Efficient Probabilistic Inference in Bayesian Networks with Multi-Valued NIN-AND Tree Local Models [1]

Yang Xiang *and* Yiting Jin

School of Computer Science, University of Guelph, Canada

Abstract

A multi-valued Non-Impeding Noisy-AND (NIN-AND) tree model has linear complexity and is more expressive than several Causal Independence Models (CIMs) for expressing Conditional Probability Tables (CPTs) in Bayesian Networks (BNs). We show that it is also more general than the well-known noisy-MAX. To exploit NIN-AND tree models in inference, we develop a sound Multiplicative Factorization (MF) of multi-valued NIN-AND tree models. We show how to apply the MF to NIN-AND tree modeled BNs, and how to compile such BNs for exact lazy inference. For BNs with sparse structures, we demonstrate experimentally significant gain of inference efficiency in both space and time.

Keywords: Bayesian networks, causal independence models, noisy-MAX, multiplicative factorization, NIN-AND tree models, NAT models.

1. Introduction

A Bayesian Network (BN) [2] quantifies the causal strength between an effect and its n causes by a CPT, with the number of parameters being exponential in n. Common CIMs such as noisy-OR [2], noisy-AND [3], noisy-MAX [4], and recursive noisy-OR [5] reduce the number of parameters to being linear in n, but are limited in expressiveness: expressing causal reinforcement only. NIN-AND tree CIMs can express both causal reinforcement and undermining as well as their recursive mixtures, with a linear number of parameters. A NIN-AND Tree (NAT) model can be binary (over binary variables only) [6] or multi-valued [7]. It is shown in this work that multi-valued NAT models are more expressive than noisy-MAX.

Although CIMs reduce the number of parameters from being exponential to linear, they cannot be used directly by common BN inference algorithms, e.g., the cluster tree method [8]. A number of techniques have been proposed to overcome the difficulty, e.g., [9, 10, 11, 12, 13]. One technique is Multiplicative Factorization (MF) [10] and tensor rank-one decomposition [12] is closely related. MF has been applied to binary NAT models [13]. However, binary NAT models are not sufficiently general, limiting the applicability. MF for multi-valued NAT models is developed in this work, removing this limitation. BNs whose CPTs are expressible by multi-valued NAT models are then considered. Applying the MF to such BNs allows significantly more efficient inference both in

[1] This article significantly extends Xiang and Jin [1] as part of the FLAIRS 2016 Proceedings.

space and in time. We demonstrate up to two orders of magnitude efficiency gain for sparse BNs experimentally.

The main contributions of this paper are the following. We show that multi-valued NAT models are strictly more general and more expressive than the well-known noisy-MAX model. We then develop MFs of four alternative NIN-AND gate models. MF of a NAT model is integrated from MFs of these gate models. We establish the exactness of the MF through a formal analysis. A computational framework that utilizes the MF for exact probabilistic inference through lazy propagation is presented. We report our experimental result, where significant efficiency gain in inference is achieved under the framework.

Sec. 2 covers background on NAT models. Sec. 3 shows that NAT models are more general than noisy-MAX. The MFs of NIN-AND gates are developed in Sec. 4. The MF of NAT models is developed in Sec. 5 with the soundness and space complexity analyzed in Sec. 6. How to apply the MF to NAT-modeled BNs is shown in Sec. 7 with experimental evaluation described in Sec. 8.

2. Multi-Valued NAT Models

We overview multi-valued NAT models [7]. A multi-valued variable e is *graded*, if it has a finite ordered domain $D_e = \{e^0, e^1, ..., e^\eta\}$ $(\eta \geq 1)$, where a higher index represents a higher intensity. For example, fever e with $D_e = \{e^0, e^1, e^2\}$ represents conditions $\{normal, low\ fever, high\ fever\}$. We refer to value e^0 as *inactive* and $e^1, , ..., e^\eta$ as *active*. We consider a set of causes and their effect, all being graded variables. Hence, we use *multi-valued* and *graded* interchangeably. We denote the effect by e and each cause by c_i $(i = 1, 2, ...)$ of domain $D_i = \{c_i^0, c_i^1, ..., c_i^m\}$. We denote a set of causes by $X = \{c_1, c_2, ...\}$. The set of *all causes* of e is denoted by C, which may include a leaky variable.

We categorize causal events from three perspectives. A causal event can be a *success* or *failure*, depending on whether e is rendered active at certain intensity. It can be *single-causal* or *multi-causal*, depending on the number of active causes. It can also be *simple* or *congregate*, depending on the range of effect values involved. A *simple single-causal success* $e^k \leftarrow c_i^j$ occurs when $c_i = c_i^j$ $(j > 0)$ caused $e = e^k$ $(k > 0)$ to occur while every other cause is inactive. The *causal probability* $P(e^k \leftarrow c_i^j)$ is

$$P(e^k \leftarrow c_i^j) = P(e^k | c_i^j, c_m^0 : \forall m \neq i). \tag{1}$$

A *congregate single-causal success* $e \geq e^k \leftarrow c_i^j$ occurs when $c_i = c_i^j$ $(j > 0)$ caused e to occur at value e^k $(k > 0)$ or higher while every other cause is inactive. Its causal probability is

$$P(e \geq e^k \leftarrow c_i^j) = P(e \geq e^k | c_i^j, c_m^0 : \forall m \neq i). \tag{2}$$

A *multi-causal success* involves a set $X = \{c_1, ..., c_n\}$ $(n > 1)$ of active causes. A *simple multi-causal success* $e^k \leftarrow c_1^{j_1}, ..., c_n^{j_n}$ (or $e^k \leftarrow \underline{x}^+$) occurs when causes in X collectively caused $e = e^k$ $(k > 0)$ to occur while every other cause $c_m \in C \setminus X$ is inactive. Its causal probability is

$$P(e^k \leftarrow c_1^{j_1}, ..., c_n^{j_n}) = P(e^k | c_1^{j_1}, ..., c_n^{j_n}, c_m^0 : c_m \in C \setminus X). \tag{3}$$

A *congregate multi-causal success* $e \geq e^k \leftarrow \underline{x}^+$ occurs when causes in X collectively caused e to occur at e^k $(k > 0)$ or higher, while every cause in $C \setminus X$ is inactive. Its causal probability is

$$P(e \geq e^k \leftarrow c_1^{j_1}, ..., c_n^{j_n}) = P(e \geq e^k | c_1^{j_1}, ..., c_n^{j_n}, c_m^0 : c_m \in C \setminus X). \tag{4}$$

A *congregate single-causal failure* $e < e^k \leftarrow c_i^j$ occurs when $c_i = c_i^j$ ($j > 0$) caused $e < e^k$ ($k > 0$) to occur while every other cause is inactive. It is a failure as c_i fails to produce the effect with intensity e^k or higher. Its causal probability is

$$P(e < e^k \leftarrow c_i^j) = P(e < e^k | c_i^j, c_m^0 : \forall m \neq i). \tag{5}$$

A *congregate multi-causal failure* $e < e^k \leftarrow c_1^{j_1}, ..., c_n^{j_n}$ (or $e < e^k \leftarrow \underline{x}^+$) occurs when a set $X = \{c_1, ..., c_n\}$ ($n > 1$) of active causes caused $e < e^k$ ($k > 0$) to occur while every other cause $c_m \in C \setminus X$ is inactive. Its causal probability is

$$P(e < e^k \leftarrow c_1^{j_1}, ..., c_n^{j_n}) = P(e < e^k | c_1^{j_1}, ..., c_n^{j_n}, c_m^0 : c_m \in C \setminus X). \tag{6}$$

When all causes are inactive, casual probability of the *null causal event* is

$$P(e^k \leftarrow \perp) = P(e^k | c_i^0 : \forall i) = \begin{cases} 1 & (k = 0) \\ 0 & (k > 0) \end{cases} \tag{7}$$

Causal probabilities can be converted between one another, e.g., between congregate and simple as follows:

$$P(e \geq e^k \leftarrow \underline{x}^+) = \sum_{j=k}^{\eta} P(e^j \leftarrow \underline{x}^+). \tag{8}$$

A multi-valued NAT model is built upon multi-valued NIN-AND gates. A *direct* gate involves disjoint sets of causes $W_1, ..., W_m$. Its input events are $e \geq e^k \leftarrow \underline{w}_1^+$, ... , $e \geq e^k \leftarrow \underline{w}_m^+$ and its output event is $e \geq e^k \leftarrow \underline{w}_1^+, ..., \underline{w}_m^+$, all being causal successes. Fig. 1 (a) shows a direct gate

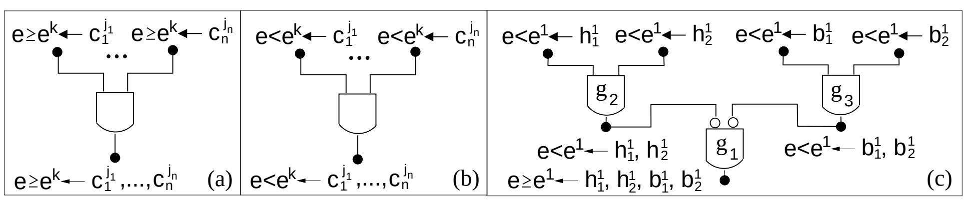

Figure 1: (a) A multi-valued direct NIN-AND gate. (b) A dual NIN-AND gate. (c) A NAT.

where each W_i is a singleton $\{c_i\}$. The causal probability of the output event satisfies

$$P(e \geq e^k \leftarrow \underline{w}_1^+, ..., \underline{w}_m^+) = \prod_{i=1}^{m} P(e \geq e^k \leftarrow \underline{w}_i^+), \tag{9}$$

where each factor can be obtained from single-causal probabilities and Eqn. (8).

Input events of a *dual* gate are $e < e^k \leftarrow \underline{w}_1^+$, ... , $e < e^k \leftarrow \underline{w}_m^+$ and its output event is $e < e^k \leftarrow \underline{w}_1^+, ..., \underline{w}_m^+$, all being causal failures, as Fig. 1 (b). The output event satisfies

$$P(e < e^k \leftarrow \underline{w}_1^+, ..., \underline{w}_m^+) = \prod_{i=1}^{m} P(e < e^k \leftarrow \underline{w}_i^+). \tag{10}$$

Causal interactions can be characterized as reinforcing or undermining as defined below.

Definition 1. *Let* $R = \{W_1, W_2, ...\}$ *be a partition of a set* X *of causes of* e*,* $R^- \subset R$*, and* $Y = \cup_{W_i \in R^-} W_i$*. Sets of causes in* R *reinforce each other relative to an active value* e^k *(*$k > 0$*) of effect, iff* $\forall R^-\ P(e \geq e^k \leftarrow \underline{y}^+) \leq P(e \geq e^k \leftarrow \underline{x}^+)$*. Sets of causes in* R *undermine each other relative to* e^k*, iff* $\forall R^-\ P(e \geq e^k \leftarrow \underline{y}^+) > P(e \geq e^k \leftarrow \underline{x}^+)$*.*

A direct gate models undermining causal interactions, and a dual gate models reinforcing [7]. A multi-valued NAT consists of multiple gates organized into a tree to express mixtures of reinforcing and undermining recursively. Causes involved in the root events of a NAT are disjoint. From the leaf up the NAT, gates at the same level are of the same type and gates at adjacent levels alternate in types. See the above reference for the general definition.

Example 1. *Fig. 1 (c) shows a NAT on surface enhancing, where* $\eta = 1$ *for all variables. Acidic surface enhancers* h_1 *and* h_2 *are more effective when used together. Basic surface enhancers* b_1 *and* b_2 *work together similarly. When both groups are applied to a product, their effectiveness* e *is reduced. In the NAT,* g_2 *and* g_3 *are dual gates, white ovals signify inverse of events, and* g_1 *is direct. From the NAT and single-causal probabilities (four of them),* $P(e^1|h_1^1, h_2^1, b_1^1, b_2^1)$ *can be obtained processing from the roots to the leaf (see [7] for detailed numerical examples).*

As shown in Example 1, each conditional probability is derived from a NAT and the relevant single-causals. For instance, $P(e^1|h_1^1, h_2^1, b_1^1, b_2^0)$ would be derived from a NAT distinct from Fig. 1 (c). Hence, a CPT $P(e|C)$ is derived from a set of NATs whose size is exponential on $|C|$. Since the entire set can be generated from a single NAT where all causes are active, a method has been developed to compute $P(e|C)$ from the NAT without explicitly generating the set [14]. With this understood, we associate a CPT below with a single NAT.

As mentioned in Example 1, each value of CPT is computed from single causal probabilities following the tree order of the NAT. On the other hand, common BN inference algorithms, e.g., the cluster tree method [8], are based on direct manipulation of conditional probabilities in CPTs (rather than single causal probabilities) following the tree order of clusters. This prevents the NAT model from being usable directly by these inference algorithms. In this work, we develop the MF of NAT models to overcome this difficulty.

3. On the Expressiveness of NAT Models

We show that a NAT model is strictly more general and expressive than the well-known noisy-MAX model (Sec. 3.1). We also clarify on some misconceptions that we perceived on NAT models (Sec. 3.2).

3.1. Equivalence of Noisy-MAX to Dual NIN-AND Gates

We show that NAT models generalize noisy-MAX [4, 15]. In particular, we show that noisy-MAX models are equivalent to multi-valued dual NIN-AND gate models, and hence are a special case of multi-valued NAT models.

Theorem 1. *Let* $X = \{c_1, ..., c_n\}$ *(*$n \geq 1$*) be a set of causes of effect* e *that interact according to noisy-MAX. Let* g *be a dual NIN-AND gate, where each input event involves exactly one* c_i *(*$i = 1, ..., n$*). Then the causal probability of the output event of* g *is identical to that of noisy-MAX.*

Proof: What qualifies as a noisy-MAX model is defined by Eqn. (36) of reference [15]. Specifically, when causes interact according to noisy-MAX, they satisfy the following in our notation, where $0 < k' \leq \eta$ and $j_i > 0$ for each i.

$$P(e \leq e^{k'} \leftarrow c_1^{j_1}, ..., c_n^{j_n}) = \prod_{i=1}^{n} P(e \leq e^{k'} \leftarrow c_i^{j_i}). \tag{11}$$

For $k' < \eta$, Eqn. (11) is identical to Eqn. (10) with $k = k' + 1$, $m = n$, and $\underline{w}_i^+ = c_i^{j_i}$ for each i. For $k' = \eta$, both noisy-MAX and NIN-AND gate result in probability 1 trivially. $\Box$

Example 2. *Consider a noisy-MAX model (Example 2 [16] with variables renamed) over* $e \in \{0, 1, 2\}$ *and* $c_1, c_2, c_3 \in \{0, 1\}$ *with the following parameters.*

$$\begin{array}{ll} P(e = 1|c_1 = 1, c_2 = 0, c_3 = 0) = 0.2, & P(e = 2|c_1 = 1, c_2 = 0, c_3 = 0) = 0.1, \\ P(e = 1|c_1 = 0, c_2 = 1, c_3 = 0) = 0.2, & P(e = 2|c_1 = 0, c_2 = 1, c_3 = 0) = 0.3, \\ P(e = 1|c_1 = 0, c_2 = 0, c_3 = 1) = 0.4, & P(e = 2|c_1 = 0, c_2 = 0, c_3 = 1) = 0.5. \end{array}$$

The CPT from the noisy-MAX is the following.

c_1	c_2	c_3	$P(e = 1\|c_1, c_2, c_3)$	$P(e = 2\|c_1, c_2, c_3)$	$P(e = 2\|c_1, c_2, c_3)$
0	0	0	1.0	0	0
0	0	1	0.1	0.4	0.5
0	1	0	0.5	0.2	0.3
0	1	1	0.05	0.3	0.65
1	0	0	0.7	0.2	0.1
1	0	1	0.07	0.38	0.55
1	1	0	0.35	0.28	0.37
1	1	1	0.035	0.28	0.685

Applying Eqn. (10) to the dual NIN-AND gate models of relevant subsets of $\{c_1, c_2, c_3\}$ *with the above single-causal probabilities produces exactly the same CPT.*

By Theorem 1, a noisy-MAX can always be expressed by a dual NIN-AND gate with exactly the same parameters. Since a dual gate models reinforcing, so does a noisy-MAX. On the other hand, a NAT can have direct gates and multiple dual gates, and can model both reinforcing and undermining as well as their recursive mixtures. Hence, NAT models are strictly more expressive than noisy-MAX models, as summarized in Corollary 1.

Corollary 1. *Let* C *be the set of all causes of an effect* e*. Then the following holds.*

1. *Whenever causes in* C *interact according to noisy-MAX, there exists a multi-valued NAT of the same parameters as noisy-MAX, such that the CPT* $P(e|C)$ *from the NAT model equals that of the noisy-MAX.*
2. *There exist multi-valued NAT models over* C *and* e*, whose CPT* $P(e|C)$ *cannot be encoded by any noisy-MAX model over* C *and* e*.*

Example 2 illustrates the first statement of Corollary 1, and the NAT in Example 1 is a demonstration of the second statement.

A NAT model of a CPT is uniquely defined by C, e, a NAT, and a set of single-causals. A distinct CPT (hence a distinct NAT model) is obtained if the NAT or the set of single-causals is modified. The number of distinct NATs given n is super-exponential in n [14] and the noisy-MAX is equivalent to exactly one of them. Single-causals are real numbers. Hence, given C and e, the number of NAT models that satisfy the second statement of Corollary 1 is infinite.

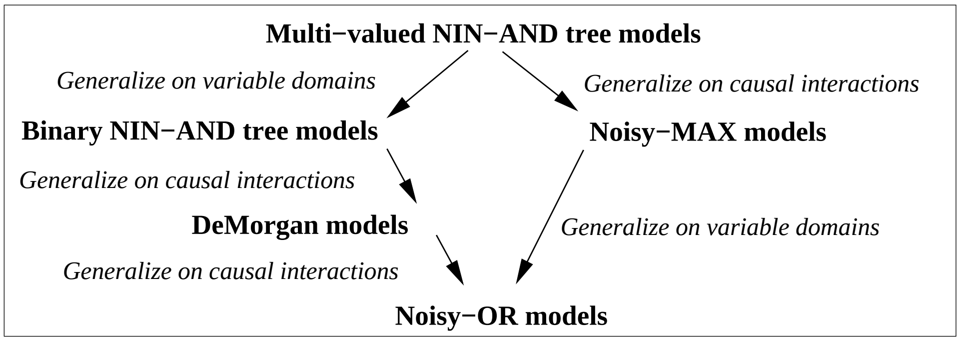

Figure 2: Relations among five classes of CIMs.

Fig. 2 provides a unified view on noisy-OR, noisy-MAX, binary NAT, multi-valued NAT models, as well as DeMorgan models [17]. It reveals that multi-valued NAT models are the generalization of noisy-OR models along two independent dimensions: domain sizes of variables and causal interactions. When generalization is only on causal interactions, it results in the binary NAT models. If generalization is only on variable domain sizes, the result is the noisy-MAX models. The DeMorgan models are intermediate between noisy-OR and binary NAT models, as has been shown [7].

3.2. Potential Misconceptions on NAT Models

An arbitrary combination of noisy-OR and noisy-AND gates is not equivalent to a NAT. This discussion refers to noisy-OR rather than noisy-MAX, since noisy-MAX degenerates to noisy-OR when all variables are binary. Limitation inherent to noisy-OR (other than being binary) cannot be overcome by noisy-MAX (for instance, the first point below).

First, a noisy-OR gate represents reinforcement, not undermining (see analysis in [6]). Second, a noisy-AND gate also represents reinforcement only, since a common noisy-AND gate is impeding (see [6]). Encoding cause variables through their complements in a noisy-AND gate does not overcome this limitation as we show below. Suppose $C = \{c_1, c_2\}$, all variables are binary, $P(e^1 \leftarrow c_1^1) = q > 0$, and $P(e^1 \leftarrow c_2^1) = r > 0$. A noisy-AND model produces the following CPT, where the last two equalities in the left equation are the impeding behavior.

$$P(e^1|c_1^0, c_2^0) = P(e^1|c_1^1, c_2^0) = P(e^1|c_1^0, c_2^1) = 0, \quad P(e^1|c_1^1, c_2^1) = q * r > 0.$$

As more active causes make the effect more likely, the interaction is reinforcing. Suppose we replace each variable value by its complement. The resultant noisy-AND has the CPT below.

$$P(e^0|c_1^1, c_2^1) = P(e^0|c_1^0, c_2^1) = P(e^0|c_1^1, c_2^0) = 0, \quad P(e^0|c_1^0, c_2^0) > 0.$$

It is equivalent to the following.

$$P(e^1|c_1^1, c_2^1) = P(e^1|c_1^0, c_2^1) = P(e^1|c_1^1, c_2^0) = 1, \;\; P(e^1|c_1^0, c_2^0) \in (0, 1).$$

The left equation makes each cause deterministic, and the right expression makes the effect active when all causes are inactive: an unintuitive semantics. If we replace only cause variables by complements, the resultant noisy-AND model produces the following CPT.

$$P(e^1|c_1^1, c_2^1) = P(e^1|c_1^0, c_2^1) = P(e^1|c_1^1, c_2^0) = 0, \;\; P(e^1|c_1^0, c_2^0) > 0.$$

The left equation makes the effect inactive when all causes are active, and the right inequality suffers from the same problem above: an unintuitive semantics again. Hence, the noisy-AND cannot model undermining.

Third, an arbitrary combination of several direct and dual NIN-AND gates does not constitute a more general representation than a NAT. When several NIN-AND gates are combined arbitrarily, the topology and causes involved in each root event are not subject to the syntactic restriction of NATs. The resultant structure does not ensure a meaningful semantics required by a causal model. On the other hand, the topology of a NAT and influences of its causes are regulated to ensure a coherent semantics.

4. MF of NIN-AND Gate Models

A NAT model reduces the space of a BN CPT from being exponential to being linear on n. However, as explained in the end of Sec. 2, NAT models cannot be directly used by common BN inference algorithms. To overcome this difficulty, we develop MFs of multi-valued NAT models. MFs for NIN-AND gates models are defined in this section, and MFs for NAT models are presented in the next section.

MFs contain auxiliary variables (in addition to effects and causes) and generalized potentials (with possibly negative values) over these variables. The potentials are so defined that their product, after marginalizing out auxiliary variables, is exactly the intended CPT. The product and marginalization operations, however, can be carried out in flexible orders. Hence, MFs trade cost of operating on potentials over auxiliary variables with flexibility of computation order. The key is to minimize the cost while ensuring exactness and order flexibility. Therefore, we develop MFs according to the criteria below.

1. MF of a NAT model is a graphical model whose graph is consistent with the NAT topology.
2. Auxiliary variables in the MF are as few as possible.
3. Domains of auxiliary variables are as small as possible.

In the following, we develop MFs for NIN-AND gates models that observe these criteria.

4.1. MF of Dual NIN-AND Gate Models

We organize MF of a dual NIN-AND gate model according to a hybrid graph.

Definition 2. *The* **MF structure** *of a NIN-AND gate model over effect $e \in \{e^0, e^1, ..., e^\eta\}$ and its causes c_i ($i = 1, ..., n$) is a hybrid graph G, whose nodes are labeled by e, c_i ($i = 1, ..., n$), and auxiliary variables d_j ($j = 1, ..., \eta$). Each link $\langle d_j, c_i \rangle$ between an auxiliary variable and a cause is a* **clink** *and is undirected. Each link (d_j, e) is directed.*

Each clink is assigned a **clink potential** *$f(d_j, c_i)$. Node e is assigned a* **family potential** *$f(d_1, ..., d_\eta, e)$ defined over its family determined by incoming directed links.*

Fig. 3 illustrates the MF structure of a dual NIN-AND gate model.

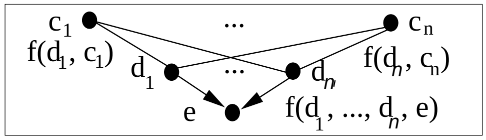

Figure 3: Hybrid graphical model for MF of a NIN-AND gate.

For example, when $\eta = 2$, $n = 2$, and $|D_i| = 3$ for $i = 1, 2$, the MF has 4 clink potentials $f(d_1, c_1)$, $f(d_2, c_1)$, $f(d_1, c_2)$, $f(d_2, c_2)$ and the family potential $f(d_1, d_2, e)$. Table 1 shows $f(d_1, c_1)$, $f(d_2, c_1)$ and $f(d_1, d_2, e)$. Each potential table starts with a double line, and the table $f(d_1, d_2, e)$ splits into two parts. Potential values are shown in f columns. Each table is *sectioned* by grouping rows according to the value(s) for one or more variables. If potential values are independent of a variable, the corresponding section is compressed, e.g., the last section of $f(d_1, c_1)$. Table 2 shows MF potentials in general. We refer to the collection of graph G and the potentials

Table 1: MF potentials $f(d_1, c_1)$, $f(d_2, c_1)$ and $f(d_1, d_2, e)$ of an example dual gate model

d_1	c_1	f
0	c_1^0	1
0	c_1^1	$P(e < e^1 \leftarrow c_1^1)$
0	c_1^2	$P(e < e^1 \leftarrow c_1^2)$
1	c_1	1

d_2	c_1	f
0	c_1^0	1
0	c_1^1	$P(e < e^2 \leftarrow c_1^1)$
0	c_1^2	$P(e < e^2 \leftarrow c_1^2)$
1	c_1	1

d_1	d_2	e	f
0	0	e	0
0	1	e^0	1
0	1	e^1	-1
0	1	e^2	0

d_1	d_2	e	f
1	0	e^0	0
1	0	e^1	1
1	0	e^2	-1
1	1	e^0	0
1	1	e^1	0
1	1	e^2	1

Table 2: The clink and family potentials $f(d_j, c_i)$ and $f(d_1, ..., d_\eta, e)$ of MDu

d_j	c_i	f
0	c_i^0	$P(e < e^j \leftarrow \perp) = 1$
0	c_i^1	$P(e < e^j \leftarrow c_i^1)$
	...	...
0	c_i^m	$P(e < e^j \leftarrow c_i^m)$
1	c_i	1

line	$(d_1, ..., d_\eta, e)$	f
1	$d_i = 0,\ \forall_{j \neq i}\ d_j = 1,\ e = e^{i-1}$	1
2	$d_i = 0,\ \forall_{j \neq i}\ d_j = 1,\ e = e^i$	-1
3	$\forall_i\ d_i = 1,\ e = e^\eta$	1
4	otherwise	0

as the MF of a Dual gate model (MDu). For analysis, we decompose the MF into the *link layer*, consisting of clinks (including end nodes) and their potentials, and the *family layer*, consisting of the directed links and the family potential. Auxiliary variables are included in both layers.

In the remainder of the paper, we frequently obtain the product of several potentials over a set of variables and then marginalize out some variables (may be none), such as the following.

$$f(e, c_1, ..., c_n) = \sum_{d_1, ..., d_\eta} f(d_1, ..., d_\eta, e) \prod_{1 \leq j \leq \eta, 1 \leq i \leq n} f(d_j, c_i).$$

We refer to the result as a *marginalized potential product* (MPP). When the relevant set of potentials is clear from context, we mention the MPP, e.g., $f(e, c_1, ..., c_n)$, without elaborating the factor potentials.

By Theorem 1, MDu is equivalent to MF of noisy-MAX [10, 11], from which Corollary 2 on the exactness of MDu follows.

Corollary 2. *Let MDu be applied to a dual NIN-AND gate model whose CPT is $P(e|c_1, ..., c_n)$. The MPP $f(e, c_1, ..., c_n)$ from potentials of the MDu satisfies $f(e, c_1, ..., c_n) = P(e|c_1, ..., c_n)$.*

MDu has the same numerical parameters as the MF of noisy-MAX, but differs from the previous work [10, 11] as follows. The MF in [10] is not defined as a graphical model. The MF in [11] is defined as a DAG model where potential assignment does not follow the family convention. We define MDu as a hybrid graphical model with a rigorous syntax, where a potential is assigned to each clink when the link is undirected or to the family when links are directed.

Def. 3 specifies a property of the link layer, where the MPP is obtained from the layer potentials. It is used later to show exactness of MF of a NAT model. It is phrased to allow both single ($k = 1$) and multiple causes. Its statement 1 abuses the notation $P(e < e^j \leftarrow c_1, ..., c_k)$ slightly for simplicity, as some or all causes listed may be inactive, e.g., the first row in Table 2 (left).

Definition 3. *[MDu link potential trait] Let $\{c_1, ..., c_k\}$ ($k \geq 1$) be a set of causes of MDu and $\{d_1, ..., d_\eta\}$ be a set of auxiliary variables. An MPP $f(d_1, ..., d_\eta, c_1, ..., c_k)$ from the link layer of the MDu satisfies the MDu link potential trait, if the following holds.*

1. *If $\exists_j\ d_j = 0$ and $\forall_{i \neq j}\ d_i = 1$, then $f(d_1, ..., d_\eta, c_1, ..., c_k) = P(e < e^j \leftarrow c_1, ..., c_k)$.*
2. *If $\forall_i\ d_i = 1$, then $f(d_1, ..., d_\eta, c_1, ..., c_k) = 1$.*

We refer to each statement in Def. 3 as a *subtrait*. We refer to the collection of statement conditions (the premise in each subtrait) as preconditions of the trait. Prop. 1 shows that the above property holds for MDu relative to a single cause.

Proposition 1. *Let c be a cause in a dual NIN-AND gate model. The product of clink potentials over c from the MDu of the model, $f(d_1, ..., d_\eta, c) = \prod_{1 \leq i \leq \eta} f(d_i, c)$, satisfies the MDu link potential trait (Def. 3) with $k = 1$.*

Proof: From Table 2 (left), the first subtrait holds, since the factor $f(d_j, c)$ is from the section where $d_j = 0$, and all other factors are from the section where $d_i = 1$ with value 1.

The second subtrait holds, since all factors are from the section where $f(d_i, c) = 1$. $\square$

Prop. 1, in fact, covers only products that satisfy preconditions in Def. 3. Prop. 2 says that these products are all that matter. The other products over c make no contribution to $f(e, c_1, ..., c_n)$ for MDu.

Proposition 2. *Let c be a cause in a dual NIN-AND gate model, and a product of clink potentials over c from the MDu of the model be $f(d_1, ..., d_\eta, c) = \prod_{1 \leq i \leq \eta} f(d_i, c)$. Then $f(d_1, ..., d_\eta, c)$ contributes to $f(e, c_1, ..., c_n)$, only if preconditions of the MDu link potential trait (Def. 3) hold.*

Proof: We have $f(e, c_1, ..., c_n) = \sum_{d_1, ..., d_\eta} f(d_1, ..., d_\eta, e) \prod_{1 \leq i \leq n} f(d_1, ..., d_\eta, c_i)$. We show that when the two preconditions do not hold for some $c = c_i$, the factor $f(d_1, ..., d_\eta, e)$ is zero, blocking the product $f(d_1, ..., d_\eta, c)$. This can be seen from Table 2 (right). If the first precondition does not hold, lines 1 and 2 are ruled out. If the second precondition does not hold, line 3 is ruled out. Hence, $f(d_1, ..., d_\eta, e)$ is zero by line 4. $\square$

4.2. MF of Direct NIN-AND Gate Models

We develop MF of a direct gate model with the same hybrid structure G in Fig. 3, but each $d_j \in \{0, 1, 2\}$. Table 3 shows 2 clink potentials and the family potential when $\eta = 2$, $n = 2$,

Table 3: MF potentials $f(d_1, c_1)$, $f(d_2, c_1)$ and $f(d_1, d_2, e)$ of an example direct gate model

d_1	c_1	f
0	c_1^0	1
0	c_1^1	$P(e \geq e^1 \leftarrow c_1^1)$
0	c_1^2	$P(e \geq e^1 \leftarrow c_1^2)$
1	c_1	1
2	c_1^0	1
2	c_1^1	0
2	c_1^2	0

d_2	c_1	f
0	c_1^0	1
0	c_1^1	$P(e \geq e^2 \leftarrow c_1^1)$
0	c_1^2	$P(e \geq e^2 \leftarrow c_1^2)$
1	c_1	1
2	c_1^0	1
2	c_1^1	0
2	c_1^2	0

d_1	d_2	e	f
0	0	e	0
0	1	e^0	-1
0	1	e^1	1
0	1	e^2	0
0	2	e	0
1	0	e^0	0
1	0	e^1	-1
1	0	e^2	1

d_1	d_2	e	f
1	1	e^0	1
1	1	e^1	0
1	1	e^2	0
1	2	e	0
2	0	e^0	1
2	0	e^1	0
2	0	e^2	-1
2	1	e	0
2	2	e	0

Table 4: The clink and family potentials $f(d_j, c_i)$ and $f(d_1, ..., d_\eta, e)$ of MDi

d_j	c_i	f
0	c_i^0	1
0	c_i^1	$P(e \geq e^j \leftarrow c_i^1)$
	...	...
0	c_i^m	$P(e \geq e^j \leftarrow c_i^m)$
1	c_i	1

d_j	c_i	f
2	c_i^0	1
2	c_i^1	0
	...	...
2	c_i^m	0

line	$(d_1, ..., d_\eta, e)$	f
1	$d_i = 0,\ \forall_{j \neq i}\ d_j = 1,\ e = e^{i-1}$	-1
2	$d_i = 0,\ \forall_{j \neq i}\ d_j = 1,\ e = e^i$	1
3	$\forall_i\ d_i = 1,\ e = e^0$	1
4	$d_1 = 2,\ \forall_{i>1}\ d_i = 0,\ e = e^0$	1
5	$d_1 = 2,\ \forall_{i>1}\ d_i = 0,\ e = e^\eta$	-1
6	otherwise	0

and $|D_i| = 3$ for $i = 1, 2$, with $f(d_1, c_2)$ and $f(d_2, c_2)$ left out. Note that $f(d_i = 0, c_j^0) = 1$ is not equal to $P(e \geq e^i \leftarrow c_j^0) = 0$. Hence, unlike MDu, $f(d_i = 0, c_i)$ is not made entirely of probabilities. We refer to the clink potential as *pseudo-probabilistic*. This is necessary as a value 0 would block other potential values in a product. To ensure exactness of MF (Theorem 2 below), each d_j increases the domain size by 1 relative to MDu. Table 4 shows MF potentials in general. We refer to the collection of graph G and the potentials as the MF of a Direct gate model (MDi). Theorem 2 shows its exactness. The space complexity is discussed in Sec. 6.

Theorem 2. *Let MDi be applied to a direct NIN-AND gate model whose CPT is* $P(e|c_1, ..., c_n)$. *The MPP* $f(e, c_1, ..., c_n)$ *from potentials of the MDi satisfies* $f(e, c_1, ..., c_n) = P(e|c_1, ..., c_n)$.

Proof: The MPP $f(e, c_1, ..., c_n) = \sum_{d_1,...,d_\eta} f(d_1, ..., d_\eta, e) \prod_{1 \leq j \leq \eta, 1 \leq i \leq n} f(d_j, c_i)$ is defined from potentials in Table 4. Consider the product $f(d_j, c_1, ..., c_n) = \prod_{1 \leq i \leq n} f(d_j, c_i)$. If any c_i is active, by Table 4 (section $d_j = 0$) and Eqn. (9), we have $f(d_j = 0, c_1, ..., c_n) = P(e \geq e^j \leftarrow \underline{x}^+)$, where $\underline{x}^+$ denotes all active c_i. Otherwise (each c_i is inactive), $f(d_j = 0, c_1, ..., c_n) = 1$ by row 1 of Table 4. For $d_j = 1$, we have $f(d_j = 1, c_1, ..., c_n) = 1$ from the last row of Table 4. For $d_j = 2$,

Table 5: Summary on $f(d_j, c_1, ..., c_n)$

	$f(d_j, c_1, ..., c_n)$	
d_j	$\exists c_i \; c_i > c_i^0$	$\forall c_i \; c_i = c_i^0$
0	(1) $P(e \geq e^j \leftarrow \underline{x}^+)$	(4) 1
1	(2) 1	(5) 1
2	(3) 0	(6) 1

we have $f(d_j = 2, c_1, ..., c_n) = 0$ if any c_i is active and $f(d_j = 2, c_1, ..., c_n) = 1$ otherwise, from Table 4. This is summarized in Table 5.

The MPP becomes $f(e, c_1, ..., c_n) = \sum_{d_1,...,d_\eta} f(d_1, ..., d_\eta, e) f(d_1, c_1, ..., c_n)...f(d_\eta, c_1, ..., c_n)$, where $f(d_1, ..., d_\eta, e)$ selects products $f(d_1, c_1, ..., c_n)...f(d_\eta, c_1, ..., c_n)$ for summation.

We consider two mutually exclusive and exhaustive cases: (1) some c_i is active, and (2) each c_i is inactive. Under case (1), we have the following intended causal probabilities, where $0 < k < \eta$ and justifications refer to Table 4 (right). Their derivations are explained after the equations.

$$f(e^0, c_1, ..., c_n) = 1 - P(e \geq e^1 \leftarrow \underline{x}^+) = P(e^0 \leftarrow \underline{x}^+) \quad \text{(lines 3 and 1)} \tag{12}$$

$$\begin{aligned} f(e^k, c_1, ..., c_n) &= P(e \geq e^k \leftarrow \underline{x}^+) - P(e \geq e^{k+1} \leftarrow \underline{x}^+) \\ &= P(e^k \leftarrow \underline{x}^+) \quad \text{(lines 2 and 1)} \end{aligned} \tag{13}$$

$$f(e^\eta, c_1, ..., c_n) = P(e \geq e^\eta \leftarrow \underline{x}^+) = P(e^\eta \leftarrow \underline{x}^+) \quad \text{(line 2)} \tag{14}$$

For Eqn. (12), all products $f(d_1, c_1, ..., c_n)...f(d_\eta, c_1, ..., c_n)$ are zeros if any $d_j = 2$ by Table 5 (3). Line 3 of Table 4 (right) yields 1. The product with $d_1 = 0$ and all other $d_j = 1$ yields $-P(e \geq e^1 \leftarrow \underline{x}^+)$ by line 1 of Table 4 (right) and Table 5 (1). All remaining products are zeros by line 6 of Table 4 (right). Hence, Eqn. (12) holds. For Eqn. (13), line 2 of Table 4 (right) and Table 5 (1) yield $P(e \geq e^k \leftarrow \underline{x}^+)$. Line 1 of Table 4 (right) and Table 5 (1) yield $-P(e \geq e^{k+1} \leftarrow \underline{x}^+)$. For Eqn. (14), line 2 of Table 4 (right) and Table 5 (1) yield $P(e \geq e^\eta \leftarrow \underline{x}^+)$.

Under case (2), all products $f(d_1, c_1, ..., c_n)...f(d_\eta, c_1, ..., c_n)$ equal 1 by Table 5 (4), (5) and (6). Hence, $f(c_1, ..., c_n, e) = \sum_{d_1,...,d_\eta} f(d_1, ..., d_\eta, e)$ is completely determined by Table 4 (right). We have the following intended causal probabilities.

$$\begin{aligned} f(e^0, c_1, ..., c_n) &= -1 + 1 + 1 = 1 = P(e^0 \leftarrow \perp) && \text{(lines 1, 3 and 4)} \\ f(e^k, c_1, ..., c_n) &= 1 - 1 = 0 = P(e^k \leftarrow \perp) && \text{(lines 2 and 1)} \\ f(e^\eta, c_1, ..., c_n) &= 1 - 1 = 0 = P(e^\eta \leftarrow \perp) && \text{(lines 2 and 5)} \end{aligned}$$

It then follows that $f(e, c_1, ..., c_n) = P(e|c_1, ..., c_n)$. □

Def. 4 specifies a property of the link layer of MDi, where the MPP is obtained from layer potentials. It is used later to show exactness of MF of a NAT model. It is phrased generally to allow both single and multiple causes.

Definition 4. *[MDi link potential trait] Let* $\{c_1, ..., c_k\}$ *(*$k \geq 1$*) be a set of causes of MDi and* $\{d_1, ..., d_\eta\}$ *be a set of auxiliary variables. An MPP* $f(d_1, ..., d_\eta, c_1, ..., c_k)$ *from the link layer of the MDi satisfies the MDi link potential trait, if the following holds.*

1. *If* $\exists_j \; d_j = 0, \forall_{i \neq j} \; d_i = 1$*, and some* c_i *is active, then*

$$f(d_1, ..., d_\eta, c_1, ..., c_k) = P(e \geq e^j \leftarrow c_1, ..., c_k).$$

2. *If* $\exists_j\ d_j = 0$*,* $\forall_{i \neq j}\ d_i = 1$*, and each* c_i *is inactive, then* $f(d_1, ..., d_\eta, c_1, ..., c_k) = 1$*.*
3. *If* $\forall_i\ d_i = 1$*, then* $f(d_1, ..., d_\eta, c_1, ..., c_k) = 1$*.*
4. *If* $d_1 = 2$*,* $\forall_{i>1}\ d_i = 0$*, and each* c_i *is inactive, then* $f(d_1, ..., d_\eta, c_1, ..., c_k) = 1$*.*

Prop. 3 below shows that the above property holds for MDi relative to a single cause.

Proposition 3. *Let* c *be a cause in a direct NIN-AND gate model. The product of clink potentials over* c *from the MDi of the model,* $f(d_1, ..., d_\eta, c) = \prod_{1 \leq i \leq \eta} f(d_i, c)$*, satisfies the MDi link potential trait (Def. 4) with* $k = 1$*.*

Proof: All subtraits follow from Table 4 (left). We refer to the 3 sections where $d_j = 0, 1, 2$ as section 0, 1 and 2, respectively. The first subtrait is derived from rows of section 0 other than the first row and section 1. The second subtrait is from the first row of section 0 and section 1.

The third subtrait follows from section 1. The fourth subtrait is derived from the first rows of sections 0 and 2. □

Prop. 3 covers effectively only products that satisfy preconditions of Def. 4. Prop. 4 says that these products are all that matter. The other products over c make no contribution to the MPP $f(e, c_1, ..., c_n)$ of MDi.

Proposition 4. *Let* c *be a cause in a direct NIN-AND gate model. The product of clink potentials over* c *from the MDi of the model,* $f(d_1, ..., d_\eta, c) = \prod_{1 \leq i \leq \eta} f(d_i, c)$*, contributes to the MPP* $f(e, c_1, ..., c_n)$ *of the MDi, only if preconditions of the MDi link potential trait (Def. 4) hold.*

Proof: We show that when the preconditions do not hold, either the factor $f(d_1, ..., d_\eta, e)$ in the MPP is zero, blocking product $f(d_1, ..., d_\eta, c)$, or $f(d_1, ..., d_\eta, c)$ is zero. This is seen from Table 4 (right). The first two preconditions combine into the condition $\exists_j\ d_j = 0 \wedge \forall_{i \neq j}\ d_i = 1$. If it does not hold, lines 1 and 2 are ruled out. If the third precondition does not hold, line 3 is ruled out.

The fourth precondition has 3 subconditions. If the first two do not hold, lines 4 and 5 are ruled out. If the first two subconditions hold but the third does not, the product $f(d_1, ..., d_\eta, c)$ is zero by Table 4 (left) from section 2 (rows other than the first).

In all other cases, $f(d_1, ..., d_\eta, e)$ is zero by line 6 of Table 4 (right). □

5. MF of NAT Models

We develop the MF of NAT models by integrate NIN-AND gate models. We explain why the stand-alone gate models presented above must be extended when they are embedded in a NAT model. We then extend them into 4 distinct gate MFs as components of a NAT MF.

5.1. Overview

A nontrivial NAT has at least two NIN-AND gates, e.g., Fig. 4 (a) and (c), where event labels are simplified and ovals into gates are omitted. The MF of a NAT model consists of a hybrid graph G and a collection of potentials defined over each undirected link and each family in G. The graph G is integrated from graphs of gate MFs according to the topology of the NAT, as shown in (b) and (d). For instance, gate g_2 in (a) induces the subgraph spanning $\{c_1, c_2, a_1, a_2, b\}$ in (b). The child variable from the MF of the leaf gate is still referred to as the *effect* variable and labeled by e, e.g., the leaf gate g_1 in (a) and (c). Child variables from MFs of other gates are referred to as *internal* variables and labeled differently. For instance, the child variable of MF for g_2 in (a) is labeled as b in (b). The graph G of the MF for a multi-valued NAT differs significantly from the graph of MF for a binary NAT [13], in that the latter is a tree while the former is multiply connected.

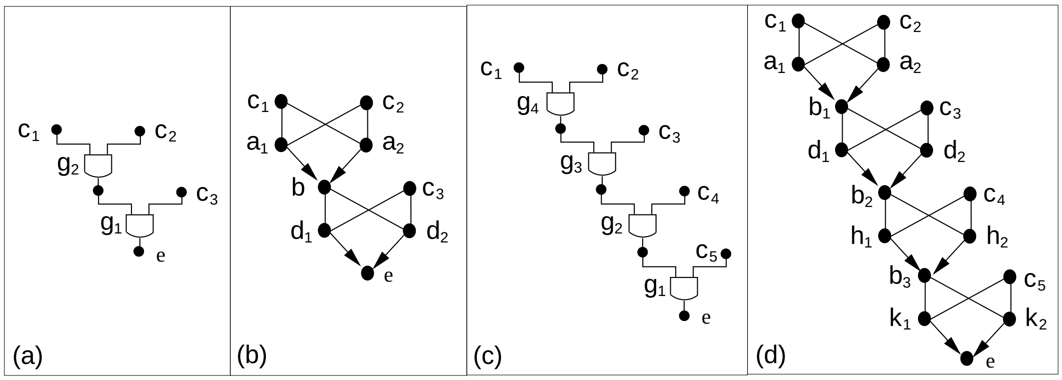

Figure 4: (a) A 2-gate NAT. (b) MF graph of (a). (c) A 4-gate NAT. (d) MF graph of (c).

Since G is integrated according to the topology of the NAT, the input-output direction of the NAT is maintained by the directed links in G. Hence, for each child node in G, its *ancestor subgraph* (containing the node and its ancestors) can be uniquely defined. For instance, the ancestor subgraph of b_2 in Fig. 4 (d) is the subgraph spanned by $\{c_1, c_2, a_1, a_2, b_1, c_3, d_1, d_2, b_2\}$. We also refer to it as the *ancestor subgraph with leaf* b_2. We refer to the collection of an ancestor subgraph with leaf x and potentials assigned to its links and families as the *subMF with leaf* x.

In graph G, the subgraph induced by a gate is identical to the graph of the MF for a standalone gate. Due to several factors, however, variable domains, link potentials, and the family potential associated with the subgraph may differ from those associated with the MF of a standalone gate. First, an undirected link, e.g., $\langle d_1, b\rangle$ in Fig. 4 (b), may connect an internal variable b. It is thus called an *ilink* and its link potential must be defined differently from that of a clink. Second, gates in a NAT are located at different levels. The leaf gate is at level 1, e.g., g_1 in (a). A gate feeding the leaf gate is at level 2, e.g., g_2 in (a). The MF of a gate must be adjusted according to its level. For instance, the MF of a dual gate at level 2 is more sophisticated than that of a leaf dual gate. This is because the former feeds into a gate at the next level while the latter is terminal.

Third, all gates at the same level have the same type (dual or direct) and gates at adjacent levels differ in types [18]. Hence, a dual gate at level 2 receives input from direct gates at level 3, and feeds into the direct leaf gate at level 1. The MF of a gate must be adjusted according to the gate from which it receives input and the gate that it feeds into.

Fourth, a gate in a NAT may receive input from both clinks and ilinks. For instance, the MF of g_1 in Fig. 4 (b) receives input from clinks $\langle c_3, d_k\rangle$ ($k = 1, 2$) as well as from ilinks $\langle b, d_k\rangle$. For the family potential over $\{d_1, d_2, e\}$ to work with both types of input uniformly, the product of potentials over $\langle c_i, a_j\rangle$, $\{a_1, a_2, b\}$, and $\langle b, d_k\rangle$, after marginalizing out $\{a_1, a_2, b\}$, must be equivalent (syntactically and semantically) to the product of potentials over $\langle c_3, d_k\rangle$. For example, assume that the leaf gate g_1 in Fig. 4 (a) is direct and g_2 is dual. Recall from Sec. 4.2 that a clink potential of a direct gate is pseudo-probabilistic. To render it probabilistic (as in Theorem 2), the domain size of auxiliary variables is increased from 2 in MDu to 3 in MDi. Here, output from MF of the dual gate g_2 is probabilistic (Sec. 4.1). To be equivalent to clink potentials over $\langle c_3, d_k\rangle$, the output needs to be rendered pseudo-probabilistic. As a result, the output variable b of g_2 needs to have a domain size of e plus one, which will affect both the family potential for the MF of gate g_2 and the ilink potentials for the MF of gate g_1.

Due to the above factors, we develop 4 distinct gate MFs for a NAT model: MDu enhanced with ilink potentials, MDi enhanced with ilink potentials, Extended MDu (EDu), and Extended MDi (EDi). A particular gate in a NAT is assigned one of the MFs depending on its gate type (dual or direct) and its level in the NAT. The rule of assignment is shown in Table 6. For example, if g_1

Table 6: Rule of MF assignment for NIN-AND gates

	gate MF	
gate level	dual	direct
3 or higher	EDu	EDi
2	EDu	MDi
1	MDu	MDi

of Fig. 4 (c) is direct, the MF assignment for the gates is (g_1: MDi; g_2: EDu; g_3: EDi; g_4: EDu). If g_1 is dual, the MF assignment is (g_1: MDu; g_2: MDi; g_3: EDu; g_4: EDi).

Each gate MF consists of 4 types of variables. We denote an input cause variable by c, an input internal variable by b, an auxiliary variable by d, and the output (child) variable by h if not e. Their domain sizes depend on the gate MF. In MDu for dual gates, b has the domain size of e and d is binary. In EDu, d is ternary, and b and h have the domain size of e plus one. In both MFs for direct gates, b has the domain size of e plus one due to input from EDu, and d is ternary. In MDi, h has the domain size of e. In EDi, h has the domain size of e plus one. This is summarized in Table 7.

Table 7: Domain sizes for variables b, d and h in a gate MF

	domain size			
var	MDu	EDu	MDi	EDi
b	$\eta+1$	$\eta+2$	$\eta+2$	$\eta+2$
d	2	3	3	3
h		$\eta+2$	$\eta+1$	$\eta+2$

In MDu, MDi and EDi, d has the same domain size as in the MF of corresponding standalone gate. For each clink $\langle c, d\rangle$, assign the clink potential accordingly. In EDu, extend the clink potential of MDu with $f(d=2,c) = (1,0,...,0)$. Below, we describe each MF focusing on ilink and family potentials. We illustrate the potentials for $|D_e| = \eta+1 = 3$ and then present them in general.

5.2. MDu

The ilink potentials $f(d_i, b)$ for $|D_e| = 3$ are shown in Table 8.

If $|D_e|$ increases by 1, each section (with the same d_i value) of $f(d_i, b)$ has another row, and another link potential $f(d_3, b)$ is needed, whose first section is $(1,1,1,0)$. The general ilink potential of MDu is shown in Table 9. The general clink and family potentials of MDu are shown in Table 2.

Table 8: The ilink potentials $f(d_1, b)$ and $f(d_2, b)$ of an example MDu, and their product

d_1	b	f
0	b^0	1
0	b^1	0
0	b^2	0
1	b	1

d_2	b	f
0	b^0	1
0	b^1	1
0	b^2	0
1	b	1

d_1	d_2	$f(d_1, d_2, b^0)$	$f(d_1, d_2, b^1)$	$f(d_1, d_2, b^2)$
0	0	1	0	0
0	1	1	0	0
1	0	1	1	0
1	1	1	1	1

Table 9: The ilink potential $f(d_j, b)$ ($j = 1, ..., \eta$) of MDu

line	(d_j, b)	f
1	$d_j = 0,\ b = b^0, ..., b^{j-1}$	1
2	$d_j = 0,\ b = b^j, ..., b^\eta$	0
3	$d_j = 1$	1

5.3. EDu

In EDu, each auxiliary variable d_j has the domain size 3. Table 10 shows ilink potentials $f(d_i, b)$ and family potential $f(d_1, ..., d_\eta, h)$ for $|D_e| = 3$. The $f(d_i, b)$ extends that of MDu (Table 8) by adding one row on each section and a new section. The row added to the first section is 1, the row added to the second section is 0, and the new section is $(0, ..., 0, 1)$.

Table 10: The ilink and family potentials $f(d_1, b)$, $f(d_2, b)$, and $f(d_1, d_2, h)$ of an example EDu

d_1	b	f
0	b^0	1
0	b^1	0
0	b^2	0
0	b^3	1
1	b^0	1
1	b^1	1
1	b^2	1
1	b^3	0
2	b^0	0
2	b^1	0
2	b^2	0
2	b^3	1

d_2	b	f
0	b^0	1
0	b^1	1
0	b^2	0
0	b^3	1
1	b^0	1
1	b^1	1
1	b^2	1
1	b^3	0
2	b^0	0
2	b^1	0
2	b^2	0
2	b^3	1

d_1	d_2	h	f
0	0	h	0
0	1	h^0	1
0	1	h^1	-1
0	1	h^2	0
0	1	h^3	0
0	2	h	0
1	0	h^0	0
1	0	h^1	1
1	0	h^2	-1
1	0	h^3	0

d_1	d_2	h	f
1	1	h^0	0
1	1	h^1	0
1	1	h^2	1
1	1	h^3	0
1	2	h	0
2	0	h^0	-1
2	0	h^1	0
2	0	h^2	1
2	0	h^3	1
2	1	h	0
2	2	h	0

The $f(d_1, ..., d_\eta, h)$ extends that of MDu (Table 2) by adding one row of 0 on each section and new sections where $d_i = 2$. The new sections are all zeros, except the section for $(d_1 = 2, d_2 = 0, ..., d_\eta = 0)$ is $(-1, 0, ..., 1, 1)$.

Each clink potential extends that of MDu (Table 2, left) with the additional section $f(d_j = 2, c_i) = (1, 0, ..., 0)$. The general clink, ilink and family potentials of EDu are shown in Table 11.

Table 11: The clink, ilink and family potentials $f(d_j, c_i)$, $f(d_j, b)$ and $f(d_1, ..., d_\eta, h)$ of EDu

d_j	c_i	f
0	c_i	$P(e < e^j \leftarrow c_i)$
1	c_i	1
2	c_i^0	1
2	$c_i > c_i^0$	0

line	(d_j, b)	f
1	$d_j = 0,\ b = b^0, ..., b^{j-1}, b^{\eta+1}$	1
2	$d_j = 1,\ b = b^0, ..., b^\eta$	1
3	$d_j = 2,\ b = b^{\eta+1}$	1
4	otherwise	0

line	$(d_1, ..., d_\eta, h)$	f
1	$d_i = 0,\ \forall_{j \neq i}\ d_j = 1,\ h = h^{i-1}$	1
2	$d_i = 0,\ \forall_{j \neq i}\ d_j = 1,\ h = h^i$	-1
3	$\forall_i\ d_i = 1,\ h = h^\eta$	1
4	$d_1 = 2,\ \forall_{i>1}\ d_i = 0,\ h = h^0$	-1
5	$d_1 = 2,\ \forall_{i>1}\ d_i = 0,\ h = h^\eta, h^{\eta+1}$	1
6	otherwise	0

5.4. *MDi*

The ilink potentials $f(d_i, b)$ for $|D_e| = 3$ are shown in Table 12. If $|D_e|$ increases by 1, each

Table 12: The ilink potentials $f(d_1, b)$ and $f(d_2, b)$ of an example MDi or EDi

d_1	b	f
0	b^0	0
0	b^1	1
0	b^2	1
0	b^3	1
1	b^0	1
1	b^1	1
1	b^2	1
1	b^3	0
2	b^0	0
2	b^1	0
2	b^2	0
2	b^3	1

d_2	b	f
0	b^0	0
0	b^1	0
0	b^2	1
0	b^3	1
1	b^0	1
1	b^1	1
1	b^2	1
1	b^3	0
2	b^0	0
2	b^1	0
2	b^2	0
2	b^3	1

Table 13: The ilink potential $f(d_j, b)$ ($j = 1, ..., \eta$) of MDi and EDi

line	(d_j, b)	f
1	$d_j = 0,\ b = b^j, ..., b^{\eta+1}$	1
2	$d_j = 1,\ b = b^0, ..., b^\eta$	1
3	$d_j = 2,\ b = b^{\eta+1}$	1
4	otherwise	0

section of $f(d_i, b)$ has another row, and another potential $f(d_3, b)$ is needed. The first section of $f(d_1, b)$ is $(0, 1, 1, 1, 1)$ and that of $f(d_3, b)$ is $(0, 0, 0, 1, 1)$. The second section of $f(d_i, b)$ is $(1, 1, 1, 1, 0)$ and the third section is $(0, 0, 0, 0, 1)$. The general ilink potential of MDi is shown in Table 13. The general clink and family potentials of MDi are shown in Table 4.

5.5. *EDi*

For $|D_e| = 3$, ilink potentials $f(d_i, b)$ are identical to those of MDi in Table 12. The family potential $f(d_1, ..., d_\eta, h)$ is shown in Table 14. Relative to MDi, the domain size of h is increased to $|D_e|+1$. The family potential is extended accordingly from that of MDi (Table 4) by adding one row of 0 on each section, except the row added to section $(d_1 = 2, d_2 = 0, ..., d_\eta = 0)$ is 1. The

Table 14: Family potential $f(d_1, d_2, h)$ of an example EDi

d_1	d_2	h	f
0	0	h	0
0	1	h^0	-1
0	1	h^1	1
0	1	h^2	0
0	1	h^3	0
0	2	h	0
1	0	h^0	0
1	0	h^1	-1
1	0	h^2	1
1	0	h^3	0
1	1	h^0	1
1	1	h^1	0
1	1	h^2	0
1	1	h^3	0
1	2	h	0
2	0	h^0	1
2	0	h^1	0
2	0	h^2	-1
2	0	h^3	1
2	1	h	0
2	2	h	0

general ilink and clink potentials of EDi are shown in Tables 13 and 4, respectively. The general family potential of EDi is shown in Table 15.

Table 15: The family potential $f(d_1, ..., d_\eta, h)$ of EDi

line	$(d_1, ..., d_\eta, h)$	f
1	$d_i = 0,\ \forall_{j \neq i}\ d_j = 1,\ h = h^{i-1}$	-1
2	$d_i = 0,\ \forall_{j \neq i}\ d_j = 1,\ h = h^{i}$	1
3	$\forall_i\ d_i = 1,\ h = h^0$	1
4	$d_1 = 2,\ \forall_{i>1}\ d_i = 0,\ h = h^0, h^{\eta+1}$	1
5	$d_1 = 2,\ \forall_{i>1}\ d_i = 0,\ h = h^{\eta}$	-1
6	otherwise	0

6. Exactness and Complexity of MF of NAT Models

We analyze main properties of the MF of NAT models. Most of this section is devoted to the exactness and the last subsection analyzes complexity. Since the MF of NAT models was integrated from four gate MFs (last section), we analyze main properties of gate MFs and then establish exactness of the MF of NAT models. When analyzing a gate MF, we must consider the case, where all inputs are cause variables (referred to as *terminal* input), and the case, where some inputs are internal from other gates. From Table 6, in a NAT model, only EDu, EDi and MDi can have terminal-only inputs. They can have inputs from other gates as well. MDu, however, must have inputs from other gates (unless the NAT is trivial: a single gate). Properties of MDu and MDi with terminal inputs only have been analyzed in Sec. 4. Here, we analyze the following in that order.

EDu 1. Output property without ilinks (terminal inputs only)
2. Output property with ilinks from EDi

EDi 1. Output property without ilinks (terminal inputs only)
2. Output property with ilinks from EDu

MDi Output property with ilinks from EDu

MDu Output property with ilinks from MDi

The existence of four distinct gate MFs has a cost in establishing exactness when they are interacting within a NAT model, as will be seen below. It is conceivable that by making domain sizes of b, d and h in MDu and MDi the same as those of EDu and EDi, (as $\eta+2$, 3, and $\eta+2$, respectively), the number of distinct gate MFs may be reduced to two and the corresponding analysis may be simpler. Following the criteria stated in Sec. 4, we have instead made each gate MF as space-efficient as possible (e.g., domain size of b in MDu is smaller than that in EDu) even though this increases sophistication in analysis.

Consider a subMF with leaf variable h. If h belongs to an EDu, we refer to the subMF as an EDu-based subMF. For instance, if leaf gate g_1 in Fig. 4 (c) is direct, the MF for gate g_2 is EDu. The subMF with leaf variable b_3 in (d) is an EDu-based subMF. It includes the ancestor subgraph with leaf b_3 and potentials associated with all undirected links and families in the subgraph. When analyzing the output property of EDu, the focus is the output property of an EDu-based subMF. The similar naming and focus of analysis apply to EDi, MDi and MDu as well.

6.1. Output Property of EDu

Our analysis aims to establish that an EDu-based subMF has the output property specified in Def. 5.

Definition 5 (EDu output potential trait). *Let $\{c_1, ..., c_k\}$ ($k > 1$) be a set of causes and h be the output variable of EDu. An MPP $f(h, c_1, ..., c_k)$ satisfies the EDu output potential trait if the following holds.*

1. *If $h \leq h^{\eta}$ and some c_i is active, then $f(h, c_1, ..., c_k) = P(e|c_1, ..., c_k)$.*
2. *If $h = h^{\eta+1}$ and some c_i is active, then $f(h, c_1, ..., c_k) = 0$.*
3. *If $h < h^{\eta}$ and each c_i is inactive, then $f(h, c_1, ..., c_k) = 0$.*
4. *If $h \geq h^{\eta}$ and each c_i is inactive, then $f(h, c_1, ..., c_k) = 1$.*

The subtraits 3 and 4 show that the above trait is pseudo-probabilistic, since $P(e < e^{\eta} \leftarrow \perp) = 1$ and $P(e \geq e^{\eta} \leftarrow \perp) = 0$. This is necessary as EDu feeds into MDi or EDi.

In the following, we first analyze an EDu-based subMF made of the EDu only (no ilinks). Theorem 3 shows that the above property holds in such a subMF. In the theorem, each value h^i of h corresponds to the value e^i of e if $0 \leq i \leq \eta$.

Theorem 3. *Let EDu be applied to a dual NIN-AND gate model whose CPT is $P(e|c_1, ..., c_n)$. The MPP $f(h, c_1, ..., c_n)$ from potentials of the EDu satisfies the EDu output potential trait (Def. 5) with $k = n$.*

Proof: The MPP $f(h, c_1, ..., c_n) = \sum_{d_1,...,d_\eta} f(d_1, ..., d_\eta, h) \prod_{1 \leq j \leq \eta, 1 \leq i \leq n} f(d_j, c_i)$ is defined from potentials in Table 11 (top left and bottom). Consider product $f(d_j, c_1, ..., c_n) = \prod_{1 \leq i \leq n} f(d_j, c_i)$. From Table 11 (top left), we have $f(d_j = 1, c_1, ..., c_n) = 1$. If some c_i is active, we have $f(d_j = 0, c_1, ..., c_n) = P(e < e^j \leftarrow \underline{x}^+)$ by Eqn. (10), where $\underline{x}^+$ denotes all active c_i, and $f(d_j = 2, c_1, ..., c_n) = 0$. If each c_i is inactive, we have $f(d_j = 0, c_1, ..., c_n) = 1$, since $P(e < e^j \leftarrow \perp) = 1$ by Eqn. (7), and $f(d_j = 2, c_1, ..., c_n) = 1$. This is summarized in Table 16. The MPP now becomes $f(h, c_1, ..., c_n) = \sum_{d_1,...,d_\eta} f(d_1, ..., d_\eta, h) f(d_1, c_1, ..., c_n) \ldots f(d_\eta, c_1, ..., c_n)$, where $f(d_1, ..., d_\eta, h)$

Table 16: Summary on $f(d_j, c_1, ..., c_n)$ ($j = 1, ..., \eta$)

	$f(d_j, c_1, ..., c_n)$	
d_j	$\exists c_i\ c_i > c_i^0$	$\forall c_i\ c_i = c_i^0$
0	(1) $P(e < e^j \leftarrow \underline{x}^+)$	(4) 1
1	(2) 1	(5) 1
2	(3) 0	(6) 1

selects products $f(d_1, c_1, ..., c_n)...f(d_\eta, c_1, ..., c_n)$ for summation. We consider the two cases: (1) some c_i is active, and (2) each c_i is inactive.

Under case (1), if $h = h^0$, $f(d_1, ..., d_\eta, h)$ is non-zero by lines 1 and 4 of Table 11 (bottom). However, $f(d_1 = 2, c_1, ..., c_n) = 0$ by Table 16 (3). From Table 16 (2) and (1), we have

$$f(h^0, c_1, ..., c_n) = f(d_1 = 0, 1, ..., 1, h^0)f(d_1 = 0, c_1, ..., c_n) = P(e < e^1 \leftarrow \underline{x}^+) = P(e^0 \leftarrow \underline{x}^+).$$

If $h = h^\eta$, $f(d_1, ..., d_\eta, h)$ is non-zero by lines 2, 3 and 5 of Table 11 (bottom). By Table 16 (3), $f(d_1 = 2, c_1, ..., c_n) = 0$. From Table 16 (2) and (1), we have

$$\begin{aligned} f(h^\eta, c_1, ..., c_n) &= f(1, ..., 1, d_\eta = 0, h^\eta)f(d_\eta = 0, c_1, ..., c_n) + f(1, ..., 1, h^\eta) \\ &= -P(e < e^\eta \leftarrow \underline{x}^+) + 1 = P(e^\eta \leftarrow \underline{x}^+). \end{aligned}$$

If $h = h^k$ with $0 < k < \eta$, $f(d_1, ..., d_\eta, h)$ is non-zero by lines 1 and 2 of Table 11 (bottom). From Table 16 (2) and (1), we have

$$\begin{aligned} f(h^k, c_1, ..., c_n) &= f(1, ..., 1, d_{k+1} = 0, 1, ..., 1, h^k)f(d_{k+1} = 0, c_1, ..., c_n) \\ &+ f(1, ..., 1, d_k = 0, 1, ..., 1, h^k)f(d_k = 0, c_1, ..., c_n) \\ &= P(e < e^{k+1} \leftarrow \underline{x}^+) - P(e < e^k \leftarrow \underline{x}^+) = P(e^k \leftarrow \underline{x}^+). \end{aligned}$$

Hence, from Eqn. (3), the subtrait 1 holds.

If $h = h^{\eta+1}$, then $f(d_1, ..., d_\eta, h)$ is non-zero by line 5 of Table 11 (bottom) where $d_1 = 2$. As $f(d_1 = 2, c_1, ..., c_n) = 0$ by Table 16 (3), we have $f(h^{\eta+1}, c_1, ..., c_n) = 0$, which is the subtrait 2.

Under case (2), $f(d_j, c_1, ..., c_n) = 1$ by Table 16 (4), (5) and (6). The MPP now becomes $f(h, c_1, ..., c_n) = \sum_{d_1,...,d_\eta} f(d_1, ..., d_\eta, h)$. For $h < h^\eta$, lines 1, 2 and 4 in Table 11 (bottom) are non-zero. If $h = h^0$, by lines 1 and 4, we have

$$f(h^0, c_1, ..., c_n) = f(d_1 = 0, 1, ..., 1, h^0) + f(d_1 = 2, 0, ..., 0, h^0) = 1 - 1 = 0.$$

If $h = h^k$ with $0 < k < \eta$, by lines 1 and 2, we have

$$f(h^k, c_1, ..., c_n) = f(1, ..., 1, d_{k+1} = 0, 1, ..., 1, h^k) + f(1, ..., 1, d_k = 0, 1, ..., 1, h^k) = 1 - 1 = 0.$$

Hence, the subtrait 3 holds.

For $h \geq h^\eta$, lines 2, 3 and 5 of Table 11 (bottom) are non-zero. If $h = h^\eta$, we have

$$f(h^\eta, c_1, ..., c_n) = f(1, ..., 1, d_\eta = 0, h^\eta) + f(1, ..., 1, h^\eta) + f(d_1 = 2, 0, ..., 0, h^\eta) = -1 + 1 + 1 = 1.$$

If $h = h^{\eta+1}$, by line 5, we have $f(h^{\eta+1}, c_1, ..., c_n) = f(d_1 = 2, 0, ..., 0, h^{\eta+1}) = 1$. Hence, the subtrait 4 holds. □

We extend the two layers of MFs of standalone gates to subMFs. For an EDu-based subMF, the family layer is the same as before, made of the family of the leaf variable and its family potential. The link layer includes the rest of the ancestor subgraph and the associated potentials. In Fig. 4 (c), if g_1 is direct, the subMF with leaf b_3 in (d) is EDu-based. Its link layer consists of the subgraph including and above nodes h_1 and h_2, as well as relevant potentials. Def. 6 summarizes a property of the link layer in terms of an MPP defined over its potentials. It consists of 3 subtraits with the first two being identical to the MDu link potential trait (Def. 3).

Definition 6 (EDu link potential trait). *Let $\{c_1, ..., c_k\}$ ($k \geq 1$) be a set of causes of an EDu-based subMF and $\{d_1, ..., d_\eta\}$ be a set of auxiliary variables of the EDu. An MPP from the link layer of the subMF $f(d_1, ..., d_\eta, c_1, ..., c_k)$ satisifies the EDu link potential trait, if the MDu link potential trait (Def. 3) holds as well as the following.*

3. *If $d_1 = 2$, $\forall_{i>1}\ d_i = 0$, and each c_i is inactive, then $f(d_1, ..., d_\eta, c_1, ..., c_k) = 1$.*

Prop. 5 shows that the above property holds for an EDu-based subMF made of the EDu only (no ilinks).

Proposition 5. *Let c be a cause in a dual NIN-AND gate model. The product of clink potentials over c from the EDu of the model, $f(d_1, ..., d_\eta, c) = \prod_{1 \leq i \leq \eta} f(d_i, c)$, satisfies the EDu link potential trait (Def. 6) with $k = 1$.*

Proof: The clink potentials are defined in Table 11 (top left). We refer to sections where $d_j = 0, 1, 2$ as sections 0, 1, 2, respectively.

The subtrait 1 holds, since the factor $f(d_j, c)$ is from section 0 and all other factors are from section 1. The subtrait 2 holds as all factors are from section 1. The subtrait 3 follows from the first row of section 2 and from section 0, where $P(e < e^i \leftarrow \perp) = 1$ ($i > 1$). □

Prop. 5, in fact, covers only products of clink potentials that satisfy preconditions of the EDu link potential trait. Prop. 6 shows that these products are all that matter. It plays a similar role as Prop. 2, but the latter only applies to MDu.

Proposition 6. *Let c be a cause in a dual NIN-AND gate model over $c_1, ..., c_n$. The product of clink potentials from its EDu $f(d_1, ..., d_\eta, c) = \prod_{1 \leq i \leq \eta} f(d_i, c)$ contributes to MPP $f(e, c_1, ..., c_n)$, only if preconditions of the EDu link potential trait (Def. 6) hold with $k = 1$.*

Proof: The $f(d_1, ..., d_\eta, c)$ contributes to $f(e, c_1, ..., c_n)$ through its product with family potential $f(d_1, ..., d_\eta, h)$ in Table 11 (bottom). We show that if the preconditions fail, one of them is 0.

If none of preconditions 1, 2 and the first two subconditions of 3 holds, the first 5 lines of Table 11 (bottom) do not apply. By line 6, $f(d_1, ..., d_\eta, h) = 0$. If $d_1 = 2$, $\forall_{i>1}\ d_i = 0$, but c is active, $f(d_1, ..., d_\eta, c) = 0$ by Table 11 (top left, last row). □

Below, we move on to analyze an EDu-based subMF with ilinks. By Table 6, each ilink of the EDu is connected to the output variable of an EDi. To evaluate the impact from the EDi, we specify an output property of an EDi-based subMF in Def. 7. We show that the property holds for an EDi-based subMF in Section 6.2.

Definition 7 (EDi output potential trait). *Let $\{c_1, ..., c_k\}$ ($k > 1$) be a set of causes of an EDi-based subMF and h be the output variable of the EDi. An MPP $f(h, c_1, ..., c_k)$ from the subMF satisfies the EDi output potential trait, if the following holds.*

1. *If $h \leq h^\eta$, then $f(h, c_1, ..., c_k) = P(e|c_1, ..., c_k)$.*
2. *If $h = h^{\eta+1}$ and each c_i is inactive, then $f(h, c_1, ..., c_k) = 1$.*
3. *If $h = h^{\eta+1}$ and some c_i is active, then $f(h, c_1, ..., c_k) = 0$.*

Prop. 7 analyzes the behavior of the link layer with the focus on one input variable. It shows that MPP from the relevant ilink potentials behaves equivalently to products of clink potentials as analyzed in Prop. 5. Note that b is the output variable of an EDi, and the above output property of EDi is assumed. In Fig. 4 (c), if gate g_1 is direct, then the MF of g_2 in (d) is EDu and Prop. 7 applies to it. Variable b in Prop. 7 refers to b_2 in (d).

Proposition 7. *Let b be the leaf variable of an EDi-based subMF over causes $c_1, ..., c_n$ and an input variable of an EDu with ilinks $\langle b, d_1\rangle, ..., \langle b, d_\eta\rangle$. Let the MPP from the subMF be $f(b, c_1, ..., c_n)$ such that the EDi output potential trait (Def. 7) holds with $k = n$ and $h = b$.*

Then the MPP $f(d_1, ..., d_\eta, c_1, ..., c_n) = \sum_b f(d_1, b)...f(d_\eta, b)f(b, c_1, ..., c_n)$ from the link layer of the EDu-based subMF satisfies the EDu link potential trait (Def. 6) with $k = n$.

Proof: From Table 6, b is the output variable of an EDi, and has a domain size of $\eta + 2$ by Table 7. Table 11 (top right) defines ilink potentials in the MPP. Rewrite $f(d_1, ..., d_\eta, c_1, ..., c_k)$ as

$$f(d_1, b^{\eta+1})...f(d_\eta, b^{\eta+1})f(b^{\eta+1}, c_1, ..., c_k) + \sum_{b \leq b^\eta} f(d_1, b)...f(d_\eta, b)f(b, c_1, ..., c_k).$$

For the EDu subtrait 1, if $\exists_j\ d_j = 0$ and $\forall_{i\neq j}\ d_i = 1$, the first term is 0, since we have $f(d_{i\neq j} = 1, b^{\eta+1}) = 0$ by Table 11 (top right, lines 2 and 4). By line 2, the second term is $\sum_{b\leq b^\eta} f(d_j = 0, b)f(b, c_1, ..., c_k)$ and it becomes $\sum_{b^0\leq b\leq b^{j-1}} f(b, c_1, ..., c_k)$ by line 1. From the EDi subtrait 1, the sum equals $P(e < e^j \leftarrow c_1, ..., c_k)$, and the EDu subtrait 1 follows.

If $\forall_i\ d_i = 1$, the first term above is 0 by Table 11 (top right, lines 2 and 4), and the second becomes $\sum_{b\leq b^\eta} f(b, c_1, ..., c_k)$ by line 2. From the EDi subtrait 1, the EDu subtrait 2 follows as the probabilities sum to one.

If $d_1 = 2$ and $\forall_{i>1}\ d_i = 0$, the second term of $f(d_1, ..., d_\eta, c_1, ..., c_k)$ is 0 by Table 11 (top right, lines 3 and 4) as $f(d_1 = 2, b \leq b^\eta) = 0$. The first term is $f(b^{\eta+1}, c_1, ..., c_k)$ by lines 1 and 3. If each c_i is inactive, the value is 1 by the EDi subtrait 2. Hence, the EDu subtrait 3 holds. $\square$

The MPPs covered in Prop. 7 include only those satisfying preconditions of the EDu link potential trait. Prop. 8 shows that those are all that matter.

Proposition 8. *Let $c_1, ..., c_n$ be the set of causes of a NAT model. Let b be the leaf variable of a subMF over causes $c_1, ..., c_m$ ($m < n$) and an input variable of an EDu with ilinks $\langle b, d_1\rangle, ..., \langle b, d_\eta\rangle$. Let the MPP from the subMF be $f(b, c_1, ..., c_m)$ such that the EDi output potential subtrait 3 (Def. 7) holds with $k = m$ and $h = b$.*

Then the MPP $f(d_1, ..., d_\eta, c_1, ..., c_m) = \sum_b f(d_1, b)...f(d_\eta, b)f(b, c_1, ..., c_m)$ contributes to the MPP $f(e, c_1, ..., c_n)$ of the NAT model, only if preconditions of the EDu link potential trait (Def. 6) holds with $k = m$.

Proof: The $f(d_1, ..., d_\eta, c_1, ..., c_m)$ contributes to $f(e, c_1, ..., c_n)$ by its product with family potential $f(d_1, ..., d_\eta, h)$ in Table 11 (bottom). We show that if the preconditions fail, one of them is 0.

If none of preconditions 1, 2 and the first two subconditions of 3 holds, the first 5 lines of Table 11 (bottom) do not apply. By line 6, $f(d_1, ..., d_\eta, h) = 0$. If $d_1 = 2$ and $\forall_{i>1}\ d_i = 0$, but some c_i is active (failing condition 3), rewrite $f(d_1, ..., d_\eta, c_1, ..., c_m)$ as

$$f(d_1, b^{\eta+1})...f(d_\eta, b^{\eta+1})f(b^{\eta+1}, c_1, ..., c_m) + \sum_{b \leq b^\eta} f(d_1, b)...f(d_\eta, b)f(b, c_1, ..., c_m).$$

The second term is 0 by Table 11 (top right, lines 3 and 4). The first term is $f(b^{\eta+1}, c_1, ..., c_m)$ by lines 1 and 3. The value equals 0 by the EDi subtrait 3. Hence, $f(d_1, ..., d_\eta, c_1, ..., c_m) = 0$. $\square$

We are now ready to show the output property of EDu when both clinks and ilinks exist.

Theorem 4. *Let $\{c_1, ..., c_n\}$ be the set of causes of an EDu-based subMF. For any input variable b of the EDu such that b is the leaf variable of an EDi-based subMF over causes $\{s_1, ..., s_m\} \subset \{c_1, ..., c_n\}$ and the MPP from the EDi-based subMF is $f(b, s_1, ..., s_m)$, the EDi output potential trait (Def. 7) holds with the set of causes being $\{s_1, ..., s_m\}$ and $h = b$.*

Then the MPP $f(h, c_1, ..., c_n)$ from the EDu-based subMF satisfies the EDu output potential trait (Def. 5) with $k = n$.

Proof: By Theorem 3, the EDu output potential trait holds when the EDu-based subMF has clinks only. By Prop. 5, this result is obtained when products of clink potentials of the subMF satisfy the EDu link potential trait. Although the trait has preconditions, by Prop. 6, those products of clink potentials obtained when the preconditions fail do not matter.

When the EDu-based subMF has ilinks, by Prop. 7, MPPs from ilink potentials of the subMF also satisfy the EDu link potential trait, as long as input EDi-based subMF potentials satisfy the EDi output potential trait. Although the trait has preconditions, by Prop. 8, those MPPs obtained when the preconditions fail do not matter. Hence, MPPs from ilink potentials of the subMF behave equivalently to products of clink potentials. Therefore, Theorem 3 can be generalized to the EDu-based subMF with arbitrary combinations of ilinks and clinks, as long as input EDi-based subMF potentials satisfy the EDi output potential trait. $\square$

6.2. Output Property of EDi

An output property for an EDi-based subMF is specified in Def. 7. Theorem 5 shows that the property holds in EDi when there are no ilinks. It is phrased in the context of a standalone direct gate, as the property remains the same whether the gate is standalone or is in a NAT. In Fig. 4 (c), if gate g_1 is dual, then the MF of g_4 in (d) is EDi and Theorem 5 applies to it. In the theorem, each value h^i of h corresponds to the value e^i of e if $0 \leq i \leq \eta$. The first subtrait says that when output variable h is restricted to the domain of e, the MPP is the exact CPT. The other subtraits are relevant when the EDi feeds into an EDu-based subMF.

Theorem 5. *Let EDi be applied to a direct NIN-AND gate model whose CPT is $P(e|c_1, ..., c_n)$. The MPP $f(h, c_1, ..., c_n)$ from the EDi satisfies the EDi output potential trait (Def. 7) with $k = n$.*

Proof: The MPP $f(h, c_1, ..., c_n) = \sum_{d_1,...,d_\eta} f(d_1, ..., d_\eta, h) \prod_{1 \le j \le \eta, 1 \le i \le n} f(d_j, c_i)$ is defined from potentials in Tables 15 and 4. Compare the MPP with that of Theorem 2 on MDi. The clink potentials (Table 4) are identical. The family potential in Table 15 differs from that in Table 4 only when $h = h^{\eta+1}$. Hence, $f(h, c_1, ..., c_n)$ differs from $f(e, c_1, ..., c_n)$ in Theorem 2 only when $h = h^{\eta+1}$, from which the subtrait 1 follows.

When $h = h^{\eta+1}$, the only non-zero family potential value is $f(d_1 = 2, 0, ..., 0, h^{\eta+1})$ by Table 15 (line 4). Consider the corresponding product $f(d_1 = 2, d_2 = 0, ..., d_\eta = 0, c_1, ..., c_n) = \prod_{1 \le j \le \eta, 1 \le i \le n} f(d_j, c_i)$. If each cause is inactive, $f(d_1 = 2, d_2 = 0, ..., d_\eta = 0, c_1, ..., c_n) = 1$ by Table 4. Hence, $f(h^{\eta+1}, c_1, ..., c_n) = 1$, which is the subtrait 2.

If some cause is active, then $f(d_1 = 2, d_2 = 0, ..., d_\eta = 0, c_1, ..., c_n) = 0$ by Table 4. Hence, $f(h^{\eta+1}, c_1, ..., c_n) = 0$, which is the subtrait 3. $\square$

Prop. 3 reveals a property of products of clink potentials in MDi. Since clink potentials of MDi and EDi are identical (Table 4), Prop. 3 is applicable to EDi as well. Prop. 9 below shows that products covered in Prop. 3 are all that matter to EDi. It plays a similar role as Prop. 4, but the latter only applies to MDi. Note that the first two preconditions of the MDi link potential trait combine into $\exists_j\ d_j = 0 \wedge \forall_{i \ne j}\ d_i = 1$.

Proposition 9. *Let c be a cause in a direct NIN-AND gate model. The product of clink potentials from its EDi $f(d_1, ..., d_\eta, c) = \prod_{1 \le i \le \eta} f(d_i, c)$ contributes to MPP $f(e, c_1, ..., c_n)$ of the EDi, only if preconditions of the MDi link potential trait (Def. 4) hold.*

Proof: The proof of Prop. 4 can be directly applied. This is justified because the proof involves only clink and family potentials. The clink potentials of EDi and MDi are identical (Table 4, left). The family potential of EDi (Table 15) differs from that of MDi (Table 4, right) only on $h = h^{\eta+1}$, and $h = h^{\eta+1}$ is irrelevant to preconditions of the MDi link potential trait. $\square$

Next, we analyze an EDi-based subMF when ilinks exist. Prop. 10 shows that the MPP from ilink potentials behaves equivalently to the product of clink potentials (Prop. 3). Note that b is the output variable of an EDu, and conditions on MPP from the subMF are consistent with those of Theorem 4 on EDu. In Fig. 4 (c), if gate g_1 is direct, then the MF of g_3 in (d) is EDi and Prop. 10 applies to it. Variable b in Prop. 10 refers to b_1 in (d).

Proposition 10. *Let b be the leaf variable of an EDu-based subMF over causes $c_1, ..., c_n$ and an input variable of an EDi with ilinks $\langle b, d_1 \rangle, ..., \langle b, d_\eta \rangle$. Let the MPP from the EDu-based subMF be $f(b, c_1, ..., c_n)$ such that the EDu output potential trait (Def. 5) holds with $k = n$ and $h = b$.*

Then the MPP $f(d_1, ..., d_\eta, c_1, ..., c_n) = \sum_b f(d_1, b)...f(d_\eta, b) f(b, c_1, ..., c_n)$ from the link layer of the EDi-based subMF satisfies the MDi link potential trait (Def. 4) with $k = n$.

Proof: Since b is the output variable of an EDu, its domain size is $\eta + 2$ by Table 7. Table 13 defines ilink potentials in the EDi MPP. Rewrite $f(d_1, ..., d_\eta, c_1, ..., c_k)$ as

$$f(d_1, b^{\eta+1})...f(d_\eta, b^{\eta+1}) f(b^{\eta+1}, c_1, ..., c_k) + \sum_{b \le b^\eta} f(d_1, b)...f(d_\eta, b) f(b, c_1, ..., c_k).$$

For the MDi subtrait 1, if $\exists_j\ d_j = 0$ and $\forall_{i \ne j}\ d_i = 1$, the first term of $f(d_1, ..., d_\eta, c_1, ..., c_k)$ is 0 by Table 13 (lines 2 and 4). By line 2, the second term is $\sum_{b \le b^\eta} f(d_j = 0, b) f(b, c_1, ..., c_k)$

and it becomes $\sum_{b^j \leq b \leq b^\eta} f(b, c_1, ..., c_k)$ by lines 1 and 4. If some c_i is active, the sum equals $P(e \geq e^j \leftarrow c_1, ..., c_k)$ by the EDu subtrait 1, and the MDi subtrait 1 follows. If each c_i is inactive, the sum equals 1 from the EDu subtraits 3 and 4. Hence, the MDi subtrait 2 follows.

If $\forall_i \ d_i = 1$, the first term of $f(d_1, ..., d_\eta, c_1, ..., c_k)$ is 0 by Table 13 (lines 2 and 4), and the second is $\sum_{b \leq b^\eta} f(b, c_1, ..., c_k)$ by line 2. If some c_i is active, the sum equals 1 according to the EDu subtrait 1. If each c_i is inactive, the sum equals 1 by the EDu subtraits 3 and 4. Hence, the MDi subtrait 3 follows.

If $d_1 = 2$ and $\forall_{i>1} \ d_i = 0$, the second term of $f(d_1, ..., d_\eta, c_1, ..., c_k)$ is 0 by Table 13 (lines 3 and 4). The first term becomes $f(b^{\eta+1}, c_1, ..., c_k)$ by lines 1 and 3. If each c_i is inactive, the value is 1 by the EDu subtrait 4. Hence, the MDi subtrait 4 follows. $\square$

The MPPs covered in Prop. 10 include only those satisfying preconditions of the EDi link potential trait. Prop. 11 shows that those are all that matter.

Proposition 11. *Let $c_1, ..., c_n$ be the set of causes of a NAT model. Let b be the leaf variable of an EDu-based subMF over causes $c_1, ..., c_m$ ($m < n$) and an input variable of an EDi with ilinks $\langle b, d_1 \rangle, ..., \langle b, d_\eta \rangle$. Let the MPP from the EDu-based subMF be $f(b, c_1, ..., c_m)$ such that the EDu output potential trait (Def. 5) holds with $k = m$ and $h = b$.*

Then the MPP $f(d_1, ..., d_\eta, c_1, ..., c_m) = \sum_b f(d_1, b)...f(d_\eta, b) f(b, c_1, ..., c_m)$ from the link layer of the EDi-based subMF contributes to the MPP $f(e, c_1, ..., c_n)$ of the NAT model, only if preconditions of the MDi link potential trait (Def. 4) hold.

Proof: The MPP $f(d_1, ..., d_\eta, c_1, ..., c_m)$ contributes to $f(e, c_1, ..., c_n)$ by product with the family potential $f(d_1, ..., d_\eta, h)$ in Table 15. We show that if the preconditions fail, one of them is 0.

The first two preconditions combine into $\exists_j \ d_j = 0 \wedge \forall_{i \neq j} \ d_i = 1$. If none of this combined condition, precondition 3, and the first two subconditions of precondition 4 holds, then the first 5 lines of Table 15 do not apply. By line 6, $f(d_1, ..., d_\eta, h) = 0$. If $d_1 = 2$ and $\forall_{i>1} \ d_i = 0$, but some c_i is active (failing the third subcondition of precondition 4), rewrite $f(d_1, ..., d_\eta, c_1, ..., c_m)$ as

$$f(d_1, b^{\eta+1})...f(d_\eta, b^{\eta+1}) f(b^{\eta+1}, c_1, ..., c_m) + \sum_{b \leq b^\eta} f(d_1, b)...f(d_\eta, b) f(b, c_1, ..., c_m).$$

The second term is 0 by Table 13 (lines 3 and 4). The first term becomes $f(b^{\eta+1}, c_1, ..., c_m)$ by lines 1 and 3. The value equals 0 by the EDu subtrait 2. Hence, $f(d_1, ..., d_\eta, c_1, ..., c_m) = 0$. $\square$

Theorem 6 shows the output property of an EDi-based subMF when both clinks and ilinks may exist.

Theorem 6. *Let $\{c_1, ..., c_n\}$ be the set of causes of an EDi-based subMF. For any input variable b of the EDi such that b is the leaf variable of an EDu-based subMF over causes $\{s_1, ..., s_m\} \subset \{c_1, ..., c_n\}$ and the MPP from the EDu-based subMF is $f(b, s_1, ..., s_m)$, the EDu output potential trait (Def. 5) holds with the set of causes being $\{s_1, ..., s_m\}$ and $h = b$. Then the MPP $f(h, c_1, ..., c_n)$ from the EDi-based subMF satisfies the EDi output potential trait (Def. 7) with $k = n$.*

Proof: By Theorem 5, the EDi output potential trait holds when the EDi-based subMF has clinks only. By Prop. 3, this result is obtained when products of clink potentials of the subMF satisfy the

EDi link potential trait. Although the trait has preconditions, by Prop. 9, those products of clink potentials obtained when the preconditions fail do not matter.

When the EDi-based subMF has ilinks, by Prop. 10, MPPs from ilink potentials of the subMF also satisfy the EDi link potential trait, as long as input EDu-based subMF potentials satisfy the EDu output potential trait. Although the trait has preconditions, by Prop. 11, those MPPs obtained when the preconditions fail do not matter. Hence, MPPs from ilink potentials of the subMF behave equivalently to products of clink potentials. Therefore, Theorem 5 can be generalized to the EDi-based subMF with arbitrary combinations of ilinks and clinks, as long as input EDu-based subMF potentials satisfy the EDu output potential trait. □

6.3. Output Property of MDi

The output property for an MDi-based subMF without ilinks has been analyzed in Theorem 2. To analyze the output property when ilinks exist, we compare with an EDi-based subMF. MDi and EDi have the same clink potentials (Table 4, left) and ilink potentials (Table 13). By Table 6, both MDi and EDi receive ilink input from EDu. In Table 7, the two only differ in that the domain size of output variable h is $\eta + 1$ in MDi and $\eta + 2$ in EDi. As a result, the family potential of EDi (Table 15) is the same as that of MDi (Table 4, right) everywhere except when $h = h^{\eta+1}$.

This comparison suggests that all analyses on EDi are applicable to MDi, except that $h = h^{\eta+1}$ is undefined under MDi and any property of EDi on $h = h^{\eta+1}$ must be ignored. Theorem 7 below establishes the output property of an MDi-based subMF. It follows directly from Theorem 6 by ignoring the statements on $h = h^{\eta+1}$. In Fig. 4 (c), if gate g_1 is direct, then the MF of g_1 in (d) is MDi and Theorem 7 applies to it. Variable b in Theorem 7 refers to b_3 in (d).

Theorem 7. *Let $\{c_1, ..., c_n\}$ be the set of causes in an MDi-based subMF. For any input variable b of the MDi such that b is the leaf variable of an EDu-based subMF over causes $\{s_1, ..., s_m\} \subset \{c_1, ..., c_n\}$ and the MPP from the EDu-based subMF is $f(b, s_1, ..., s_m)$, the EDu output potential trait (Def. 5) holds with the set of causes being $\{s_1, ..., s_m\}$ and $h = b$. Then the MPP from the MDi-based subMF satisfies $f(h, c_1, ..., c_n) = P(e|c_1, ..., c_n)$.*

6.4. Output Property of MDu

The output property for an MDu-based subMF without ilinks is stated in Corollary 2. To analyze the output property when ilinks exist, Prop. 12 shows that MPPs from ilink potentials behave equivalently to products of clink potentials (Prop. 1). By Table 6, an MDu only receives input from MDi. Note that b and e share the domain size.

Proposition 12. *Let b be the leaf variable of an MDi-based subMF over causes $c_1, ..., c_n$ and an input variable of an MDu with ilinks $\langle b, d_1\rangle, ..., \langle b, d_\eta\rangle$. Let the MPP from the MDi-based subMF be $f(b, c_1, ..., c_n) = P(e|c_1, ..., c_n)$. Then the MPP*

$$f(d_1, ..., d_\eta, c_1, ..., c_n) = \sum_b f(d_1, b)...f(d_\eta, b)f(b, c_1, ..., c_n)$$

from the MDu-based subMF satisfies the MDu link potential trait (Def. 3) with $k = n$.

Proof: For the subtrait 1, suppose $\exists_j \; d_j = 0$ and $\forall_{i \neq j} \; d_i = 1$. Consider the case $d_1 = 0$. From line 2 in Table 9, $f(d_1 = 0, b) = 0$ for all $b \geq b^1$. Hence, all terms in the summation of the proposition where $b \geq b^1$ are zeros. The only non-zero term is for $b = b^0$. From line 1 in Table 9, we have $f(d_1 = 0, b^0) = 1$. From line 3, $f(d_j = 1, b^0) = 1$ for all $j > 1$. This yields

$$\begin{aligned} & f(d_1 = 0, d_2 = 1, ..., d_\eta = 1, c_1, ..., c_n) = f(b^0, c_1, ..., c_n) \\ = \; & P(e^0 | c_1, ..., c_n) = P(e < e^1 \leftarrow c_1, ..., c_n). \end{aligned}$$

Next, consider the case $d_j = 0$ ($j > 1$). From line 2 of Table 9, all terms of the summation where $b \geq b^j$ are zeros. The non-zero terms are those where $b < b^j$. This yields

$$\begin{aligned} & f(d_1 = 1, ..., d_{j-1} = 1, d_j = 0, d_{j+1} = 1, ..., d_\eta = 1, c_1, ..., c_n) \\ = \; & f(b^0, c_1, ..., c_n) + ... + f(b^{j-1}, c_1, ..., c_n) \\ = \; & P(e^0 | c_1, ..., c_n) + ... + P(e^{j-1} | c_1, ..., c_n) = P(e < e^j \leftarrow c_1, ..., c_n). \end{aligned}$$

Hence, the subtrait 1 holds.

If $\forall_i \; d_i = 1$, we have $f(d_i, b) = 1$ for all i, by Table 9 (line 3). That is,

$$f(d_1, ..., d_\eta, c_1, ..., c_n) = P(e^0 | c_1, ..., c_n) + ... + P(e^\eta | c_1, ..., c_n) = 1.$$

Hence, the subtrait 2 holds. $\square$

The MPPs covered in Prop. 12 include only those satisfying preconditions of the MDu link potential trait. Prop. 13 shows that those are all that matter.

Proposition 13. *Let $c_1, ..., c_n$ be the set of causes of a NAT model. Let b be the leaf variable of an MDi-based subMF over causes $c_1, ..., c_m$ ($m < n$) and an input variable of an MDu with ilinks $\langle b, d_1 \rangle, ..., \langle b, d_\eta \rangle$. Let the MPP from the MDi-based subMF be $f(b, c_1, ..., c_m) = P(e | c_1, ..., c_m)$. Then the MPP $f(d_1, ..., d_\eta, c_1, ..., c_m) = \sum_b f(d_1, b) ... f(d_\eta, b) f(b, c_1, ..., c_m)$ from the link layer of the MDu-based subMF contributes to the MPP $f(e, c_1, ..., c_n)$ of the NAT model, only if preconditions of the MDu link potential trait (Def. 3) hold.*

Proof: The MPP $f(d_1, ..., d_\eta, c_1, ..., c_m)$ contributes to $f(e, c_1, ..., c_n)$ by product with the family potential $f(d_1, ..., d_\eta, e)$ in Table 2 (right). If none of the two preconditions holds, then the first 3 lines of Table 2 (right) do not apply. By line 4, $f(d_1, ..., d_\eta, e) = 0$. $\square$

Theorem 8 establishes the output property of an MDu-based subMF when both clinks and ilinks may exist. From Table 6, MDu can only be applied to the leaf gate of a NAT model. Hence, an MDu-based subMF is the MF of the NAT model in its entirety.

Theorem 8. *Let $\{c_1, ..., c_n\}$ be a set of causes of an MDu-based subMF. For any input variable b of the MDu such that b is the leaf variable of an MDi-based subMF over causes $\{s_1, ..., s_m\} \subset \{c_1, ..., c_n\}$, the MPP from the MDi-based subMF is $f(b, s_1, ..., s_m) = P(e | s_1, ..., s_m)$. Then the MPP from the MDu-based subMF satisfies $f(e, c_1, ..., c_n) = P(e | c_1, ..., c_n)$.*

Proof: From Corollary 2, the condition on $f(e, c_1, ..., c_n)$ holds when the MDu-based subMF has no ilinks. By Prop. 1, this result is obtained when products of clink potentials of the subMF satisfy the MDu link potential trait. Although the trait has preconditions, by Prop. 2, those products obtained when the preconditions do not hold do not matter.

When the MDu-based subMF has ilinks, by Prop. 12, MPPs from ilink potentials and input MDi-based subMFs also satisfy the MDu link potential trait, as long as the MPP from each MDi-based subMF is probabilistically exact. Although the trait has preconditions, by Prop. 13, those MPPs obtained when the preconditions do not hold do not matter. Hence, MPPs from ilink potentials and input MDi-based subMFs behave equivalently to products of clink potentials. Therefore, Corollary 2 can be generalized to the MDu-based subMF with arbitrary combinations of ilinks and clinks, as long as the MPP from each MDi-based subMF satisfies $f(b, s_1, ..., s_m) = P(e|s_1, ..., s_m)$. □

6.5. Soundness of MF of NAT Models

We establish the exactness of NAT MF in Theorem 9.

Theorem 9. *Let MF be applied to a NAT model over causes $c_1, ..., c_n$ by applying MDu, EDu, MDi, and EDi to appropriate gates. Then the MPP from all potentials of the MF satisfies $f(e, c_1, ..., c_n) = P(e|c_1, ..., c_n)$.*

Proof: We prove by induction on the maximum level L of the NAT. If $L = 1$, the NAT has a single gate. The statement holds by Corollary 2 if the gate is dual and by Theorem 2 if the gate is direct.

If $L = 2$ and the leaf gate is dual, then apply MDi to all direct gates at level 2 and apply MDu to the leaf gate. Since every gate at level 2 is terminal (all inputs are single-causal events), by Theorem 2, the MPP is the exact CPT over causes in the MDi-based subMF. The subMF feeds into the MDu of the leaf gate and satisfies the condition of Theorem 8. Hence, the statement holds by Theorem 8.

If $L = 2$ and the leaf gate is direct, then apply EDu to all dual gates at level 2 and apply MDi to the leaf gate. Since every gate at level 2 is terminal, by Theorem 3, the MPP from the EDu-based subMF satisfies the EDu output potential trait. The subMF feeds into the MDi of the leaf gate and satisfies the condition of Theorem 7. Hence, the statement holds by Theorem 7.

If $L = 3$ and the leaf gate is dual, apply EDu to each terminal dual gate at level 3 and the MPP from the EDu-based subMF satisfies the EDu output potential trait by Theorem 3. The subMF feeds into a direct gate at level 2 with MDi and satisfies the input condition of Theorem 7. By Theorem 7, the MPP from the MDi-based subMF is the exact CPT over its causes. From the analysis above on $L = 2$ with the dual leaf gate, the statement holds.

If $L = 3$ and the leaf gate is direct, apply EDi to each terminal direct gate at level 3 and the MPP from the EDi-based subMF satisfies the EDi output potential trait by Theorem 5. The subMF feeds into a dual gate at level 2 with EDu and satisfies the input condition of Theorem 4. By Theorem 4, the MPP from the EDu-based subMF satisfies the EDu output potential trait. From the analysis above on $L = 2$ with the direct leaf gate, the statement holds.

Assume that the statement holds when $L = k \geq 3$. Consider a NAT where $L = k + 1$. If the terminal gates at level $k + 1$ are dual, apply EDu to each and the MPP from the EDu-based subMF satisfies the EDu output potential trait by Theorem 3. The subMF feeds into a direct gate at level k with EDi. By Prop. 10, relative to the direct gate, the subMF behaves equivalently to a terminal input single-causal event. As a result, the MF of the NAT behaves equivalently to one where $L = k$. By the inductive hypothesis, the statement holds for $L = k + 1$.

If the terminal gates at level $k+1$ are direct, apply EDi to each and the MPP from the EDi-based subMF satisfies the EDi output potential trait by Theorem 5. The subMF feeds into a dual gate at level k with EDu. By Prop. 7, relative to the dual gate, the subMF behaves equivalently to a terminal input single-causal event. As a result, the MF of the NAT behaves equivalently to one where $L = k$. By the inductive hypothesis, the statement holds for $L = k+1$. $\square$

6.6. *Space Complexity of MF of NAT Models*

For space complexity of a gate MF, the size of a clink potential is $O(3(m+1))$, where $m+1$ bounds the domain size of a cause. The size of an ilink potential is $O(3(\eta+2))$. The size of a family potential is $O(3^\eta(\eta+2))$. Denoting the number of inputs to a gate by n', the gate MF has $n'\eta$ link potentials and one family potential. Assuming $m = \eta$, the space complexity of the gate MF is $O(3n'\,\eta\,(\eta+2) + 3^\eta(\eta+2)) = O((3n'\,\eta + 3^\eta)\,\eta)$.

Suppose that a NAT has k gates. Then its MF takes $O((3n'\,\eta+3^\eta)\,\eta\,k)$ space. The product $n'\,k$ counts the number of non-auxilary nodes (non-d_j) in the graph of the MF and hence $n'k < 2n$. We also have $k < n$. Hence, the space complexity of the MF of a NAT model is $O(6\,n\,\eta^2 + 3^\eta\,n\,\eta) \approx O(n\,\eta\,(6\,\eta + 3^\eta))$.

The existence of 3^η in the above complexity raises the question whether the MF of a NAT model is more efficient than the corresponding CPT as η grows. We consider this below. To concentrate on the main factors, we evaluate space complexity of a NAT MF by $n\,\eta\,3^\eta$ and that of a full CPT by η^n. We define their ratio as function $g(\eta,n) = \eta^n/(n\,\eta\,3^\eta)$. The derivative of $g(\eta,n)$ with respect to n is as follows.

$$g'(\eta,n) = (\eta\,3^\eta)^{-1}(\eta^n/n)'.$$

$$\text{Since}\quad (\eta^n/n)' = (\eta^n\,n^{-1})' = (\eta^n)'\,n^{-1} + (n^{-1})'\,\eta^n = ln(\eta)\,\eta^n\,n^{-1} + (-1)n^{-2}\,\eta^n$$
$$= ln(\eta)\,\eta^n\,n^{-1} - n^{-2}\,\eta^n = n\,ln(\eta)\,\eta^n\,n^{-2} - \eta^n\,n^{-2} = (n\,ln(\eta)-1)\,\eta^n\,n^{-2},$$
$$\text{we have}\quad g'(\eta,n) = (\eta\,3^\eta)^{-1}\,(n\,ln(\eta)-1)\,\eta^n\,n^{-2} = (n\,ln(\eta)-1)\,\eta^{n-1}\,n^{-2}\,3^{-\eta}.$$

When $\eta \geq 3$ and $n \geq 3$, we have $(n\,ln(\eta)-1) > 0$ and $g'(\eta,n) > 0$. Hence, $g(\eta,n)$ is monotonically increasing for $\eta \geq 3$ and $n \geq 3$. In other words, for each domain size $\eta \geq 3$, as n grows beyond some value, the NAT MF is always more efficient than a full CPT. For each η, the n value at which the space of full CPT exceeds that of NAT MF is shown in Fig. 5.

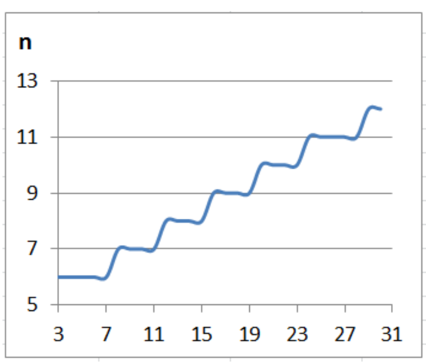

Figure 5: For each η value, the curve shows the n value, where the space of full CPT exceeds that of NAT MF.

For instance, for η between 3 and 7 inclusive, the space of full CPT exceeds that of NAT MF as long as $n \geq 6$. For η between 8 and 11, the condition holds as long as $n \geq 7$. For $\eta = 30$, the same occurs as long as $n \geq 12$. In summary, NAT MF pays off for moderate and large n values.

7. Lazy Propagation with NAT-Modeled Bayesian Networks

We present a framework in which MFs of NAT models can be utilized for exact inference with BNs through lazy propagation. The necessary background on lazy propagation is first introduced. The framework and its exactness are then presented.

7.1. Overview of Lazy Propagation

Lazy propagation (LP) [19] is an inference method based on a junction tree (JT) compiled from a BN over a set V of variables. Each cluster (a subset of variables) in the JT is assigned a set of CPTs from the BN, but these CPTs are not multiplied as commonly performed [8]. We refer to these CPTs as potentials, refer to the cluster of current focus by C, and refer to the set of potentials at C by β. The product of potentials in all clusters, denoted by $B(V)$, equals the product of CPTs of the BN, which in turn equals the joint probability distribution (JPD) of the BN.

The intersection of two adjacent clusters C and C' is their separator S and it is associated with two buffers. One stores message from C' to C and the other from C to C'. For the given cluster C and separator S, we refer to the two buffers as *in-buffer* and *out-buffer*, respectively, relative to C. The atomic operation at a cluster C relative to an adjacent cluster C' computes a set of potentials over their separator as the message and sends it to the out-buffer with C'. The message is so computed that the size of each resultant potential is kept as small as possible. LP proceeds by atomic operations performed at each JT separator in two passes. In the first pass, messages flow towards an arbitrarily specified cluster. In the second pass, messages flow away from the cluster. Prop. 14 summarizes the effect of LP. It says that after LP, the product of the local potentials at each cluster and potentials in its in-buffers is the exact marginal of the JPD of the BN.

Proposition 14 (Proposition 3.4 in [20]). *Let LP be performed in a JT. For any cluster C with the potential set β and the in-buffer message β_i from the adjacent separator R_i, denote the product of potentials in β as $\beta(C)$ and the product of potentials in β_i as $\beta_i(R_i)$. Then, we have $\beta(C)\ \prod_i \beta_i(R_i) = const \sum_{V \setminus C} B(V)$, where $const$ is a constant.*

We denote the potential product at cluster C after LP by $\alpha(C) = \beta(C)\ \prod_i \beta_i(R_i)$ and refer to $\alpha(C)$ as the *post-LP cluster belief* at C.

7.2. Lazy Propagation with NAT-Modeled BNs

We present how to use MFs of NAT models to improve inference efficiency with BNs. Consider a BN over a set V of variables with the DAG D. Each root of D is assigned a prior, collected in a set PR. Each single-parent non-root is assigned a CPT, collected in a set PS. We assume that the family of each multi-parent non-root can be expressed as a NAT model, collected in set Ψ. Then, $\Gamma = (V, D, PR, PS, \Psi)$ is a *NAT modeled BN* (NATBN). For efficient inference, we compile a NATBN into a JT representation by Alg. 1.

Algorithm 1.
Input: a NAT modeled BN $\Gamma = (V, D, PR, PS, \Psi)$

1 *get the skeleton* G *of* D *by dropping direction of links;*
2 *for each multi-parent family* $\{e, c_1, ..., c_n\}$ *in* D*, modify* G *as follows:*
3   *replace subgraph spanned by* $\{e, c_1, ..., c_n\}$ *with the MF graph* SG *of the NAT model in* Ψ*;*
4   *for each family in* SG*, connect members pairwise and drop direction of links;*
5 *triangulate* G *into a chordal graph* G'*;*
6 *construct a junction tree* T *from maximum cliques of* G'*;*
7 *assign each potential in* $PR \cup PS \cup \Psi$ *to a domain-containing cluster in* T*;*
8 *return* T*;*

The DAG of a trivial BN is shown in Fig. 6 (a). Suppose the family $\{e, c_1, ..., c_5\}$ is modeled by

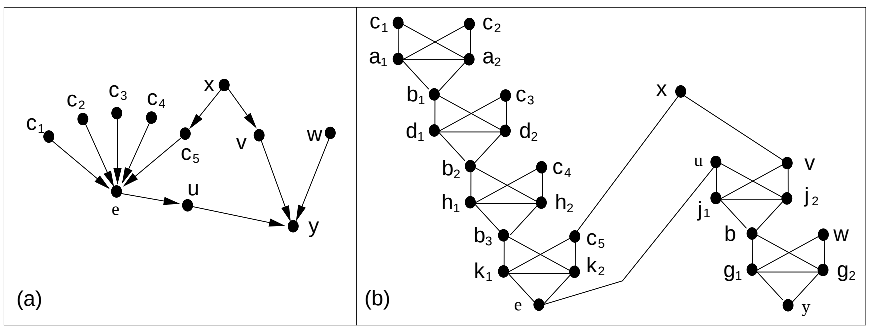

Figure 6: (a) A NATBN. (b) Undirected structure of compiled NATBN.

the NAT in Fig. 4 (c), and the family $\{y, u, v, w\}$ is modeled by the NAT in Fig. 4 (a). The graph G after lines 1 to 4 is shown in Fig. 6 (b). The compilation does not include moralization of D, and potentials assigned to each JT cluster are not multiplied. We refer to T as the JT of Multiplicatively Factorized NAT modeled BN (JTMFNB), to which LP is directly applicable. Note that the set of variables in T includes those in V as well as auxiliary and output variables from MF.

Given a NATBN $\Gamma = (V, D, PR, PS, \Psi)$, we define its *peer BN* $\Gamma' = (V, D, CP)$, where CP is a set of CPTs one per node in D. If the node has less than two parents, its CPT is the one in PR or PS. Otherwise, the CPT is defined by the NAT model in Ψ. Theorem 10 establishes exactness of LP when applied to a JTMFNB. It says that after LP is performed in a JTMFNB, the post-LP belief at every cluster, with auxiliary and output variables from MF being marginalized out, is exactly the marginal probability over the cluster.

Theorem 10. *Let* $\Gamma = (V, D, PR, PS, \Psi)$ *be a NATBN,* T *be the JTMFNB from* Γ*,* U *be the set of all variables in* T*, and LP be performed in* T*. Let* C *be any cluster of* T *and* $\alpha(C)$ *be the post-LP cluster belief. Let* $\Gamma' = (V, D, CP)$ *be the peer BN of* Γ *and* $P'(V)$ *be the product of CPTs in* CP*. Then,* $\sum_{U \setminus V} \alpha(C) = const \sum_{V \setminus C} P'(V)$*.*

Proof: By the chain rule of BNs, $P'(V)$ is the JPD of the peer BN Γ'. By the definition of peer BN Γ', $P'(V)$ is also the JPD of the NATBN Γ. Let $P(U)$ be the product of all potentials assigned to clusters in T before LP. By Theorem 9, $\sum_{U \setminus V} P(U) = P'(V)$. Marginalizing variables in $V \setminus C$

from both sides, we get $\sum_{V\setminus C}\sum_{U\setminus V} P(U) = \sum_{V\setminus C} P'(V)$. Switching the order of marginalizations, the lefthand side becomes $\sum_{U\setminus V}(\sum_{V\setminus C} P(U)) = const \sum_{U\setminus V} \alpha(C)$ by Prop. 14. Hence, $\sum_{U\setminus V} \alpha(C) = const \sum_{V\setminus C} P'(V)$. $\square$

8. Experimental Evaluation

A collection of 140 NATBNs are simulated, divided into 4 groups of 35 each. Each NATBN contains 100 ternary variables. For NATBNs in the same group, the number n of causes per NAT model is identically upper-bounded. The bounds are 5, 7, 9 and 11, respectively. All NATBNs have the same density (5% more links than singly-connected). Each is compiled into a JTMFNB.

A *peer BN* is derived from each NATBN, where each multi-parent variable is assigned the CPT computed from the corresponding NAT model. Peer BNs are compiled for lazy inference normally, which provides a golden standard for soundness and a baseline for efficiency.

For each NATBN and its peer BN, 5 randomly chosen variables are observed and posteriors for all variables are computed. For each NATBN, exactly the same posteriors are obtained from the JTMFNB and its peer BN, which empirically confirms soundness of the MF.

The performance is summarized in Table 17. Each row summarizes for one group of NATBNs. The space efficiency of JTMFNB and peer BN is shown by the size of the state space of the JT (with sample mean and standard deviation). The time efficiency is shown by lazy inference time.

Table 17: Experimental Results

	Peer BN State Space		JTMFNB State Space		Peer BN Time (ms)		JTMFNB Time (ms)	
n	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$
5	11070.8	590.1	9742.7	1317.9	63.8	12.0	31.3	0.5
7	25951.4	3800.3	10546.0	1570.3	212.5	65.7	30.3	3.5
9	80061.9	6076.6	11189.8	2721.7	1117.9	749.9	33.1	7.3
11	575750.3	37149.6	10996.1	1550.2	12160.8	7658.3	30.7	2.5

As n grows from 5 to 11, peer BN JTs grow in space by 52 times, while JTMFNBs grow only 1.1 times. The runtime with peer BN JTs grows by 193 times, while lazy inference with JTMFNBs takes about the same time. For $n = 11$, inference with JTMFNBs is about 400 times faster than peer BN JTs. The experiment shows that the MF of NAT models allows significant improvement in space and time efficiency for sparse NAT modeled BNs.

9. Conclusion

The main contributions of this work are the following. We show that a multi-valued, dual NIN-AND gate is equivalent to noisy-MAX, and hence, a multi-valued NAT model is strictly more expressive than noisy-MAX. We developed the MF of multi-valued NAT models. It extends the MF of binary NAT models in that the MF graph of a binary NAT model is a tree while the MF graph of a multi-valued NAT model is multiply connected. The space complexity of the MF $O(n\eta(6m+3^{\eta}))$ is linear on the number of causes n, and the domain size of causes m, and is exponential on the domain size of effect $\eta+1$ with a constant base 3. We proposed a scheme to multiplicatively factorize NATBNs and compile them for lazy inference. This scheme is more powerful than lazy inference based on MF of noisy-MAX [11], since NATBNs are strictly more expressive than noisy-MAX modeled BNs. We experimentally demonstrated that JTMFNBs compiled from sparse NATBNs allow exact lazy inference that is significantly more efficient in both space and time.

This work opens a promising direction along which significantly less computational resource is necessary for probabilistic reasoning with BNs, making them deployable on pervasive computing devices.

Acknowledgement

Financial support through a Discovery Grant from NSERC, Canada to the first author is acknowledged.

References

[1] Y. Xiang, Y. Jin, Multiplicative factorization of multi-valued NIN-AND tree models, in: Z. Markov, I. Russell (Eds.), Proc. 29th Inter. Florida Artificial Intelligence Research Society Conf., AAAI Press, 2016, pp. 680–685.

[2] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, 1988.

[3] S. Galan, F. Diez, Modeling dynamic causal interaction with Bayesian networks: temporal noisy gates, in: Proc. 2nd Inter. Workshop on Causal Networks, 2000, pp. 1–5.

[4] M. Henrion, Some practical issues in constructing belief networks, in: L. Kanal, T. Levitt, J. Lemmer (Eds.), Uncertainty in Artificial Intelligence 3, Elsevier Science Publishers, 1989, pp. 161–173.

[5] J. Lemmer, D. Gossink, Recursive noisy OR - a rule for estimating complex probabilistic interactions, IEEE Trans. on System, Man and Cybernetics, Part B 34 (6) (2004) 2252–2261.

[6] Y. Xiang, N. Jia, Modeling causal reinforcement and undermining for efficient CPT elicitation, IEEE Trans. Knowledge and Data Engineering 19 (12) (2007) 1708–1718.

[7] Y. Xiang, Non-impeding noisy-AND tree causal models over multi-valued variables, International J. Approximate Reasoning 53 (7) (2012) 988–1002.

[8] F. Jensen, S. Lauritzen, K. Olesen, Bayesian updating in causal probabilistic networks by local computations, Computational Statistics Quarterly (4) (1990) 269–282.

[9] N. Zhang, D. Poole, Exploiting causal independence in bayesian network inference, J. Artificial Intelligence Research 5 (1996) 301–328.

[10] M. Takikawa, B. D'Ambrosio, Multiplicative factorization of noisy-max, in: Proc. 15th Conf. Uncertainty in Artificial Intelligence, 1999, pp. 622–630.

[11] A. Madsen, B. D'Ambrosio, A factorized representation of independence of causal influence and lazy propagation, Inter. J. Uncertainty, Fuzziness and Knowledge-Based Systems 8 (2) (2000) 151–166.

[12] P. Savicky, J. Vomlel, Exploiting tensor rank-one decomposition in probabilistic inference, Kybernetika 43 (5) (2007) 747–764.

[13] Y. Xiang, Bayesian network inference with NIN-AND tree models, in: A. Cano, M. Gomez-Olmedo, T. Nielsen (Eds.), Proc. 6th European Workshop on Probabilistic Graphical Models, Granada, 2012, pp. 363–370.

[14] Y. Xiang, Acquisition and computation issues with NIN-AND tree models, in: P. Myllymaki, T. Roos, T. Jaakkola (Eds.), Proc. 5th European Workshop on Probabilistic Graphical Models, Finland, 2010, pp. 281–289.

[15] F. Diez, Parameter adjustment in Bayes networks: The generalized noisy OR-gate, in: D. Heckerman, A. Mamdani (Eds.), Proc. 9th Conf. on Uncertainty in Artificial Intelligence, Morgan Kaufmann, 1993, pp. 99–105.

[16] W. Li, P. Poupart, P. van Beek, Exploiting structure in weighted model counting approaches to probabilistic inference, J. Artificial Intelligence Research 40 (2011) 729–765.

[17] P. Maaskant, M. Druzdzel, An independence of causal interactions model for opposing influences, in: M. Jaeger, T. Nielsen (Eds.), Proc. 4th European Workshop on Probabilistic Graphical Models, Hirtshals, Denmark, 2008, pp. 185–192.

[18] Y. Xiang, J. Zhu, Y. Li, Enumerating unlabeled and root labeled trees for causal model acquisition, in: Y. Gao, N. Japkowicz (Eds.), Advances in Artificial Intelligence, LNAI 5549, Springer, 2009, pp. 158–170.

[19] A. Madsen, F. Jensen, Lazy propagation: A junction tree inference algorithm based on lazy evaluation, Artificial Intelligence 113 (1-2) (1999) 203–245.

[20] G. Shafer, Probabilistic Expert Systems, Society for Industrial and Applied Mathematics, Philadelphia, 1996.