Nested Canalizing Functions and Their Networks

C. Kadelka\textsuperscript{a,1,*}, Y. Li\textsuperscript{b,2}, J.O. Adeyeye\textsuperscript{b,2}, R. Laubenbacher\textsuperscript{c,d,1}

\textsuperscript{a}Virginia Bioinformatics Institute & Department of Mathematics, Virginia Tech, Blacksburg, VA 24061, USA
\textsuperscript{b}Department of Mathematics, Winston-Salem State University, NC 27110, USA
\textsuperscript{c}Center for Quantitative Medicine, University of Connecticut Health Center, Farmington, CT 06030, USA
\textsuperscript{d}Jackson Laboratory for Genomic Medicine, Farmington, CT 06030, USA

Abstract

The concept of a nested canalizing Boolean function has been studied over the last decade in the context of understanding the regulatory logic of molecular interaction networks, such as gene regulatory networks. Such networks are predominantly governed by nested canalizing functions. Derrida values are frequently used to analyze the robustness of a Boolean network to perturbations. This paper introduces closed formulas for the calculation of Derrida values of networks governed by Boolean nested canalizing functions, which previously required extensive simulations. Recently, the concept of nested canalizing functions has been generalized to include multistate functions, and a recursive formula has been derived for their number, as a function of the number of variables. This paper contains a detailed analysis of the class of nested canalizing functions over an arbitrary finite field. In addition, the concept of nested canalization is further generalized and closed formulas for the number of such generalized functions, as well as for the number of equivalence classes under permutation of variables, are derived. The latter is motivated by the fact that two nested canalizing functions that differ only by a permutation of the variables share many important properties.

Keywords: Nested canalizing function, Derrida value, Discrete dynamical system, Formula, Multistate

1. Introduction

S. Kauffman introduced canalizing Boolean functions as appropriate rules in Boolean network models of gene regulatory networks \cite{8}. In \cite{7}, a formula for the number of canalizing Boolean functions was presented. More recently, a subclass of these functions, so-called nested canalizing functions (NCFs), was introduced \cite{9}. The interest in studying these came from dynamic stability properties of networks constructed from such functions. Later, canalizing functions were generalized to include functions that take values in an arbitrary finite set, and a closed formula for the number of such functions was obtained \cite{12}. A multistate version of nested canalizing functions has been introduced in \cite{15,16}. It was shown that networks whose dynamics are controlled by multistate nested canalizing functions have stability properties similar to the Boolean case, namely large attractor basins and short limit cycles. An analysis of published Boolean and multistate models of molecular regulatory networks revealed that the large majority of regulatory rules is canalizing, with most

\*Corresponding author

Email addresses: ckadelka@vt.edu (C. Kadelka), liyu@wssu.edu (Y. Li), adeyeyej@wssu.edu (J.O. Adeyeye), laubenbacher@uchc.edu (R. Laubenbacher)

\textsuperscript{1}Supported by NSF Grant CMMI-0908201

\textsuperscript{2}Supported by US DoD # W911NF-11-10166
rules being in fact nested canalizing \[3, 10, 15, 18\]. Thus, nested canalizing rules and the properties of networks governed by them are important to study because of their relevance in systems biology. Furthermore, they also play a role in computational science; it was shown that the class of Boolean NCFs is identical to the class of so-called unate cascade Boolean functions, which has been studied extensively in engineering and computer science \[4\]. This class, in turn, corresponds exactly to the class of Boolean functions with corresponding binary decision diagrams of shortest average path length \[1\]. Thus, a more detailed mathematical study of NCFs has applications to problems in engineering and computer science as well.

This paper is ordered as follows. In Section 2 we introduce the concept of (nested) canalization. In Section 3 we present a way to calculate the values of the Derrida plot for networks governed by Boolean NCFs. This plot is a popular tool to evaluate the robustness of molecular interaction networks and its calculation previously required extensive simulations \[6\]. A comprehensive analysis of Boolean NCFs was given in \[11\]. In the second part of this paper, we conduct a similar analysis for the multistate case. In Section 4 we derive a unique polynomial representation of NCFs, as well as explicit formulas for the number of NCFs and equivalence classes of NCFs for a given number of variables. Lastly, in Section 5 we describe a generalization of the concept of NCFs. We finish with the conclusion in Section 6.

2. Definitions and Notation

In this section we review some concepts and definitions from \[15, 16\] to introduce the computational concept of canalization. Let \(\mathbb{F} = \mathbb{F}_p\) be a finite field with \(p\) elements, where \(p\) is prime. Note that in earlier work \(\mathbb{F}\) only had to be a finite set. The more stringent requirement on \(\mathbb{F}\) allows the use of a wider range of mathematical tools, while all results that were previously discovered for finite sets remain valid. In particular, network models can still be represented as polynomial dynamical systems \[20\].

**Definition 2.1.** A function \(f(x_1, x_2, \ldots, x_n)\) is essential in the variable \(x_i\) if there exist \(r, s \in \mathbb{F}\) and \((x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \in \mathbb{F}^{n-1}\) such that
\[
f(x_1, \ldots, x_{i-1}, r, x_{i+1}, \ldots, x_n) \neq f(x_1, \ldots, x_{i-1}, s, x_{i+1}, \ldots, x_n).
\]

**Definition 2.2.** \[12\] Given \(a, b \in \mathbb{F}\) and \(i \in \{1, \ldots, n\}\), a function \(f(x_1, x_2, \ldots, x_n)\) is \(< i : a : b >\) canalizing if for all \((x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \in \mathbb{F}^{n-1}\)
\[
f(x_1, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_n) = b.
\]

We call \(x_i\) the canalizing variable of \(f\). The set \(S\) of all \(a \in \mathbb{F}\) such that \(f\) is \(< i : a : b >\) canalizing for some \(b\) will be called the canalizing input set with respect to \(x_i\), and \(b\) will be called the canalized output with respect to \(x_i\) and \(a\).

We now assume that \(\mathbb{F} = \{0, 1, \ldots, p - 1\}\) is ordered, in the natural order \(0 < 1 < \cdots < p - 1\). A proper subset \(S\) of \(\mathbb{F}\) is called an interval if and only if \(S = \{0, \ldots, j\}\) or \(S^c = \mathbb{F} - S = \{0, \ldots, j\}\), where \(0 \leq j < p - 1\). Hence, a proper subset \(S\) is an interval if and only if \(S^c\) is an interval.

**Definition 2.3.** \[16\] Let \(f\) be a function in \(n\) variables and \(S_i\) be intervals of \(\mathbb{F}\), \(i = 1, \ldots, n\). Let \(\sigma\) be a permutation of the set \(\{1, 2, \ldots, n\}\). Then \(f\) is a nested canalizing function (NCF) in the variable order \(x_{\sigma(1)}, \ldots, x_{\sigma(n)}\) with canalizing input sets \(S_1, \ldots, S_n\) and canalized output values
Note that all variables appearing in the definition of an NCF must be essential. A constant function \( V \) positions, and let \( I \) \( \cup \mathcal{J} \Rightarrow \mathcal{J} \) simulations. We derive direct formulas for the Derrida value in networks governed by NCFs.

In this section we formally define the concept of Derrida values in the framework of Boolean discrete environment changes but also flexible enough to allow adaptation \cite{5, 19}. Many biological systems seem to remain of similar size, the system is in the critical regime. A chaotic network possesses long limit cycles. Lastly, if the perturbation amplifies over time, the system is in the chaotic regime. A chaotic network possesses long limit cycles. Lastly, if the perturbation remains of similar size, the system is in the critical regime. Many biological systems seem to work in this critical regime; they must be robust enough to withstand perturbations caused by environmental changes but also flexible enough to allow adaptation \cite{3] [19].

In this section we formally define the concept of Derrida values in the framework of Boolean discrete dynamical systems. Until now, the calculation of Derrida values has required extensive Monte Carlo simulations. We derive direct formulas for the Derrida value in networks governed by NCFs.

**Definition 3.1.** Let \( F = (f_i)_{i=1}^n \) be a synchronously updated Boolean discrete dynamical system. Let \( I(i) \) be the set of essential variables of \( f_i \). We call \( k_i = |I(i)| \) the connectivity of \( f_i \). Moreover, let \( x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \mathbb{F}_2^n \) be two system configurations that differ at \( m \in \{1, \ldots, n\} \) positions, and let \( V = \{i | x_i \neq y_i\} \) be the set of positions where they differ. Lastly, let \( J(i) = I(i) \cap V \) be the set of essential variables of node \( i \) where \( x \) and \( y \) differ. Then for an initial perturbation of size \( m \), the Derrida value, \( D(m) \), is defined as the average size of the perturbation after one update,

\[
D(m) = \mathbb{E} \left[ d(F(x), F(y)) | d(x, y) = m \right],
\]

where \( d : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1, \ldots, n\} \) is the Hamming distance (the standard \( \ell^1 \) metric).

**Theorem 3.2.** The Derrida value can be calculated as the sum of the node-specific probabilities that \( x \) and \( y \) differ after the update,

\[
D(m) = \sum_{i=1}^n \mathbb{P}(f_i(x) \neq f_i(y)) = \sum_{i=1}^n \sum_{c=0}^m \mathbb{P}(|J(i)| = c) q(c, f_i),
\]
where
\[ q(c, f_i) = P((f_i(x))i \neq (f_i(y))i \mid |I(i)| = k_i, |J(i)| = c) \]
describes the probability that the ith component of x and y varies after the update, given the update function and the number of essential variables where x and y differ are known. |J(i)| follows a hypergeometric distribution,
\[ P(|J(i)| = c) = \frac{\binom{m}{c} \binom{n-m}{k_c-c}}{\binom{n}{k_c}} = \frac{\binom{k_i}{c} \binom{n-k_i}{m-c}}{\binom{n}{m}} \]

Proof. Let \( x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \mathbb{F}_2^n \) be two system configurations that differ at m positions. Since \( x, y \) are synchronously updated, the update of each component is independent from the update of other components. This implies that the Derrida value is simply the sum of the probabilities that \( x_i \) and \( y_i \) differ after the update, \( i = 1, \ldots, n \). Conditioning with respect to \( |J(i)| \) leads to the second equality. \( J(i) \) is the intersection of two sets so that its magnitude \( |J(i)| \) follows a hypergeometric distribution.

\[ \square \]

**Lemma 3.3.** Let \( f \) be an NCF with \( k \) essential variables, and let \( c \) be the number of essential variables where two system configurations \( x \) and \( y \) differ. Then the q-value in Theorem 3.2 can be evaluated,

\[ q(c, f) = \begin{cases} \frac{1}{2} & \text{if } c = 1 \\ \frac{1}{2^n} \sum_{j=1}^{k-c+1} \binom{k_j}{c-1} \left( \frac{1}{2} \right)^j & \text{if } 1 < c \leq k \end{cases} \]

Proof. Boolean NCFs correspond to the class of Boolean functions with corresponding shortest binary decision diagram [1]. Let \( f \) be an NCF as in Definition 2.3 with \( k \) essential variables. The value of half the states in \( f \)'s truth table is determined by only one variable, for a quarter of the states two variables matter, for an eighth three variables play a role, etc.

(i) If \( x_{\sigma(1)} \neq y_{\sigma(1)} \), w.l.o.g. we can assume \( x_{\sigma(1)} \in S_1 \), which means \( f(x) = b_1 \) and \( f(y) = b_j \) for some \( j \in \{2, \ldots, k+1\} \). Since \( P(b_1 = b_j) = \frac{1}{2} \), \( f(x) = f(y) \) with probability \( \frac{1}{2} \) as well. If on the other hand \( x_{\sigma(1)} = y_{\sigma(1)} \), either \( f(x) = f(y) = b_1 \) or \( \sigma(2) \) is required for the evaluation, both with probability \( \frac{1}{2} \). Only in the latter case, may \( f(x) \) and \( f(y) \) differ. If \( x_{\sigma(2)} \neq y_{\sigma(2)} \), w.l.o.g. we can again assume \( x_{\sigma(2)} \in S_2 \) and obtain with the same argument as before that \( f(x) = f(y) \) with probability \( \frac{1}{2} \). If \( x_{\sigma(2)} = y_{\sigma(2)} \), the decision moves to the third most important variable, etc.

In case the least important variable, \( \sigma(k) \), is required for the evaluation and \( x_{\sigma(k)} = y_{\sigma(k)} \), then \( f(x) \) and \( f(y) \) must differ because \( \sigma(k) \) would not be an essential variable otherwise \( b_k \neq b_{k+1} \), by Definition 2.3. Therefore,

\[ P(f(x) \neq f(y)) = \begin{cases} \frac{1}{2} & \text{if } x_{\sigma(1)} \neq y_{\sigma(1)} \\ \frac{1}{4} & \text{if } x_{\sigma(1)} = y_{\sigma(1)}, x_{\sigma(2)} \neq y_{\sigma(2)} \\ \frac{1}{8} & \text{if } x_{\sigma(1)} = y_{\sigma(1)}, x_{\sigma(2)} = y_{\sigma(2)}, x_{\sigma(3)} \neq y_{\sigma(3)} \\ \vdots & \quad \vdots \\ \frac{1}{2^{k-1}} & \text{if } x_{\sigma(1)} = y_{\sigma(1)}, \ldots, x_{\sigma(k-2)} = y_{\sigma(k-2)}, x_{\sigma(k-1)} \neq y_{\sigma(k-1)} \\ \frac{1}{2^k} & \text{if } x_{\sigma(1)} = y_{\sigma(1)}, \ldots, x_{\sigma(k-2)} = y_{\sigma(k-2)}, x_{\sigma(k-1)} = y_{\sigma(k-1)}, x_{\sigma(k)} \neq y_{\sigma(k)} \end{cases} \]

(ii) There are \( \binom{k}{c} \) c-subsets in the set \( \{1, \ldots, k\} \). Of these c-subsets, \( \binom{k-1}{c-1} \) contain 1 as its lowest element, \( \binom{k-2}{c-1} \) contain 2 as its lowest element, etc., and finally, \( \binom{c-1}{c-1} = 1 \) contain \( k - c + 1 \) as its
lowest element. Thus if two configurations \( x \) and \( y \) differ at \( c \) essential variables, there are \( \binom{k-j}{c-1} \) possibilities that they first differ at \( \sigma(j) \).

(iii) If \( c > 1 \), \( x \) and \( y \) will already first differ before \( \sigma(k) \). The least important variable is therefore never needed to calculate the probability that \( f(x) \) and \( f(y) \) differ. From (i) and (ii), we get

\[
q(c, f) = \mathbb{P}(f(x) \neq f(y)) = \frac{1}{k} \sum_{j=1}^{k-c+1} \binom{k-j}{c-1} \left( \frac{1}{2} \right)^j.
\]

If \( c = 1 \), \( x \) and \( y \) might differ at \( \sigma(k) \) for the first time and we get

\[
q(c, f) = \mathbb{P}(f(x) \neq f(y)) = \frac{1}{k} \sum_{j=1}^{k-c+1} \binom{k-j}{c-1} \left( \frac{1}{2} \right)^j.
\]

Lemma 3.3 enables us to calculate average Derrida values for a system of NCFs. The NCFs are however only characterized by the number of their essential variables, which is why this theorem only yields an average Derrida value. We can also distinguish NCFs by their Hamming weight (number of 1’s in the truth table), and find a formula for the Derrida values of a system of NCFs, specified by their Hamming weights.

**Lemma 3.4.** Let \( f \) be an NCF with \( k \) essential variables and Hamming weight \( w \). Let \( c \) be the number of essential variables where two system configurations \( x \) and \( y \) differ. Then the \( q \)-value in Theorem 3.2 can be evaluated,

\[
q(c, f) = \frac{1}{k} \sum_{j=1}^{k-c+1} \binom{k-j}{c-1} \left( \frac{1}{2} \right)^j p_j,
\]

where \( p_0 = \frac{w}{2^k}, \ p_{j+1} = 1 - |1 - 2p_j| \) for \( 0 \leq j \leq k - c \)

**Proof.** Let \( f \) be an NCF with \( k \) essential variables and Hamming weight \( w \).

Let \( p_j = \mathbb{P}(f(x) \neq f(y) \mid x_{\sigma(1)} = y_{\sigma(1)}, \ldots, x_{\sigma(j-1)} = y_{\sigma(j-1)}, x_{\sigma(j)} \neq y_{\sigma(j)}) \) for \( 1 \leq j \leq k - c + 1 \).

If \( \frac{w}{2^k} > \frac{1}{2} \), then \( b_1 = 1 \) and if \( \frac{w}{2^k} < \frac{1}{2} \), then \( b_1 = 0 \). Therefore,

\[
p_1 = \begin{cases} 2 \frac{w}{2^k} & \text{if } \frac{w}{2^k} < \frac{1}{2} \\ 2(1 - \frac{w}{2^k}) & \text{if } \frac{w}{2^k} > \frac{1}{2} \end{cases}
\]

More general, \( p_j < \frac{1}{2} \) means that \( b_{j+1} = b_j \), whereas \( p_j > \frac{1}{2} \) means that \( b_{j+1} \neq b_j \). Therefore,

\[
p_{j+1} = \begin{cases} 2p_j & \text{if } p_j < \frac{1}{2} \\ 2(1 - p_j) & \text{if } p_j > \frac{1}{2} \end{cases}
\]

\[= 1 - |1 - 2p_j|\]
and for conclusiveness we set $p_0 = \frac{m}{2^n}$.

As in the previous proof,

$$q(c, f) = \sum_{j=1}^{k-c+1} \mathbb{P}(x_{\sigma(1)} = y_{\sigma(1)}, \ldots, x_{\sigma(j-1)} = y_{\sigma(j)}, x_{\sigma(j)} \neq y_{\sigma(j)} \mid c) \cdot \mathbb{P}(f(x) \neq f(y) \mid x_{\sigma(1)} = y_{\sigma(1)}, \ldots, x_{\sigma(j-1)} = y_{\sigma(j-1)}, x_{\sigma(j)} \neq y_{\sigma(j)})$$

$$= \sum_{j=1}^{k-c+1} \frac{(k-j)}{c} \frac{1}{2}^{j-1} p_j$$

Lemma 3.4 allows the exact calculation of the Derrida plot for a system of Boolean NCFs of any Hamming weight. Especially for large systems, this offers a huge improvement, since the time required to approximate the Derrida plot through simulations increases exponentially in the number of regulators. Figure 1 depicts the Derrida values for networks of $n = 100$ genes, which are governed by NCFs with $k = 5$ regulators and different Hamming weights. The calculation of all the 800 plotted values took less than a second on a regular desktop computer. In networks of NCFs with Hamming weight 1, 3, 29 and 31, small perturbations vanish in average over time. These networks therefore operate in the stable regime. Networks of NCFs with Hamming weight 5, 7, 25 and 27 operate very close to the critical regime, while networks of NCFs with Hamming weight between 11 and 23 operate in the chaotic regime. Surprisingly, networks of NCFs with Hamming weight 11, 13, 19, and 21 are more chaotic than those governed by almost balanced NCFs of Hamming weight 15 and 17. One possible explanation for this observation may be the layer number (see Definition 4.6 in [11] or Definition 4.3 in this paper). NCFs with $k = 5$ and Hamming weight 15 or 17 consist of two layers, while NCFs with Hamming weight 13 and 19 (11 and 21) have three (four) different layers. The number of layers seems to be positively correlated with the Derrida value for small perturbations.

Gene regulatory networks are stochastic in nature. A recently introduced generalization of Boolean networks, called Stochastic Discrete Dynamical Systems (SDDS), captures this inherent stochasticity by assigning gene-specific activation and degradation probabilities, which describe how likely a specific concentration change happens at a given update step [17]. This framework allows modeling different reaction speeds, while preserving the simplicity of a Boolean network model.

**Definition 3.5.** A Boolean stochastic discrete dynamical system is a collection of $n$ triplets $F = (f_i, p_i^\uparrow, p_i^\downarrow)_{i=1}^n$ where

- $f_i : \{0, 1\}^n \rightarrow \{0, 1\}$ is the update function for $x_i$ for all $i = 1, \ldots, n$
- $p_i^\uparrow \in (0, 1]$ is the activation propensity for $x_i$. $\mathbb{P}(x_i \rightarrow 1 | x_i = 0) = \begin{cases} p_i^\uparrow & \text{if } f_i(x) = 1 \\ 0 & \text{if } f_i(x) = 0 \end{cases}$
- $p_i^\downarrow \in (0, 1]$ is the degradation propensity for $x_i$. $\mathbb{P}(x_i \rightarrow 0 | x_i = 1) = \begin{cases} p_i^\downarrow & \text{if } f_i(x) = 0 \\ 0 & \text{if } f_i(x) = 1 \end{cases}$

**Theorem 3.6.** The Derrida value for an SDDS $F = (f_i, p_i^\uparrow, p_i^\downarrow)_{i=1}^n$ can be calculated as

$$D(m) = \sum_{i=1}^n \sum_{c=0}^m \mathbb{P}(|J(i)| = c) \left( \frac{m}{n} (q(c, f_i)r_1 + (1 - q(c, f_i))r_2) + \frac{n-m}{n} (q(c, f_i)r_3 + (1 - q(c, f_i))r_4) \right)$$
Here, $J(i)$ and $q(c, f_i)$ are defined as in Theorem 3.2 and

$$r_1 = 1 - \frac{1}{2}(p_i^+ + p_i^-) + p_i^+ p_i^-$$

$$r_2 = 1 - \frac{1}{2}(p_i^+ + p_i^-)$$

$$r_3 = \frac{1}{2}(p_i^+ + p_i^-)$$

$$r_4 = \frac{1}{2}(p_i^+ (1 - p_i^+) + p_i^- (1 - p_i^-))$$

Proof. Let $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \mathbb{F}_2^n$ be two system configurations that differ at $m$ positions. For each node $i \in \{1, \ldots, n\}$, define three events

$A_i = \{x_i \neq y_i\}$,

$B_i = \{f_i(x) \neq f_i(y) \text{ before applying } p_i^+, p_i^-\}$,

$C_i = \{f_i(x) \neq f_i(y) \text{ after applying } p_i^+, p_i^-\}$.

Then, since $A_i$ is independent from $B_i$, we have

$$D(m) = \sum_{i=1}^{n} P(C_i)$$

$$= \sum_{i=1}^{n} \sum_{c=0}^{m} P(|j(i)| = c) \left( P(C_i | A_i, B_i) P(B_i | c) P(A_i) + P(C_i | A_i, \neg B_i) P(\neg B_i | c) P(A_i) + \ldots \right.$$

$$+ P(C_i | \neg A_i, B_i) P(B_i | c) P(\neg A_i) + P(C_i | \neg A_i, \neg B_i) P(\neg B_i | c) P(\neg A_i) \right)$$

Figure 1: The Derrida values are plotted for networks of $n = 100$ genes, each governed by a nested canalizing function with $k = 5$ regulators. The Hamming weight of the nested canalizing functions varies.
\[
= \sum_{i=1}^{n} \sum_{c=0}^{m} \mathbb{P}(|J(i)| = c) \left( \mathbb{P}(A_i) \mathbb{P}(C_i|A_i, B_i) \mathbb{P}(B_i|c) + \mathbb{P}(C_i|A_i, -B_i) \mathbb{P}(-B_i|c) \right) + \ldots \\
+ \mathbb{P}(-A_i) \left( \mathbb{P}(C_i|-A_i, B_i) \mathbb{P}(B_i|c) + \mathbb{P}(C_i|-A_i, -B_i) \mathbb{P}(-B_i|c) \right)
\]

If \(x\) and \(y\) differ at \(m\) out of \(n\) positions, \(\mathbb{P}(A_i) = \frac{m}{n}\).

\(\mathbb{P}(B_i|c) = q(c, f_i)\) depends on the choice of the update function. If NCFs are chosen, Lemma \ref{lemma:necessary} and Lemma \ref{lemma:3} yield the probability.

Lastly, the probability that \(x\) and \(y\) differ after applying the propensity probabilities needs to be calculated. If \(x_i \neq y_i\) and \(f_i(x) \neq f_i(y)\) before applying \(p_i^\uparrow, p_i^\downarrow\), assume w.l.o.g. that \(x_i = 0, y_i = 1\). Then, either \(f_i(x) = 0, f_i(y) = 1\) or \(f_i(x) = 1, f_i(y) = 0\), both with probability \(\frac{1}{2}\). In the first case, there is no change in values so that the propensity probabilities play no role and \(f_i(x) \neq f_i(y)\) after applying \(p_i^\uparrow, p_i^\downarrow\) with probability 1. In the second case, \(f_i(x)\) and \(f_i(y)\) only differ after applying \(p_i^\uparrow, p_i^\downarrow\), if both updates happen (probability \(p_i^\uparrow p_i^\downarrow\)) or neither update happens (probability \((1 - p_i^\uparrow)(1 - p_i^\downarrow)\)). That means,

\[
\mathbb{P}(C_i|A_i, B_i) = \frac{1}{2} \cdot 1 + \frac{1}{2}(p_i^\uparrow p_i^\downarrow + (1 - p_i^\uparrow)(1 - p_i^\downarrow)) = 1 - \frac{1}{2}(p_i^\uparrow + p_i^\downarrow) + p_i^\uparrow p_i^\downarrow
\]

Similarly, we can derive

\[
\mathbb{P}(C_i|A_i, -B_i) = 1 - \frac{1}{2}(p_i^\uparrow + p_i^\downarrow)
\]

\[
\mathbb{P}(C_i|\neg A_i, B_i) = \frac{1}{2}(p_i^\uparrow + p_i^\downarrow)
\]

\[
\mathbb{P}(C_i|\neg A_i, -B_i) = \frac{1}{2}(p_i^\uparrow (1 - p_i^\downarrow) + p_i^\downarrow (1 - p_i^\uparrow))
\]

\(\square\)

4. Characterization of Nested Canalizing Functions

In this section we first derive a unique polynomial representation of multistate NCFs, which allows a further categorization of this class of functions. Then, we use this polynomial representation to derive explicit formulas for the number of NCFs and for the number of equivalence classes of NCFs. In the Boolean case, the extended monomial plays an important role in determining a novel polynomial form of NCFs \cite{11}. In the multistate case, the product of indicator functions, which was used in \cite{16}, will take over this role.

**Definition 4.1.** Given a proper subset \(S\) of \(\mathbb{F}\), the indicator function (of \(S^c\)) is defined as

\[
Q_S(x) = \begin{cases} 
0 & x \in S, \\
1 & x \in S^c.
\end{cases}
\]

The following theorem states the main result of this section. It gives an algebraic characterization of nested canalizing functions.

**Theorem 4.2.** Given \(n \geq 2\), the function \(f(x_1, \ldots, x_n)\) is nested canalizing if and only if it can be uniquely written as

\[
f(x_1, \ldots, x_n) = M_1(M_2(\cdots (M_{r-1}(B_{r+1}M_r + B_r) + B_{r-1}) \cdots ) + B_2) + B_1,
\]

where each \(M_i\) is a product of indicator functions of a set of disjoint variables. More precisely, \(M_i = \prod_{j=1}^{k_i} Q_{S_{ij}}(x_{ij}), i = 1, \ldots, r, k_i \geq 1\) for \(i = 1, \ldots, r, k_1 + \cdots + k_r = n, B_2, \ldots, B_{r+1} \neq 0, B_1 \in \mathbb{F}, \{i_j|j = 1, \ldots, k_i, i = 1, \ldots, r\} = \{1, \ldots, n\}, \) and if \(k_r = 1\), then \(B_{r+1} + B_r \neq 0\).
Proof. First, let $b_i = \sum_{j=1}^{i} B_j$. Then it is straightforward to check that any function written as in Equation 4.1 is a $\{\sigma^* : S^* : \beta^*\}$ NCF, where $\sigma^*(x_1, \ldots, x_n) = (x_1^{k_1}, \ldots, x_{k_1}^{k_2}, \ldots, x_{k_r}^{k_r})$, $S^* = (S_1^{k_1}, \ldots, S_{k_1}^{k_2}, \ldots, S_{k_r}^{k_r})$, and $\beta^* = (b_1, b_2, \ldots, b_{n+1})$.

Second, suppose $f$ is a $\{\sigma : S : \beta\}$ NCF, where $S = (S_1, S_2, \ldots, S_n)$ and $\beta = (b_1, b_2, \ldots, b_{n+1})$ with $b_n \neq b_{n+1}$. Then there exist $k_i, i = 1, \ldots, r$, $k_1 + \cdots + k_r = n$, $k_i \geq 1$, such that $b_1 = \cdots = b_{k_1} = C_1, b_{k_1+1} = \cdots = b_{k_1+k_2} = C_2, \ldots, b_{k_1+\cdots+k_{r-1}+1} = \cdots = b_n = C_r, b_{n+1} = C_{r+1}$, and $C_j \neq C_{j+1}, j = 1, \ldots, r$. Let $B_1 := C_1, B_2 := C_2 - C_1, \ldots, B_{r+1} = C_{r+1} - C_r$. Hence, $B_1 \in \mathbb{F}, B_2, \ldots, B_{r+1} \in \mathbb{F} - \{0\}$, and $M_1(M_2(\cdots(M_{r+1}(M_r + B_r) + B_{r-1})\cdots) + B_2) + B_1$ equals $f$, which shows that any NCF can be written as in Equation 4.1.

Finally, we need to show that each NCF has a unique polynomial representation. Without loss of generality, let $\sigma$ be the identity permutation, i.e., let $f$ be nested canalizing in the variable order $x_1, \ldots, x_n$. Besides, let $f$ be written as in Equation 4.1. Then all the variables of $M_1$, $x_1, \ldots, x_{k_1}$, are canalizing variables of $f$ with common canalized output $B_1$. We will now show that $f$ has no other canalizing variables to prove the uniqueness of $M_1$ and $B_1$. If $x_1 \in S_1^{k_1}, \ldots, x_{k_1} \in S_1^{k_1}$, then all the variables of $M_2$, $x_{k_1+1}, \ldots, x_{k_2}$, are canalizing variables of the subfunction $f_1 = M_2(\cdots(M_{r+1}(M_r + B_r) + B_{r-1})\cdots) + B_2$.

Because each NCF can be uniquely written in the form of Equation 4.1, the number $r$ is uniquely determined by $f$, and can be used to specify the class of NCFs.

Definition 4.3. For an NCF $f$, written in the form of Equation 4.1, let the number $r$ be called its layer number. Essential variables of $M_1$ are called most dominant variables (canalizing variables), and are part of the first layer of $f$. Essential variables of $M_2$ are called second most dominant variables, and are part of the second layer, etc.

Remark 4.4. In the Boolean case ($p = 2$), each $M_i$ in Theorem 4.2 is an extended monomial, and, since $B_{r+1} + B_r = 1 + 1 = 0$, $k_r$ is greater than 1. Thus, Theorem 4.2 reduces to its Boolean version, already stated as Theorem 4.2 in [11]. On the other hand, if $B_{r+1} + B_r$ would be zero, then $k_r = 1$ is impossible since the function could be uniquely written in $r - 1$ layers (see comments after Lemma 4.6).

Remark 4.5. Equation 4.1 allows the use of Corollary 4.8 in [11]. The layer number of any NCF can be determined by counting the number of changes in the canalized output values. For example, if $p = 3$ and if $f$ is nested canalizing with canalized output values $\beta = (1, 1, 1, 0, 0, 0, 2, 0, 0, 2, 2, 1)$ ($n = 11$), then the layer number of $f$ is 5.

Lemma 4.6. Let $a, b$ be any nonzero elements of $\mathbb{F}$, and let $S$ be any interval of $\mathbb{F}$. The number of different functions $f = bQ_S(x) + a$, which cannot be written as $cQ_{S'}(x)$, where $c \neq 0$ and $S'$ is an interval of $\mathbb{F}$, is $(p - 1)^2(p - 2)$.

Proof. If a function $f(x) = bQ_S(x) + a$ can be written as $cQ_{S'}(x)$, then

$$f = bQ_S(x) + a = \begin{cases} a & x \in S \\ a + b & x \in S^c \end{cases} = \begin{cases} 0 & x \in S' \\ c & x \in S'^c \end{cases} = cQ_{S'}(x).$$
Since $a$ and $c$ are nonzero, $a + b = 0 \iff a = -b$ must hold for such a function. Since $F$ contains $p - 1$ nonzero numbers, there are $p - 1$ choices for $b$ and $p - 2$ choices for $a$, to obtain a function that cannot be written as $cQ_{S'}(x)$. Moreover, there are $2(p - 1)$ different intervals $S$, but only half of them lead to a different function since every function can be expressed in two different ways:

$$bQ_S(x) + a = b(1 - Q_{S'}(x)) + a = -bQ_{S'}(x) + (a + b).$$

Thus, there are $(p - 1)^2(p - 2)$ different functions $f = bQ_S(x) + a$ that cannot be written as $cQ_{S'}(x)$. □

This proof also shows that $bQ_S(x) + a = cQ_{S'}(x)$ for some $c \neq 0$ and some interval $S'$ of $F$, if and only if $a + b = 0$.

**Lemma 4.7.** Given $a, b \neq 0$ and intervals $S_i$, $i = 1, \ldots, k$ with $k \geq 2$, then

1. $f(x) = f(x_1, \ldots, x_k) = b \prod_{j=1}^{k} Q_{S_j}(x_j) + a$ cannot be written as $c \prod_{j=1}^{k} Q_{S'_j}(x_j)$, where $c \neq 0$ and all $S'_j$ are intervals, $j = 1, \ldots, k$.
2. There are $2^k(p - 1)^{k+2}$ different functions of the form $b \prod_{j=1}^{k} Q_{S_j}(x_j) + a$.

**Proof.** (1) Assume a function $f(x) = b \prod_{j=1}^{k} Q_{S_j}(x_j) + a$ can be written as $c \prod_{j=1}^{k} Q_{S'_j}(x_j)$, then

$$b \prod_{j=1}^{k} Q_{S_j}(x_j) + a = \begin{cases} a & \exists j : x_j \in S_j \\ a + b & \forall j : x_j \in S'_j \end{cases}$$

$$= \begin{cases} a & x \in (S_1 \times \mathbb{F}^{k-1}) \cup \ldots \cup (\mathbb{F}^{k-1} \times S_k) \\ a + b & x \in S'_1 \times \ldots \times S'_k \\ 0 & x \in (S'_1 \times \mathbb{F}^{k-1}) \cup \ldots \cup (\mathbb{F}^{k-1} \times S'_k) \\ c & x \in S'_1 \times \ldots \times S'_k \end{cases} = c \prod_{j=1}^{k} Q_{S'_j}(x_j).$$

Since $a, c \neq 0$, $a + b = 0$ must hold. Hence,

$$S'_1 \times \ldots \times S'_k = (S'_1 \times \mathbb{F}^{k-1}) \cup \ldots \cup (\mathbb{F}^{k-1} \times S'_k).$$

This last statement is, however, impossible. Thus, there is no function $f(x) = b \prod_{j=1}^{k} Q_{S_j}(x_j) + a$ that can be written as $c \prod_{j=1}^{k} Q_{S'_j}(x_j)$.

(2) The nonzero constants $a, b$ can be arbitrarily chosen, with $p - 1$ choices each. Contrary to the previous lemma, each choice of intervals $S_1, \ldots, S_k$ leads to a different function because $S'_1 \times \ldots \times S'_k \neq (S_1 \times \mathbb{F}^{k-1}) \cup \ldots \cup (\mathbb{F}^{k-1} \times S_k)$ and because $a \neq a + b$. For each interval there are $2(p - 1)$ choices, which is why altogether there are $(p - 1)(p - 1)(2(p - 1))^k = 2^k(p - 1)^{k+2}$ different functions of the form $b \prod_{j=1}^{k} Q_{S_j}(x_j) + a$. □

Let $\mathbb{NCF}(n)$ denote the set of all nested canalizing functions in $n$ variables.

**Theorem 4.8.** For $n \geq 2$, the number of nested canalizing functions is given by

$$|\mathbb{NCF}(n)| = 2^{n-1}p(p - 2) \sum_{r=2}^{n} (p - 1)^{n-r-1} \sum_{k_1 + \ldots + k_{r-1} = n-1 \atop k_i \geq 1, i = 1, \ldots, r-1} \frac{n!}{k_1!k_2!\ldots k_{r-1}!} + 2^n p \sum_{r=1}^{n-1} (p - 1)^{n+r} \sum_{k_1 + \ldots + k_r = n \atop k_i \geq 1, i = 1, \ldots, r-2, k_r \geq 2} \frac{n!}{k_1!k_2!\ldots k_r!}.$$
Proof. If \( r = 1 \), then \( f = B_2M_1 + B_1 \). Similar to Lemma \([4, 7]\) the number of such functions is \((2(p - 1))^n(p - 1)p = 2^n(p - 1)^{n+1}p\), since \( B_1 \in \mathbb{F} \) can be arbitrarily chosen, and \( k_1 = n \).

For \( r > 1 \), Equation \([4, 1]\) yields that for each choice of \( k_1, \ldots, k_r, k_i \geq 1, \ i = 1, \ldots, r \), there are \((2(p - 1))^{k_j} (n - k_1 - \cdots - k_{j-1})\) ways to form \( M_j, j = 1, \ldots, r \). For those NCFs with \( k_r = 1 \), by Lemma \([4, 6]\) there are \((p - 1)^2(p - 2)\) different functions of the form \( B_{r+1}M_r + B_r \) with \( B_r, B_{r+1} \neq 0 \). For the remaining NCFs, i.e. those with \( k_r > 1 \), Lemma \([4, 7]\) yields that there are \((p - 1)^2(2(p - 1))^{k_r}\) ways to form \( B_{r+1}M_r + B_r \), with \( B_r, B_{r+1} \neq 0 \).

Note that there are \( p - 1 \) choices for each \( B_i, 2 \leq r \leq B_{r-1} \), \( p \) choices for \( B_1 \), and \( 2(p - 1) \) choices for each canalizing input interval. Hence, the total number of NCFs with \( r > 1, k_r = 1 \), can be given by

\[
N_1 = \sum_{r=2}^{n} \sum_{k_1 + \cdots + k_{r-1} = n-1 \atop k_i \geq 1, i=1, \ldots, r-1} (2(p - 1))^{k_1+\cdots+k_{r-1}} \binom{n}{k_1} \binom{n-k_1}{k_2} \cdots \binom{n-k_1-\cdots-k_{r-2}}{k_{r-1}} (p-1)^2(p-1)^{r-2}p
\]

Similarly, the total number of NCFs with \( r > 1, k_r > 1 \) is

\[
N_2 = \sum_{r=2}^{n-1} \sum_{k_1 + \cdots + k_{r-1} = n \atop k_i \geq 1, i=1, \ldots, r-1, k_r \geq 2} (2(p - 1))^{k_1+\cdots+k_{r-1}} \binom{n}{k_1} \binom{n-k_1}{k_2} \cdots \binom{n-k_1-\cdots-k_{r-1}}{k_{r}} (p-1)^2(p-1)^{r-2}p
\]

By combining all three groups of NCFs, the total number of NCFs in \( n \) variables is

\[
|\text{NCF}(n)| = 2^n(p - 1)^{n+1}p + N_1 + N_2
\]

\[
= 2^{n-1}p(p - 2) \sum_{r=2}^{n} (p - 1)^{n+r-1} \sum_{k_1 + \cdots + k_{r-1} = n-1 \atop k_i \geq 1, i=1, \ldots, r-1, k_r \geq 2} \frac{n!}{k_1!k_2!\cdots k_{r-1}!}
\]

\[
+ 2^n \sum_{r=1}^{n-1} (p - 1)^{n+r} \sum_{k_1 + \cdots + k_{r-1} = n \atop k_i \geq 1, i=1, \ldots, r-1, k_r \geq 2} \frac{n!}{k_1!k_2!\cdots k_{r-1}!}
\]
respectively; when \( p = 5 \) and \( n = 2, 3, 4 \), we get 5120, 547840, 78561280, respectively. These results are consistent with those in [16].

It has been shown in [16] that the number of multistate NCFs can be calculated recursively. Thus, by equating (4.2) to the recursive relation, we have

**Corollary 4.9.** For the nonlinear recursive sequence

\[
a_2 = 4(p-1)^4, a_n = \sum_{r=2}^{n-1} \binom{n}{r} 2^{r-1}(p-1)^r a_{n-r+1} + 2^{n-1}(p-1)^{n+1}(2 + n(p-2)), n \geq 3
\]

it holds that

\[
|\text{NCF}(n)| = pa_n,
\]

and the explicit solution for \( a_n \) is given by

\[
a_n = 2^{n-1}(p-2) \sum_{r=2}^{n} (p-1)^{n+r-1} \frac{n!}{k_1!k_2! \cdots k_{r-1}!} \sum_{k_1+\cdots+k_{r-1}=n-1} \sum_{k_i \geq 1, i=1,\ldots,r-1} \frac{n!}{k_1!k_2! \cdots k_r!} \sum_{k_i \geq 1, i=1,\ldots,r-1, k_r \geq 2}
\]

**Definition 4.10.** Given two functions \( f(x_1, \ldots, x_n) \) and \( g(x_1, \ldots, x_n) \) over \( \mathbb{F} \). We call \( f \) and \( g \) permutation equivalent if there exists a permutation \( \sigma \) such that \( f(x_1, \ldots, x_n) = g(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) \).

Equivalent functions share many properties. For example, two equivalent Boolean nested canalizing functions have the same sensitivity and the same average sensitivity [11, 13]. We are interested in the number of different equivalence classes of NCFs. To find this, we need the following combinatorial result.

**Lemma 4.11.** [2] (Page 70) Given \( n, r \) and \( s_i, i = 1, \ldots, r \) and \( s = s_1 + \cdots + s_r \leq n \). Then the number of integer solution of the equation \( k_1 + \cdots + k_r = n \), where \( k_i \geq s_i \), is

\[
\sum_{k_1+\cdots+k_r=n \atop k_i \geq s_i, s=s_1+\cdots+s_r \leq n} 1 = \binom{r+n-s-1}{r-1}.
\]

**Theorem 4.12.** For \( n \geq 2 \), the number of different equivalence classes of NCFs is

\[
N = 2^{n-1}(p-1)^{n+1}p^n.
\]

**Proof.** The number of different equivalent classes of NCFs equals the number of different NCFs with a fixed canalizing variable order \( \sigma \) in Equation 4.1. Thus, we can follow the same enumerative schedule as we did in the proof of Theorem 4.8 except that we do not consider the permutation of the variables. Hence, we get
\[
N = 2^{n-1}p(p-2)\sum_{r=2}^{n} (p-1)^{n+r-1} \sum_{k_1+\cdots+k_{r-1}=n-1 \atop k_i \geq 1, i=1,\ldots, r-1} 1 \\
+ 2^n p \sum_{r=1}^{n-1} (p-1)^{n+r} \sum_{k_1+\cdots+k_r=n \atop k_i \geq 1, i=1,\ldots, r, k_r \geq 2} 1 \\
= 2^{n-1}p(p-1)^{n+1}\left[(p-2)\sum_{r=2}^{n} (p-1)^{r-2}\binom{n-2}{r-2} + 2\sum_{r=1}^{n-1}(p-1)^{r-1}\binom{n-2}{r-1}\right] \\
= 2^{n-1}p(p-1)^{n+1}\left[(p-2)p^{n-2} + 2p^{n-2}\right] \\
= 2^{n-1}(p-1)^{n+1}p^2,
\]
where we used Lemma 4.11 to eradicate the inner sums in the first equality and the binomial theorem to simplify the sums in the second equality.

5. Generalization of NCFs

We can generalize the concept of nested canalizing functions to any finite field \(\mathbb{F}_q\), where \(q\) is a power of a prime. Moreover, the canalizing sets do not need to be restricted to intervals containing exactly one endpoint but can be any proper subset of \(\mathbb{F}_q\). Within this setting, we obtain similar results as in the previous section. The proofs are the same as in Section 4, we just replace the number of elements \(p\) in the finite field by \(q\), the number of nonzero elements, \(p-1\) by \(q-1\), and the number of intervals, \(2p-2\), by the number of proper subsets \(2^q - 2\), respectively.

**Definition 5.1.** Let \(f\) be a function in \(n\) variables over \(\mathbb{F}_q\), and let \(S_i\) be proper subsets of \(\mathbb{F}_q\), \(i=1,\ldots,n\). Then NCFs are defined in the same way as in Definition 2.3.

This definition generalizes the concept of NCFs as defined in [15]. However, NCFs defined in this way can still be uniquely represented as a polynomial as discussed above. The following theorem is a generalized version of Theorem 4.2 with the same arguments used to prove it.

**Theorem 5.2.** Given \(n \geq 2\), the function \(f : \mathbb{F}_q^n \to \mathbb{F}_q\) is nested canalizing if and only if it can be uniquely written as

\[
f(x_1, \ldots, x_n) = M_1(M_2(\cdots (M_{r-1}(B_{r+1}M_r + B_r) + B_{r-1})\cdots) + B_2) + B_1,
\]

where each \(M_i\) is a product of indicator functions of a set of disjoint variables. More precisely, we have that \(M_i = \prod_{j=1}^{k_i} (Q_{S_{i,j}}(x_{i,j})), i = 1,\ldots,r, k_i \geq 1\) for \(i = 1,\ldots,r, k_1 + \cdots + k_r = n, B_2,\ldots,B_{r+1} \neq 0, B_1 \in \mathbb{F}, \{i|j = 1,\ldots,k_i, i = 1,\ldots,r\} = \{1,\ldots,n\}, each S_i, i = 1,\ldots,n is any proper subset of \(\mathbb{F}_q\), and, if \(k_r = 1\), then \(B_{r+1} + B_r \neq 0\).

As in the previous section, we can count the number of NCFs and get
Theorem 5.3. For $n \geq 2$, the number of NCFs over $\mathbb{F}_q$, where NCFs are defined as in Definition 5.1, is given by
\[
|\text{NCF}(n)| = 2^{n-1}q(q-2)(2^{q-1} - 1)^n \sum_{r=2}^{n} (q-1)^{r-1} \sum_{\substack{k_1 + \ldots + k_{r-1} = n-1 \\ k_i \geq 1, i=1,\ldots,r-1}} \frac{n!}{k_1!k_2! \cdots k_{r-1}!}
+ 2^n q(2^{q-1} - 1)^n \sum_{r=1}^{n-1} (q-1)^{r} \sum_{\substack{k_1 + \ldots + k_r = n \\ k_i \geq 1, i=1,\ldots,r-1, k_r \geq 2}} \frac{n!}{k_1!k_2! \cdots k_r!}
\]

Theorem 5.4. The number of different equivalence classes of NCFs over $\mathbb{F}_q$, where NCFs are defined as in Definition 5.1, is
\[
N = 2^{n-1}(q - 1)q^n(2^{q-1} - 1)^n.
\]

6. Discussion

In this paper, we introduced explicit formulas for the calculation of Derrida values of networks governed by Boolean NCFs. Moreover, we generalized the concept of nested canalization to finite fields without any restriction on the canalizing input sets. By using the product of indicator functions, a generalization of extended monomials, we successfully generalized the characterization of Boolean NCFs [11] to the multistate case. Besides, we discussed the permutational equivalence of NCFs, and obtained a number of the different equivalence classes.

The main contributions of the paper are threefold. First, we introduced formulas for the calculation of the Derrida value of (stochastic) discrete dynamical systems governed by Boolean NCFs, which are the first of their kind, and abolish the need for simulations in the calculation of the Derrida plot. The presented formulas only work for the Boolean case. By extending the definition of Hamming distance and Derrida values in a straight-forward manner, we found similar, though more complicated formulas for the multistate case (unpublished work). We excluded these results from the paper, since the Derrida plot is in practice only used for Boolean systems. Second, we established a closed form formula for the number of multistate nested canalizing functions as defined in [15], improving on the recursive formula given there, and introduced a new closed form formula for the number of equivalence classes of NCFs under permutation equivalence. Third, we established a very general definition of multistate nested canalizing functions, for which many of the analytical results remain true.

7. References


