the encoding function $\mathcal{U}$ has to be such that the typical n-sequences of one $A_i^n \subset (f^n)^{-1}(y_i)$ ($y_i \in \{y_1, \ldots, y_M\}$) are used for encoding typical k-sequences of *only one* $(g^k)^{-1}(z)$, z typical k-sequence of $g^k(X'^k)$. Thus a decoding will be *deterministic* if:

$$f^n(\mathcal{U}(x'_1)) = f^n(\mathcal{U}(x'_2)) \Rightarrow g^k(x'_1) = g^k(x'_2)$$

It can also be required that $\mathcal{V}$ be an injection in order to forbid the decoder to be able to reliably compute the (non-injective) function g^k. If the function $\mathcal{V}$ is injective then the typical k-sequences of a $(g^k)^{-1}(z)$, z typical k-sequence of $g^k(X'^k)$, are encoded in typical n-sequences of one and only one A_i^n . Thus, if x'_1 and x'_2 are two typical k-sequences of X'^k:

$$f^n(\mathcal{U}(x'_1)) = f^n(\mathcal{U}(x'_2)) \Leftrightarrow g^k(x'_1) = g^k(x'_2)$$

In this case the deterministic decoding is said to be *injective*, figure 4.

The model fulfills the constraints identified above. The encoder implements an injection and thus cannot be the desired function g nor f followed

[5] This important assumption is motivated below.

$$(g^k)^{-1}(z) = \left\{ \begin{array}{c} x'_1 \\ \vdots \\ \vdots \\ \vdots \\ x'_m \end{array} \right\} \xrightarrow{\mathcal{U}} \left\{ \begin{array}{l} A_i^n = \left\{ \begin{array}{c} x_{i_1} \\ \vdots \\ x_{i_N} \end{array} \right\} \underset{1-\epsilon}{\xrightarrow{F^n}} \Gamma_i \to y_i \\ \vdots \\ A_j^n = \left\{ \begin{array}{c} x_{j_1} \\ \vdots \\ x_{j_M} \end{array} \right\} \underset{1-\epsilon}{\xrightarrow{F^n}} \Gamma_j \to y_j \end{array} \right\} \xrightarrow{\mathcal{V}} z$$

Figure 3: Deterministic decoding

$$(g^k)^{-1}(z) = \left\{ \begin{array}{c} x'_1 \\ \vdots \\ x'_m \end{array} \right\} \xrightarrow{\mathcal{U}} A_i^n = \left\{ \begin{array}{c} x_{i_1} \\ \vdots \\ \vdots \\ x_{i_N} \end{array} \right\} \underset{1-\epsilon}{\xrightarrow{F^n}} \Gamma_i \to y_i \xrightarrow{\mathcal{V}} z$$

Figure 4: Injective decoding

by a encoding for transmission (if f and g are not injective). The same comment applies to an injective decoding step $\mathcal{V}$.

Definition 18. *Let $g : A'^{\mathcal{I}} \to B'^{\mathcal{I}}$ and $f : A^{\mathcal{I}} \to C^{\mathcal{I}}$ be measurable weakly causal functions. Let $k \geq N_g$ and $n \geq N_f$. Let A'^k_ϵ and B'^k_ϵ be respectively the sets of ϵ-typical k-sequences of A'^k and of B'^k. An* encoding-decoding $(1-\epsilon)$-compatible with *(g^k,f^n) is a pair $(\mathcal{U}, \mathcal{V})$ where:*

- $\mathcal{U} : A'^k_\epsilon \to A^n$ *is a function called the* encoding function
- $\mathcal{V} : B^n \to B'^k_\epsilon$ *is a function called the* decoding function *or the* deterministic decoding
- $\forall (x'_1, x'_2) \in A'^k_\epsilon \times A'^k_\epsilon, f^n(\mathcal{U}(x'_1)) = f^n(\mathcal{U}(x'_2)) \Rightarrow g^k(x'_1) = g^k(x'_2)$

If $\forall (x'_1, x'_2) \in A'^k_\epsilon \times A'^k_\epsilon, f^n(\mathcal{U}(x'_1)) = f^n(\mathcal{U}(x'_2)) \Leftrightarrow g^k(x'_1) = g^k(x'_2)$, then the decoding is injective*.*

7.2. Achievability of an encoding rate

Definition 19. *Let $[A', X']$ be a source on a finite alphabet A'. Let $g : A'^{\mathcal{I}} \to B'^{\mathcal{I}}$ be a measurable weakly causal function and $[A, f \times F, B \times C]$ be* *a noisy computation such that f is a weakly causal function. Let $R > 0$. The encoding rate R is said* achievable *if there exists a source $[A, X_0]$ and a sequence $\left(\left\{(A_i^{n_j}, \Gamma_i^{n_j}), i = 1, \cdots, M_j\right\}\right)_{j\geq 1}$ of $(M_j, n_j, \epsilon_j, \lambda_j)$-Feinstein codes for the hookups $[[A^{n_j}, X_0^{n_j}]; [A^{n_j}, f^{n_j} \times F^{n_j}, B^{n_j} \times C^{n_j}]]$, $j \geq 1$, $n_j \geq N_f$, such that:*

1. $\lim_{j\to\infty} \epsilon_j = 0$
2. $\lim_{j\to\infty} \lambda_j = 0$
3. $M_j = \lfloor e^{n_j(R - \overline{H}(X_0|f(X_0))} \rfloor$
4. *there exists an encoding-decoding $(\mathcal{U}_j, \mathcal{V}_j)$ $(1-\epsilon)$-compatible with (g^{k_j}, f^{n_j}) such that $\mathcal{U}_j(A'^{k_j}_{\epsilon}) \subset \bigcup_{i=1}^{M_j} A_i^{n_j}$ and $R.n_j = \overline{H}(X').k_j$*

Remark: $\mathbb{Q}$ being dense in $\mathbb{R}$, $\frac{R}{H(X')}$ is assumed rational.

Proposition 2. *Let $g : A'^{\mathcal{I}} \to B'^{\mathcal{I}}$ be a measurable weakly causal function defining an AMS and ergodic deterministic channel. Let $[A, f \times F, B \times C]$ be an AMS and ergodic noisy computation such that f is a weakly causal function. Let $[A', X']$ be an AMS and ergodic source. Let $R > 0$. Then*

$$R < C_f(F) \implies R \text{ is achievable}$$

Assume that $\overline{H}(X'|g(X')) \neq 0$. If

$$R < \sup_{P_X \in \mathfrak{P}_{ae}(A^{\mathcal{I}})} \left[min \left(\frac{\overline{H}(X|f(X))}{\overline{H}(X'|g(X'))}.\overline{H}(X'), \overline{B}(X, f, F) \right) \right]$$

then R is achievable for an injective decoding.

Proof. Let $(\epsilon_j)_{j\in\mathbb{N}}$ and $(\lambda_j)_{j\in\mathbb{N}}$ be two sequences of positive real numbers such that $\lim_{j\to\infty} \epsilon_j = 0$ and $\lim_{j\to\infty} \lambda_j = 0$.

Let $R > 0$ be such that $R < C_f(F)$. Then there exists a source $[A, X_0]$, $P_{X_0} \in \mathfrak{P}_{ae}(A^{\mathcal{I}})$, and two positive real numbers R' and δ such that $R < R' - \delta < R' < \overline{B}(X_0, f, F) \leq C_f(F)$.

For a fixed j, let $\delta' > 0$ be such that $\lambda_j - \delta' > 0$. From Proposition 1, there exists an integer $n(\epsilon_j, \lambda_j) \geq N_f$ such that, for any $n_j \geq n(\epsilon_j, \lambda_j)$, there exists a $(M'_j, n_j, \epsilon_j, \lambda_j - \delta')$-Feinstein code for the hookup $[[A^{n_j}, X_0^{n_j}]; [A^{n_j}, f^{n_j} \times F^{n_j}, B^{n_j} \times C^{n_j}]]$ where $M'_j = \lfloor e^{n_j(R' - \overline{H}(X_0|f(X_0))} \rfloor$. n_j can be chosen large enough to also have $k_j \geq N_g$. Let $(\tilde{A}_i^{n_j}, \Gamma_i^{n_j})_{i=1,\cdots,M'_j}$ be this code for a given n_j. Then:

$$P_{X_0^{n_j}|f^{n_j}(X_0^{n_j})}(\tilde{A}_i^{n_j}|y_i) > 1 - (\lambda_j - \delta')$$

Let $\epsilon' > 0$. For n_j large enough, the subset $[(f^{n_j})^{-1}(y_i)]_{typ}$ of n_j-sequences ϵ'-conditionally typical given y_j has a conditional probability given y_i:

$$P_{X_0^{n_j}|f^{n_j}(X_0^{n_j})}([(f^{n_j})^{-1}(y_i)]_{typ}|y_i) > 1 - \delta'$$

Thus, setting $A_i^{n_j} = \tilde{A}_i^{n_j} \cap [(f^{n_j})^{-1}(y_i)]_{typ}$:

$$P_{X_0^{n_j}|f^{n_j}(X_0^{n_j})}(A_i^{n_j}|y_i) > 1 - \lambda_j$$

Obviously, for any $x \in A_i^{n_j}$

$$P_{F^{n_j}(X_0^{n_j})|X_0^{n_j}}((\Gamma_i^{n_j})^c|x) \leq \epsilon_j$$

This implies that $(A_i^{n_j}, \Gamma_i^{n_j})_{i=1,\cdots,M'_j}$ is a $(M'_j, n_j, \epsilon_j, \lambda_j)$-Feinstein code for the hookup $[[A^{n_j}, X_0^{n_j}]; [A^{n_j}, f^{n_j} \times F^{n_j}, B^{n_j} \times C^{n_j}]]$ and each $A_i^{n_j}$ contains only ϵ'-conditionally typical sequences given y_i.

Let α_1 the number of ϵ'-typical (given z) sequences within $(g^{k_j})^{-1}(z)$ (z typical k-sequence). Let α_2 the number of ϵ'-typical (given y_i) sequences within $A_i^{n_j}$. Thanks to conditional AEP (theorem 2), for n_j (and thus $k_j = \frac{R}{\overline{H}(X')} n_j$) large enough:

$$(1 - \lambda_j) e^{k_j(\overline{H}(X'|g(X')) - \epsilon')} \leq \alpha_1 \leq e^{k_j(\overline{H}(X'|g(X')) + \epsilon')}$$

and

$$(1 - \lambda_j) e^{n_j(\overline{H}(X_0|f(X_0)) - \epsilon')} \leq \alpha_2 \leq e^{n_j(\overline{H}(X_0|f(X_0)) + \epsilon')}$$

An encoding-decoding $(1-\epsilon')$-compatible with (g^{k_j}, f^{n_j}) using the $(A_i^{n_j})_{i=1,\ldots,M'_j}$ will require that the encoding of the conditionally typical sequences of one inverse image $(g^{k_j})^{-1}(z)$ consume a number (i.e. $\lceil \alpha_1/\alpha_2 \rceil$) of $A_i^{n_j}$ which is upper bounded by[6]:

$$\left\lceil \frac{e^{k_j(\overline{H}(X'|g(X')) + \epsilon')}}{(1 - \lambda_j) e^{n_j(\overline{H}(X_0|f(X_0)) - \epsilon')}} \right\rceil$$

If N is the total number of $A_i^{n_j}$ needed to encode all the inverse images $(g^{k_j})^{-1}(z)$ then:

$$N \leq \left\lceil \frac{e^{k_j(\overline{H}(X'|g(X')) + \epsilon')}}{(1 - \lambda_j) e^{n(\overline{H}(X_0|f(X_0)) - \epsilon')}} \right\rceil . e^{k_j(\overline{H}(g(X') + \epsilon')}$$

[6] The constraint is due to the determinism of the decoding. The encoding may or may not be injective: the constraint allows injective encoding but does not make it mandatory.

This leads to:

$$\begin{aligned} N &\leq \left[\frac{e^{k_j(\overline{H}(X'|g(X'))+\epsilon')}}{(1-\lambda_j)e^{n_j(\overline{H}(X_0|f(X_0))-\epsilon')}}+1\right].e^{k_j(\overline{H}(g(X')+\epsilon')} \\ &\leq \left[\frac{e^{k_j(\overline{H}(X'|g(X'))+\epsilon'+\epsilon'')}}{e^{n_j(\overline{H}(X_0|f(X_0))-\epsilon')}}+1\right].e^{k_j(\overline{H}(g(X')+\epsilon')} \end{aligned}$$

where $e^{k\epsilon''}=1/(1-\lambda_j)$. For any $\epsilon'''>0$, for k_j and n_j large enough:

$$\frac{e^{k_j(\overline{H}(X'|g(X'))+\epsilon'+\epsilon'')}}{e^{n_j(\overline{H}(X_0|f(X_0))-\epsilon')}}+1 \leq \frac{e^{k_j(\overline{H}(X'|g(X'))+\epsilon'+\epsilon'')}}{e^{n_j(\overline{H}(X_0|f(X_0))-\epsilon')}}.e^{k_j\epsilon'''}$$

giving:

$$\begin{aligned} N &\leq \frac{e^{k_j(\overline{H}(X'|g(X'))+\epsilon'+\epsilon''+\epsilon''')}}{e^{n_j(\overline{H}(X_0|f(X_0))-\epsilon')}}.e^{k_j(\overline{H}(g(X')+\epsilon')} \\ &\leq \frac{e^{k_j(\overline{H}(X')+2\epsilon'+\epsilon''+\epsilon''')}}{e^{n_j(\overline{H}(X_0|f(X_0))-\epsilon')}} \end{aligned}$$

Since $k_j.\overline{H}(X')=n_j.R$:

$$\begin{aligned} N &\leq \frac{e^{n_j(R+\frac{\epsilon'+2\epsilon''+\epsilon'''}{\overline{H}(X')})}}{e^{n_j(\overline{H}(X_0|f(X_0))-\epsilon')}} \\ &\leq e^{n_j(R-\overline{H}(X_0|f(X_0))+(2\epsilon'+\epsilon''+\epsilon''')/\overline{H}(X')-\epsilon)} \end{aligned}$$

for suitably chosen ϵ', ϵ'' and ϵ''' and large enough k_j and n_j. Hence:

$$\begin{aligned} N &\leq e^{n_j(R'-\overline{H}(X_0|f(X_0)))} \\ &\leq \lfloor e^{n_j(R'-\overline{H}(X_0|f(X_0)))}\rfloor \end{aligned}$$

since N is an integer. Thus

$$N \leq M'_j$$

In words, there are enough $A_i^{n_j}$ in the Feinstein code to encode the inverse images $(g^{k_j})^{-1}(z^{k_j})$. This closes the first step of the proof.

An injective decoding will be possible if one inverse image $(g^{k_j})^{-1}(z)$ can be encoded using only one $A_i^{n_j}$. For n_j and k_j large enough, a sufficient condition is:

$$e^{k_j\overline{H}(X'|g(X'))} < e^{n_j\overline{H}(X_0|f(X_0))}$$

which is equivalent to:

$$\frac{k_j}{n_j} < \frac{\overline{H}(X_0|f(X_0))}{\overline{H}(X'|g(X'))}$$

Since $\frac{k_j}{n_j} = \frac{R}{\overline{H}(X')}$:

$$R < \frac{\overline{H}(X_0|f(X_0))}{\overline{H}(X'|g(X'))}.\overline{H}(X')$$

The condition

$$R < \sup_{P_X \in \mathfrak{P}_{ae}(A^{\mathcal{I}})} \left[min \left(\frac{\overline{H}(X|f(X))}{\overline{H}(X'|g(X'))}.\overline{H}(X'), \overline{B}(X,f,F) \right) \right]$$

will thus ensure that an injective encoding of X' on a Feinstein code and allowing an injective decoding is possible. □

Proposition 2 does not rely on the assumption of an injective encoding: it also holds for an injective encoding. Thus such an assumption does not narrow the context of Proposition 2, it makes the model compliant with the constraints expressed above.

The use of conditional typical sequences and of an expected function f is worth some comments. Designing a code is finding a set of pairs $\{(A_i^n, \Gamma_i), i = 1, \cdots M\}$ such that the "image" of A_i^n (set of typical sequences) by the "noisy function" F^n falls within Γ_i with high probability. The approach followed here has been to determine this code thanks to a function f called the expected function. A dual approach could be to choose a code and then to determine a function f such that $A_i^n \subset f^{-1}(y_i)$ for a given collection of typical n-sequences y_i and such that the typical n-sequences of X^n belonging to A_i^n are conditionally typical given y_i.

Seeking a relevant concept of capacity requires to associate to a code a cost measure which is a rate of encoding. For channel coding and fixed block encoding, the rate determines the number of usable input blocks for a given block size. For computation coding, two cardinalities are needed: the number of sets A_i^n and the number of elements in each A_i^n. A convenient consequence of determining the A_i^n's by a function f^n is that, for large n, the A_i^n's have almost the same number of well identified elements, that is the n-sequences conditionally typical given y_i. A unique number, i.e. $\overline{H}(X|f(X))$, gives the common size (i.e. $\simeq e^{n\overline{H}(X|f(X))}$) of precisely determined subsets A_i^n's. Moreover, such a subset A_i^n has almost the same probability measure as the entire $(f^n)^{-1}(y_i)$. The method gives a balanced code which is well characterized by a unique number (the typical input rate) which represents the two cardinalities mentioned above.

7.3. The converse

Proposition 3. *Let $g : A'^{\mathcal{I}} \to B'^{\mathcal{I}}$ be a measurable weakly causal function defining an AMS and ergodic deterministic channel. Let $[A, f \times F, B \times C]$ be an AMS and ergodic noisy computation. Assume that f is a weakly causal function. Let $[A', X']$ be an AMS and ergodic source. Let $R > 0$. If $R > C_f(F)$, R is not achievable.*

Proof. $R = \frac{k}{n}\overline{H}(X') > C_f(F)$ then

$$\forall X \in \mathfrak{P}_{ae}(A^{\mathcal{I}}), R = \frac{k}{n}\overline{H}(X') > \overline{H}(X|f(X)) + \overline{I}(f(X); F(X))$$

Let $X_0 \in \mathfrak{P}_{ae}(A^{\mathcal{I}})$.

Using the same notations (dropping the subscript j) as in the proof of Proposition 2, if the decoding is deterministic then the encoding of the conditionally typical sequences of an inverse image $(g^k)^{-1}(z^k)$ uses a number of A_i^n (or equivalently of $(f^n)^{-1}(y_i)$) which is lower bounded by:

$$\left\lceil \frac{(1-\lambda)e^{k(\overline{H}(X'|g(X'))-\epsilon')}}{e^{n(\overline{H}(X_0|f(X_0))+\epsilon')}} \right\rceil$$

If N is the total number A_i^n (or equivalently of inverse images $(f^n)^{-1}(y_i)$ or equivalently the number of typical sequences of $f^n(\mathcal{U}(X'^k))$) needed to encode all the inverse images $(g^k)^{-1}(z)$ (z typical k-sequences) then:

$$\begin{aligned} N &\geq \left\lceil \frac{(1-\lambda)e^{k(\overline{H}(X'|g(X'))-\epsilon')}}{e^{n(\overline{H}(X_0|f(X_0))+\epsilon')}} \right\rceil .(1-\lambda)e^{k(\overline{H}(g(X'))-\epsilon')} \\ &\geq e^{k(\overline{H}(X')-2\epsilon')-n(\overline{H}(X_0'|f(X_0))+\epsilon')}(1-\lambda)^2 \\ &\geq e^{n.\frac{R}{\overline{H}(X')}(\overline{H}(X')-2\epsilon')-n(\overline{H}(X_0|f(X_0))+\epsilon')}(1-\lambda)^2 \\ &\geq e^{n.(R-\frac{2\epsilon'}{\overline{H}(X')}-\overline{H}(X_0|f(X_0))+\epsilon')}(1-\lambda)^2 \\ &> e^{n(\overline{I}(f(X_0;F(X_0)))+\epsilon'')} \end{aligned} \tag{5}$$

for suitably chosen λ, ϵ' and ϵ''. This is possible to find such λ, ϵ' and ϵ'' since $R - \overline{H}(X_0|f(X_0)) > \overline{I}(f(X_0); F(X_0))$.

Hence, for any ϵ'', if n is large enough, (5) holds.

Assume that, thanks to deterministic encoding and decoding of rate R', an ergodic and AMS random process Y^l (associated to a source $[B, Y]$) is to be obtained from $F^n(X^n)$ (i.e. $R' = \frac{l}{n}\overline{H}(Y)$): $Y^l \to f^n(X^n) \to F^n(X^n) \to \widehat{Y^l}$ is a Markov Chain (where $\widehat{Y^l}$ is the "estimate" of Y^l after decoding).

This implies (thanks to Data Processing Inequality) that:

$$\begin{aligned} I(Y^l;\widehat{Y^l}) &\leq I(f^n(X^n);F^n(X^n)) \\ &\Rightarrow H(Y^l) - H(Y^l|\widehat{Y^l}) \leq I(f^n(X^n);F^n(X^n)) \end{aligned}$$

Let $P_e(l) = P(\widehat{Y^l} \neq Y^l)$. By Fano's inequality:

$$\begin{aligned} & H(Y^l) - (H_2(P_e(l)) + l.P_e(l).log(|B|)) \leq I(f^n(X^n);F^n(X^n)) \\ & \quad \Rightarrow \frac{H(Y^l)}{l} \leq \frac{I(f^n(X^n);F^n(X^n))}{l} + \frac{(H_2(P_e(l)) + l.P_e(l).log(|B|))}{l} \\ & \Rightarrow \frac{H(Y^l)}{l} \leq \frac{I(f^n(X^n);F^n(X^n))}{n}.\frac{\overline{H}(Y)}{R'} + \frac{(H_2(P_e(l)) + l.P_e(l).log(|B|))}{l} \end{aligned}$$

If the error probability $P_e(l)$ vanishes (i.e. $\lim_{l\to\infty} P_e(l) = 0$) then, necessarily:

$$R' \leq \overline{I}(f(X);F(X))$$

Hence, given $\epsilon'' > 0$, for large enough l, the number $N_Y(l)$ of typical sequences of the source $[Y, B']$ necessarily verifies

$$N_Y(l) \leq e^{n+\overline{I}(f(X);F(X))\epsilon''}$$

This is nothing else than the classical proof of the converse of the Channel Coding Theorem.

Assuming there exists an injection ϕ associating to typical sequences of $f^n(\mathcal{U}(X'^k))$ typical sequences of $f^n(X^n)$, it is always possible to find an injection ψ associating to typical sequences of $f^n(\mathcal{U}(X'^k))$ typical sequences of Y^l and an injection δ associating to typical sequences of Y^l typical sequences of Y^l $f^n(X^n)$ such that $\psi = \delta \circ \phi$. To obtain an asymptotically perfect estimation of $f^n(\mathcal{U}(X'^k))$, it is necessary to have

$$N \leq N_Y(l) \leq e^{n+\overline{I}(f(X);F(X))\epsilon''}$$

which is impossible by (5).

Then, by (5), the error probability cannot tends to 0 (the encoding of X' by X_0 requires more inverse images $(f^n)^{-1}(y^n)$ or A_i^n than "available").

In this proof, the use of an "ancillary" ergodic AMS source $[B, Y]$ is motivated by the fact that $f^n(\mathcal{U}(X'^k))$ cannot be assumed to be the prefix of an ergodic AMS process $f(\mathcal{U}^*(X'))$. □

7.4. Reliable computation and noisy encoding

The model of reliable computation assumes that the encoder and the decoder are perfectly reliable. This assumption could be justified by quoting from [25] *"The computation system [model] was devised for the sole purpose of studying the relation of information theory and reliable automata"*. Moreover, it could be argued that if the complexity of the computation device is of a much greater magnitude than that of the encoder and decoder then the unreliability of the encoder and decoder have almost no impact on the overall reliability of the computation and thus can be neglected. For complex systems, this is quite realistic.

In any case, it is impossible to overcome the fact that the reliability reached is at the best the reliability of the final decoding device. The only way is to built a intrinsically reliable enough decoder (for example thanks to gate redundancy).

A noisy encoder is a noisy computation itself and thus can be handled from the point of view of "cascaded noisy computations". To allow downstream reliable computation, a noisy encoder has to possess intrinsic performance which can be expressed as follows.

For any $\epsilon > 0$ and $\lambda > 0$, it should exist $k(\epsilon, \lambda)$ such that for any $k > k(\epsilon, \lambda)$ and for any $z \in {C'}^k$ there must exist a set $A'_z \subset (g^k)^{-1}(z)$ and a set $A_z \subset \cup_{y \in \mathcal{V}^{-1}(z)} (f^n)^{-1}(y)$ for which:

$$\begin{aligned} P_{X^n|X'^k}(A_z^c|x') &< \epsilon \ \forall x' \in A'_z && (6) \\ P_{X'^k|g(X'^k)}(A'_z|z) &> 1 - \lambda && (7) \end{aligned}$$

This just states that encoding $x' \in (g^k)^{-1}(z)$ by any randomly chosen $x \in \cup_{y \in \mathcal{V}^{-1}(z)} (f^n)^{-1}(y)$ will preserve the input code for the noisy computation and thus allows correction through the decoding. Stated another way, the encoding errors remain compatible with the code of the hookup $[[A, X^n]; [A, f^n \times F^n), B \times C]]$.

(6) and (7) characterize $(A_z, A'_z)_{z \in C'^k}$ as a (M, k, ϵ, δ)-Feinstein code for the hookup $[[A', X'^k]; [A', (u^k, \mathcal{U}), A \times A]]$ where u stands for the perfect encoding function, $\mathcal{U}$ for the noisy one and $M = |C'^k|$.

8. Example: Computation with noisy inputs

In the special case where a computation is noisy due to the input device, it is possible to assert that correcting the results of a function applied to noisy input will offer a better input rate than correcting inputs before applying the function. This case is simply captured by a noisy computation $[A, f \times F, B \times B)$ where:

- f is the expected perfect function
- $[A, F, B]$ is a cascade of a channel $[A, \nu, A]$ and of $[A, f, B]$ (i.e. $F \equiv \nu f$)

$$\begin{tikzcd}
\begin{array}{c}\text{Input source}\\ {[A,X]}\end{array} \arrow[r, "\nu"] \arrow[dr, dotted, "f"] & \begin{array}{c}\text{Noisy input source}\\ {[A,Y]}\end{array} \arrow[r, "f"] & \begin{array}{c}\text{Noisy result}\\ {[B,f(Y)]}\end{array} \\
 & \begin{array}{c}\text{Expected result}\\ {[B,f(X)]}\end{array} &
\end{tikzcd}$$

Figure 5: Computation with noisy inputs

The following lemma asserts that the capacity of the channel ν is lower or equal than the capacity of the noisy function $f \times (\nu f)$. Then it is possible to obtain a greater input rate for $f \times (\nu f)$ than for ν.

Lemma 17. *Let $[A, f \times F, B \times B]$ be an AMS and ergodic noisy computation such that*

- *f is weakly causal*
- *$[A, F, B]$ is a cascade of an AMS and ergodic channel $[A, \nu, A]$ and of $[A, f, B]$ (i.e. $F \equiv \nu f$)*

Let C_ν denote the capacity of the channel $[A, \nu, A]$ and $C_f(F)$ denote the capacity of the noisy computation $[A, f \times F, B \times B]$. Then

$$C_\nu \leq C_f(F)$$

Proof. Let X and Y two random variables with values in a finite set A. Let $f : A \to B$ a measurable function from A to a finite set B. Then

$$\begin{aligned}
H(X|Y) &= \sum_{(a_1,a_2)\in A^2} P_{X,Y}(a_1,a_2) log\left(\frac{P_Y(a_2)}{P_{X,Y}(a_1,a_2)}\right) \\
&= \sum_{(b_1,b_2)\in B^2} \sum_{(a_1,a_2)\in(f^{-1}(b_1)\times f^{-1}(b_2)} P_{X,Y}(a_1,a_2) log\left(\frac{P_Y(a_2)}{P_{X,Y}(a_1,a_2)}\right)
\end{aligned}$$

By the log-sum inequality:

$$\begin{aligned}
H(X|Y) &\geq \sum_{(b_1,b_2)\in B^2} \left(\sum_{(a_1,a_2)\in(f^{-1}(b_1)\times f^{-1}(b_2)} P_{X,Y}(a_1,a_2)\right). \\
&\quad log\left(\frac{\sum_{(a_1,a_2)\in(f^{-1}(b_1)\times f^{-1}(b_2)} P_Y(a_2)}{\sum_{(a_1,a_2)\in(f^{-1}(b_1)\times f^{-1}(b_2)} P_{X,Y}(a_1,a_2)}\right) \\
&\geq \sum_{(b_1,b_2)\in B^2} P_{f(X),f(Y)}(b_1,b_2) log\left(\frac{P_{f(Y)}(b_2)\sum_{(a_1)\in(f^{-1}(b_1)} 1}{P_{f(X),f(Y)}(b_1,b_2)}\right) \\
&\geq \sum_{(b_1,b_2)\in B^2} P_{f(X),f(Y)}(b_1,b_2) log\left(\frac{P_{f(Y)}(b_2)}{P_{f(X),f(Y)}(b_1,b_2)}\right) \\
&\geq H(f(X)|f(Y))
\end{aligned}$$

Let $[A,X]$ and $[A,Y]$ be two AMS and ergodic sources and $f : A^{\mathcal{I}} \to B^{\mathcal{I}}$ a weakly causal measurable function. Then for any n large enough, $H(X^n|Y^n) \geq H(f^n(X^n)|f^n(Y^n))$. Taking the limit when n tends to infinity:

$$\overline{H}(X|Y) \geq \overline{H}(f(X)|f(Y))$$

Let $[A, f \times F, B \times B]$ be an AMS and ergodic noisy computation such that

- f is weakly causal
- $[A,F,B]$ is a cascade of an AMS and ergodic channel $[A,\nu,A]$ and of $[A,f,B]$ (i.e. $F \equiv \nu f$)

Let $[A,X]$ be an AMS and ergodic source and let and $[A,Y]$ denote the output of the channel ν when fed by the source X.

$$\begin{aligned}
C_\nu - C_f(F) &= \sup_X(\overline{H}(X) - \overline{H}(X|Y)) - \sup_X(\overline{H}(X) - \overline{H}(f(X)|f(Y))) \\
&= \sup_X(\overline{H}(X) - \overline{H}(X|Y)) + \inf_X(-\overline{H}(X) + \overline{H}(f(X)|f(Y)))
\end{aligned}$$

Then, for any source X':

$$C_\nu - C_f(F) \leq \sup_X (\overline{H}(X) - \overline{H}(X|Y)) - \overline{H}(X') + \overline{H}(f(X')|f(Y')))$$

Thus, for any $\epsilon > 0$, there exists a source X_0 such that:

$$C_\nu - C_f(F) \leq \epsilon + \overline{H}(X_0) - \overline{H}(X_0|Y_0) - \overline{H}(X_0) + \overline{H}(f(X_0)|f(Y_0)))$$

Since $\overline{H}(X_0|Y_0) \geq \overline{H}(f(X_0)|f(Y_0))$, for any $\epsilon > 0$

$$C_\nu - C_f(F) \leq \epsilon$$

Then $C_\nu \leq C_f(F)$

□

As an example, assume that:

- A is a set of binary words of some finite length, f is such that, for all $x \in A^{\mathcal{I}}$, $f(x) = \phi(x_0) \cdots \phi(x_i) \cdots$ where ϕ is a Turing-computable function from A to $B = \phi(A)$; then f is obviously stationary and weakly causal
- the noisy channel $[A, F, C]$ is made of a stationary and ergodic noisy communication channel $[A, \nu, A]$ (e.g. a memoryless channel) followed by a perfect instance of f. I.e. $B = C$ and $[A, F, B]$ is a cascade νf.

This models an infinite sequence of computations of a computable function ϕ on inputs acquired through a noisy channel. Then, thanks to the following lemmas (Lemma 18 and Lemma 19, proved in Appendix), Propositions 1, 2 and 3 and Lemma 17 hold.

Lemma 18. *If $[A, \nu, A]$ is a stationary ergodic channel and $f : A^{\mathcal{I}} \to B^{\mathcal{I}}$ is a stationary function (i.e. $f(T_A x) = T_B f(x)$ for all $x \in A^{\mathcal{I}}$) then the cascade $[A, \nu f, B]$ is a stationary and ergodic channel $[A, F, B]$ ($F \equiv \nu f$).*

Lemma 19. *If f is a stationary function (i.e. $f(T_A x) = T_B f(x)$ for all $x \in A^{\mathcal{I}}$) and F is stationary and ergodic then $[A, f \times F, B \times C]$ is stationary and ergodic.*

Part III
Ergodic and AMS noisy computations

9. Channel Products

In this section, it is proved that a channel product $[A, \nu_1 \times \nu_2, B \times C]$ is respectively stationary, AMS (w.r.t. a stationary source), recurrent or output weakly mixing if and only if both channels $[A, \nu_1, B]$ and $[A, \nu_2, C]$ are respectively stationary, AMS (w.r.t. a stationary source), recurrent or output weakly mixing. The section ends with corollaries dedicated to noisy computations.

9.1. Stationary and AMS Channel Products

Proposition 4. *The channel product $[A, \nu_1 \times \nu_2, B \times C]$ is stationary w.r.t a stationary source distribution μ if and only if $[A, \nu_1, B]$ and $[A, \nu_2, C]$ are stationary w.r.t μ.*

Proof. If the channel product is stationary w.r.t μ then the hookup $\mu(\nu_1 \times \nu_2)$ is stationary. Thus the marginals $\mu\nu_1$ and $\mu\nu_2$ are stationary. Then $[A, \nu_1, B]$ and $[A, \nu_2, C]$ are stationary w.r.t μ.

Assume that $[A, \nu_1, B]$ and $[A, \nu_2, C]$ are stationary w.r.t μ. $\forall E \in \mathcal{B}_{A^\mathcal{I}}$, $\forall G \in \mathcal{B}_{B^\mathcal{I}}$ and $\forall H \in \mathcal{B}_{C^\mathcal{I}}$:

$$\mu(\nu_1 \times \nu_2)(E \times G \times H) = \int_E \nu_1(x, G).\nu_2(x, H) d\mu$$

Since μ is stationary, by a change of variable:

$$\mu(\nu_1 \times \nu_2)(E \times G \times H) = \int_{T_A^{-1}E} \nu_1(T_A x, G)\nu_2(T_A x, H) d\mu$$

Thanks to lemma 3 (p. 13)

$$\mu(\nu_1 \times \nu_2)(E \times G \times H) = \int_{T_A^{-1}E} \nu_1(x, T_B^{-1}G)\nu_2(x, T_C^{-1}H) d\mu$$

Then $\mu(\nu_1 \times \nu_2)(E \times G \times H) = \mu(\nu_1 \times \nu_2)(T_A^{-1}E \times T_B^{-1}G \times T_C^{-1}H)$. Stationarity on the semi-algebra of rectangles implies stationarity on the product σ-field ([14], Remark 3, p. 77) then $\mu(\nu_1 \times \nu_2)$ is stationary. □

Proposition 5. *The channel product $[A, \nu_1 \times \nu_2, B \times C]$ is AMS w.r.t a stationary source distribution μ if and only if $[A, \nu_1, B]$ and $[A, \nu_2, C]$ are AMS w.r.t μ. In this case, $\mu(\nu_1 \times \nu_2) \ll^a \mu(S\overline{\nu}_{1_\mu} \times S\overline{\nu}_{2_\mu})$.*

Proof of Proposition 5 relies on the following lemma:

Lemma 20. *Let μ_1, μ_2, resp. η_1, η_2, be probabilities on a space $(\Omega, \mathcal{B}_\Omega)$, resp. $(\Lambda, \mathcal{B}_\Lambda)$. Let $\mu_1 \times \eta_1$ and $\mu_2 \times \eta_2$ denote the product probability on $(\Omega \times \Lambda, \mathcal{B}_{\Omega\times\Lambda})$. Then:*

1. *$\mu_1 \times \eta_1 \ll \mu_2 \times \eta_2$ if and only if $\mu_1 \ll \mu_2$ and $\eta_1 \ll \eta_2$*
2. *assuming that Ω and Λ are sequence spaces and that μ_2 and η_2 are stationary, $\mu_1 \times \eta_1 \ll^a \mu_2 \times \eta_2$ if and only if $\mu_1 \ll^a \mu_2$ and $\eta_1 \ll^a \eta_2$*

An obvious consequence is that $\mu_1 \times \eta_1$ is AMS if and only if μ_2 and η_2 are AMS.

Proof of Lemma 20. μ_1 and η_1 (resp. μ_2 and η_2) are marginals of $\mu_1 \times \eta_1$ (resp. of $\mu_2 \times \eta_2$). It is thus obvious that

- $\mu_1 \times \eta_1 \ll \mu_2 \times \eta_2 \Rightarrow \mu_1 \ll \mu_2$ and $\eta_1 \ll \eta_2$
- $\mu_1 \times \eta_1 \ll^a \mu_2 \times \eta_2 \Rightarrow \mu_1 \ll^a \mu_2$ and $\eta_1 \ll^a \eta_2$

Assume that $\mu_1 \ll \mu_2$ and $\eta_1 \ll \eta_2$. Let E be an event of the product σ-field $\mathcal{B}_{\Omega\times\Lambda}$ such that $(\mu_2 \times \eta_2)(E) = 0$. $\mu_2 \times \eta_2$ is the product of the two probabilities μ_2 and η_2 then:

$$(\mu_2 \times \eta_2)(E) = \int \eta_2(E_x) d\mu_2$$

where $E_x = \{y \in B^{\mathcal{I}} / (x, y) \in E\}$ denotes the section of E at x. $(\mu_2 \times \eta_2)(E) = 0$ then

$$\eta_2(E_x) = 0 \ \mu_2\text{-a.e.}$$

$\mu_1 \ll \mu_2$ and $\eta_1 \ll \eta_2$, this implies that $\eta_1(E_x) = 0$ μ_1-a.e. Then

$$\mu_1 \times \eta_1(E) = \int \eta_1(E_x) d\mu_1 = 0$$

Hence $\mu_1 \times \eta_1 \ll \mu_2 \times \eta_2$.

Assume now that μ_2 and η_2 are stationary and that $\mu_1 \ll^a \mu_2$ and $\eta_1 \ll^a \eta_2$. From Lemma 1 (p. 9), this implies that $\mu_{1_\infty} \ll \mu_{2_\infty}$ and $\eta_{1_\infty} \ll \eta_{2_\infty}$. Thus, from above, $\mu_{1_\infty} \times \eta_{1_\infty} \ll \mu_{2_\infty} \times \eta_{2_\infty}$. Thanks to Lemma 4 (p. 14), for any probabilities μ and η, $\mu_\infty \times \eta_\infty = (\mu \times \eta)_\infty$. Then, by Lemma 1 (p. 9) since $\mu_2 \times \eta_2$ is stationary, $\mu_1 \times \eta_1 \ll^a \mu_2 \times \eta_2$. □

Proof of Proposition 5. If the channel product is AMS w.r.t μ then $\mu(\nu_1 \times \nu_2)$ is AMS then the marginals $\mu\nu_1$ and $\mu\nu_2$ are AMS. Then $[A, \nu_1, B]$ and $[A, \nu_2, C]$ are AMS w.r.t μ.

Assume that $[A, \nu_1, B]$ and $[A, \nu_2, C]$ are AMS w.r.t μ. Let $\overline{\nu}_{1_\mu}$ and $\overline{\nu}_{2_\mu}$ respectively denote the stationary means of ν_1 and ν_2 w.r.t μ.

By Lemma 6 (p. 17) $S\overline{\nu}_{1_\mu}$ and $S\overline{\nu}_{2_\mu}$ are stationary w.r.t μ. Then, by Proposition 4 (p. 52), the channel product $S\overline{\nu}_{1_\mu} \times S\overline{\nu}_{2_\mu}$ is stationary w.r.t. μ, i.e., $\mu(S\overline{\nu}_{1_\mu} \times S\overline{\nu}_{2_\mu})$ is stationary.

Thanks to Proposition 10 (p. 19):

$$\nu_1(x, .) \ll^a S\overline{\nu}_{1_\mu}(x, .) \text{ and } \nu_2(x, .) \ll^a S\overline{\nu}_{2_\mu}(x, .)\ \mu\text{-a.e.}$$

Since, $S\overline{\nu}_{1_\mu}(x, .)$ and $S\overline{\nu}_{2_\mu}(x, .)$ are stationary, by Lemma 1 (p. 9), this implies that

$$(\nu_1(x, .))_\infty \ll (S\overline{\nu}_{1_\mu}(x, .))_\infty \text{ and } (\nu_2(x, .))_\infty \ll (S\overline{\nu}_{2_\mu}(x, .))_\infty\ \mu\text{-a.e.}$$

By Lemma 20 (p. 53):

$$(\nu_1(x, .))_\infty \times (\nu_2(x, .))_\infty \ll (S\overline{\nu}_{1_\mu}(x, .))_\infty \times (S\overline{\nu}_{2_\mu}(x, .))_\infty\ \mu\text{-a.e.}$$

Thanks to Lemma 4 (p. 14), $(\nu_1(x, .))_\infty \times (\nu_2(x, .))_\infty = (\nu_1(x, .) \times \nu_2(x, .))_\infty$ and $(S\overline{\nu}_{1_\mu}(x, .))_\infty \times (S\overline{\nu}_{2_\mu}(x, .))_\infty = (S\overline{\nu}_{1_\mu}(x, .) \times S\overline{\nu}_{2_\mu}(x, .))_\infty$, then, by Lemma 9 (p. 19):

$$\mu_\infty(\nu_1 \times \nu_2)_\infty \ll \mu_\infty(S\overline{\nu}_{1_\mu} \times S\overline{\nu}_{2_\mu})_\infty$$

Thus, by Lemma 1 (p. 9), since $\mu(S\overline{\nu}_{1_\mu} \times S\overline{\nu}_{2_\mu})$ is stationary:

$$\mu(\nu_1 \times \nu_2) \ll^a \mu(S\overline{\nu}_{1_\mu} \times S\overline{\nu}_{2_\mu})$$

Hence $\mu(\nu_1 \times \nu_2)$ is AMS. □

9.2. Recurrent Channel Product

Proposition 6. *Let $[A, \nu_1, B]$ and $[A, \nu_2, C]$ be two channels and μ be the distribution of a recurrent source $[A, X]$. The channel product $[A, \nu_1 \times \nu_2, B \times C]$ is recurrent w.r.t μ if and only if $[A, \nu_1, B]$ and $[A, \nu_2, C]$ are both recurrent w.r.t. μ.*

Proof. The "only if " part is obvious. Assume that $[A, \nu_1, B]$ and $[A, \nu_2, B]$ are recurrent w.r.t μ. By Lemma 5 (p. 14), $\nu_1(x, .)$ and $\nu_2(x, .)$ are recurrent μ-a.e. Then, $\forall G \in \mathcal{B}_{B^{\mathcal{I}}}$, $\forall H \in \mathcal{B}_{C^{\mathcal{I}}}$:

$$\begin{aligned}(\nu_1 \times \nu_2)(x, G \times H \setminus \cup_{i\geq 1}(T_{BC}^{-i} G \times H) = \\ (\nu_1 \times \nu_2)(x, G \times H \cap \bigcap_{i\geq 1}(T_B^{-i} G \times T_C^{-i} H^c) \\ + (\nu_1 \times \nu_2)(x, G \times H \cap \bigcap_{i\geq 1}(T_B^{-i} G^c \times T_C^{-i} H) \\ + (\nu_1 \times \nu_2)(x, G \times H \cap \bigcap_{i\geq 1}(T_B^{-i} G^c \times T_C^{-i} H^c)\end{aligned}$$

$$(\nu_1 \times \nu_2)(x, G \times H \cap \bigcap_{i\geq 1}(T_B^{-i} G \times T_B^{-i} H^c) \leq \nu_2(x, H \setminus \cup_{i\geq 1} T_C^{-i} H) = 0 \ \mu\text{-a.e.}$$

since $\nu_2(x, .)$ is recurrent μ-a.e.

$$(\nu_1 \times \nu_2)(x, G \times H \cap \bigcap_{i\geq 1}(T_B^{-i} G^c \times T_C^{-i} H) \leq \nu_1(x, G \setminus \cup_{i\geq 1} T_B^{-i} G = 0) \ \mu\text{-a.e.}$$

since $\nu_1(x, .)$ is recurrent μ-a.e.

$$\begin{aligned}(\nu_1 \times \nu_2)(x, G \times H \cap \bigcap_{i\geq 1}(T_B^{-i} G^c \times T_C^{-i} H^c) = \\ \nu_1(x, G \setminus \cup_{i\geq 1}(T_B^{-i} G)).\nu_2(x, H \setminus \cup_{i\geq 1}(T_C^{-i} H)) = 0 \ \mu\text{-a.e.}\end{aligned}$$

since $\nu_1(x, .)$ and $\nu_2(x, .)$ are recurrent μ-a.e.

Then the product $(\nu_1 \times \nu_2)(x, .)$ is recurrent on rectangles μ-a.e. Then, by Lemma 5 (p. 14), $(\nu_1 \times \nu_2)(x, .)$ is recurrent μ-a.e. Thanks to Lemma 5 (p. 14), $[A, \nu_1 \times \nu_2, B \times C]$ is recurrent w.r.t μ. □

9.3. Output Weakly Mixing Channel Product

It is shown that the product of output weakly mixing stationary channels is output weakly mixing. This property is of interest because it implies ergodicity for a stationary channel thanks to the following lemma which is Lemma 2.7 of [10].

Lemma 21. *Let $[A, \nu, B]$ be a channel stationary w.r.t. an ergodic and stationary source distribution μ. If $[A, \nu, B]$ is output weakly mixing then $[A, \nu, B]$ is ergodic w.r.t. μ*

Proposition 7. *Let $[A, \nu_1, B]$ and $[A, \nu_2, C]$ be channels stationary w.r.t. a stationary source distribution μ. $[A, \nu_1 \times \nu_2, B \times C]$ is output weakly mixing w.r.t. μ if and only if $[A, \nu_1, B]$ and $[A, \nu_2, C]$ are output weakly mixing w.r.t. μ.*

The proof relies of the following lemmas.

Lemma 22. *Let (a_n), (b_n), (c_n) and (d_n) be bounded sequences of real numbers such that $\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} |a_i - b_i| = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} |c_i - d_i| = 0$. Then $\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} |a_i.c_i - b_i.d_i| = 0$.*

Proof.

$$\frac{1}{n}\sum_{i=0}^{n-1} |a_i.c_i - b_i.d_i| = \frac{1}{n}\sum_{i=0}^{n-1} |a_i.c_i - b_i.c_i + b_i.c_i - b_i.d_i|$$

$$\leq \frac{1}{n}\sum_{i=0}^{n-1} |c_i|.|a_i - b_i| + \frac{1}{n}\sum_{i=0}^{n-1} |b_i|.|c_i - d_i|$$

b_i , c_i are bounded then the RHS tends to 0 when $n \to \infty$. □

Lemma 23. *If a probability μ is AMS and if $\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} |\mu(F \cap T_A^{-i}G) - \mu(F).\mu(T_A^{-i}G)| = 0$ for any sets F and G of a generating field, then μ is weakly mixing.*

Proof. This is Lemma 8.1 of [9] (page 246). □

Proof of Proposition 7. Assume that $[A, \nu_1 \times \nu_2, B \times C]$ is output weakly mixing. Then, by definition, $\forall E, E' \in \mathcal{B}_{B^{\mathcal{I}} \times C^{\mathcal{I}}}$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \Big|\nu_1 \times \nu_2(x, T_{BC}^{-i}E \bigcap E') - \nu_1 \times \nu_2(x, T_{BC}^{-i}E).\nu_1 \times \nu_2(x, E')\Big| = 0$$

Then choosing $E = G \times C^{\mathcal{I}}$ and $E' = G' \times C^{\mathcal{I}}$ where $G, G' \in \mathcal{B}_{B^{\mathcal{I}}}$ (resp. $E = B^{\mathcal{I}} \times H$ and $E' = B^{\mathcal{I}} \times H'$ where $H, H' \in \mathcal{B}_{C^{\mathcal{I}}}$) proves that ν_1 (resp. ν_2) is output weakly mixing.

Assume that $[A, \nu_1, B]$ and $[A, \nu_2, C]$ are output weakly mixing. Then (all statements are μ-a.e.):

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \big|\nu_1(x, T_B^{-i}G \cap G') - \nu_1(x, T_B^{-i}G).\nu_1(x, G')\big| = 0$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\left|\nu_2(x,{T_C}^{-i}H\cap H')-\nu_2(x,{T_C}^{-i}H).\nu_2(x,H')\right|=0$$

Then, by Lemma 22 (p. 56), with

$$a_n=\nu_1(x,{T_B}^{-n}G\cap G'),\ b_n=\nu_1(x,{T_B}^{-n}G).\nu_1(x,G')$$

and

$$c_n=\nu_2(x,{T_C}^{-n}H\cap H'),\ d_n=\nu_2(x,{T_C}^{-i}H).\nu_2(x,H')$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\Big|\nu_1\times\nu_2(x,T_{BC}^{-i}(G\times H)\bigcap(G'\times H')) \\ -\nu_1\times\nu_2(x,T_{BC}^{-i}(G\times H)).\nu_1\times\nu_2(x,(G'\times H'))\Big|=0$$

Thus, $\nu_1\times\nu_2(x,.)$ is weakly mixing on the semi-algebra of the rectangles $\mathcal{B}_{B^{\mathcal{I}}}\times\mathcal{B}_{C^{\mathcal{I}}}$.

Any element of the algebra generated by the rectangles is a finite union of disjoint rectangles. Let E and E' two elements of this algebra. Then

$$E=\cup_{k=1}^{K}G_k\times H_k,\ E'=\cup_{l=1}^{L}G'_l\times H'_l$$

The unions are disjoint then:

$$(\nu_1\times\nu_2)(x,(T_{BC}^{-i}E)\cap E')=\sum_{k=1}^{K}\sum_{l=1}^{L}(\nu_1\times\nu_2)(x,(T_{BC}^{-i}G_k\times H_k)\cap G'_l\times H'_l)$$

and

$$(\nu_1\times\nu_2)(x,T_{BC}^{-i}E).(\nu_1\times\nu_2)(x,E')= \\ \sum_{k=1}^{K}\sum_{l=1}^{L}(\nu_1\times\nu_2)(x,T_{BC}^{-i}G_k).(\nu_1\times\nu_2)(x,G'_l\times H'_l)$$

Then

$$\frac{1}{n}\sum_{i=0}^{n-1}|(\nu_1\times\nu_2)(x,(T_{BC}^{-i}E)\cap E')-(\nu_1\times\nu_2)(x,T_{BC}^{-i}E).(\nu_1\times\nu_2)(x,E')|\leq \\ \sum_{k=1}^{K}\sum_{l=1}^{L}\frac{1}{n}\sum_{i=0}^{n-1}|(\nu_1\times\nu_2)(x,(T_{BC}^{-i}G_k\times H_k)\cap G'_l\times H'_l)- \\ (\nu_1\times\nu_2)(x,T_{BC}^{-i}G_k).(\nu_1\times\nu_2)(x,G'_l\times H'_l)|$$

This obviously implies that $\nu_1 \times \nu_2(x,.)$ is weakly mixing on the algebra generated by $\mathcal{B}_{B^{\mathcal{I}}} \times \mathcal{B}_{C^{\mathcal{I}}}$.

Moreover, since $\nu_1 \times \nu_2$ is a stationary channel (Proposition 4, p. 52), $\nu_1 \times \nu_2(x,.)$ is AMS μ-a.e. (Lemma 7, p. 17), by Lemma 23 (p. 56), $\nu_1 \times \nu_2(x,.)$ is weakly mixing on the σ-field $\mathcal{B}_{B^{\mathcal{I}} \times C^{\mathcal{I}}}$. □

9.4. Noisy computations

The next corollary immediately follows from Proposition 4 (p. 52) and Proposition 5 (p. 52).

Corollary 3. *A noisy computation $[A, f \times F, B \times C]$ is AMS (resp. stationary) w.r.t a stationary source distribution μ if and only if the deterministic channel $[A, f, B]$ and the random channel $[A, F, C]$ are both AMS (resp. stationary) w.r.t μ.*

The next corollary immediately follows from Proposition 6 (p. 54).

Corollary 4. *A noisy computation $[A, f \times F, B \times C]$ is recurrent w.r.t. a recurrent source distribution μ if and only if the deterministic channel $[A, f, B]$ and the random channel $[A, F, C]$ are both recurrent w.r.t. μ.*

The next corollary immediately follows from Proposition 7 (p. 56).

Corollary 5. *A noisy computation $[A, f \times F, B \times C]$, stationary w.r.t. a stationary source distribution μ, is output weakly mixing w.r.t μ if and only if the deterministic channel $[A, f, B]$ and the random channel $[A, F, C]$ are both output weakly mixing w.r.t. μ.*

10. Sufficient Conditions for Ergodicity of Channel Products and Noisy Computations

The objective is to identify conditions on channels to get an ergodic channel product. It is obvious that ergodicity of ν_1 and ν_2 is a necessary condition for $\nu_1 \times \nu_2$ to be ergodic. It cannot be a sufficient condition since the product of ergodic probabilities may not be ergodic.

In this section, it is shown that, if the product $[A, S\overline{\nu_1}_{\overline{\mu}} \times S\overline{\nu_2}_{\overline{\mu}}, B \times C]$ is ergodic w.r.t to the stationary mean $\overline{\mu}$ of an AMS source distribution μ, the product of the two AMS (w.r.t. $\overline{\mu}$) channels $[A, \nu_1, B]$ and $[A, \nu_2, C]$ is ergodic w.r.t. μ.

It should be noticed that, thanks to Lemma 2 (p. 9), asymptotic mean stationarity and ergodicity jointly imply recurrence and, thus, ergodic AMS sources and channels are ergodic R-AMS. This leads to the fact that the AMS probabilities, channels and (by Proposition 8 below) channel products under consideration are dominated and not only asymptotically dominated by their stationary means.

10.1. Channel products

Proposition 8. *Let $[A, X]$ be an AMS source with distribution μ. Let $[A, \nu_1, B]$ and $[A, \nu_2, C]$ be two channels AMS w.r.t the stationary mean $\overline{\mu}$ of μ. Then ν_1, ν_2 and $\nu_1 \times \nu_2$ are AMS w.r.t. μ.*

If, in addition, μ is ergodic (or equivalently $\overline{\mu}$ is ergodic) and if $[A, S\overline{\nu_1}_{\overline{\mu}} \times S\overline{\nu_2}_{\overline{\mu}}, B \times C]$ is ergodic w.r.t. $\overline{\mu}$ then $[A, \nu_1 \times \nu_2, B \times C]$ is ergodic w.r.t. μ and w.r.t. $\overline{\mu}$ (where $\overline{\nu_1}_{\overline{\mu}}$ and $\overline{\nu_2}_{\overline{\mu}}$ are respectively the stationary means of ν_1 and of ν_2 w.r.t to $\overline{\mu}$).

In this case

$$\overline{(\nu_1 \times \nu_2)}_{\mu} = S\overline{\nu_1}_{\mu} \times S\overline{\nu_2}_{\mu} = \overline{(\nu_1 \times \nu_2)}_{\overline{\mu}} = S\overline{\nu_1}_{\overline{\mu}} \times S\overline{\nu_2}_{\overline{\mu}} \ \overline{\mu}\text{-a.e.}$$

If $[A, S\overline{\nu_1}_{\overline{\mu}}, B]$ and $[A, S\overline{\nu_2}_{\overline{\mu}}, C]$ are output weakly mixing w.r.t. $\overline{\mu}$ then $[A, \nu_1 \times \nu_2, B \times C]$ is ergodic w.r.t. μ.

Proof. μ is AMS then $\mu \ll^a \overline{\mu}$. By Lemma 1 (p. 9), $\mu_\infty \ll \overline{\mu}_\infty$. Then, by Lemma 8 (p. 19), $\mu_\infty \nu_{1_\infty} \ll \overline{\mu}_\infty \nu_{1_\infty}$. Since ν_1 is AMS w.r.t. $\overline{\mu}$, and thanks to Lemma 1 (p. 9), $\overline{\mu}_\infty \nu_{1_\infty} \ll \overline{\mu}_\infty \overline{\nu}_{1_{\overline{\mu}_\infty}}$. Then $\mu_\infty \nu_{1_\infty} \ll \overline{\mu}_\infty \overline{\nu}_{1_{\overline{\mu}_\infty}}$. This implies that ν_1 is AMS w.r.t. μ. The same holds for ν_2.

ν_1 and ν_2 are AMS w.r.t. the stationary probability $\overline{\mu}$ then, thanks to Proposition 5 (p. 52), $\nu_1 \times \nu_2$ is AMS w.r.t. $\overline{\mu}$ and (from above) w.r.t. μ.

Assume now that, in addition, μ (or equivalently $\overline{\mu}$) is ergodic. Thanks to Lemma 2 (p. 9), μ is R-AMS and then $\mu \ll \overline{\mu}$. By Lemma 8 (p. 19), this implies that $\mu(\nu_1 \times \nu_2) \ll \overline{\mu}(\nu_1 \times \nu_2)$.

Tanks to Proposition 5 (p. 52),

$$\mu(\nu_1 \times \nu_2) \ll \overline{\mu}(\nu_1 \times \nu_2) \ll^a \overline{\mu}(S\overline{\nu}_{1_{\overline{\mu}}} \times S\overline{\nu}_{2_{\overline{\mu}}})$$

Then, if $\overline{\mu}(S\overline{\nu}_{1_{\overline{\mu}}} \times S\overline{\nu}_{2_{\overline{\mu}}})$ is ergodic, $\mu(\nu_1 \times \nu_2)$ and $\overline{\mu}(\nu_1 \times \nu_2)$ are ergodic.

$\overline{\mu}(\nu_1 \times \nu_2)$ is ergodic then $\overline{\mu}\nu_1$ is ergodic: ν_1 is AMS and ergodic w.r.t $\overline{\mu}$. This implies that $\overline{\mu}\nu_1 \ll \overline{\mu}\overline{\nu}_{1_{\overline{\mu}}}$ i.e.

$$\nu_1(x, .) \ll \overline{\nu}_{1_{\overline{\mu}}}(x, .)\ \overline{\mu}\text{-a.e.}$$

But $\overline{\nu}_{1_{\overline{\mu}}}(x, .) \ll S\overline{\nu}_{1_{\overline{\mu}}}(x, .)$ $\overline{\mu}$-a.e.. Then

$$\nu_1(x, .) \ll S\overline{\nu}_{1_{\overline{\mu}}}(x, .)\ \overline{\mu}\text{-a.e.}$$

The same holds for ν_2:

$$\nu_2(x, .) \ll S\overline{\nu}_{2_{\overline{\mu}}}(x, .)\ \overline{\mu}\text{-a.e.}$$

Then, by Lemma 20 (p. 53), since $S\overline{\nu}_{2_{\overline{\mu}}}(x, .)$ and $S\overline{\nu}_{2_{\overline{\mu}}}(x, .)$ are stationary:

$$\nu_1(x, .) \times \nu_2(x, .) \ll S\overline{\nu}_{1_{\overline{\mu}}}(x, .) \times S\overline{\nu}_{2_{\overline{\mu}}}(x, .)\ \overline{\mu}\text{-a.e.}$$

This implies that:

$$\overline{\mu}(\nu_1 \times \nu_2) \ll^a \overline{\mu}(S\overline{\nu}_{1_{\overline{\mu}}} \times S\overline{\nu}_{2_{\overline{\mu}}})$$

and thus

$$\mu(\nu_1 \times \nu_2) \ll \overline{\mu}(\nu_1 \times \nu_2) \ll \overline{\mu}(S\overline{\nu}_{1_{\overline{\mu}}} \times S\overline{\nu}_{2_{\overline{\mu}}})$$

From [9], an AMS probability is ergodic if and only if its stationary mean is ergodic. Then $\overline{\mu}\overline{(\nu_1 \times \nu_2)}_{\overline{\mu}}$ and $\overline{\mu}(S\overline{\nu_1}_{\overline{\mu}} \times S\overline{\nu_2}_{\overline{\mu}})$ are two ergodic stationary probabilities on the same space. They are thus equal or mutually singular. Since they both dominate the same probability (i.e. $\mu(\nu_1 \times \nu_2)$), they are equal. Then

$$\overline{(\nu_1 \times \nu_2)}_{\overline{\mu}} = S\overline{\nu_1}_{\overline{\mu}} \times S\overline{\nu_2}_{\overline{\mu}}\ \overline{\mu}\text{-a.e.}$$

Similarly, $\overline{\mu}\ \overline{\nu_1}_{\mu}$, $\overline{\mu}S\overline{\nu_1}_{\mu}$, $\overline{\mu}\ \overline{\nu_1}_{\overline{\mu}}$, $\overline{\mu}S\overline{\nu_1}_{\overline{\mu}}$ are stationary and ergodic and they dominate the same probability $\mu\nu_1$ (which is AMS ergodic and thus recurrent). They are thus equal. The same holds for ν_2. Then

$$\overline{\nu_1}_{\mu} = S\overline{\nu_1}_{\mu} = \overline{\nu_1}_{\overline{\mu}} = S\overline{\nu_1}_{\overline{\mu}}\ \overline{\mu}\text{-a.e.}$$

and

$$\overline{\nu_2}_\mu = S\overline{\nu_2}_\mu = \overline{\nu_2}_{\overline{\mu}} = S\overline{\nu_2}_{\overline{\mu}} \ \overline{\mu}\text{-a.e.}$$

Finally, if $[A, S\overline{\nu_1}_{\overline{\mu}}, B]$ and $[A, S\overline{\nu_2}_{\overline{\mu}}, C]$ are output weakly mixing w.r.t. $\overline{\mu}$ then, by Proposition 7 (p. 56), $[A, S\overline{\nu_1}_{\overline{\mu}} \times S\overline{\nu_2}_{\overline{\mu}}, B \times C]$ is output weakly mixing w.r.t. $\overline{\mu}$ and thus ergodic w.r.t. $\overline{\mu}$. This implies, from above, that $[A, \nu_1 \times \nu_2, B \times C]$ is ergodic w.r.t. μ. □

10.2. Noisy computations

The next corollary immediately follows from Proposition 8 (p. 59).

Corollary 6. *Let $[A, f \times F, B \times C]$ be a noisy computation AMS w.r.t. the stationary mean $\overline{\mu}$ of an ergodic AMS source distribution μ.*

If the channel product $[A, S\overline{f}_{\overline{\mu}} \times S\overline{F}_{\overline{\mu}}, B \times C]$ is ergodic w.r.t. $\overline{\mu}$ then the noisy computation $[A, f \times F, B \times C]$ is ergodic w.r.t. μ and w.r.t $\overline{\mu}$.

If $[A, S\overline{f}_{\overline{\mu}}, B]$ and $[A, S\overline{F}_{\overline{\mu}}, B]$ are output weakly mixing w.r.t. $\overline{\mu}$ then the noisy computation $[A, f \times F, B \times C]$ is ergodic w.r.t. μ.

11. Noisy computations and computable functions

In this section, a way to instantiate the noisy computation model to the case of computable functions is proposed. First, it is shown that a function on finite length symbol sequences (thus which can be a computable function) f induces a deterministic channel f^* (equivalently a measurable function on infinite length sequences). This deterministic channel f^* has the obvious property that, for any input sequence x, the output sequence $f^*(x)$ is made the concatenation of a finite length sequence and of an infinite sequence of blank symbols.

Secondly, it is shown that any measurable function g on infinite length sequences (i.e., deterministic channel) possesses stability properties (AMS, mixing) with respect to sources which produce, almost surely, input sequences x such that $g(x)$ is made the concatenation of a finite length sequence and of an infinite sequence of blank symbols. This allows to conclude that the deterministic channel f^*, induced by a computable function f, is AMS and mixing w.r.t any AMS and ergodic source μ.

Then, given an AMS and ergodic source distribution μ, for any random channel $[A, F, C]$ AMS w.r.t. $\overline{\mu}$ such that $[A, \overline{F}_{\overline{\mu}}, C]$ is output weakly mixing w.r.t. $\overline{\mu}$, the noisy computation $[A, f^* \times F, B \times C]$ is AMS and ergodic w.r.t μ. In other words, for a computable function f, it is sufficient to focus on stability properties of the noisy realization F under concern in order to apply the coding results given above.

This section ends with very simple examples.

11.1. Linking the noisy computation model to a formal computational model and its noisy version

When dealing with a function f computable according to a formal computational model such as Turing machines, Moore automata, ..., it will be necessary to associate to f, which is a function from a set of finite length sequences A^+ to a set of finite length sequences B^+, a function f^* from $A^{\mathbb{N}}$ to $B^{\mathbb{N}}$ defining a deterministic channel $[A, f^*, B]$. A finite length sequence of $u \in A^+$, on which f is defined, will be associated to an infinite length sequence x by concatenating an infinite sequence of blank symbols to u. This association is obviously a measurable function ϕ. This can be viewed as writing a finite length string on a semi-infinite tape with cells initially containing blank symbols. Reading the output of such a computing formal device can be viewed as extracting the prefix u of an infinite length sequence made of u

concatenated to an infinite length sequence of blank symbols. This can also be associated to a measurable function ψ.

Given two measurable finite alphabets A and B, let $(A^+, \mathcal{P}(A^+))$ and $(B^+, \mathcal{P}(B^+))$ be the corresponding measurable spaces of finite length sequences, $\mathcal{P}(A^+)$ being the set of subsets of A^+. If u is a finite string of symbols and v is a finite or countable string of symbols, uv denotes the concatenation of u and v, $|u|$ is the length of u. If s is a symbol, s^∞ denotes the countable string $s \cdots s \cdots$. A^+ and B^+ are assumed to contain an empty string denoted ϵ for both sets ($|\epsilon| = 0$).

Let L be a subset of A^+ such that $\forall x \in L, x_{|x|} \neq \alpha$ and $f : L \to B^+$ be a function. By convention, $f(\epsilon) = \epsilon$. Let $\alpha \in A$ called the blank symbol of A and $\beta \in B$ called the blank symbol of B.

Let $\psi_\alpha^A : A^\mathbb{N} \to A^+$ be the function such that, $\forall x = x_1 \cdots x_n \cdots \in A^\mathbb{N}$,

$$\psi_\alpha^A(x) = \begin{cases} x_1 \cdots x_n & \text{if } x = x_1 \cdots x_n \alpha^\infty \text{ and } x_1 \cdots x_n \in L \setminus \{\epsilon\} \\ \epsilon & \text{else} \end{cases}$$

Let $\phi_\alpha^A : A^+ \to A^\mathbb{N}$ be the function such that, $\forall x = x_1 \cdots x_{|x|} \in A^+$,

$$\phi_\alpha^A(x) = x\alpha^\infty$$

Let $\psi_\beta^B : B^\mathbb{N} \to B^+$ be the function such that, $\forall y = y_1 \cdots y_n \cdots \in B^\mathbb{N}$,

$$\psi_\beta^B(y) = \begin{cases} y_1 \cdots y_n & \text{if } y = y_1 \cdots y_n \beta^\infty, y_n \neq \beta \\ \epsilon & \text{else} \end{cases}$$

Let $\phi_\beta^B : B^+ \to B^\mathbb{N}$ be the function such that, $\forall y = y_1 \cdots y_{|y|} \in B^+$,

$$\phi_\beta^B(y) = y\beta^\infty$$

Lemma 24. *The function $f^* = \phi_\beta^B \circ f \circ \psi_\alpha^A$ is $(\mathcal{B}_{A^\mathbb{N}}, \mathcal{B}_{B^\mathbb{N}})$-measurable and thus induces a deterministic channel $[A, f^*, B]$. Moreover:*

- *if $x = x_0 \dots x_n \alpha^\infty$ where $x_0 \dots x_n \in L$ then $f^*(x) = f(x_0 \dots x_n)\beta^\infty$, else $f^*(x) = \beta^\infty$*
- *for any $x_0 \dots x_n \in L$, $f(x_0 \dots x_n) = \psi_\beta^B \circ f^* \circ \phi_\alpha^A(x_0 \dots x_n)$*
- *for any $x_0 \dots x_n \notin L$, $\psi_\beta^B \circ f^* \circ \phi_\alpha^A(x_0 \dots x_n) = \epsilon$*
- *$\forall x \in A^\mathbb{N}, \exists N_x$ such that $\forall n > N_x, (f^*(x))_n = \beta$*

$$\begin{tikzcd}
(A^+,\mathcal{P}(A^+)) \arrow[r, "f"] \arrow[d, shift left, "\phi_\alpha^A"] & (B^+,\mathcal{P}(B^+)) \arrow[d, shift left, "\phi_\beta^B"] \\
(A^{\mathbb{N}},\mathcal{B}_{A^{\mathbb{N}}}) \arrow[r, "f^*"] \arrow[u, shift left, "\psi_\alpha^A"] & (B^{\mathbb{N}},\mathcal{B}_{B^{\mathbb{N}}}) \arrow[u, shift left, "\psi_\beta^B"]
\end{tikzcd}$$

Figure 6: A function f on finite sequences and the induced deterministic channel $[A, f^*, B]$

Proof. Obviously a function $f : L \to B^+$ is $(\mathcal{P}(A^+), \mathcal{P}(B^+))$-measurable, ϕ_α^A is $(\mathcal{P}(A^+), \mathcal{B}_{A^{\mathbb{N}}})$-measurable and ϕ_β^B is $(\mathcal{P}(B^+), \mathcal{B}_{B^{\mathbb{N}}})$-measurable.

It remains to prove that ψ_α^A is $(\mathcal{B}_{A^{\mathbb{N}}}, \mathcal{P}(A^+),)$-measurable (the proof also applies to ψ_β^B).

Let $x_1 \cdots x_n \in A^+$. Then

$$\psi_\alpha^{-1}(x_1 \cdots x_n) = \{x_1 \cdots x_n \alpha^\infty\} = c_1^n(x_1 \cdots x_n) \bigcap \cap_{i>n} c_i^i(\alpha)$$

or

$$\psi_\alpha^{-1}(x_1 \cdots x_n) = \emptyset$$

where $c_1^n(x_1 \cdots x_n) = \{a \in A^{\mathbb{N}}, a_i = x_i, i = 1, \cdots, n\}$ is the thin cylinder induced by $(x_1 \cdots x_n)$.

$\psi_\alpha^{-1}(x_1 \cdots x_n)$ is thus a countable intersection of thin cylinders which are measurable. Then $\psi_\alpha^{-1}(x_1 \cdots x_n)$ is measurable. Any subset E of A^+ is countable then, $\forall E \neq \{\alpha\}$, $\psi_\alpha^{-1}(E) = \cup_{x \in E} \psi^{-1}(x)$ is measurable.

The last statements of the proposition are immediate from the definition of the functions ϕ_α^A, ϕ_β^B, ψ_α^A and ψ_β^B □

Let $D_\beta(f^*)$ be the set:

$$D_\beta(f^*) = \{x \in A^{\mathbb{N}} / \exists N_x \in \mathbb{N} \text{ such that } \forall n > N_x, (f^*(x))_n = \beta\}$$

From Lemma 24 (p. 63), $D_\beta(f^*) = A^{\mathbb{N}}$. Thus, for any probability μ on $(A^{\mathbb{N}}, \mathcal{B}_{A^{\mathbb{N}}})$, $\mu(D_\beta(f^*)) = 1$.

Associated to a formal computational model (e.g., circuits made of binary gates from a complete basis), a noisy computational model can be considered in order to capture unreliability of an actual implementation of a function (e.g., circuits made of noisy binary gates, [16]). It is also relevant to consider a noisy implementation of a computable function as a channel taking infinite length sequences as inputs and producing infinite length sequences as outputs. An input finite sequence u can be considered as an infinite one i.e. $u\alpha^\infty$ (see above). It the noisy device halts on an input, the finite output sequence v can also be considered as the infinite one $v\beta^\infty$. If the noisy device doest not halt on the input $u\alpha^\infty$ then its output is an infinite sequence z.

The choice of a peculiar noisy computational model is in fact the choice of probabilistic assumptions intended to capture the nature of faults of a technology in use. A noisy implementation of a function, in such a noisy computational model, can also be described as a random channel whose probabilistic properties are a consequence of the probabilistic assumptions at the base of the noisy computational model. For example, in the case of circuits based on noisy boolean gates, if the gates behave independently and have constant probability of failure, a circuit made of such noisy gates will behave as a stationary memoryless random channel.

To summarize, the realization F of a computable function f built according to a noisy computational model induces a noisy computation $[A, f^* \times F, B \times C]$ where the characteristics of the random channel $[A, F, C]$ depends on the noisy computational model. Moreover, for any input source with distribution μ, $\mu(D_\beta(f^*)) = 1$.

11.2. Stability properties of a deterministic channel f^* w.r.t a source μ such that $\mu(D_\beta(f^*)) = 1$

It is shown below that, if f^* is a $(\mathcal{B}_{A^\mathbb{N}}, \mathcal{B}_{B^\mathbb{N}})$-measurable function, the corresponding deterministic channel $[A, f^*, B]$ possesses some stability properties w.r.t probabilities μ such that $\mu(D_\beta(f^*)) = 1$ (β is a fixed symbol of B).

In the following, the shift T_B is assumed non-invertible: $\forall G \in \mathcal{B}_{B^\mathbb{N}}$, $T_B^{-1}(G) = B \times G$.

Lemma 25. *Let $f^* : A^\mathbb{N} \to B^\mathbb{N}$ be a $(\mathcal{B}_{A^\mathbb{N}}, \mathcal{B}_{B^\mathbb{N}})$-measurable function such that $D_\beta(f^*)$ is non-empty. Let $\{f^*(x, .), x \in A^\mathbb{N}\}$ denote the probability kernel defining the deterministic channel $[A, f^*, B]$. Then*

1. *for any $x \in D_\beta(f^*)$, $(f^*(x, .))_\infty$ is a stationary probability on the space $(B^\mathbb{N}, (\mathcal{B}_{B^\mathbb{N}})_\infty, T_B)$*
2. *if μ is a stationary probability on $(A^\mathbb{N}, \mathcal{B}_{A^\mathbb{N}}, T_A)$ such that $\mu(D_\beta(f^*)) = 1$ then the channel $[A, f^*, B]$ is AMS w.r.t. μ*
3. *if μ is a R-AMS probability on $(A^\mathbb{N}, \mathcal{B}_{A^\mathbb{N}}, T_A)$ such that $\overline{\mu}(D_\beta(f^*)) = 1$ then the channel $[A, f^*, B]$ is AMS w.r.t. μ*
4. *if μ is a recurrent probability on $(A^\mathbb{N}, \mathcal{B}_{A^\mathbb{N}})$ such that $\mu(D_\beta(f^*)) = 1$ then the channel $[A, f^*, B]$ is recurrent w.r.t. μ*

5. *if μ is an R-AMS probability on $(A^{\mathbb{N}}, \mathcal{B}_{A^{\mathbb{N}}})$ such that $\overline{\mu}(D_\beta(f^*)) = 1$ then the channel $[A, f^*, B]$ is R-AMS w.r.t. μ*
6. *if μ is an AMS probability on $(A^{\mathbb{N}}, \mathcal{B}_{A^{\mathbb{N}}}, T_A)$ such that $\overline{\mu}(D_\beta(f^*)) = 1$ then the channel $[A, f^*, B]$ is AMS w.r.t. μ*

Proof. 1. Let $x \in D_\beta(f)$. Then $f^*(x) = y_1 \cdots y_{N_x} \beta \cdots \beta \cdots$.
Let $G \in (\mathcal{B}_{B^{\mathbb{N}}})_\infty$. Then, $\forall i \geq 0$, there exists $G_i \in \mathcal{B}_{B^{\mathbb{N}}}$ such that $G = T_B^{-i} G_i$. Since T_B is the non-invertible shift, $T_B^{-1} G_i = B \times G_i$. Then $G = B^i \times G_i$.
In particular, there exists $G_{N_x} \in \mathcal{B}_{B^{\mathbb{N}}}$ such that $G = B^{N_x} \times G_{N_x}$ and $T_B^{-1} G = B^{N_x+1} \times G_{N_x}$. In fact $G_{N_x+1} = B \times G_{N_x}$ i.e. $G_{N_x+1} = T_B^{-1} G_{N_x}$.
f^* defines a deterministic channel then $f^*(x, G) = 1$ if and only if $f^*(x) = y_1 \cdots y_{N_x} \beta \cdots \beta \cdots \in G = B^{N_x} \times G_{N_x}$. $y_1 \cdots y_{N_x} \in B^{N_x}$ is obviously always true, then $f^*(x, G) = 1$ if and only if $\beta^\infty = \beta \cdots \beta \cdots \in G_{N_x}$. Moreover $\beta^\infty \in G_{N_x} \Rightarrow \beta^\infty \in B \times G_{N_x}$.Thus $f^*(x, G) = 1$ implies $\beta^\infty \in B^{N_x+1} \times G_{N_x} = G_{N_x+1}$ which implies that $f^*(x) \in T_B^{-1} G$ then $f^*(x, T_B^{-1} G) = 1$.
If $f^*(x, G) = 0$ then $f^*(x) = y_1 \cdots y_{N_x} \beta \cdots \beta \cdots \notin G = B^{N_x} G_{N_x}$. This implies that $\beta^\infty \notin G_{N_x}$ and thus $\beta^\infty \notin B \times G_{N_x}$. Hence $f^*(x) \notin T_B^{-1} G$ which is equivalent to $f^*(x, T_B^{-1} G) = 0$.
Thus, in any case,$\forall G \in (\mathcal{B}_{B^{\mathbb{N}}})_\infty$, $f^*(x, G) = f^*(x, T_B^{-1} G)$.

2. by 1), $\forall x \in D_\beta(f^*)$, $f^*(x, .)_\infty$ is stationary thus R-AMS. Then, there exits a stationnary probability η such that $f^*(x, .)_\infty \ll \eta_\infty$ (e.g. let η' be a stationary probability on $(B^{\mathbb{N}}, \mathcal{B}_{B^{\mathbb{N}}})$, set $\eta = \eta'$ on $\mathcal{B}_{B^{\mathbb{N}}} \setminus (\mathcal{B}_{B^{\mathbb{N}}})_\infty$ and $\eta = f^*(x, .)_\infty$ on $(\mathcal{B}_{B^{\mathbb{N}}})_\infty)$. Then, by Lemma 1 (p. 9), $f^*(x, .)$ is AMS. Since $\mu(D_\beta(f^*)) = 1$, $f^*(x, .)$ is AMS μ-a.e. Thanks to Lemma 11 (p. 19), this implies that f^* is AMS w.r.t. μ.
3. μ is R-AMS thus μ is dominated by its stationary mean: $\mu \ll \overline{\mu}$. Then, by Lemma 8 (p. 19), $\mu f^* \ll \overline{\mu} f^*$. By 2), f^* is AMS w.r.t $\overline{\mu}$: $\overline{\mu} f^* \ll^a \overline{\overline{\mu} f^*}_{\overline{\mu}}$. Hence $\mu f^* \ll^a \overline{\overline{\mu} f^*}_{\overline{\mu}}$ i.e. μf^* is AMS. f^* is thus AMS w.r.t. μ.
4. $\forall G \in \mathcal{B}_{B^{\mathbb{N}}}$, $T_B^{-i} G = B^i \times G$. Let $G \in \mathcal{B}_{B^{\mathbb{N}}}$ be such that $T_B^{-1} G \subset G$. Then $\forall i$, $T_B^{-i} G = B^i \times G \subset G$. This implies that $\forall i$ there exists $G_i \in \mathcal{B}_{B^{\mathbb{N}}}$ such that $G = B^i \times G_i$, then $G \in (\mathcal{B}_{B^{\mathbb{N}}})_\infty$. Thus, by 1), $f^*(x, G) = f^*(x, T_B^{-1} G)$. Then $f^*(x, .)$ is incompressible, equivalently recurrent.
5. obvious from 3) and 4)

6. μ is AMS then $\mu \ll^a \overline{\mu}$. Then by Lemma 1 (p. 9), $\mu_\infty \ll \overline{\mu}_\infty$. By Lemma 8 (p. 19), $\mu_\infty f^*_\infty \ll \overline{\mu}_\infty f^*_\infty$. By 2), $\overline{\mu} f^*$ is AMS thus $\overline{\mu}_\infty f^*_\infty \ll \overline{\mu}_\infty \overline{f^*}_{\overline{\mu}_\infty}$. This implies that $\mu_\infty f^*_\infty \ll \overline{\mu}_\infty \overline{f^*}_{\overline{\mu}_\infty}$ and thus (since is $\overline{\mu}\,\overline{f^*}_{\overline{\mu}}$ is stationary), by Lemma 1 (p. 9), $\mu f^* \ll^a \overline{\mu}\,\overline{f^*}_{\overline{\mu}}$. Hence, f^* is AMS w.r.t. μ

□

Lemma 26. *Let μ be the distribution of a stationary and ergodic source $[A, X]$. Let $f^* : A^{\mathbb{N}} \to B^{\mathbb{N}}$ be a $(\mathcal{B}_{A^{\mathbb{N}}}, \mathcal{B}_{B^{\mathbb{N}}})$-measurable function such that $\mu(D_\beta(f^*)) = 1$. Since, by Lemma 25 (p. 65), the stationary mean $[A, \overline{f^*}_\mu, B]$ exists, then*

1. *$\forall G \in \mathcal{B}_{B^{\mathbb{N}}}$, $\overline{f^*}_\mu(x, G) = 1$ if and only if $\beta^\infty \in G$ and $\overline{f^*}_\mu(x, G) = 0$ if and only if $\beta^\infty \notin G$, μ-a.e.*
2. *$\forall G, G' \in \mathcal{B}_{B^{\mathbb{N}}}$, $\overline{f^*}_\mu(x, G \cap G') = \overline{f^*}_\mu(x, G).\overline{f^*}_\mu(x, G')$ μ-a.e.*
3. *$\overline{f^*}_\mu(x, .)$ is stationary μ-a.e. and consequently $\overline{f^*}_\mu(x, .) = S\overline{f^*}_\mu(x, .)$ μ-a.e. and $[A, \overline{f^*}_\mu, B] = [A, S\overline{f^*}_\mu, B]$*
4. *$\overline{f^*}_\mu(x, .)$ is strongly mixing μ-a.e.*

Proof. From Lemma 25 (p. 65), the deterministic channel $[A, f^*, B]$ is R-AMS w.r.t. μ. Then the channels $[A, \overline{f^*}_\mu, B]$ and $[A, S\overline{f^*}_\mu, B]$ exist.

1. Since $\overline{\mu f^*} = \mu\overline{f^*}_\mu$, $\forall E \in \mathcal{B}_{A^{\mathbb{N}}}$ and $\forall G \in \mathcal{B}_{B^{\mathbb{N}}}$

$$\int_E \overline{f^*}_\mu(x, G) d\mu = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \int f^*(x, T_B^{-i} G) 1_E(T_A^i x) d\mu \qquad (8)$$

$\forall i \in \mathbb{N}$, $f^*(x, T_B^{-i} G) = 1_G(T_B^i f^*(x))$ and $\forall x \in D_\beta(f^*)$, $f^*(x) = y_1 \cdots y_{N_x} \beta^\infty$. Then $\forall x \in D_\beta(f^*)$ and $\forall i > N_x$,

$$f^*(x, T_B^{-i} G) = 1 \text{ if and only if } \beta^\infty \in G$$

and

$$f^*(x, T_B^{-i} G) = 0 \text{ if and only if } \beta^\infty \notin G$$

Moreover

$$\forall i, j > N_x, f^*(x, T_B^{-i} G) = f^*(x, T_B^{-j} G)$$

For $n > N_x + 1$, let

$$\Delta_n(x, E, G) = \frac{1}{n} \sum_{i=0}^{n-1} f^*(x, T_B^{-i} G) 1_E(T_A^i x)$$

Then

$$\Delta_n(x, E, G) = \\ \frac{1}{n} \sum_{i=0}^{N_x} f^*(x, T_B^{-i} G) 1_E(T_A^i x) + \frac{1}{n} \sum_{i=N_x+1}^{n-1} f^*(x, T_B^{-i} G) 1_E(T_A^i x)$$

$\forall x \in D_\beta(f^*)$, $\forall n > N_x$ and $\forall G$ such that $\beta^\infty \notin G$

$$\Delta_n(x, E, G) = \frac{1}{n} \sum_{i=0}^{N_x} f^*(x, T_B^{-i} G) 1_E(T_A^i x)$$

Thus

$$\forall x \in D_\beta(f^*), \forall G \text{ such that } \beta^\infty \notin G, \ \lim_{n \to \infty} \Delta_n(x, E, G) = 0$$

$\forall x \in D_\beta(f^*)$, $\forall n > N_x$ and $\forall G$ such that $\beta^\infty \in G$

$$\Delta_n(x, E, G) = \frac{1}{n} \sum_{i=0}^{N_x} f^*(x, T_B^{-i} G) 1_E(T_A^i x) + \frac{1}{n} \sum_{i=N_x+1}^{n-1} 1_E(T_A^i x)$$

Thus

$$\lim_{n \to \infty} \Delta_n(x, E, G) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=N_x+1}^{n-1} 1_E(T_A^i x) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_E(T_A^i x)$$

Since μ is stationary and ergodic:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_E(T_A^i x) = \mu(E) \ \mu\text{-a.e.}$$

Thus, since $\mu(D_\beta(f^*)) = 1$, $\forall G$ such that $\beta^\infty \in G$,

$$\lim_{n \to \infty} \Delta_n(x, E, G) = \mu(E) \ \mu\text{-a.e.}$$

Thanks to the Bounded Convergence Theorem and (8)

$$\int_E \overline{f^*}_\mu(x,G)d\mu = \int \lim_{n\to\infty} \Delta_n(x,E,G)d\mu$$

$\mu(D_\beta(f^*)) = 1$, then

$$\int_E \overline{f^*}_\mu(x,G)d\mu = \int_{D_\beta(f^*))} \lim_{n\to\infty} \Delta_n(x,E,G)d\mu$$

Thus, if $\beta^\infty \in G$, $\forall E \in \mathcal{B}_{A^\mathbb{N}}$

$$\int_E \overline{f^*}_\mu(x,G)d\mu = \int_{D_\beta(f^*))} \mu(E)d\mu = \mu(E)\mu(D_\beta(f^*)) = \mu(E)$$

Then $\overline{f^*}_\mu(x,G) = 1$ μ-a.e.
If $\beta^\infty \notin G$, $\forall E \in \mathcal{B}_{A^\mathbb{N}}$

$$\int_E \overline{f^*}_\mu(x,G)d\mu = 0$$

Then $\overline{f^*}_\mu(x,G) = 0$ μ-a.e.

2. obvious from 1).
3. Let $x \in D_\beta(f^*)$. Let $G \in \mathcal{B}_{A^\mathbb{N}}$. Since $T_B^{-1}G = B \times G$, $\beta^\infty \in G \Leftrightarrow \beta^\infty \in T_B^{-1}G$. Thus, from 1)

$$\overline{f^*}_\mu(x,G) = 1 \Leftrightarrow \overline{f^*}_\mu(x,T_B^{-1}G) = 1$$

And $\beta^\infty \notin G \Leftrightarrow \beta^\infty \notin T_B^{-1}G$. Thus

$$\overline{f^*}_\mu(x,G) = 0 \Leftrightarrow \overline{f^*}_\mu(x,T_B^{-1}G) = 0$$

Then, $\forall x \in D_\beta(f^*)$, $\overline{f^*}_\mu(x,.)$ is stationary and thus is equal to its stationary mean $S\overline{f^*}_\mu(x,.)$. Since $\mu(D_\beta(f^*)) = 1$, $\overline{f^*}_\mu(x,.) = S\overline{f^*}_\mu(x,.)$ μ-a.e.
4. By 2), $|\overline{f^*}_\mu(x, G \cap T_B^{-n}G') - \overline{f^*}_\mu(x,G).\overline{f^*}_\mu(x,T_B^{-n}G')| = 0$ μ-a.e. Then, obviously, $\overline{f^*}_\mu(x,.)$ is strongly mixing, μ-a.e.

□

Proposition 9. *Let μ be the distribution of an AMS and ergodic source $[A, X]$. Let $f^* : A^{\mathbb{N}} \to B^{\mathbb{N}}$ be a $(\mathcal{B}_{A^{\mathbb{N}}}, \mathcal{B}_{B^{\mathbb{N}}})$-measurable function such that $\overline{\mu}(D_\beta(f^*)) = 1$. Then*

- *the channel $[A, f^*, B]$ is R-AMS and ergodic w.r.t. μ*
- *$[A, \overline{f^*}_\mu, B] = [A, \overline{f^*}_{\overline{\mu}}, B] = [A, S\overline{f^*}_\mu, B] = [A, S\overline{f^*}_{\overline{\mu}}, B]$ mod. $\overline{\mu}$ and mod. μ*
- *$\overline{f^*}_\mu(x, .)$ is an R-AMS and strongly mixing probability on $(B^{\mathbb{N}}, \mathcal{B}_{B^{\mathbb{N}}})$ $\overline{\mu}$-a.e. and μ-a.e.*

Proof. μ is AMS and ergodic thus, by Lemma 2 (p. 9), also recurrent. Since $\overline{\mu}(D_\beta(f^*)) = 1$, thanks to Lemma 25 (p. 65), f^* is R-AMS w.r.t. μ and w.r.t. $\overline{\mu}$.

$\overline{\mu}$ is stationary and ergodic, then, by Lemma 26 (p. 67), $\overline{f^*}_{\overline{\mu}}(x, .)$ is strongly (then weakly) mixing w.r.t. $\overline{\mu}$. By Lemma 21 (p. 55), this implies that $\overline{\mu}\overline{f^*}_{\overline{\mu}}$ is ergodic.

Meanwhile $\mu \ll \overline{\mu}$. Then, by Lemma 9 (p. 19), $\mu f^* \ll \overline{\mu} f^*$. Since f^* is R-AMS w.r.t. $\overline{\mu}$, $\overline{\mu} f^* \ll \overline{\overline{\mu} f^*}_{\overline{\mu}}$. This implies that

$$\mu f^* \ll \overline{\mu}\overline{f^*}_{\overline{\mu}}$$

Then μf^* is ergodic i.e. f^* is ergodic w.r.t. μ.

From Proposition 8 (p. 59) and its proof, $\overline{f^*}_\mu = S\overline{f^*}_\mu = \overline{f^*}_{\overline{\mu}} = S\overline{f^*}_{\overline{\mu}}$.

This implies that $\overline{f^*}_\mu(x, .)$ is strongly mixing probability on $(B^{\mathbb{N}}, \mathcal{B}_{B^{\mathbb{N}}})$ $\overline{\mu}$-a.e. and μ-a.e. □

11.3. An AMS and output weakly mixing random channel makes a noisy computation AMS and ergodic.

Proposition 9 (p. 70) straightforwardly leads to the following corollary which gives a sufficient condition of ergodicity for noisy computations relying only on properties of the input source and of the random channel $[A, F, C]$.

Corollary 7. *Let $[A, f^* \times F, B \times C]$ be a noisy computation and μ be the distribution of an AMS and ergodic source. If*

- $\overline{\mu}(D_\beta(f^*)) = 1$
- *and $[A, F, C]$ is AMS w.r.t. $\overline{\mu}$*

- *and $[A, \overline{F}_{\overline{\mu}}, C]$ is output weakly mixing w.r.t. $\overline{\mu}$*

then the noisy computation $[A, f^* \times F, B \times C]$ is AMS and ergodic w.r.t μ.

Proof. From Lemma 25 (p. 65), f^* is R-AMS w.r.t. $\overline{\mu}$. Since F is AMS w.r.t. $\overline{\mu}$, by Proposition 5 (p. 52), the noisy computation $[A, f^* \times F, B \times C]$ is AMS w.r.t. $\overline{\mu}$.

From Proposition 9 (p. 70), $[A, \overline{f^*}_\mu, B]$ is output weakly mixing w.r.t. $\overline{\mu}$. Since $[A, \overline{F}_{\overline{\mu}}, C]$ is output weakly mixing w.r.t. $\overline{\mu}$, by Corollary 6 (p. 61), the noisy computation $[A, f^* \times F, B \times C]$ is ergodic w.r.t μ

□

Let $f : L \to B^+$ be a computable function. Let $[A, f^*, B]$ be the induced deterministic channel (c.f. 11.1) and μ be an AMS and ergodic source. Since $D_\beta(f^*) = A^{\mathbb{N}}$, $\mu(D_\beta(f^*)) = \overline{\mu}(D_\beta(f^*)) = 1$. For any random channel $[A, F, C]$ AMS w.r.t. $\overline{\mu}$ such that $[A, \overline{F}_{\overline{\mu}}, C]$ is output weakly mixing w.r.t. $\overline{\mu}$, the noisy computation $[A, f^* \times F, B \times C]$ is AMS and ergodic w.r.t μ.

In other words, for a function on finite length sequences (thus a computable function) f, it is sufficient to focus on stability properties of the noisy realization F under concern in order to apply the coding results given above.

11.4. Examples

11.4.1. Design error

Let $f : L_f \to B^+$ be a Turing computable function. A design error (e.g., software bug) in the realization of f can be viewed as a discrepancy between the algorithm specified by f and the algorithm obtained through the (imperfect) design process. The obtained algorithm is also a Turing computable function $g : L_g \to B^+$. Indeed, g is based on the same computational model as f.

According to section 11.1), f and g both determine deterministic channels $[A, f^*, B]$ and $[A, g^*, B]$. The computation of f by g is thus modeled by the noisy computation $[A, f^* \times g^*, B \times B]$.

By Proposition 9 (p. 70), $[A, g^*, B]$ is AMS w.r.t μ and $[A, \overline{g^*_{\overline{\mu}}}, B]$ is output weakly mixing w.r.t $\overline{\mu}$ for any AMS ergodic probability μ. According to Corollary 7 (p. 70), the noisy computation $[A, f^* \times g^*, B \times B]$ is AMS and ergodic w.r.t. any AMS and ergodic probability μ.

11.5. Noisy circuit

Let $A = \{0,1\}^k$ and $B = \{0,1\}^{k'}$. Let C be a circuit made of boolean gates and implementing a function $\phi : A \to B$. Let $f : A^+ \to B^+$ be the function such that $f(x_0x_1 \cdots x_{n-1}) = \phi(x_0)\phi(x_1)\cdots\phi(x_{n-1})$. f induces a deterministic channel $[A, f^*, B]$.

Let $\tilde{C}$ be a noisy realization of C made of noisy boolean gates. Each noisy gate produces a correct output with probability $1 - \xi$. Errors of different gates are assumed to be independent and the noisy circuit is assumed to be memoryless. Then successive uses of $\tilde{C}$ induces a memoryless and stationary random channel $[A, F, B]$. The computation of f by successive uses of $\tilde{C}$ is thus modeled by the noisy computation $[A, f^* \times F, B \times B]$.

The channel $[A, F, B]$ is stationary and memoryless thus, for any AMS and ergodic source with distribution μ, $\overline{F_{\overline{\mu}}} = F$. Moreover, being stationary and memoryless, $[A, F, B]$ is output weakly mixing w.r.t. $\overline{\mu}$. Then, according to Corollary 7 (p. 70), the noisy computation $[A, f^* \times F, B \times B]$ is AMS and ergodic w.r.t. any AMS and ergodic probability μ.

Assuming that $\tilde{C}$ is exactly C made from noisy gates (they have the same structure), $\tilde{C}$ and C have the same size (number of gates). Let apply the model of reliable computation of Section 7 to f^* ($g = f^*$) and to the noisy computation $[A, f^* \times F, B \times B]$. Then, for large enough k and n, it is possible to obtain $f(x_0x_1 \cdots x_{k-1}) = \phi(x_0)\phi(x_1)\cdots\phi(x_{k-1})$ from a sequence of n uses (or copies) of $\tilde{C}$ with an arbitrarily small error probability under the condition that:

$$R = \frac{k}{n}\overline{H}(X') < C_{f^*}(F)$$

where $\overline{H}(X')$ is the entropy of the input source.

Then instead of k uses or copies of C (i.e., a circuit of size $k.size(C)$), the reliable computation relies on n uses or copies of $\tilde{C}$ (i.e., a circuit of size $n.size(\tilde{C})$). In addition to the encoder and decoder sizes, the "cost of reliability" is given by an increasing factor of circuit size Λ which is

$$\Lambda = \frac{n.size(\tilde{C})}{k.size(C)} = \frac{n}{k}$$

Assuming that the rate R is close to capacity $C_{f^*}(F)$ ($C_{f^*}(F)$ assumed strictly positive) then

$$\frac{C_{f^*}(F) - \epsilon}{\overline{H}(X')} \leq \frac{k}{n} < \frac{C_{f^*}(F)}{\overline{H}(X')}$$

Then, for any $\epsilon > 0$, it is possible to build a noisy circuit F^n (i.e., n copies of $\tilde{C}$) such that

$$\frac{\overline{H}(X')}{C_{f^*}(F)} < \Lambda \leq \frac{\overline{H}(X')}{C_{f^*}(F) - \epsilon}$$

and giving an arbitrarily low error probability for large k and n.

This very simple example also complies with the general reliable computation model given in [21]. Although the model of reliable computation given in the present paper and the one of [21] have different goals, they possess the common characteristic to clearly identify an encoding phase, a (noisy) computation phase and a decoding phase. Connections between these two approaches are certainly worth to be further investigated.

12. Further work

In addition to try to extend other classical results from the channel case to the computation case (e.g., related to error probability), some specific questions regarding noisy computation could be addressed.

Relevant models of noisy Turing machines remain to be explored and analyzed regarding stability properties (ergodicity, asymptotic mean stationarity, ...). To the author's best knowledge, very few models of noisy Turing machines have been proposed and studied ([2], [3]). The examples given above are very simple and involve a noisy computing device F which is stationary and memoryless. If F is a Turing machine whose behavior is altered by (e.g.) random perturbations of the transition function, stationarity and memorylessness will not hold anymore.

Instantiation of the noisy and reliable computation models to peculiar formal computational models and their noisy counterpart deserves some work, one goal being to assess key parameters such as the cost of reliability. In particular, connections with the domain of reliable circuits built from noisy gates could be of significant interest. As sketch in the very simple example given above, it might be possible to obtain asymptotics of parameters such as the blow-up of circuits.

The question of noisy encoding has only been skimmed over but deserves in depth analysis through the point of view of cascaded noisy computations. This has probably some links with asymptotic properties of restoring organs ([7]).

Appendix A. Proof of Lemma 15

Proof of Lemma 15 is based on the following lemma.

Lemma 27. *Let A and C be standard alphabets and B be a countable standard alphabet. Let $[A, f \times F, B \times C]$ be a noisy computation. Let $[A, X]$ be a source. Let $\epsilon > 0$ and $\lambda > 0$. Let $\tilde{B}$ be a measurable subset of $B^{\mathcal{I}}$. Assume there exist a set $\left\{y_k \in \tilde{B}, k = 1 \cdots M\right\}$, a collection of measurable disjoint sets $\Gamma_k, k = 1, ..., M$ members of $\mathcal{B}_{C^{\mathcal{I}}}$ and a collection of measurable sets $A_k \subset f^{-1}(y_k), k = 1, \cdots, M$ members of $\mathcal{B}_{A^{\mathcal{I}}}$ such that:*

$$\forall k = 1, \cdots, M \ P_{X|f(X)}(A_k|y_k) > 1 - \lambda \text{ and } \forall x \in A_k \ P_{F(X)|X}\left(\Gamma_k^c|x\right) \leq \epsilon$$

If the set $\left\{y_k \in \tilde{B}, k = 1 \cdots M\right\}$ is maximal, meaning there do not exist any other $y_{M+1} \in \tilde{B}$, Γ_{M+1} and A_{M+1} complying with the given properties, then $\forall y_0 \in \tilde{B} \setminus \{y_1, \cdots, y_M\}$ there exits $A_0 \subset f^{-1}(y_0)$ such that:

$$P_{X|f(X)}(f^{-1}(y_0) \setminus A_0|y_0) > \lambda \tag{A.1}$$

and

$$\forall x \in f^{-1}(y_0) \setminus A_0 \ P_{F(X)|X}\left(\bigcup_{k=1}^{M} \Gamma_k|x\right) > \epsilon \tag{A.2}$$

Proof of Lemma 27. If $\tilde{B} \setminus \{y_1, \cdots, y_M\} = \emptyset$, the result is trivial.

Let $y_0 \in \tilde{B} \setminus \{y_1, \cdots, y_M\}$. Assume that there is no $A_0 \subset f^{-1}(y_0)$ for which the two statements (A.1) and (A.2) both hold. This implies that for *any* subset A_0 of $f^{-1}(y_0)$:

$$P_{X|f(X)}(f^{-1}(y_0) \setminus A_0|y_0) \leq \lambda \tag{A.3}$$

or

$$\exists x \in f^{-1}(y_0) \setminus A_0 \ P_{F(X)|X}\left(\bigcup_{k=1}^{M} \Gamma_k|x\right) \leq \epsilon \tag{A.4}$$

$A_0 = \{x \in f^{-1}(y_0)/P_{F(X)|X}\left(\bigcup_{k=1}^{M} \Gamma_k|x\right) \leq \epsilon\}$ is a subset of $f^{-1}(y_0)$ for which (A.4) does not hold, then (A.3) hold: $P_{X|f(X)}(f^{-1}(y_0) \setminus A_0|y_0) \leq \lambda$. Then

$$P_{X|f(X)}(A_0|y_0) > 1 - \lambda \text{ and } \forall x \in A_0 \ P_{F(X)|X}\left(\bigcup_{k=1}^{M} \Gamma_k|x\right) \leq \epsilon$$

Thus setting $\Gamma_0 = C_{\mathcal{I}} \setminus \bigcup_{k=1}^{M} \Gamma_k$, y_0 can be added to the set $\{y_1, \cdots, y_M\}$, contradicting the maximality of this set. □

Proof of Lemma 15. Since $f(X) \to X \to F(X)$ is a Markov Chain (from [10], Chapter 2):

$$\begin{aligned} P_{F(X)|fX)}\left(\bigcup_{k=1}^{M} \Gamma_k | y_0\right) &= \int P_{F(X)|X}\left(\bigcup_{k=1}^{M} \Gamma_k | x\right) dP_{X|f(X)}(x|y_0) \\ &\geq \int_{f^{-1}(y_0)\setminus A_0} P_{F(X)|X}\left(\bigcup_{k=1}^{M} \Gamma_k | x\right) dP_{X|f(X)}(x|y_0) \end{aligned}$$

Then, by Lemma 27

$$P_{F(X)|fX)}\left(\bigcup_{k=1}^{M} \Gamma_k | y_0\right) > \lambda\epsilon \tag{A.5}$$

proving statement 1).

$$\begin{aligned} P_{F(X)}\left(\bigcup_{k=1}^{M} \Gamma_k\right) &= \int_{\{y_1,\cdots,y_M\}} P_{F(X)|f(X)}\left(\bigcup_{k=1}^{M} \Gamma_k | y\right) dP_{f(X)} \\ &\quad + \int_{\{y_1,\cdots,y_M\}^c} P_{F(X)|f(X)}\left(\bigcup_{k=1}^{M} \Gamma_k | y\right) dP_{f(X)} \end{aligned}$$

Since $f(X) \to X \to F(X)$ is a Markov chain:

$$\begin{aligned} P_{F(X)|f(X)}\left(\bigcup_{k=1}^{M} \Gamma_k | y\right) = \int P_{F(X)|X}\left(\bigcup_{k=1}^{M} \Gamma_k | x\right) dP_{X|f(X)} \\ \geq \int_{A_k} P_{F(X)|X}\left(\bigcup_{k=1}^{M} \Gamma_k | x\right) dP_{X|f(X)} \end{aligned}$$

By 1), $P_{F(X)|f(X)}\left(\bigcup_{k=1}^{M}\Gamma_k|y\right) \geq \lambda\epsilon$ for any $y \notin \{y_1, \cdots, y_M\}$. Then

$$\begin{array}{rcl} P_{F(X)}\left(\bigcup_{k=1}^{M}\Gamma_k\right) & > & \int_{\{y_1,\cdots,y_M\}}\int_{A_k} P_{F(X)|X}\left(\Gamma_k|x\right) dP_{X|f(X)} dP_{f(X)} \\ & & +\lambda\epsilon\int_{\tilde{B}\backslash\{y_1,\cdots,y_M\}} dP_{f(X)} \\ & > & \int_{\{y_1,\cdots,y_M\}} (1-\epsilon)P_{X|f(X)}(A_k|y_k)dP_{f(X)} \\ & & +\lambda\epsilon\int_{\tilde{B}\backslash\{y_1,\cdots,y_M\}} dP_{f(X)} \\ & > & \int_{\{y_1,\cdots,y_M\}} (1-\epsilon)(1-\lambda)dP_{f(X)} \\ & & +\lambda\epsilon\int_{\tilde{B}\backslash\{y_1,\cdots,y_M\}} dP_{f(X)} \end{array}$$

$$\Rightarrow P_{F(X)}\left(\bigcup_{k=1}^{M}\Gamma_k\right) > min((1-\lambda)(1-\epsilon), \lambda\epsilon).P_{f(X)}(\tilde{B})$$

proving statement 2).

Finally y_1, $A_1 = f^{-1}(y_1)$ and $C_{\mathcal{I}}$ fulfill the properties. Thus $M \geq 1$. □

Proof of Lemma 18. A cascade of stationary channels is a stationary channel (see [10]). Then it remains to prove that $F = \nu f$ is ergodic. Let $[A, X]$ be a stationary and ergodic source. Let $O \in \mathcal{B}_{A^{\mathcal{I}}\times B^{\mathcal{I}}}$ be an invariant event: $(T_A \times T_B)^{-1}O = O$.

Let $\tilde{O} = \left\{(x, y) \in A^{\mathcal{I}} \times A^{\mathcal{I}}/(x, f(y)) \in O\right\}$.

$\tilde{O}_x = f^{-1}(O_x)$ since :

$$\begin{array}{rcl} \tilde{O}_x & = & \left\{y \in A^{\mathcal{I}}/(x, y) \in \tilde{O}\right\} \\ & = & \left\{y \in A^{\mathcal{I}}/(x, f(y)) \in O\right\} \\ & = & \left\{y \in A^{\mathcal{I}}/f(y) \in O_x\right\} \end{array}$$

This implies $P_{XF(X)}(O) = P_{XY}(\tilde{O})$ since:

$$\begin{aligned}
P_{XF(X)}(O) &= \int P_{F(X)|X}(O_x|x)dP_X \\
&= \int\int P_{f(Y)|Y}(O_x|y)dP_{Y|X}dP_X \\
&= \int\int 1_{O_x}(f(y))dP_{Y|X}dP_X \\
&= \int P_{Y|X}(f^{-1}(O_x)|x)dP_X \\
&= \int P_{Y|X}(\tilde{O}_x)|x)dP_X \\
&= P_{XY}(\tilde{O})
\end{aligned}$$

O is $(T_A \times T_B)$-invariant then $\tilde{O}$ is $(T_A \times T_A)$-invariant since

$$\begin{aligned}
&(T_A \times T_A)^{-1}\tilde{O} = \\
&\qquad \Big\{(x,y) \in A^{\mathcal{I}} \times A^{\mathcal{I}} / (T_A x, T_A y) \in \tilde{O}\Big\} = \\
&\qquad\qquad \big\{(x,y) \in A^{\mathcal{I}} \times A^{\mathcal{I}} / (x, f(T_A y)) \in O\big\}
\end{aligned}$$

By stationarity of f:

$$\begin{aligned}
&(T_A \times T_A)^{-1}\tilde{O} = \\
&\qquad \big\{(x,y) \in A^{\mathcal{I}} \times A^{\mathcal{I}} / (x, T_B f(y)) \in O\big\} = \\
&\qquad\qquad \big\{(x,y) \in A^{\mathcal{I}} \times A^{\mathcal{I}} / (x, f(y)) \in (T_A \times T_B)^{-1}(O)\big\}
\end{aligned}$$

By invariance of O:

$$\begin{aligned}
(T_A \times T_C)^{-1}\tilde{O} &= \big\{(x,y) \in A^{\mathcal{I}} \times A^{\mathcal{I}} / (x, f(y)) \in O\big\} \\
&= \tilde{O}
\end{aligned}$$

P_{XY} is ergodic then $P_{XF(X)}(O) = P_{XY}(\tilde{O})$ is either 1 or 0. Thus $(X, F(X))$ is an ergodic process. □

Proof of Lemma 19. Let $[A, X]$ be a stationary source. f is a stationary function then f induces a stationary deterministic channel $[A, f, B]$. $[A, f, B]$ is thus an ergodic channel (a deterministic stationary channel is ergodic, see

[10]). $[A, F, C]$ is a stationary and ergodic channel. Then, all these assumptions imply that, for any stationary source $[A, X]$, $(X, f(X))$ and $(X, F(X))$ are stationary and ergodic processes. To show that $[A, f \times F, B \times C]$ is stationary and ergodic, it remains to prove that $(f(X), F(X))$ is stationary and ergodic.

f being a function, $F(X) \to X \to f(X)$ is a Markov chain, then $\forall G \in \mathcal{B}_{B^{\mathcal{I}}}$:

$$P_{f(X)|XF(X)}(G|xz) = P_{f(X)|X}(G|x) \; P_{XF(X)}\text{-a.e.}$$

then

$$P_{f(X)|XF(X)}(T_B^{-1}G|xz) = P_{f(X)|X}(T_B^{-1}G|x) \; P_{XF(X)}\text{-a.e.}$$

By stationarity of the deterministic channel $[A, f, B]$:

$$P_{f(X)|XF(X)}(T_B^{-1}G|xz) = P_{f(X)|X}(G|T_A x) \; P_{XF(X)}\text{-a.e.}$$

Using once more the fact that $F(X) \to X \to f(X)$ is a Markov chain:

$$P_{f(X)|XF(X)}(T_B^{-1}G|xz) = P_{f(X)|XF(X)}(G|T_A x T_B z) \; P_{XF(X)}\text{-a.e.}$$

Then the channel $(X, F(X)) \to f(X)$ is stationary. Since $(X, F(X))$ is a stationary process, this implies that $(X, f(X), F(X))$ is a stationary process and so is $(f(X), F(X))$.

Assume in addition that $[A, X]$ is ergodic. Let $O \in \mathcal{B}_{B^{\mathcal{I}} \times C^{\mathcal{I}}}$ be an invariant event: $(T_B \times T_C)^{-1} O = O$. Let $\tilde{O} = \{(x, z) \in A^{\mathcal{I}} \times C^{\mathcal{I}} / (f(x), z) \in O\}$.

$\tilde{O}_z = f^{-1}(O_z)$ since :

$$\begin{aligned} \tilde{O}_z &= \left\{x \in A^{\mathcal{I}} / (x, z) \in \tilde{O}\right\} \\ &= \{x \in A^{\mathcal{I}} / (f(x), z) \in O\} \\ &= \{x \in A^{\mathcal{I}} / f(x) \in O_z\} \end{aligned}$$

This implies $P_{f(X)F(X)}(O) = P_{XF(X)}(\tilde{O})$ since:

$$\begin{aligned} P_{f(X)F(X)}(O) &= \int P_{f(X)|F(X)}(O_z/z) dP_{F(X)} \\ &= \int P_{X|F(X)}(f^{-1}(O_z)/z) dP_{F(X)} \\ &= \int P_{X|F(X)}(\tilde{O}_z/z) dP_{F(X)} \end{aligned}$$

O is $(T_B \times T_C)$-invariant then $\tilde{O}$ is $(T_A \times T_C)$-invariant since

$$\begin{aligned}
&(T_A \times T_C)^{-1}\tilde{O} = \\
&\qquad \Big\{(x,z) \in A^{\mathcal{I}} \times C^{\mathcal{I}} / (T_A x, T_C z) \in \tilde{O}\Big\} = \\
&\qquad\qquad \Big\{(x,z) \in A^{\mathcal{I}} \times C^{\mathcal{I}} / (f(T_A x), T_C z) \in O\Big\}
\end{aligned}$$

By stationarity of f:

$$\begin{aligned}
&(T_A \times T_C)^{-1}\tilde{O} = \\
&\qquad \big\{(x,z) \in A^{\mathcal{I}} \times C^{\mathcal{I}} / (T_B f(x), T_C z) \in O\big\} = \\
&\qquad\qquad \big\{(x,z) \in A^{\mathcal{I}} \times C^{\mathcal{I}} / (f(x), z) \in (T_B \times T_C)^{-1}(O)\big\}
\end{aligned}$$

By invariance of O:

$$\begin{aligned}
(T_A \times T_C)^{-1}\tilde{O} &= \big\{(x,z) \in A^{\mathcal{I}} \times C^{\mathcal{I}} / (f(x), z) \in O\big\} \\
&= \tilde{O}
\end{aligned}$$

$P_{XF(X)}$ is ergodic then $P_{f(X)F(X)}(O) = P_{XF(X)}(\tilde{O})$ is either 1 or 0. Thus $(f(X), F(X))$ is an ergodic process. □

References

[1] Ahlswede, R., november 1984. Improvements of winograd's result on computation in the presence of noise. IEEE Transactions on information theory IT-30 (6).

[2] Asarin, E., Collins, P., 2005. Noisy turing machines. Automata, Language and Programming - Lecture Notes in Computer Science 3580.

[3] Çapuni, I., Gács, P., 2012. A turing machine resisting isolated bursts of faults. In: SOFSEM 2012: Theory and practice of computer science. pp. 165–176.

[4] Elias, P., october 1958. Computation in the presence of noise. IBM Journal.

[5] Faden, A. M., 1985. The existence of regular conditional probabilities: Necessary and sufficient conditions. The Annals of Probability 13.

[6] Fontana, R. J., Gray, R. M., Kieffer, J. C., may 1981. Asymptotically mean stationary channels. IEEE Transactions on Information Theory 27.

[7] Gacs, P., 2005. Reliable computation. Tech. rep., Boston University.

[8] Girardin, V., 2005. On the different extensions of the ergodic theorem of information theory, in: Recent advances in applied probability. Springer, pp. 163–179.

[9] Gray, R. M., 2009. Probability, Random Processes and Ergodic Properties. Springer.

[10] Gray, R. M., 2011. Entropy and Information Theory, 2nd edition. Springer.

[11] Gray, R. M., Kieffer, J. C., 1980. Asymptotically mean stationary measures. The Annals of Probability 8.

[12] Hadjicostis, C. N., Verghese, G. C., january 2005. Coding approaches to fault tolerance in linear dynamic systems. IEEE Transactions on Information Theory 51 (1), 210–228.

[13] Kadota, T., november 1970. Generalization of Feinstein's fundamental lemma. IEEE Transactions on information theory 16.

[14] Kakihara, Y., 1999. Abstract Methods in Information Theory. World Scientific.

[15] Nazer, B., Gastpar, M., october 2007. Computation over multiple-access channels. IEEE Transactions on information theory 53 (10).

[16] Neumann, J. V., 1956. Probabilistics logics and the synthesis of reliable organisms from unreliable components. Automata studies.

[17] Nicolaidis, M., 1993. Efficient implementations of self-checking adders and alus. In: International Symposium on Fault Tolerant Computing. pp. 586–595.

[18] Peterson, W., Rabin, M., april 1959. On codes for checking logical operations. IBM Journal.

[19] Rachlin, E., Savage, J. E., 2008. A framework for coded computation. In: ISIT 2008.

[20] Rao, T., Fujiwara, E., 1989. Error-control coding for computer systems. Prentice-Hall.

[21] Romashchenko, A., 2006. Reliable computations based on locally decodable codes. Lecture notes in computer sciences 3884.

[22] Simon, F., september 2010. Capacity of a noisy function. In: Information Theory Workshop - Dublin.

[23] Simon, F., october 2011. On the capacity of noisy computation. In: Information Theory Workshop - Paraty.

[24] Spielman, D. A., 1996. Highly fault-tolerant parallel computation. In: Annual Symposium Foundations of Computer Science. pp. 154–160.

[25] Winograd, S., october 1962. Coding for logical operations. IBM Journal.

[26] Winograd, S., Cowan, J., 1963. Reliable computation in presence of noise. The MIT Press.