Theoretical Computer Science 410 (2009) 1618–1628

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Holographic algorithms: The power of dimensionality resolved

Jin-Yi Cai [a], Pinyan Lu [b,*]

[a] University of Wisconsin-Madison, United States

[b] Tsinghua University and Microsoft Research Asia, China

ARTICLE INFO

Keywords:
Holographic algorithms
Bases
Dimension

ABSTRACT

Valiant's theory of holographic algorithms is a novel methodology to achieve exponential speed-ups in computation. A fundamental parameter in holographic algorithms is the dimension of the linear basis vectors. We completely resolve the problem of the power of higher dimensional bases. We prove that 2-dimensional bases are universal for holographic algorithms.

1. Introduction

Complexity theory has learned a great deal about the nature of efficient computation. However if the ultimate goal is to gain a fundamental understanding such as what differentiates polynomial time from exponential time, we are still a way off. In fact, in the last 20 years, the most spectacular advances in the field have come from discovering new and surprising ways to do efficient computations. The theory of holographic algorithms introduced recently by Valiant [13] is one such new methodology which gives polynomial time algorithms to some problems which seem to require exponential time.

To describe this theory requires some background. At the top level it is a method to represent computational information in a superposition of linear vectors, somewhat analogous to quantum computing. This information is manipulated algebraically, but in a purely classical way. Then via a beautiful theorem called the Holant Theorem [13], which expresses essentially an invariance of tensor contraction under basis transformations [2], this computation is reduced to the computation of perfect matchings in planar graphs. It so happens that counting perfect matchings for planar graphs is computable in polynomial time by the elegant FKT method [8–10]. Thus we obtain a polynomial time algorithm. The whole exercise can be thought of as an elaborate scheme to introduce a custom made process of exponential cancellations. The end result is a polynomial time evaluation of an exponential sum which expresses the desired computation.

On a more technical level, there are two main ingredients in the design of a holographic algorithm. First, a collection of planar matchgates. Second, a choice of linear basis vectors, through which the computation is expressed and interpreted. Typically there are two basis vectors n and p in dimension 2, which represent the bit values 0 and 1 respectively, and their tensor product will represent a combination of 0-1 bits. It is the superpositions of these vectors in the tensor product space that are manipulated by a holographic algorithm in the computation. This superposition gives rise to exponential sized aggregates with which massive cancellations take place. In this sense holographic algorithms are more akin to quantum algorithms than to classical algorithms in their design and operation.

No polynomial time algorithms were known previously for any of the problems in [13,2,1,14], and some minor variations are NP-hard. These problems may also appear quite restricted. Here is a case in point. Valiant showed [14] that the problem $\#_7$Pl-Rtw-Mon-3CNF is solvable in P by this method. This problem is a restrictive Satisfiability counting problem. Given a planar read-twice monotone 3CNF formula, it counts the number of satisfying assignments, modulo 7. However, it is known

* Corresponding author.
E-mail addresses: jyc@cs.wisc.edu (J.-Y. Cai), lpy@mails.tsinghua.edu.cn (P. Lu).

0304-3975/$ – see front matter
doi:10.1016/j.tcs.2008.12.047

that even for this restricted class of Boolean formulae, the counting problem without the modulo 7 is #P-complete. Also, the counting problem modulo 2 (denoted as $\#_2$Pl-Rtw-Mon-3CNF) is $\oplus$P-complete (thus NP-hard by randomised reductions). The ultimate power of this theory is unclear.

It is then natural to ask, whether holographic algorithms will bring about a collapse of complexity classes. Regarding conjectures such as P $\neq$ NP undogmatically, it is incumbent for us to gain a systematic understanding of the capabilities of holographic algorithms. This brings us closer to the fundamental reason why these algorithms are fascinating—its implication for complexity theory. The fact that some of these problems such as $\#_7$Pl-Rtw-Mon-3CNF might appear a little contrived is beside the point. When potential algorithmic approaches to P vs. NP were surveyed, these algorithms were not part of the repertoire; presumably the same "intuition" for P $\neq$ NP would have applied equally to $\#_7$Pl-Rtw-Mon-3CNF and to $\#_2$Pl-Rtw-Mon-3CNF.

In holographic algorithms, since the underlying computation is ultimately reduced to perfect matchings, the linear basis vectors which express the computation are necessarily of dimension 2^k, for some integer k. This k is called the size of the basis. Most holographic algorithms so far [13,2,1,14] use bases of size 1. Surprisingly Valiant's algorithm for $\#_7$Pl-Rtw-Mon-3CNF used a basis of size 2. Utilising bases of a higher dimension has always been a theoretical possibility, which may further extend the reach of holographic algorithms. Valiant's algorithm makes it realistic.

It turns out that for $\#_7$Pl-Rtw-Mon-3CNF one can design another holographic algorithm with a basis of size 1 [4]. Subsequently we have proved [6] the surprising result that any basis of size 2 can always be replaced by a suitable basis of size 1 in a holographic algorithm. In this paper we completely resolve the problem of whether bases of higher dimensions are more powerful. They are not.

Our starting point is a theorem from [6] concerning degenerate tensors of matchgates. For bases of size 2 we were able to find explicit constructions of certain gadgets from scratch. But this approach encountered major difficulties for arbitrary size k. The underlying reason for this is that for larger matchgates there is a set of exponential sized algebraic constraints called matchgate identities [12,1,3] which control their realisability. This additional constraint is absent for small matchgates. The difficulty is finally overcome by deriving a tensor theoretic decomposition. This reveals an internal structure for non-degenerate matchgate tensors. We discover that for any basis of size k, except in a degenerate case, there is an embedded basis of size 1. To overcome the difficulty of realisability, we make use of the given matchgates on a basis of size k, and "fold" these matchgates onto themselves to get new matchgates on the embedded basis of size 1. These give geometric realizations, by planar graphs, of those tensors in the decomposition which were defined purely algebraically. Thus we are able to completely bypass matchgate identities here. In the process, we gain a substantial understanding of the structure of a general holographic algorithm on a basis of size k.

This paper is organised as follows. In Section 2, we give a brief summary of background information. In Section 3, we give a structural theorem for valid bases, the tensor theoretic decomposition, and prove two key theorems for the realisability of generators. In Section 4, we prove a realisability theorem for recognizers. This leads to the main theorem. In Section 5, we give an overall picture of the landscape of holographic algorithms after the structural understanding from this work.

2. Background

Let $G = (V, E, W)$ be a weighted undirected planar graph. A *generator matchgate* Γ is a tuple (G, X) where $X \subseteq V$ is a set of external *output* nodes. A *recognizer matchgate* Γ' is a tuple (G, Y) where $Y \subseteq V$ is a set of external *input* nodes. The external nodes are ordered counter-clockwise on the external face. Γ (or Γ') is called an odd (resp. even) matchgate if it has an odd (resp. even) number of nodes.

Each matchgate is assigned a *signature* tensor. A generator Γ with n output nodes is assigned a contravariant tensor $\mathbf{G}$ of type $\binom{n}{0}$. Under the standard basis, it takes the form $\underline{G}$ with 2^n entries, where

$$\underline{G}^{i_1 i_2 \ldots i_n} = \text{PerfMatch}(G - Z).$$

Here PerfMatch is the sum of all weighted perfect matchings, and Z is the subset of the output nodes having the characteristic sequence $\chi_Z = i_1 i_2 \ldots i_n$. $\underline{G}$ is called the standard signature of the generator Γ. We can view $\underline{G}$ as a column vector (whose entries are ordered lexicographically according to χ_Z).

Similarly a recognizer $\Gamma' = (G', Y)$ with n input nodes is assigned a covariant tensor $\mathbf{R}$ of type $\binom{0}{n}$. Under the standard basis, it takes the form $\underline{R}$ with 2^n entries, where

$$\underline{R}_{i_1 i_2 \ldots i_n} = \text{PerfMatch}(G' - Z),$$

where Z is the subset of the input nodes having $\chi_Z = i_1 i_2 \ldots i_n$. $\underline{R}$ is called the standard signature of the recogniser Γ'. We can view $\underline{R}$ as a row vector (again with entries ordered lexicographically).

Because of the parity constraint of perfect matchings, half of all entries of a standard signature $\underline{G}$ (or $\underline{R}$) are zero. Therefore, we can use a tensor in V_0^{n-1} (or V_{n-1}^0) to represent all the information contained in $\underline{G}$ (or $\underline{R}$). More precisely, we have the following definition (we only need for the generators).

Definition 2.1. If a generator matchgate $\varGamma$ with arity n is even (resp. odd), a condensed standard signature $\underset{\sim}{G}$ of $\varGamma$ is a tensor in V_0^{n-1}, and $\underset{\sim}{G}^{\alpha} = \underline{G}^{\alpha b}$ (resp. $\underset{\sim}{G}^{\alpha} = \underline{G}^{\alpha \overline{b}}$), where $\underline{G}$ is the standard signature of $\varGamma$, $\alpha \in \{0, 1\}^{n-1}$ and $b = \oplus\alpha$ is the sum of the bits of α mod 2, i.e., the parity of the Hamming weight of α, and $\overline{b}$ denotes flipping the bit b.

We will consider matchgate tensors under a basis transformation. When the basis consists of two vectors of dimension 2, the arity of the matchgate is its number of external nodes. However more generally we will consider a basis T consisting of 2 vectors $(\mathbf{t}_0, \mathbf{t}_1)$ (also denoted as $(\mathbf{n}, \mathbf{p})$), each of dimension 2^k (size k). In this case, we assume the matchgate has kn external nodes, and these external nodes are grouped in blocks of k nodes each, and each block encodes a superposition of the two basis vectors n and p and therefore stands for one bit. In this case, the arity of the matchgate is n. We use the following notation: $T = (t_i^{\alpha}) = [n^{\alpha}, p^{\alpha}]$, where $i \in \{0, 1\}$ and $\alpha \in \{0, 1\}^k$. We follow the convention that upper index α is for row and lower index i is for column (see [7]). We assume $\mathrm{rank}(T) = 2$ in the following discussion because a basis of $\mathrm{rank}(T) \le 1$ is useless. Under a basis T, we can talk about non-standard signatures (or simply signatures).

Definition 2.2. The contravariant tensor $\mathbf{G}$ of a generator $\varGamma$ has signature G under basis T iff $\underline{G} = T^{\otimes n}G$ is the standard signature of the generator $\varGamma$.

Definition 2.3. The covariant tensor $\mathbf{R}$ of a recogniser $\varGamma'$ has signature R under basis T iff $R = \underline{R}T^{\otimes n}$, where $\underline{R}$ is the standard signature of the recogniser $\varGamma'$.

We have

$$\underline{G}^{\alpha_1\alpha_2\cdots\alpha_n} = \sum_{i_1, i_2, \ldots, i_n \in \{0,1\}} G^{i_1 i_2 \cdots i_n} t_{i_1}^{\alpha_1} t_{i_2}^{\alpha_2} \cdots t_{i_n}^{\alpha_n} \text{ (where } \alpha_j \in \{0, 1\}^k, \text{ for } j = 1, 2, \ldots, n). \tag{1}$$

$$R_{i_1 i_2 \cdots i_n} = \sum_{\alpha_1, \alpha_2, \ldots, \alpha_n \in \{0,1\}^k} \underline{R}_{\alpha_1\alpha_2\cdots\alpha_n} t_{i_1}^{\alpha_1} t_{i_2}^{\alpha_2} \cdots t_{i_n}^{\alpha_n} \text{ (where } i_j \in \{0, 1\} \text{ for } j = 1, 2, \ldots, n). \tag{2}$$

Definition 2.4. A contravariant tensor $\mathbf{G} \in V_0^n$ (resp. a covariant tensor $\mathbf{R} \in V_n^0$) is realisable on a basis T iff there exists a generator $\varGamma$ (resp. a recogniser $\varGamma'$) such that G (resp. R) is the signature of $\varGamma$ (resp. $\varGamma'$) under basis T.

A *matchgrid* $\varOmega = (A, B, C)$ is a weighted planar graph consisting of a disjoint union of: a set of g generators $A = (A_1, \ldots, A_g)$, a set of r recognisers $B = (B_1, \ldots, B_r)$, and a set of f connecting edges $C = (C_1, \ldots, C_f)$, where each C_i edge has weight 1 and joins an output node of a generator with an input node of a recogniser, such that every input and output node in every constituent matchgate has exactly one such incident connecting edge.

Let $G(A_i, T)$ be the signature of generator A_i under the basis T and $R(B_j, T)$ be the signature of recogniser B_j under the basis T. And Let $G = \bigotimes_{i=1}^{g} G(A_i, T)$ and $R = \bigotimes_{j=1}^{r} R(B_j, T)$. Then $\mathrm{Holant}(\varOmega)$ is defined to be the contraction of these two product tensors, where the corresponding indices match up according to the f connecting edges in C. We note that for a holographic algorithm to use a basis of size $k > 1$, each matchgate of arity n in the matchgrid has kn external nodes, grouped in blocks of k nodes each. These k nodes are connected in a block-wise fashion between matchgates, where the combinations of tensor products of the 2^k-dimensional basis vectors are interpreted as truth values.

Valiant's Holant Theorem is

Theorem 2.1 (*Valiant*). *For any matchgrid $\varOmega$ over any basis T, let G be its underlying weighted graph, then*

$$\mathrm{Holant}(\varOmega) = \mathrm{PerfMatch}(G).$$

We illustrate these concepts by the problem #Pl-Rtw-Mon-3CNF (counting without mod) from Section 1. Given a planar 3CNF formula φ as a planar graph G_{φ} where variables and clauses are represented by vertices. For each variable x we try to find a generator G with signature $G^{00} = 1, G^{01} = 0, G^{10} = 0, G^{11} = 1$, or $(1, 0, 0, 1)^{\mathrm{T}}$ for short. This is indeed realisable as the standard signature of a matchgate which consists of a path of length 3 and all weights 1. Note that when we remove exactly one of the two external nodes we have 3 vertices left and therefore the value of PerfMatch is 0. If we remove both or none of the two external nodes we get the value 1. We can replace the vertex for x, which is read-twice in the planar formula, by this generator G. This signature $(1, 0, 0, 1)^{\mathrm{T}}$ corresponds to a truth assignment: its outputs will be a consistent assignment of either 0 or 1. We also wish to find a recogniser R with 3 inputs having signature $(0, 1, 1, 1, 1, 1, 1, 1)^{\mathrm{T}}$. This signature corresponds to a Boolean OR. The matchgrid is formed by connecting the generator outputs to the recogniser inputs as given in G_{φ}. If this recogniser exists, we would have shown #Pl-Rtw-Mon-3CNF $\in$ P, and therefore $\mathrm{P}^{\#\mathrm{P}} = \mathrm{P}$. The conclusion is even $\mathrm{P}^{\#\mathrm{P}} = \mathrm{NC}_2$.

It turns out that a recogniser with the *standard* signature $(0, 1, 1, 1, 1, 1, 1, 1)^{\mathrm{T}}$ does not exist. However, under a suitable basis this signature is in fact realisable by a recogniser. Indeed it is simultaneously realisable together with a generator having the signature $(1, 0, 0, 1)^{\mathrm{T}}$, *over the field* $\mathbf{Z}_7$ (but not $\mathbf{Q}$). This gives the surprising result that $\#_7$Pl-Rtw-Mon-3CNF $\in$ P. The basis of size 2 used by Valiant in [14] is $n = (1, 1, 2, 1)^{\mathrm{T}}, p = (2, 3, 6, 2)^{\mathrm{T}}$. Written in this basis, the signature $(1, 0, 0, 1)^{\mathrm{T}}$ stands for $1\mathbf{n} \otimes \mathbf{n} + 0\mathbf{n} \otimes \mathbf{p} + 0\mathbf{p} \otimes \mathbf{n} + 1\mathbf{p} \otimes \mathbf{p}$ which has dimension $4^2 = 16$. The one for $(0, 1, 1, 1, 1, 1, 1, 1)^{\mathrm{T}}$ has dimension $4^3 = 64$. They happen to be realisable by matchgates with 4 and 6 external nodes respectively. The external nodes are grouped in blocks of size 2.

The first crucial insight is to isolate certain degenerate bases.

Definition 2.5. A basis T is *degenerate* iff $t^{\alpha} = (t_0^{\alpha}, t_1^{\alpha}) = 0$ for all $\mathrm{wt}(\alpha)$ even (or for all $\mathrm{wt}(\alpha)$ odd), where $\mathrm{wt}(\alpha)$ is the Hamming weight of α.

Definition 2.6. A generator tensor $G \in V_0^n$ ($\dim(V) = 2$) is *degenerate* iff it has the following form (where $G_i \in V$ is an arity 1 tensor):

$$G = G_1 \otimes G_2 \otimes \cdots \otimes G_n. \tag{3}$$

Degenerate generators can be completely decoupled. A holographic algorithm that uses only degenerate generators has no connections between its various components and hence is essentially trivial.

In [6], we proved the following theorem. The proof uses matchgate identities.

Theorem 2.2. *If a basis T is degenerate and $\mathrm{rank}(T) = 2$, then every generator $G \in V_0^n$ realisable on the basis T is degenerate.*

3. Valid bases

Definition 3.1. A basis T is *valid* iff there exists some non-degenerate generator realisable on T.

Our starting point is a careful study of the structure of high dimensional valid bases. From Theorem 2.2 we have

Corollary 3.1. *A valid basis is non-degenerate.*

Theorem 3.1. *For every valid basis $T = [\mathbf{n}, \mathbf{p}]$, (n^{α}, p^{α}) and (n^{β}, p^{β}) are linearly dependent, for all $\mathrm{wt}(\alpha), \mathrm{wt}(\beta)$ having the same parity.*

Proof. Since $T = [\mathbf{n}, \mathbf{p}]$ is valid, by definition, there exists a non-degenerate generator G which is realisable on T. From Corollary 3.1, we know that $T = [\mathbf{n}, \mathbf{p}]$ is non-degenerate.

Let α_0, β_0 be two arbitrary indices of even weight and α_1, β_1 be two arbitrary indices of odd weight. Let $T_0 = \left[\begin{pmatrix} n^{\alpha_0} \\ n^{\beta_0} \end{pmatrix}, \begin{pmatrix} p^{\alpha_0} \\ p^{\beta_0} \end{pmatrix}\right]$ and $T_1 = \left[\begin{pmatrix} n^{\alpha_1} \\ n^{\beta_1} \end{pmatrix}, \begin{pmatrix} p^{\alpha_1} \\ p^{\beta_1} \end{pmatrix}\right]$. Then we need to prove $\det(T_0) = \det(T_1) = 0$.

For all $i_1, i_2, \ldots, i_n \in \{0, 1\}$, recall that

$$(T_0^{\otimes n} G)^{i_1 i_2 \cdots i_n} = G^{\gamma_1 \gamma_2 \cdots \gamma_n}, \text{ where } \gamma_j = \begin{cases} \alpha_0, & i_j = 0, \\ \beta_0, & i_j = 1. \end{cases}$$

Similarly for T_1,

$$(T_1^{\otimes n} G)^{i_1 i_2 \cdots i_n} = G^{\gamma_1 \gamma_2 \cdots \gamma_n}, \text{ where } \gamma_j = \begin{cases} \alpha_1, & i_j = 0, \\ \beta_1, & i_j = 1. \end{cases}$$

According to the parity of the arity n and the parity of the matchgate realising G, we have 4 cases:

Case 1: even n and odd matchgate

From the parity constraint, we have $T_0^{\otimes n} G = 0$ and $T_1^{\otimes n} G = 0$. Since $G \not\equiv 0$ (i.e., G is not identically 0), we have $\det(T_0) = \det(T_1) = 0$. Note that $\det(T^{\otimes n}) = (\det(T))^{n2^{n-1}}$.

Case 2: odd n and odd matchgate

From the parity constraint, we have $T_0^{\otimes n} G = 0$. Since $G \not\equiv 0$, we have $\det(T_0) = 0$. Since the basis is non-degenerate, from the definition, there exists an α such that $\mathrm{wt}(\alpha)$ is even and $(n^{\alpha}, p^{\alpha}) \neq (0, 0)$.

From the parity constraint, for all $t \in [n] = \{1, \ldots, n\}$, we have

$$(T_1^{\otimes(t-1)} \otimes (n^{\alpha}, p^{\alpha}) \otimes T_1^{\otimes(n-t)}) G = 0. \tag{4}$$

Let G_t be the tensor of type V_0^{n-1} defined by

$$G_t^{i_1 i_2 \ldots i_{n-1}} = n^{\alpha} G^{i_1 i_2 \ldots i_{t-1} 0 i_t i_{t+1} \ldots i_{n-1}} + p^{\alpha} G^{i_1 i_2 \ldots i_{t-1} 1 i_t i_{t+1} \ldots i_{n-1}},$$

where $i_1, i_2, \ldots, i_{n-1} = 0, 1$. Then Eq. (4) translates to $T_1^{\otimes(n-1)} G_t = 0$.

If $\forall t \in [n]$ we have $G_t \equiv 0$, then we claim G is symmetric and degenerate. To see this, first suppose $p^{\alpha} \neq 0$. Then for all $i_1, i_2, \ldots, i_n = 0, 1$, $G^{i_1 i_2 \ldots i_n} = G^{00\ldots 0}(-n^{\alpha}/p^{\alpha})^{\mathrm{wt}(i_1 i_2 \ldots i_n)}$. This is clearly symmetric, and degenerate by (3). The proof is similar if $n^{\alpha} \neq 0$. Since by assumption $(n^{\alpha}, p^{\alpha}) \neq (0, 0)$, it follows that G is degenerate. This is a contradiction.

Therefore there exists some $t \in [n]$ such that $G_t \not\equiv 0$. Then from $T_1^{\otimes(n-1)} G_t = 0$, we have $\det(T_1) = 0$.

Case 3: odd n and even matchgate

This is similar to Case 2. We apply the argument for T_0 to T_1, and apply the argument for T_1 to T_0.

Case 4: even n and even matchgate

This case is also similar to Case 2 and Case 3. We simply apply the same argument for T_1 as in Case 2 and the same argument for T_0 as in Case 3. $\square$

From this theorem, we know that for any valid basis $T = [n^\alpha, p^\alpha]$ (where $\alpha \in \{0,1\}^k$), there exist non-zero vectors $(n^{\alpha_0}, p^{\alpha_0})$, and $(n^{\alpha_1}, p^{\alpha_1})$, where $\alpha_0, \alpha_1 \in \{0,1\}^k$, and $\mathrm{wt}(\alpha_0)$ is even and $\mathrm{wt}(\alpha_1)$ is odd, such that every other (n^α, p^α) is a scalar multiple of one of these two vectors (the one with the same parity). More precisely, we define $\widehat{n}^b = n^{\alpha_b}$ and $\widehat{p}^b = p^{\alpha_b}$ for $b = 0, 1$, then there exist λ^α for all $\alpha \in \{0,1\}^k$, such that $(n^\alpha, p^\alpha) = \lambda^\alpha(\widehat{n}^{\oplus\alpha}, \widehat{p}^{\oplus\alpha})$, where $\oplus\alpha$ is the parity of $\mathrm{wt}(\alpha)$.

Note that $(\widehat{n}^0, \widehat{p}^0)$, $(\widehat{n}^1, \widehat{p}^1)$ are linearly independent, otherwise $\mathrm{rank}(T) < 2$. Therefore each is determined up to a scalar multiplier. This justifies the following definition:

Definition 3.2. We call $\widehat{T} = \left[\begin{pmatrix}\widehat{n}^0\\ \widehat{n}^1\end{pmatrix}, \begin{pmatrix}\widehat{p}^0\\ \widehat{p}^1\end{pmatrix}\right]$ an embedded size 1 basis of T.

Now suppose a non-degenerate generator G is realisable on a valid basis $T = [n^\alpha, p^\alpha]$, (where $\alpha \in \{0,1\}^k$), and $\widehat{T} = (\widehat{t}_i^\alpha)$ is an embedded size 1 basis of T.

Substituting $(t_0^\alpha, t_1^\alpha) = \lambda^\alpha(\widehat{t}_0^{\oplus\alpha}, \widehat{t}_1^{\oplus\alpha})$ in (1), we have

$$\begin{aligned}\underline{G}^{\alpha_1\alpha_2\cdots\alpha_n} &= \sum_{i_1,i_2,\cdots,i_n\in\{0,1\}} G^{i_1i_2\cdots i_n} t_{i_1}^{\alpha_1} t_{i_2}^{\alpha_2}\cdots t_{i_n}^{\alpha_n}\\ &= \sum_{i_1,i_2,\cdots,i_n\in\{0,1\}} G^{i_1i_2\cdots i_n}\lambda^{\alpha_1}\widehat{t}_{i_1}^{\oplus\alpha_1}\lambda^{\alpha_2}\widehat{t}_{i_2}^{\oplus\alpha_2}\cdots\lambda^{\alpha_n}\widehat{t}_{i_n}^{\oplus\alpha_n}\\ &= \lambda^{\alpha_1}\lambda^{\alpha_2}\cdots\lambda^{\alpha_n}\sum_{i_1,i_2,\cdots,i_n\in\{0,1\}} G^{i_1i_2\cdots i_n}\widehat{t}_{i_1}^{\oplus\alpha_1}\widehat{t}_{i_2}^{\oplus\alpha_2}\cdots\widehat{t}_{i_n}^{\oplus\alpha_n}.\end{aligned}$$

We define a tensor $\widehat{G} \in V_0^n$ as follows: For $j_1, j_2, \ldots, j_n \in \{0,1\}$,

$$\widehat{G}^{j_1j_2\cdots j_n} = \sum_{i_1,i_2,\cdots,i_n\in\{0,1\}} G^{i_1i_2\cdots i_n}\widehat{t}_{i_1}^{j_1}\widehat{t}_{i_2}^{j_2}\cdots\widehat{t}_{i_n}^{j_n}. \tag{5}$$

Then we have

$$\underline{G}^{\alpha_1\alpha_2\cdots\alpha_n} = \lambda^{\alpha_1}\lambda^{\alpha_2}\cdots\lambda^{\alpha_n}\widehat{G}^{\oplus\alpha_1\oplus\alpha_2\cdots\oplus\alpha_n}. \tag{6}$$

Starting with any non-degenerate G which is realisable on a valid basis T, we defined its embedded size 1 basis $\widehat{T}$, (λ^α) and $\widehat{G}$ by (5). But we note that (5) and (6) are satisfied for every generator (we only need one non-degenerate G to establish $\widehat{T}$). Then regarding (6) we have the following key theorems:

Theorem 3.2. *(λ^α) (where $\alpha \in \{0,1\}^k$) is a condensed signature of some generator matchgate with arity $k+1$.*

Theorem 3.3. *$\widehat{G}$ is a standard signature of some generator matchgate with arity n.*

Put Theorems 3.1–3.3 together, we have both a necessary and sufficient condition for a basis to be valid.

The proofs of Theorems 3.2 and 3.3 are both constructive. We make one more definition. Since the basis T is non-degenerate, there exist β_0 and β_1, such that $\mathrm{wt}(\beta_0)$ is even, $\mathrm{wt}(\beta_1)$ is odd, and $\lambda^{\beta_0}\lambda^{\beta_1} \neq 0$. We also assume β_0 and β_1 is such a pair with minimum Hamming distance. To simplify notations in the following proof, we assume $\beta_0 = 00\cdots0$ and $\beta_1 = 11\cdots100\cdots0$ (where there are a 1s, a is odd). This simplifying assumption is without loss of generality; see the remarks after the proof.

Let $c_0 = \lambda^{\beta_0} = \lambda^{00\cdots000\cdots0}$ and $c_1 = \lambda^{\beta_1} = \lambda^{11\cdots100\cdots0}$. In this setting, for any pattern γ strictly between β_0 and β_1 (if any), if $\alpha_r = \gamma$ for some $r \in [n]$, then by (6)

$$\underline{G}^{\alpha_1\alpha_2\cdots\alpha_n} = 0. \tag{7}$$

Since G is realisable on T, $\underline{G}$ is the standard signature of some matchgate Γ with arity nk. For convenience, we label its $((i-1)k+j)$-th external node by a pair of integers (i,j), (where $i \in [n], j \in [k]$) (see Fig. 1.) Our constructions for Theorems 3.2 and 3.3 both start from Γ. In Fig. 1, we omit all internal structures of Γ (edges and internal nodes). We use dashed rectangle to group a block of k external nodes and the following modifications will be done block-wise. But note that these dashed rectangles are not necessarily separate parts geometrically. The modifications preserve planarity because these external nodes are all in the outer face and in the given order.

Proof of Theorem 3.3. For every $i \in [n]$, do the following modifications to the k nodes (i,j) of the i-th block of external nodes in Γ, where $j \in [k]$ (see Fig. 3):

- Connect (i,l) with $(i,l+1)$ by an edge of weight 1, for $l = 2, 4, \ldots, a-1$.
- Add two new nodes i' and i''.
- Connect $(i,1)$ and i'' by an edge of weight $1/c_1$.
- Connect i'' and i' by an edge of weight $1/c_0$.

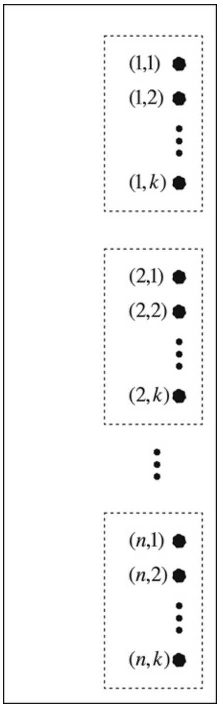

Fig. 1. Generator Matchgate Γ. We omit all the internal structures. All the kn external nodes are labelled by a pair of integers and they are all on the outer face.

After all these modifications, viewing the n nodes i' (one node stemming from each block, $i \in [n]$) as external nodes and all other nodes as internal nodes, we have a matchgate $\widehat{\Gamma}$ with arity n. Now we prove that $\widehat{G}$ is the standard signature of this matchgate $\widehat{\Gamma}$.

Denote the standard signature of $\widehat{\Gamma}$ temporarily as $(\widehat{\Gamma}^{j_1 j_2 \cdots j_n})$. For an arbitrary pattern $j_1 j_2 \cdots j_n \in \{0, 1\}^n$, we compute the value $\widehat{\Gamma}^{j_1 j_2 \cdots j_n}$. For $r \in [n]$, there are two cases:

- Case 1: $j_r = 0$. In this case, we keep the external node r'. Any perfect matching will take the edge (r'', r'), this contributes a factor of $1/c_0$. As a result, the node $(r, 1)$ must match with some node in the original Γ. And from (7), the only possible non-zero pattern of this block of $\underline{G}$ is $\beta_0 = 00 \ldots 0$. (This means that the perfect matchings will not take any of the new weight 1 edges.)
- Case 2: $j_r = 1$. In this case, we remove the external node r'. Any perfect matching will take the edge between $(r, 1)$ and r'', this contributes a factor of $1/c_1$. As a result, the node $(r, 1)$ will be removed from the original Γ. And from (7), the only possible non-zero pattern of this block of $\underline{G}$ is β_1. (This means that the perfect matchings will take all of the new weight 1 edges.)

To sum up,

$$\widehat{\Gamma}^{j_1 j_2 \cdots j_n} = \frac{1}{c_{j_1}} \frac{1}{c_{j_2}} \cdots \frac{1}{c_{j_n}} \underline{G}^{\beta_{j_1} \beta_{j_2} \cdots \beta_{j_n}}.$$

Together with (6), we know this is exactly $\widehat{G}$. This completes the proof. $\square$

Remark: Now we justify the simplifying assumption regarding the forms of β_0 and β_1. One can always add an extra edge at an external node to flip the bit from 1 to 0 to "move" β_0 to the all 0 vector. Also if the 1s in β_1 are not at the first a bit positions, the proof can still go through in the same way, except in the Figures we need to connect $(a-1)/2$ pairs of external nodes where the bit 1 occurs in a planar fashion. This can be done easily from top to bottom, two nodes at a time. As the remaining external nodes of the original matchgate Γ are no longer considered external nodes, the fact that they may no longer be placed on the outer face of the planar embedding of the matchgate constitutes no difficulty.

Before we prove Theorem 3.2, we have the following claim.

Claim 1. *For any standard signature with more than one non-zero entries, there exist two non-zero entries G^{α} and G^{β} such that the Hamming distance between α and β is 2.*

This claim follows easily from the equivalence theorems of planar matchgate signatures and general matchgate characters [1,3]. Basically, by flipping bits we may assume one of the non-zero entry is at $11\ldots1$. The bit flippings preserve Hamming distances. Then it was proved in [1,3] that in this case, a planar matchgate signature can be realised by the Pfaffians of various submatrices of a skew-symmetric matrix of a weighted (not necessarily planar) graph. This graph serves as a (not necessarily planar) matchgate whose character [11], which is defined by Pfaffians, is equal to the signature of the planar matchgate. By normalising one non-zero entry at $11\ldots1$, this signature entry corresponds to the 0-order Pfaffian. Then another signature entry being non-zero implies that there is some submatrix with a non-zero Pfaffian, which implies that the matrix is non-zero. matrix entry is equal to a 2×2 Pfaffian, which corresponds to non-zero signature entry with Hamming weight $n-2$.

Proof of Theorem 3.2. Since G is non-degenerate, at least two $\widehat{G}$'s do not vanish. By Claim 1, for notational simplicity we assume $G_0=\widehat{G}^{00j_3j_4\cdots j_n}\neq 0$ and $G_1=\widehat{G}^{11j_3j_4\cdots j_n}\neq 0$. Other cases can be proved similarly. We are given the planar matchgate Γ with standard signature $\underline{G}$. We carry out the following transformations of Γ:

- Do nothing to the first block. However, for convenience, we rename the first k nodes as $1', 2', \ldots, k'$.
- Change the second block as in Fig. 4, where $g_0=G_0\lambda^{\beta_0}\lambda^{\beta_{j_3}}\cdots\lambda^{\beta_{j_n}}$ and $g_1=G_1\lambda^{\beta_1}\lambda^{\beta_{j_3}}\cdots\lambda^{\beta_{j_n}}$. Note that $g_0, g_1\neq 0$. It has a new external node $(k+1)'$.
- For $i\geq 3$ and $j_i=0$, do nothing to the i-th block.
- For $i\geq 3$ and $j_i=1$, change the i-th block as in Fig. 5.

After all these changes, we will consider the $k+1$ nodes i' (where $i\in[k+1]$, the first k nodes all stem from the first block, and $(k+1)'$ stems from the second block) as the new external nodes and all other nodes as internal nodes. In this way we obtain a planar matchgate Γ_λ with arity $k+1$. Now we prove that λ^α is the condensed standard signature of Γ_λ.

First we show that Γ_λ is an even matchgate. Denote the standard signature of Γ_λ with $(\Gamma_\lambda^\alpha)_\alpha$. Let x be the number of nodes in Γ and $y=\mathrm{wt}(j_3j_4\cdots j_n)$. Since

$$\underline{G}^{\beta_0\beta_0\beta_{j_3}\beta_{j_4}\cdots\beta_{j_n}}=\lambda^{\beta_0}\lambda^{\beta_0}\lambda^{\beta_{j_3}}\lambda^{\beta_{j_4}}\cdots\lambda^{\beta_{j_n}}\widehat{G}^{00j_3\cdots j_n}\neq 0,$$

we know $x-ya$ is even. Given that a is odd, we can count mod 2, and get $x+y+2\equiv x-ya\equiv 0 \bmod 2$. Since $x+y+2$ is exactly the number of nodes in Γ_λ, we know Γ_λ is an even matchgate.

For $\alpha\in\{0,1\}^k$ and $\mathrm{wt}(\alpha)$ is even, we consider $\Gamma_\lambda^{\alpha 0}$ at the $(k+1)$-bit pattern $\alpha 0$. Consider each block in turn in Γ. The first block clearly should be given the k-bit pattern α. The only possible non-zero value concerning the second block is to take the edge $(2'', (k+1)')$ with weight $1/g_0$, and assign the all-0 pattern β_0 to $(2,1), (2,2), \ldots, (2,k)$. This follows from (7). Similarly for the i-th block, where $i\geq 3$, we must assign the pattern β_{j_i}. Hence, applying (6) we get,

$$\Gamma_\lambda^{\alpha 0}=\frac{1}{g_0}\underline{G}^{\alpha\beta_0\beta_{j_3}\beta_{j_4}\cdots\beta_{j_n}}=\frac{1}{g_0}\lambda^\alpha\lambda^{\beta_0}\lambda^{\beta_{j_3}}\lambda^{\beta_{j_4}}\cdots\lambda^{\beta_{j_n}}G_0=\lambda^\alpha.$$

Similarly, for $\alpha\in\{0,1\}^k$ and $\mathrm{wt}(\alpha)$ is odd,

$$\Gamma_\lambda^{\alpha 1}=\frac{1}{g_1}\underline{G}^{\alpha\beta_1\beta_{j_3}\beta_{j_4}\cdots\beta_{j_n}}=\frac{1}{g_1}\lambda^\alpha\lambda^{\beta_1}\lambda^{\beta_{j_3}}\lambda^{\beta_{j_4}}\cdots\lambda^{\beta_{j_n}}G_1=\lambda^\alpha.$$

This completes the proof. $\square$

4. Collapse theorem

By (5) and Theorem 3.3, we have

Theorem 4.1. *If a generator is realisable on a valid basis T, then it is also realisable on its embedded size 1 basis $\widehat{T}$.*

Now we prove the collapse result on the recogniser side.

Theorem 4.2. *If a recogniser is realisable on a valid basis T, then it is also realisable on its embedded size 1 basis $\widehat{T}$.*

Proof. Since T is a valid basis, from Section 3, we have its embedded size 1 basis $\widehat{T}$, and the tensor (λ^α). By the proof of Theorem 3.2 we have an even matchgate Γ_λ whose condensed signature is λ^α.

Let Γ' be a matchgate realising $\underline{R}$, where $R=\underline{R}T^{\otimes n}$. Γ' has kn external nodes (see Fig. 2).

For every block of k nodes in Γ', we use the matchgate Γ_λ from Section 3 to extend Γ' to get a new matchgate $\widehat{\Gamma'}$ of arity n (see Fig. 6).

The idea is that, for each block of k external nodes in Γ', we take one copy of Γ_λ and fold it around so that in a planar fashion its first k external nodes are connected to the k external nodes in Γ' in this block. The $(k+1)$-st external node of this copy of Γ_λ becomes a new external node of $\widehat{\Gamma'}$. Altogether $\widehat{\Gamma'}$ has n external nodes $1^*, 2^*, \ldots, n^*$.

Since Γ_λ is an even matchgate, when the node i^* is either left in (set to 0) or taken out (set to 1), the only possible non-zero patterns within the i-th copy of Γ_λ are all $\alpha_i\in\{0,1\}^k$ with the same parity.

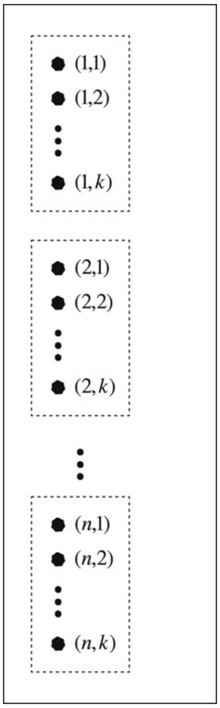

Fig. 2. Recogniser matchgate Γ'. We omit all the internal structures. All the kn external nodes are labelled by a pair of integers and they are all on the outer face.

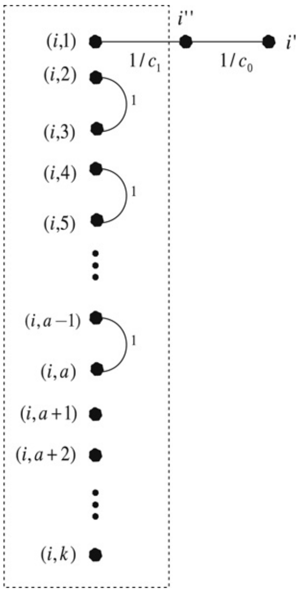

Fig. 3. Modify the i-th block of Γ to get the i-th external node of $\widehat{\Gamma}$.

It follows that the following exponential sum holds, for all $i_1, i_2, \dots, i_n \in \{0, 1\}$:

$$\widehat{\underline{R}}_{i_1 i_2 \dots i_n} = \sum_{\oplus \alpha_r = i_r} \underline{R}_{\alpha_1 \alpha_2 \cdots \alpha_n} \lambda^{\alpha_1} \lambda^{\alpha_2} \cdots \lambda^{\alpha_n}.$$

where $\widehat{\underline{R}}$ is the standard signature of $\widehat{\Gamma}'$, and $\underline{R}$ is the standard signature of Γ'.

We want to prove that $\widehat{\underline{R}}$ in the basis $\widehat{T} = (\widehat{t}_l^j) = \left[\begin{pmatrix} \widehat{n}^0 \\ \widehat{n}^1 \end{pmatrix}, \begin{pmatrix} \widehat{p}^0 \\ \widehat{p}^1 \end{pmatrix} \right]$ and $\underline{R}$ in the basis $T = (t_l^\alpha)$ give the same recognizer R. Recall that $t_l^\alpha = \lambda^\alpha \widehat{t}_l^{\oplus \alpha}$. Now from (2) we have

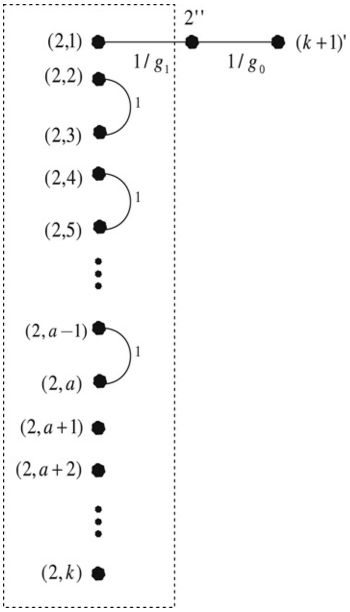

Fig. 4. Modify the second block of Γ to get the $(k+1)$-th external node of Γ_λ.

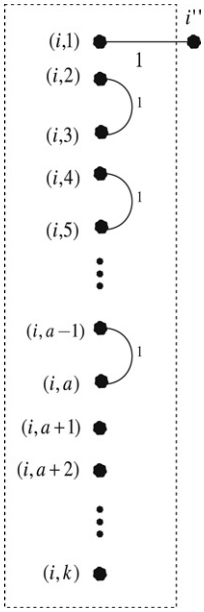

Fig. 5. Modify the i-th block of Γ when $j_i = 1$. All the nodes are viewed as internal in Γ_λ.

$$
\begin{aligned}
R_{l_1 l_2 \cdots l_n} &= \sum_{\alpha_r \in \{0,1\}^k} \underline{R}_{\alpha_1 \alpha_2 \cdots \alpha_n} t_{l_1}^{\alpha_1} t_{l_2}^{\alpha_2} \cdots t_{l_n}^{\alpha_n} \\
&= \sum_{i_r \in \{0,1\}} \sum_{\oplus \alpha_r = i_r} \underline{R}_{\alpha_1 \alpha_2 \cdots \alpha_n} t_{l_1}^{\alpha_1} t_{l_2}^{\alpha_2} \cdots t_{l_n}^{\alpha_n} \\
&= \sum_{i_r \in \{0,1\}} \sum_{\oplus \alpha_r = i_r} \underline{R}_{\alpha_1 \alpha_2 \cdots \alpha_n} \lambda^{\alpha_1} \widehat{t}_{l_1}^{\oplus \alpha_1} \lambda^{\alpha_2} \widehat{t}_{l_2}^{\oplus \alpha_2} \cdots \lambda^{\alpha_n} \widehat{t}_{l_n}^{\oplus \alpha_n} \\
&= \sum_{i_r \in \{0,1\}} \widehat{t}_{l_1}^{i_1} \widehat{t}_{l_2}^{i_2} \cdots \widehat{t}_{l_n}^{i_n} \sum_{\oplus \alpha_r = i_r} \underline{R}_{\alpha_1 \alpha_2 \cdots \alpha_n} \lambda^{\alpha_1} \lambda^{\alpha_2} \cdots \lambda^{\alpha_n} \\
&= \sum_{i_r \in \{0,1\}} \widehat{t}_{l_1}^{i_1} \widehat{t}_{l_2}^{i_2} \cdots \widehat{t}_{l_n}^{i_n} \widehat{\underline{R}}_{i_1 i_2 \cdots i_n}.
\end{aligned}
$$

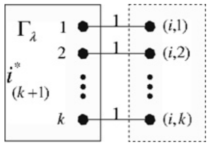

Fig. 6. Extend the i-th block of recogniser Γ' by a copy of Γ_λ. We rename the $(k+1)$-th node of this copy of Γ_λ as i^*, which is the i-th external node of the new recogniser $\widehat{\Gamma'}$.

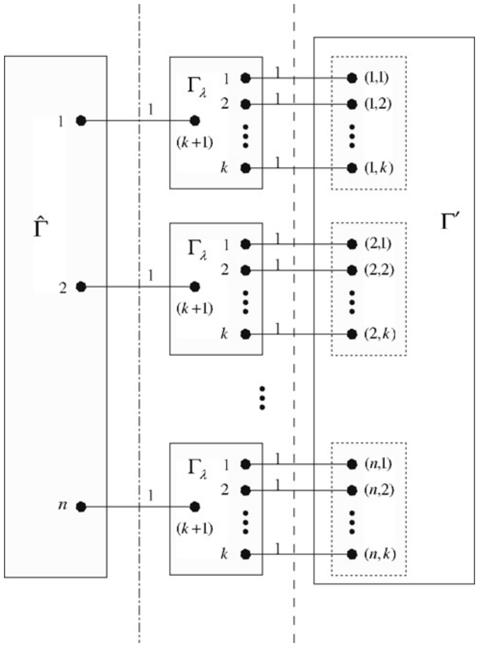

Fig. 7. This figure gives an overall picture of our collapse result. When separate the graph from the dashed line (— — —), we have the original generator Γ (left) and recogniser Γ' (right) in a size k basis. When separate the graph from the dashdotted line (— · — · —) we have the new generator $\widehat{\Gamma}$ (left) and recogniser $\widehat{\Gamma'}$ (right) in a size 1 basis.

The last equation shows that R is also the signature of $\widehat{\Gamma'}$ under basis $\widehat{T}$. This completes the proof. □

Together from Theorems 4.1 and 4.2, we have the following main theorem:

Theorem 4.3 (*Bases Collapse Theorem*). *Any holographic algorithm on a basis of any size which employs at least one non-degenerate generator can be efficiently transformed to a holographic algorithm in a basis of size* 1. *More precisely, if generators $G_1, G_2, \ldots, G_s$ and recognisers $R_1, R_2, \ldots, R_t$ are simultaneously realisable on a basis T of any size, and not all generators are degenerate, then all the generators and recognisers are simultaneously realisable on a basis $\widehat{T}$ of size* 1, *which is the embedded basis of T.*

Proof. Suppose generators $G_1, G_2, \ldots, G_s$ and recognisers $R_1, R_2, \ldots, R_t$ are simultaneously realisable on the size k basis T. Since some G_i is not degenerate, we know that T is valid. Let $\widehat{T}$ be the embedded size 1 basis of T. From Theorem 4.1, all the generators $G_1, G_2, \ldots, G_s$ are realisable on $\widehat{T}$. From Theorem 4.2, all the recognisers $R_1, R_2, \ldots, R_t$ are also realisable on $\widehat{T}$. This completes the proof. □

We remark that a holographic algorithm which only uses degenerate generators is trivial.

5. Conclusion and discussion

In this section, we give an overall picture of our collapse theorem. The decomposition (6) is pregnant with structural information. In Theorems 3.2 and 3.3, we modified the original generator matchgate Γ to obtain $\widehat{\Gamma}$ and Γ_λ respectively. These are the geometric realisations of the individual components in (6). The information of each generator Γ is now contained in $\widehat{\Gamma}$. If we extend every external node of $\widehat{\Gamma}$ by a copy of Γ_λ to encompass everything to the left of the dashed line in Fig. 7, and view the remaining k external nodes of each copy of Γ_λ as external (overall we have nk external nodes), we will have a matchgate with exactly the same signature as the original Γ. Therefore we used $n+1$ copies of the modified Γ to reconstruct a functionally equivalent Γ. It may be a little more complicated than the original one, but it has a clear structure.

When we connect to the recogniser Γ' as in Fig. 7 (we only draw one generator and one recogniser), we can compute the Holant across the interface represented by the dashed line. This is functionally equivalent to the original matchgrid. In the size k basis T, the generator Γ and the recogniser Γ' have signatures G and R, which have some combinatorial interpretations. Instead the new matchgrid computes the Holant across the interface represented by the dashdotted line. We view all the Γ_λ's as part of recognisers rather than generators. Note that every generator undergoes the same transformation. The embedded basis $\widehat{T}$ is defined from T, and (λ^α) is the same for every generator (we only need one non-degenerate generator to prove the existence of $\widehat{T}$ and define (λ^α) and Γ_λ).

The new recognisers $\widehat{\Gamma'}$ are constructed by "folding" copies of Γ_λ and then connecting to the given recognisers Γ'. This is done in Theorem 4.2. After that we can compute the Holant in the interface represented by the dashdotted line, where every bundle has only one edge. The value of the Holant will not change by the Holant Theorem. More importantly, each new generator $\widehat{\Gamma}$ and recogniser $\widehat{\Gamma'}$ in the size 1 basis $\widehat{T}$ will also have the same signatures G and R respectively, which preserve the original combinatorial interpretations.

By our construction, the size of each new matchgate will increase by at most a factor of $n + 1$. Actually the new overall matchgrid may have smaller size because they have fewer external nodes. This follows from the general realisability theorems of [1,3]. More importantly, our result shows that what can be computed in P-time by holographic algorithms in arbitrary dimensional bases can also be done with bases of size 1. This rules out infinitely many theoretical possibilities. Regarding holographic algorithms over size 1 basis, we have already built a substantial theory, e.g., a polynomial time decision procedure for the realisability question of desired signatures [5]. Therefore we believe the resolution of the power of arbitrary bases is an important step towards the understanding of the ultimate capability of holographic algorithms.

References

[1] J.-Y. Cai, Vinay Choudhary, Some results on matchgates and holographic algorithms, in: Proceedings of ICALP 2006, Part I, in: Lecture Notes in Computer Science, vol. 4051, 2006, pp. 703–714. Also available at Electronic Colloquium on Computational Complexity TR06-048.
[2] J-Y. Cai, Vinay Choudhary, Valiant's holant theorem and matchgate tensors (Extended Abstract), in: Proceedings of TAMC 2006, in: Lecture Notes in Computer Science, vol. 3959, 2006, pp. 248–261. Also available at Electronic Colloquium on Computational Complexity Report TR05-118.
[3] J-Y. Cai, Vinay Choudhary, Pinyan Lu, On the theory of matchgate computations. in: The proceedings of IEEE Conference on Computational Complexity 2007, pp 305–318.
[4] J-Y. Cai, Pinyan Lu, On symmetric signatures in holographic algorithms, in: Proceedings of STACS 2007, in: LNCS, vol. 4393, 2007, pp. 429–440. Also available at Electronic Colloquium on Computational Complexity Report TR06-135.
[5] J-Y. Cai, Pinyan Lu, Holographic algorithms: From art to science. in: Proceedings of STOC 2007, pp 401–410.
[6] J-Y. Cai, Pinyan Lu, Bases collapse in holographic algorithms. in: Proceedings of IEEE Conference on Computational Complexity 2007, pp 292–304.
[7] C.T.J. Dodson, T. Poston, Tensor Geometry, second ed., in: Graduate Texts in Mathematics, vol. 130, Springer-Verlag, New York, 1991.
[8] P.W. Kasteleyn, The statistics of dimers on a lattice, Physica 27 (1961) 1209–1225.
[9] P.W. Kasteleyn, Graph theory and crystal physics, in: F. Harary (Ed.), Graph Theory and Theoretical Physics, Academic Press, London, 1967, pp. 43–110.
[10] H.N.V. Temperley, M.E. Fisher, Dimer problem in statistical mechanics — An exact result, Philosophical Magazine 6 (1961) 1061–1063.
[11] L.G. Valiant, Quantum circuits that can be simulated classically in polynomial time, SIAM J. Comput. 31 (4) (2002) 1229–1254.
[12] L.G. Valiant, Expressiveness of matchgates, Theoret. Comput. Sci. 281 (1) (2002) 457–471.
[13] L.G. Valiant, Holographic algorithms (Extended Abstract), in: Proc. 45th IEEE Symposium on Foundations of Computer Science, 2004, pp. 306–315. A more detailed version appeared in Electronic Colloquium on Computational Complexity Report TR05-099.
[14] L.G. Valiant, Accidental algorithms. in: Proc. 47th Annual IEEE Symposium on Foundations of Computer Science, 2006, pp. 509–517.