Capacity of a noisy function

Francois Simon

Capacity of a Noisy Function

François Simon
Institut TELECOM ; Telecom SudParis ; CITI
9 rue Charles Fourier, 91011 EVRY Cedex, France
Email: francois.simon@it-sudparis.eu

***Abstract*—This paper presents an extension of the memoryless channel coding theorem to *noisy functions*, i.e. unreliable computing devices without internal states. It is shown that the concepts of *equivocation* and *capacity* can be defined for *noisy computations* in the simple case of *memoryless noisy functions*. Capacity is the upper bound of *input rates* allowing reliable computation, i.e. decodability of noisy outputs into expected outputs. The proposed concepts are generalizations of these known for channels: the capacity of a noisy implementation of a bijective function has the same expression as the capacity of a communication channel. A lemma similar to Feinstein's one is stated and demonstrated. A model of reliable computation of a function thanks to a noisy device is proposed. A coding theorem is stated and demonstrated.**

I. Introduction and related work

Reliable computation in the presence of noise has been the subject of numerous works. Either for practical objectives (e.g, self-checking circuits, [1]) or for theoretical purposes. Recent references (see for example, [2], [3], [4], [5]) continue to extend the stream opened by Von Neumann's seminal paper ([6]). 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. Meanwhile, quite surprisingly, none of such works explicitly targets the identification of a capacity for noisy computations.

The concept of capacity has been thoroughly studied for data transmission. It has not been the case for computation and relatively few results have been obtained.

Except more recently ([7], see below), to the author's best knowledge, the main attempt to address the question of a noisy computation capacity came from Winograd and Cowan in their 1963 monograph ([8]). In [8], the entropy $H(X|F(X))$ of the input source conditioned by the noisy computation output is assessed as a noise measure. As it is the equivocation between the *noisy output* and the *input*, this quantity is not relevant as the equivocation due to the sole noise, except in 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. Due to this restriction, [8] did not completely succeed in proposing a noisy computation capacity ([8], theorem 6.3, pages 47-48).

Noisy computation capacity is also considered in reliable reconstruction of a function of sources over a multiple access channel. A definition of noisy computation capacity is established by Nazer and Gastpar in [7] 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 encoding process which perfectly performs a computation equivalent to the desired function. The encoder outputs are then transmitted through a noisy MAC to a decoder (see proofs of Theorems 1 and 2 of [7]). This models a noisy computation as a perfect computation followed by a noisy transmission of the result and, thus, does not cover in full generality noisy computation of functions.

The present paper proposes a generalization of the notions of equivocation and capacity to *noisy finite unary functions*. n-ary functions can be modeled as unary ones by concatenating n input values in one "meta"-input and thus modeling a *joint coding* of operands. In the proposed model, the encoder and decoder do not compute the expected function and the noisy computing device is not considered as being only a noisy transmission channel. It is assumed that sources are i.i.d. and noisy functions memoryless. Bold characters indicate vectors (sequences of symbols, of random values).

II. Noisy Function

A noisy function F implementing a "perfect" finite function f can be modeled as follows:

Definition 1: Let f be a function from the finite input set I to the finite output set O. If a relation F from I to O is such that there exists a family of probability laws $(P(F(x_j)|x_i))_{x_i,x_j \in I}$ then F is a *noisy function* implementing the function f.

The following theorem gives a first characterization of the "noise" impinging on the computation of a function f thanks to F. The theorem is stated without proof as it directly derives from AEP ([9], pages 51-53). X denotes an i.i.d. source on I.

Theorem 1: For any couple of positive real numbers ϵ and δ, there exists an integer $n(\epsilon, \delta)$ such that, for any n greater than $n(\epsilon, \delta)$, the set $\mathbf{F^n}((\mathbf{f^n})^{-1}(\mathbf{y_0}))$, where $\mathbf{y_0}$ is a n-sequence of $\mathbf{f^n}(\mathbf{X^n})$, is split into two subsets:

1) a set of n-sequences (negligible given $\mathbf{y_0}$) with a total conditional (i.e., given $\mathbf{y_0}$) probability less than ϵ,
2) a set of n-sequences (typical given $\mathbf{y_0}$) whose number ν is such that:

$$(1-\epsilon)2^{n(H(F(X)|f(X))-\delta)} \leq \nu \leq 2^{n(H(F(X)|f(X))+\delta)}$$

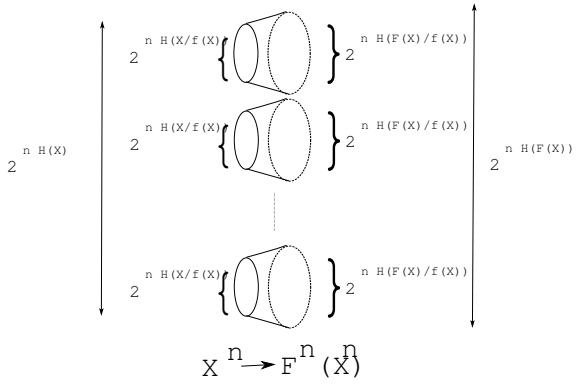

Fig. 1. Decodable sets of noisy outputs.

If $\mathbf{y_i}$ is such a typical (given $\mathbf{y_0}$) n-sequence:

$$2^{-n(H(F(X)|f(X))+\delta)} \leq p(\mathbf{y_i}|\mathbf{y_0}) \leq 2^{-n(H(F(X)|f(X))-\delta)}$$

Moreover, if $\mathbf{y_0}$ is a typical n-sequence of $\mathbf{f^n}(\mathbf{X^n})$ then the set $\mathbf{F^n}((\mathbf{f^n})^{-1}(\mathbf{y_0}))$ contains ν' typical n-sequences of $\mathbf{F^n}(\mathbf{X^n})$ which are typical given $\mathbf{y_0}$ where:

$$2^{n(H(F(X)|f(X))-\delta)} \leq \nu' \leq 2^{n(H(F(X)|f(X))+\delta)}$$

III. Capacity of a Noisy Function

Defining a notion of capacity for noisy functions 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. Decodability can be achieved by selecting inverse images $(\mathbf{f^n})^{-1}(\mathbf{y_i}), i = 1, \cdots, N, (\mathbf{y_i}$ typical n-sequences of $\mathbf{f^n}(\mathbf{X^n}))$, whose images, by the noisy function, do not overlap "too much". Informally, one can choose at most $2^{nH(F(X))}/2^{n(H(F(X)|f(X)))}$ inverse images whose noisy images by $\mathbf{F^n}$ do not overlap (for large n). Each inverse image offers $2^{nH(X|f(X))}$ different possible input n-sequences. So, at most, $2^{n(H(X|f(X))+H(F(X))-H(F(X)|f(X)))}$ input n-sequences (among the $2^{nH(X)}$ possible ones) can be used in order to reach unambiguous decodability. Figure 1 illustrates this. It will be shown below that this upper bound is reachable.

This justifies the definition of the *capacity of the noisy function F with respect to the function f* as:

$$C_f(F) = \max_{X \in \mathcal{S}(I)} [H(X|f(X)) + H(F(X)) - H(F(X)|f(X))]$$

where $\mathcal{S}(I)$ the set of i.i.d. sources on I. Hence

$$C_f(F) = \max_{X \in \mathcal{S}(I)} [H(X|f(X)) + H(f(X)) - H(f(X)|F(X))]$$

Since f is a function, $H(X) = H(f(X)) + H(X|f(X))$. We get $C_F(f) = \max_{X \in \mathcal{S}(I)} (H(X) - H(f(X)|F(X)))$

The *equivocation of the noisy function F w.r.t f* is the quantity $H(f(X)|F(X))$. If f is a bijective function, $H(f(X)|F(X)) = H(X|F(X))$. This is the equivocation of a communication channel $X \to F(X)$. The following definition is then a generalization of that of the channel capacity.

Definition 2: Let F be a noisy implementation of the function f. The *capacity of the noisy function Fw.r.t the function f* is:

$$C_f(F) = \max_{X \in \mathcal{S}(I)} (H(X) - H(f(X)|F(X)))$$

IV. A Coding Theorem

We first state and prove a coding lemma (similar to Feinstein's lemma and restricted to the simple case of memoryless processes). The proof is inspired by [10].

A. A Coding lemma

Lemma 1: For any triple of positive real numbers $(\epsilon, \delta, \lambda)$, $\epsilon, \delta < \frac{1}{2}$, there exists $n(\epsilon, \delta, \lambda)$ such that, for any $n \geq n(\epsilon, \delta, \lambda)$, it is possible to pick up N typical n-sequences of $\mathbf{f^n}(\mathbf{X^n})$, i.e., $\mathbf{y_1}, \ldots, \mathbf{y_N}$, and N sets $\mathbf{B_1}, \ldots, \mathbf{B_N}$, where $\mathbf{B_i}$ is a set of typical n-sequences of $\mathbf{F^n}(\mathbf{X^n})$, which are conditionally typical given $\mathbf{y_i}$ such that:

1) $\forall i = 1, ..., N$, $\exists \mathbf{A_i} \subset (\mathbf{f^n})^{-1}(\mathbf{y_i})$ such that $\forall \mathbf{x_n} \in \mathbf{A_i}$, $p(\mathbf{F^n}(\mathbf{X^n}) \in \mathbf{B_i}|\mathbf{x_n}) \geq 1 - \lambda$ and $\sum_{\mathbf{x_n} \in \mathbf{A_i}} p(\mathbf{x_n}|\mathbf{y_i}) \geq 1 - \epsilon$
2) $\forall i \neq j, \mathbf{B_i} \cap \mathbf{B_j} = \emptyset$
3) $2^{n(H(F(X))-H(F(X)|f(X))-\delta)} \leq N \leq 2^{n(H(F(X))-H(F(X)|f(X))+\delta)}$ and $N \leq$
4) $\forall i = 1, ..., N$, if ν is the number of conditionally typical (given $\mathbf{y_i}$) n-sequences of $\mathbf{X_n}$ belonging to $\mathbf{A_i}$, then:

$$(1 - 2\epsilon)2^{n(H(X|f(X))-\delta)} \leq \nu \leq 2^{n(H(X|f(X))+\delta)}$$

Definition 3: The set $\{(\mathbf{A_i}, \mathbf{B_i}), i = 1, \cdots, N\}$ is called a *code of size N and of length n* for the noisy function F to λ-reliably compute the function f. The sets $\{\mathbf{A_i}, i = 1, \cdots, N\}$ and $\{\mathbf{B_i}, i = 1, \cdots, N\}$ are respectively the *input code* and the *output code*.

Before proving lemma 1, we need the following lemma:

Lemma 2: Let $\{(\mathbf{y_i}, \mathbf{B_i}), i = 1, \cdots, N\}$ be a maximal set of $(\mathbf{y_i}, \mathbf{B_i})$ satisfying statements 1) and 2) of lemma 1. I.e., $\mathbf{y_{N+1}}$ and $\mathbf{B_{N+1}}$, for which 1) and 2) both hold, cannot be found. Let $\mathbf{y_0}$ be a typical n-sequence of $\mathbf{f^n}(\mathbf{X^n})$ not belonging to $\{\mathbf{y_i}, i = 1, \cdots, N\}$. Then, there exists $\mathbf{A_0}$, a subset of $(\mathbf{f^n})^{-1}(\mathbf{y_0})$, such that:

$$\forall \mathbf{x_n} \in (\mathbf{f^n})^{-1}(\mathbf{y_0}) \backslash \mathbf{A_0}, p(\mathbf{F^n}(\mathbf{X^n}) \in \bigcup_{i=1}^{N} \mathbf{B_i}|\mathbf{x_n}) > \lambda$$

$$\text{and} \sum_{\mathbf{x_n} \in (\mathbf{f^n})^{-1}(\mathbf{y_0}) \backslash \mathbf{A_0}} p(\mathbf{x_n}|\mathbf{y_0}) > \epsilon$$

Proof of lemma 2:

Suppose no such $\mathbf{A_0}$ exists for $\mathbf{y_0}$. This implies that for any subset $\mathbf{A_0}$ of $(\mathbf{f^n})^{-1}(\mathbf{y_0})$ either

$$\exists \mathbf{x_n} \in (\mathbf{f^n})^{-1}(\mathbf{y_0}) \backslash \mathbf{A_0}, p(\mathbf{F^n}(\mathbf{X^n}) \in \bigcup_{i=1}^{N} \mathbf{B_i}|\mathbf{x_n}) \leq \lambda \quad (1)$$

$$\text{or} \sum_{\mathbf{x_n} \in (\mathbf{f^n})^{-1}(\mathbf{y_0}) \backslash \mathbf{A_0}} p(\mathbf{x_n}|\mathbf{y_0}) \leq \epsilon \quad (2)$$

$(\mathbf{f^n})^{-1}(\mathbf{y_0})\backslash\{\mathbf{x_n} \in (\mathbf{f^n})^{-1}(\mathbf{y_0}) / p(\mathbf{F^n}(\mathbf{X^n}) \in \bigcup_{i=1}^{N} \mathbf{B_i}|\mathbf{x_n}) > \lambda\}$ is a subset of $(\mathbf{f^n})^{-1}(\mathbf{y_0})$ and does not satisfy inequality (1). So it must satisfy inequality (2) and therefore has a total conditional probability (given $\mathbf{y_0}$) at least $1-\epsilon$.

Thus we can add the couple $\left(\mathbf{y_0}, \{typical\ n-sequences\ of\ \mathbf{F^n}(\mathbf{X^n})\}\backslash\bigcup_{i=1}^{N}\mathbf{B_i}\right)$ to the collection $(\mathbf{y_i}, \mathbf{B_i})_{i=1,...,N}$, contradicting the maximality of the chosen collection. ∎

Proof of lemma 1:

The lower bound of N given by statement 3), under the constraints given by statements 1) and 2), will be proved by using a lower bound of $P(\mathbf{F^n}(\mathbf{X^n}) \in \bigcup_{i=1}^{N}\mathbf{B_i})$.

As the $\mathbf{B_i}$'s are disjoint sets and, from Theorem 1 (obviously $\mathbf{B_i}$ is included in $\mathbf{F^n}\left((\mathbf{f^n})^{-1}(\mathbf{y_i})\right)$):$\forall i = 1, ..., N\ card(\mathbf{B_i}) \leq 2^{n(H(F(X)|f(X))+\delta)}$

$$\Rightarrow card(\bigcup_{i=1}^{N}\mathbf{B_i}) \leq N.2^{n(H(F(X)|f(X))+\delta)}$$

The probability of any element (typical n-sequence) of $\mathbf{B_i}$ is lower than $2^{-n(H(F(X))-\delta)}$. Hence:

$$P(\mathbf{F^n}(\mathbf{X^n}) \in \bigcup_{i=1}^{N}\mathbf{B_i}) \leq N.2^{-n(H(F(X))-H(F(X)|f(X))-2\delta)}$$

$$N \geq P(\mathbf{F^n}(\mathbf{X^n}) \in \bigcup_{i=1}^{N}\mathbf{B_i}).2^{n(H(F(X))-H(F(X)|f(X))-2\delta)} \tag{3}$$

Knowing a suitable lower bound of $P(\mathbf{F^n}(\mathbf{X^n}) \in \bigcup_{i=1}^{N}\mathbf{B_i})$, (3) will give a lower bound of N. The next step is devoted to determine this lower bound.

Let $\mathbf{y_i}$ be an n-sequence of $\mathbf{f^n}(\mathbf{X^n})$. $\mathbf{f^n}$ being a function: $p(\mathbf{y_i}) = \sum_{\mathbf{x_n}\in(\mathbf{f^n})^{-1}(\mathbf{y_i})} p(\mathbf{x_n})$. So, for any event ω, $p(\omega|\mathbf{y_i}) = \sum_{\mathbf{x_n}\in(\mathbf{f^n})^{-1}(\mathbf{y_i})} p(\omega|\mathbf{x_n}).p(\mathbf{x_n}|\mathbf{y_i})$

If ω is the event $\mathbf{F^n}(\mathbf{X^n}) \in \bigcup_{i=1}^{N}\mathbf{B_i}$ and $\mathbf{y_i} = \mathbf{y_0} \notin \{\mathbf{y_1}, \cdots, \mathbf{y_N}\}$, by lemma 2, $\exists \mathbf{A_0} \subset (\mathbf{f^n})^{-1}(\mathbf{y_0})$ such that:

$$\begin{aligned} p(\omega|\mathbf{y_0}) &= \sum_{\mathbf{x_n}\in(\mathbf{f^n})^{-1}(\mathbf{y_0})} p(\omega|\mathbf{x_n}).p(\mathbf{x_n}|\mathbf{y_0}) \\ &\geq \sum_{\mathbf{x_n}\in(\mathbf{f^n})^{-1}(\mathbf{y_0})\backslash\mathbf{A_0}} p(\omega|\mathbf{x_n}).p(\mathbf{x_n}|\mathbf{y_0}) \\ &\geq \lambda \sum_{\mathbf{x_n}\in(\mathbf{f^n})^{-1}(\mathbf{y_0})\backslash\mathbf{A_0}} p(\mathbf{x_n}|\mathbf{y_0}) \geq \lambda\epsilon \\ \Rightarrow \quad & p(\mathbf{F^n}(\mathbf{X^n}) \in \bigcup_{i=1}^{N}\mathbf{B_i}|\mathbf{y_0}) \geq \lambda\epsilon \end{aligned} \tag{4}$$

Inequality (4), thus, holds for any typical n-sequence $\mathbf{y_0}$ of $\mathbf{f^n}(\mathbf{X^n})$ which is not in $\{\mathbf{y_1}, \cdots, \mathbf{y_N}\}$.

If ω is the event $\mathbf{F^n}(\mathbf{X^n}) \in \bigcup_{i=1}^{N}\mathbf{B_i}$, by inequality (4)

$$\begin{aligned} p(\omega) &= \sum_{i=1}^{N} p(\omega|\mathbf{y_i}).p(\mathbf{y_i}) + \sum_{\mathbf{y_0}\notin\{\mathbf{y_1},...,\mathbf{y_N}\}} p(\omega|\mathbf{y_0}).p(\mathbf{y_0}) \\ &\geq \sum_{i=1}^{N} p(\omega|\mathbf{y_i}).p(\mathbf{y_i}) + \lambda\epsilon. \sum_{\mathbf{y_0}\notin\{\mathbf{y_1},...,\mathbf{y_N}\}} p(\mathbf{y_0}) \end{aligned}$$

$$\begin{aligned} \text{but } p(\omega|\mathbf{y_i}) &= \sum_{\mathbf{x_n}\in(\mathbf{f^n})^{-1}(\mathbf{y_i})} p(\omega|\mathbf{x_n}).p(\mathbf{x_n}|\mathbf{y_i}) \\ &\geq \sum_{\mathbf{x_n}\in\mathbf{A_i}} (1-\lambda).p(\mathbf{x_n}|\mathbf{y_i}) \geq (1-\lambda).(1-\epsilon) \end{aligned}$$

The last inequality is given by the constraint expressed in statement 2). Thus we get:

$$\begin{aligned} p(\omega) &\geq (1-\lambda)(1-\epsilon)\sum_{i=1}^{N} p(\mathbf{y_i}) + \lambda\epsilon \sum_{\mathbf{y_0}\notin\{\mathbf{y_1},...,\mathbf{y_N}\}} p(\mathbf{y_0}) \\ &\geq \min((1-\lambda)(1-\epsilon), \lambda\epsilon). \sum_{\substack{\mathbf{y_n}\ typical \\ n-sequence \\ of\ \mathbf{f_n}(\mathbf{X_n})}} p(\mathbf{y_n}) \\ &\geq \min((1-\lambda)(1-\epsilon), \lambda\epsilon).(1-\epsilon) \end{aligned}$$

We can assume that $\epsilon \leq \frac{1}{2}$ and $\lambda \leq \frac{1}{2}$. So

$$p(\mathbf{F^n}(\mathbf{X^n}) \in \bigcup_{i=1}^{N}\mathbf{B_i}) \geq \lambda\epsilon.(1-\epsilon)$$

Giving the lower bound looked for. By (3), we get:

$$N \geq \lambda\epsilon(1-\epsilon)2^{n(H(F(X))-H(F(X)|f(X))-2\delta)}$$

For any positive ϵ, λ, δ_0, a δ small enough and a n large enough can be chosen such that $\lambda\epsilon(1-\epsilon)2^{-2n\delta} \geq 2^{-n\delta_0}$ to get

$$N \geq 2^{n(H(F(X))-H(F(X)|f(X))-\delta_0)}$$

The same kind of path can be followed to upper bound N. Since the $\mathbf{B_i}$'s are sets of typical n-sequences of $\mathbf{F^n}(\mathbf{X^n})$, $card(\bigcup_{i=1}^{N}\mathbf{B_i}) \leq 2^{n(H(F(X))+\delta)}$.

Since B_i is a set of conditionally typical (given $\mathbf{y_i}$) n-sequences, from Theorem 1, for all $i = 1, ..., N$:

$$card(\mathbf{B_i}).2^{-n(H(F(X)|f(X))-\delta)} \geq p(\mathbf{F^n}(\mathbf{X^n}) \in \mathbf{B_i}|\mathbf{y_i})$$

If ω denotes the event $\mathbf{F^n}(\mathbf{X^n}) \in \mathbf{B_i}$:

$$\begin{aligned} p(\omega|\mathbf{y_i}) &= \sum_{\mathbf{x_n}\in(\mathbf{f^n})^{-1}(\mathbf{y_i})} p\left(\omega|\mathbf{x_n}\right).p(\mathbf{x_n}|\mathbf{y_i}) \\ &\geq \sum_{\mathbf{x_n}\in\mathbf{A_i}} p\left(\omega|\mathbf{x_n}\right).p(\mathbf{x_n}|\mathbf{y_i}) \\ &\geq (1-\lambda)\sum_{\mathbf{x_n}\in\mathbf{A_i}} p(\mathbf{x_n}|\mathbf{y_i}) \geq (1-\lambda)(1-\epsilon) \end{aligned}$$

Thus $card(\bigcup_{i=1}^{N}\mathbf{B_i}) \geq N.(1-\lambda)(1-\epsilon)2^{n(H(F(X)|f(X))-\delta)}$

This leads to

$$N.(1-\lambda)(1-\epsilon)2^{n(H(F(X)|f(X))-\delta)} \leq 2^{n(H(F(X))+\delta)}$$

$$\Rightarrow N \leq \frac{1}{(1-\lambda)(1-\epsilon)}2^{n(H(F(X))-H(F(X)|f(X))+2\delta)}$$

For any positive real numbers δ_0, λ and ϵ, a small enough δ and a large enough n can be found to get:

$$N \leq 2^{n(H(F(X))-H(F(X)|f(X))+\delta_0)}$$

This closes the proof of statements 1), 2) and 3).

As:

$$\forall i = 1, ..., N \sum_{\mathbf{x_n} \in \mathbf{A_i}} p(\mathbf{x_n}|\mathbf{y_i}) \geq 1 - \epsilon$$

the Shannon-McMillan theorem extended to conditional entropies leads to statement 4):

$$(1 - 2\epsilon)2^{n(H(X|f(X))-\delta)} \leq \nu \leq 2^{n(H(X|f(X))+\delta)}$$

■

B. A coding theorem for Noisy Functions

The model of the complete process to reliably compute a finite function $g : I' \to O'$ acting on a i.i.d. source X', thanks to a noisy implementation F of a finite function $f : I \to O$ can be viewed as:

- **encoding**: let $\mathbf{X_0}^\mathbf{n}$ be the n^{th} extension of an i.i.d. source for which we have a maximal code $(\mathbf{A_i}, \mathbf{B_i})_{i=1,\cdots,N}$ allowing to λ-reliably compute $\mathbf{f^n}(\mathbf{X_0}^\mathbf{n})$ by $\mathbf{F^n}(\mathbf{X_0}^\mathbf{n})$ (cf lemma 1 and definition 3) ; a typical k-sequence $\mathbf{x'}$ of $\mathbf{X'^k}$ is encoded into a typical n-sequence of $\mathbf{X_0}^\mathbf{n}$ by a injective function, say $\mathcal{U}$, such that $\mathcal{U}(\mathbf{x'}) \in \mathbf{A_i}$ for some $i = 1, \cdots, N$
- **computation of the noisy function**: $\mathbf{F^n}$ is applied to $\mathcal{U}(\mathbf{x'})$ producing a typical n-sequence $\mathbf{F^n}(\mathcal{U}(\mathbf{x'}))$ of $\mathbf{F^n}(\mathbf{X_0}^\mathbf{n})$ where $\mathbf{F^n}(\mathcal{U}(\mathbf{x'}))$ belongs to a given $\mathbf{B_i}$ (with high probability)
- **decoding**: the first step is to associate to $\mathbf{F^n}(\mathcal{U}(\mathbf{x'}))$ the typical n-sequence $\mathbf{y_i}$ of $\mathbf{f^n}(\mathbf{X_0}^\mathbf{n})$ corresponding to $\mathbf{B_i}$ (cf lemma 1), the second step is to apply to $\mathbf{y_i}$ a function $\mathcal{V} : \{\mathbf{y_1}, \ldots, \mathbf{y_N}\} \to \{\text{typical k-sequences of } \mathbf{g^k}(\mathbf{X'^k})\}$ such that $\mathcal{V}(\mathbf{y_i}) = \mathbf{g^k}(\mathbf{x'})$

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

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 $\mathbf{A_i} \subset (\mathbf{f^n})^{-1}(\mathbf{y_i})$ ($\mathbf{y_i} \in \{\mathbf{y_1}, \ldots, \mathbf{y_N}\}$) are used for encoding typical k-sequences of *only one* $(\mathbf{g^k})^{-1}(\mathbf{z})$, $\mathbf{z}$ typical k-sequence of $\mathbf{f^k}(\mathbf{X'^k})$.

We also require that $\mathcal{V}$ be an injection (as we have required from $\mathcal{U}$). The typical k-sequences of a $\mathbf{g^k}^{-1}(\mathbf{z})$, $\mathbf{z}$ typical k-sequence of $\mathbf{g^k}(\mathbf{X'^k})$, are encoded in typical n-sequences of one and only one $\mathbf{A_i} \subset (\mathbf{f^n})^{-1}(\mathbf{y_i})$ ($\mathbf{y_i} \in \{\mathbf{y_1}, \ldots, \mathbf{y_N}\}$). So, if $\mathbf{x'_1}$ and $\mathbf{x'_2}$ are two typical k-sequences of $\mathbf{X'^k}$:

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

Definition 4: With the notations given above, the ratio $R = \frac{k.H(X')}{n}$ is called the *encoding input rate*. An input rate R is said to be *achievable with respect to the function* f if there exists a sequence of codes of size n such that the maximal probability of decoding error tends to 0 as n tends to infinity.

Theorem 2: If $R < C_f(F)$, then R is achievable w.r.t f. Conversely, if $R > C_f(F)$, there is no code such that the error probability tends to 0 as $n \to \infty$

Proof: Only a sketch of the proof is given.

The proof of achievability is conducted in two steps. First, it is shown that the injective encoding of typical k-sequences of a set $(\mathbf{g^k})^{-1}(\mathbf{z})$ on typical n-sequences belonging to $\mathbf{A_i}$ is possible for suitably chosen k and n (lossless coding). Secondly, it is shown that, at input rates below capacity and for k and n suitably chosen, the sets $\mathbf{A_i}$ are almost as many as the sets $(\mathbf{g^k})^{-1}(\mathbf{z})$. This will prove that it is possible to find a code fulfilling the encoding and decoding constraints. Thus, considering only the typical sequences of $\mathbf{X'^k}$, the maximal error probability will be upper bounded by λ.

Let $\delta'' > 0$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, there exist k and n such that:

$$\frac{H(X'|g(X'))}{H(X_0|f(X_0))} < \frac{n}{k} < \frac{H(X'|g(X')) + \delta''}{H(X_0|f(X_0))}$$

Moreover, k and n can be chosen as large as needed. Thus:

$$\frac{k.H(X'|g('X))}{H(X_0|f(X_0))} < n < \frac{k.(H(X'|g(X')) + \delta'')}{H(X_0|f(X_0))} \quad (5)$$

We can choose $\delta, \delta' > 0$ and $0 < \epsilon < 1/2$ small enough for:

$$\frac{k.(H(X'|g(X')) + \delta) - log(1-2\epsilon)}{H(X_0|f(X_0)) - \delta'} < n < \frac{k.(H(X'|g(X')) + \delta + \delta'')}{H(X_0|f(X_0)) + \delta'}$$

giving

$$k.(H(X'|g(X')) + \delta) < log(1-2\epsilon) + n.(H(X_0|f(X_0)) - \delta') < n.(H(X_0|f(X_0)) + \delta') < k.(H(X'|g(X')) + \delta + \delta'')$$

If ν_1 is the number of typical k-sequences of $(\mathbf{g^k})^{-1}(\mathbf{z})$ and ν_2 is the number of typical n-sequences in an $\mathbf{A_i}$, we have:

$$\nu_1 < 2^{k.(H(X'|g(X'))+\delta)} < (1-2\epsilon)2^{n.(H(X_0|f(X_0))-\delta')} < \nu_2 < 2^{n.(H(X_0|f(X_0))+\delta')} < 2^{k.(H(X'|g(X'))+\delta+\delta'')}$$

It is thus possible to find an injection from the set of typical k-sequences of $(\mathbf{g^k})^{-1}(\mathbf{z})$ on $\mathbf{A_i}$. This shows the first step.

Assume that $R = kH(X')/n < C_f(F) = H(X_0) - H(f(X_0)|F(X_0))$, X_0 being a source achieving capacity. So

$$k(H(g(X')) + H(X'|g(X'))) < n.(H(f(X_0)) - H(f(X_0)|F(X_0))) + n.H(X_0|f(X_0))$$

By (5), $n.H(X_0|f(X_0)) - k.H(X'|g(X')) < k.\delta''$ thus

$$kH(g(X')) < n.(H(f(X_0)) - H(f(X_0)|F(X_0))) + k.\delta''$$

$\epsilon_1, \delta''' > 0$ can be chosen small enough in order to get:

$$2^{k(H(g(X'))+\delta''')} < 2^{n.(H(f(X_0))-H(f(X_0)|F(X_0))+\frac{k}{n}.\delta''-\epsilon_1)}$$

If ν_3 is the number of typical k-sequences of $\mathbf{g^k}(\mathbf{X'^k})$ and N is the size of the code (i.e., the number of $(\mathbf{A_i}, \mathbf{B_i})$), we have (by AEP and Lemma 1):

$$\nu_3 < 2^{k(H(g(X'))+\delta''')} < 2^{n.(H(f(X_0))-H(f(X_0)|F(X_0))+\frac{k}{n}.\delta''-\epsilon_1)} < N$$

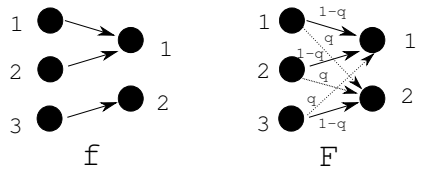

Fig. 2. Example

This ends the proof of step 2 of the "achievability" part.

Assume now that $R > C_f(F)$. As before, it is possible to find an injection from the set of typical k-sequences of $\mathbf{g^k}^{-1}(\mathbf{z})$ on $\mathbf{A_i}$. In other words, inequality (5) holds. Since $R > C_F(f)$:

$$kH(g(X')) > n.(H(f(X_0)) - H(f(X_0)|F(X_0))) + n.H(X_0|f(X_0)) - k.H(X'|g(X'))$$

$$(5) \Rightarrow n.H(X_0|f(X_0)) - k.H(X'|g(X')) > 0$$

So $kH(g(X')) > n.(H(f(X_0)) - H(f(X_0)|F(X_0)))$

We can find $\delta > 0$ small enough such that

$$2^{k(H(g(X'))-\delta)} > 2^{n.(H(f(X_0))-H(f(X_0)|F(X_0))+\delta)}$$

This implies that $\nu_3 > N$. This means that the encoding requires to use sets $\mathbf{A_0}$ which do not belong to the input code. By lemma 2, the error probability does not vanish.

Non typical (i.e., negligible) k-sequences and n-sequences have not been considered in the sketch of the demonstration as being of vanishing total probability. ∎

In the proof of the converse part, we could relax the assumptions on $\mathcal{V}$: even if $\mathcal{V}$ is non injective, at input rates above capacity, the error probability does not vanish.

V. Example

Let $f : \{1,2,3\} \to \{1,2\}$ be the function such that $f(1) = f(2) = 1$ and $f(3) = 2$. f is the simplest non trivial discrete function which is non injective. Let F be a noisy implementation of f such that $P(F(x) = f(x)) = 1 - q$, $q \neq 1/2$ (figure 2). Let α be the value $2^{\frac{1}{1-2q}}$. The capacity $C_f(F)$ is given by the expression:

$$C = \frac{(1+\alpha)(1-q)-1}{(1+\alpha)(1-2q)} + q\log(q) + (1-q)\log(1-q) - \frac{\alpha}{1+\alpha}\log(\alpha) + \log(1+\alpha)$$

For example, if $q = 0.1$, then $C_f(F) = 1.16235$. If we compute the *channel capacity* of the noisy function F (considered as a *transmission channel*), we obtain 0.531 which is lower than $C_f(F)$. This is totally consistent with the intuitive interpretation of equivocation: the amount of information to add to a noisy result to retrieve the input of the non-invertible function is likely to be larger than the amount of information we must add to a noisy result to obtain the correct one.

VI. Conclusion

The coding theorem (as the companion lemma) allows to state that information redundancy through coding is a possible way to reliably compute when given a noisy device. This noisy device needs no embedded (i.e., gate) redundancy to reach reliability. In addition to encoding and decoding, the "price to pay" is that only suitably selected input blocks among all the possible ones are used. In practical contexts, the noisy apparatus has to be built in accordance with a companion code leading to efficient encoding and decoding. Moreover, coding can efficiently be complemented with gate redundancy: the more gate redundancy, the less noise and, thus, a better capacity. Gate redundancy and information redundancy are two levers on which reliability can be built.

While the definition of F does not relate to f, the capacity of F depends upon the desired function f. According to the function f to be "extracted" from F, the capacity will be more or less. The limit case is when $H(f(X)|F(X))$ is minimum (i.e., 0) which corresponds to the availability of perfect decoding : $f(X)$ is a deterministic function of $F(X)$. This is possible if F is itself deterministic, thus either perfect or "totally" noisy.

The constraints on the encoding and the decoding processes participate to forbid the model to reduce to a coder computing the desired function f and transmitting the result through a noisy channel whose output is given to a decoder. If f is non injective, the injective coder cannot handle the same computation. The same holds for the decoder thanks to the two step paradigm. Moreover, Lemma 1 expresses the "computing capacity" the noisy function F possesses with respect to f for a given source.

The proposed model of reliable computation involves two perfect functions g and f. This is intended to capture major real cases, for example, a self-checking adder designed to handle operands in a residue code. The desired function g is a regular arithmetic addition while the actual unreliable circuit implements an regular arithmetic addition aside a modulo-adder acting on the residues ([11]).

References

[1] M. Nicolaidis, "Efficient implementations of self-checking adders and ALUs," in *International Symposium on Fault Tolerant Computing*, 1993.

[2] D. A. Spielman, "Highly fault-tolerant parallel computation," in *Annual Symposium Foundations of Computer Science*, 1996, pp. 154–160.

[3] P. Gacs, "Reliable computation," Boston University, Tech. Rep., 2005.

[4] C. N. Hadjicostis and G. C. Verghese, "Coding approaches to fault tolerance in linear dynamic systems," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 210–228, january 2005.

[5] E. Rachlin and J. E. Savage, "A framework for coded computation," in *ISIT 2008*, 2008.

[6] J. Von Neumann, "Probabilistics logics and the synthesis of reliable organisms from unreliable components," *Automata studies*, 1956.

[7] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Transactions on information theory*, vol. 53, no. 10, october 2007.

[8] S. Winograd and J. Cowan, *Reliable computation in the presence of noise*. The MIT Press, 1963.

[9] T. J. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[10] J. Wolfowitz, *Coding theorems of information theory*. Springer, 1978.

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