

Dipole Codes Attractively Encode Glue Functions

Dhananjay Ipparthy* Massimo Mastrangeli[†]
Andrew Winslow[‡]

Abstract

Dipole words are sequences of magnetic dipoles, in which alike elements repel and opposite elements attract. Magnetic dipoles contrast with more general sets of bonding types, called *glues*, in which pairwise bonding strength is specified by a *glue function*. We prove that every glue function g has a set of dipole words, called a *dipole code*, that *attractively encodes* g : the pairwise attractions (positive or non-positive bond strength) between the words are identical to those of g . Moreover, we give such word sets of asymptotically optimal length. Similar results are obtained for a commonly used subclass of glue functions.

1 Introduction

Self-assembly is the autonomous organization of components into structures without supervision [1]. Here we consider controlling self-assembly using fixed arrangements of magnetic dipoles, specifically leveraging their ability to attract and repel according to their spatial configurations. Some

*IRIDIA, CoDE, Université libre de Bruxelles, Brussels, Belgium.
dhananjay.ipparthi@ulb.ac.be

[†]Max Planck Institute for Intelligent Systems, Stuttgart, Germany.
mastrangeli@is.mpg.de

[‡]Département d'Informatique, Université libre de Bruxelles, Brussels, Belgium.
awinslow@ulb.ac.be

previous microscale [2, 3, 4] and mesoscale [5, 6, 7] self-assembling systems have used the capillary effects of surface tension as alternative bonding mechanisms. More recently, molecular recognition of mesoscale components via surface chemistries has also been used [8, 9, 10]. However, magnets are among the most common sources of interaction force in micro- and mesoscale assembly, and they are used in both *active* components that change bonding behavior [11, 12, 13] and in *passive* components whose behavior is fixed [14, 15, 16, 17].

A primary limitation of dipole-based bonding is the limited number of interactions between dipoles: alike poles repel, while opposite poles attract. Frameworks by Bhalla et al. [18, 14] and Majumder and Reif [19] describe an approach for obtaining more complex behaviors by arranging sequences of dipoles along the boundaries of components. These *dipole codes* are used to obtain many distinct bonding sites, called *glues*, that interact only with a unique complementary glue. Bhalla et al. [18, 14, 20] have demonstrated that dipole codes also work experimentally.

The use of *DNA codes* is ubiquitous in DNA-based nanoscale self-assembly, where sequences of repeating nucleotides from the alphabet $\{\mathbf{A}, \mathbf{T}, \mathbf{C}, \mathbf{G}\}$ have been used experimentally to form dozens [21, 22, 23] or even hundreds [24, 25] of glues. Their theoretical study also is extensive, as seen in several surveys [26, 27, 28, 29, 30]. However, there are several practical aspects that differentiate the design of DNA codes from dipole codes.

For instance, the elasticity of DNA requires that codes must disallow multiple portions of a single code to bond [31, 32], while chemistry requires that the codes must have balanced occurrences of letters \mathbf{A}, \mathbf{T} and \mathbf{G}, \mathbf{C} [33, 34, 35]. On the other hand, dipole codes have only a single pair of bonding letters, and inconsistency of mesoscale mixing allows codes to bond with even a single dipole pair [14].

Theoretical work on DNA-based systems has also demonstrated that the addition of a repelling force to systems with many glues increases computational power [36, 37] and efficiency [38, 39]. Thus the construction of large numbers of glues with magnetic dipole sequences gives access to yet more techniques for controlling assembly.

Our contribution The frameworks of both Bhalla et al. [18, 14] and Majumder and Reif [19] have several drawbacks. First, neither formalizes how to obtain sets of magnetic dipole sequences that encode the behavior of a

desired number of glues, nor glues with complex pairwise interactions. Second, their sequences require precise control of the system’s *temperature*, the amount of force necessary for a bond to be irreversible, as well as additional component geometry. Both are needed to prevent undesired bonds caused by dipole sequence pairs that only partially match or are misaligned.¹ Our work addresses both drawbacks, giving magnetic dipole sequences that encode pairwise bonding behaviors of arbitrarily many glues at fixed temperature and without misalignment.

Section 2 begins by giving a formal model of magnetic dipole sequences, called *dipole words*, and the net forces between them. The pairwise bonding behaviors of a set of glues are defined by a *glue function*, and we define what it means for a set of dipole words to *encode* a glue function and thus be a *dipole code*. These definitions allow for the possibility of misaligned or weak bonds. They require that the dipole words work at fixed temperature (a pair can bond if the net attractive force is positive) and that all misaligned bonds are non-attractive.

Section 3 contains a “warmup” dipole word set construction that encodes *canonical* glue functions, where each glue only bonds to itself or to a unique complementary glue. For any such function over k glues, this construction gives a dipole code of length $O(k)$ that encodes it.

Section 4 improves this construction by extending it to all glue functions, allowing for *flexible glues* (see [40]) that bond to many others. Section 5 improves on length of dipole codes for canonical glue functions only, obtaining length- $O(\log k)$ codes. For both of these results, we also prove that the word lengths of the second and third constructions are asymptotically optimal. Finally, Section 6 poses several remaining open problems.

2 Definitions

Letters and words A *letter* is a symbol x in the *alphabet* $\Sigma = \{0, 1, \diamond\}$. A *dipole word* or simply a *word* is a sequence of letters, and the *length* of a word W , denoted $|W|$, is the number of letters in W . For an integer $i \in \{1, 2, \dots, n\}$, $W[i]$ refers to the i th letter of W and $W[-i]$ refers to the i th from the last letter of W . The *reverse* of a word W , written \widetilde{W} , is the letters of W in reverse order.

¹Section 5.2 and Figure 11 of [18] and Section 4 of [19] discuss these difficulties.

A *subword* of W is a contiguous sequence of letters in W . For integers $1 \leq i, j \leq |W|$ with $i \leq j$, $W[i..j]$ denotes the subword of W from $W[i]$ to $W[j]$, inclusive. As shorthand, $W[-i..] = W[-i..|W|]$ and $W[..j] = W[1..j]$.

Forces The *force* of a pair of letters x, y is defined by function

$$f(x, y) = \begin{cases} 1 & : \{x, y\} = \{0, 1\} \\ -1 & : \{x, y\} \in \{\{0\}, \{1\}\} \\ 0 & : \text{otherwise} \end{cases}$$

and a pair of letters x, y is *attracted* or *repelled* provided $f(x, y) = 1$ or $f(x, y) = -1$, respectively. Similarly, for words X, Y with $|X| = |Y|$, the *force* of the pair X, Y , denoted $f(X, Y)$, is defined as

$$f(X, Y) = \sum_{i=1}^{|X|} f(X[i], Y[i])$$

and the pair X, Y is *attracted* provided $f(X, Y) > 0$.

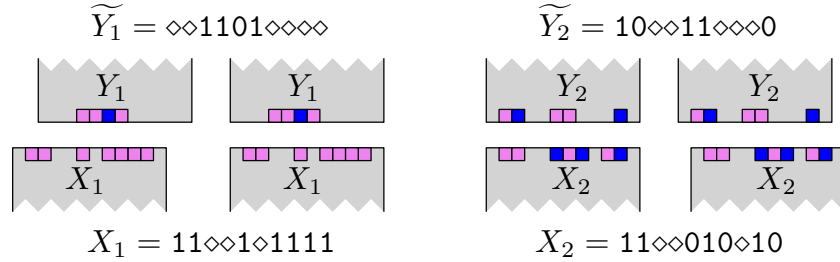


Figure 1: Left: the pair X_1, \widetilde{Y}_1 is attracted, since $f(X_1, \widetilde{Y}_1) = 1 > 0$, and aligned. Right: the pair X_2, \widetilde{Y}_2 is not attracted, since $f(X_2, \widetilde{Y}_2) = -3 + 2 = -1 \not> 0$, and not aligned, since $f(X_2[.. - 2], \widetilde{Y}_2[2..]) = 3 > 0$.

Alignment For a pair X, Y that corresponds to sequences of magnetic dipoles in self-assembling components, it may be the case that $X[i..]$ and $Y[.. - i]$ are attracted, possibly causing unwanted and “misaligned” attachments. A pair X, Y with $|X| = |Y|$ are *aligned* provided that for $A \in \{Y, \widetilde{Y}\}$, $f(X, Y) \leq 0$ and:

$$\max(\{f(X[..i], A[-i..]), f(X[-i..], A[..i]) : i \in \mathbb{N}_{|X|-1}\}) \leq 0$$

The inclusion of both of Y and \widetilde{Y} is to account for the possibility of two units attaching after being rotated ($A = \widetilde{Y}$) or reflected ($A = Y$). The inclusion of $f(X, Y) \leq 0$ is to capture the fact that self-assembly systems often allow rotation without reflection, but not the converse.

Glue functions Let \mathbb{N}_k denote $\{1, 2, \dots, k\}$. A k -glue function is a function $g : \mathbb{N}_k^2 \rightarrow \mathbb{Z}$ such that $g(i, j) = g(j, i)$. The *bond graph* of a glue function g is the graph $(V, E) = (\mathbb{N}_k, \{(i, j) : g(i, j) > 0\})$. The following k -glue function is the *canonical unsigned k -glue function*:

$$g(i, j) = \begin{cases} 1 & : i = j \\ 0 & : \text{otherwise} \end{cases}$$

Similarly, the following $2k$ -glue function is the *canonical signed $2k$ -glue function*:

$$g(i, j) = \begin{cases} 1 & : \{i, j\} = \{2a - 1, 2a\}, a \in \mathbb{N}_k \\ 0 & : \text{otherwise} \end{cases}$$

Encoding A set of common-length words $\mathcal{W} = W_1, W_2, \dots, W_k$ *encodes* a k -glue function g provided that for every $i, j \in \mathbb{N}_k$, W_i, W_j are aligned and $f(W_i, \widetilde{W}_j) = g(i, j)$. Similarly, \mathcal{W} *attractively encodes* g provided that for all i, j , $f(W_i, \widetilde{W}_j) > 0$ if and only if $g(i, j) > 0$. Such an encoding set is called a *dipole code* and the length of the words of the set is the code's *length*.

3 A First Encoding Result

Theorem 3.1. *For any k , there exists a length- $(6k + 14)$ dipole code that attractively encodes the canonical signed $2k$ -glue function.*

Proof. The encoding dipole word set \mathcal{W} consists of the following pairs for all $a \in \mathbb{N}_k$.

- $W_{2a-1} = \diamond^{2k+5} 11(\diamond 1)^{a-1}(1\diamond)(\diamond 1)^{k-a} 1\diamond^{2k+6}$
- $W_{2a} = 1^{2k+6} 0(\diamond\diamond)^{k-a}(1\diamond)(\diamond\diamond)^{a-1}\diamond 1^{2k+5}$

Each word has length $6k + 14$. Establishing the encoding requires two steps: proving that every pair of words W_i, W_j is aligned, and proving that $f(W_i, W_j) > 0$ if and only if $i = 2a - 1, j = 2a$ for some $a \in \mathbb{N}_k$.

The very few 0 letters mandate that words with positive force have almost no repelling letter pairs. The long repeated sections of 1 on both ends of W_{2a} words, and 1 letters in the centers of all words prevent misalignment by overlapping to form repelled letter pairs. The pattern of 1 and \diamond letters in the middle of each word forms at least one repelled letter pair when aligned with any other word, except in complementary pairs W_{2a-1} and W_{2a} , where the patterns have no force.

Alignment Let $W_i, W_j \in \mathcal{W}$. It is easily observed that $f(W_i, W_j) \leq 0$. Now suppose, for the sake of contradiction, that for some $l < |W_i| = |W_j|$ it is the case that $f(W_i[..l], W_j[-l..]) > 0$ or $f(W_i[..l], \widetilde{W}_j[-l..]) > 0$.

It cannot be that both i and j are odd, since then W_i and W_j contain no 0 letters and any subwords must have non-positive force. Without loss of generality, assume i is even. Since the first and last $2k + 6$ letters of W_i are in $\{\diamond, 1\}$, $l \geq 2k + 7$ and thus 1^{2k+6} is a subword of $W_i[..l]$.

If j is even, then every subword of length $2k + 6$ in W_j contains at least two 1 letters. So the subword 1^{2k+6} of $W_i[..l]$ must be involved in at least two repelled letter pairs. Also, W_i and W_j contain two 0 letters total, so any pair of subwords have at most two attracted letter pairs. So the force between any pair of subwords of W_i, W_j is at most $2 - 2 = 0$.

If j is odd, then every subword of W_j with length $2k + 6$ has at least one 1 letter, except $W[-(2k + 6)..]$. Then since $l \geq 2k + 7$, the subword 1^{2k+6} of $W_i[..l]$ is involved in a repelled letter pair. Also, W_i and W_j contain one 0 letter total, so any pair of subwords have force at most $1 - 1 = 0$.

Force Clearly $f(W_i, \widetilde{W}_j) \leq 0$ if i, j have the same parity. So all that remains is to verify that for every $a, b \in \mathbb{N}_k$, it is the case that $f(W_{2a-1}, \widetilde{W}_{2b}) > 0$ if and only if $a = b$. Observe that $W_{2a-1}[2k+8..4k+7] = (\diamond 1)^{a-1}(\mathbf{1}\diamond)(\diamond 1)^{k-a}$ and similarly $\widetilde{W}_{2b}[2k+8..4k+7] = (\diamond\diamond)^{b-1}(\diamond 1)(\diamond\diamond)^{k-b}$. The remaining portions of W_{2a-1} and \widetilde{W}_{2b} have force $f(\diamond^{2k+5}\mathbf{1}\mathbf{1}, \mathbf{1}^{2k+5}\diamond\diamond) + f(\mathbf{1}\diamond^{2k+6}, \mathbf{0}\mathbf{1}^{2k+6}) = 1$. So for all $a, b \in \mathbb{N}_k$, it follows that $f(W_{2a-1}, \widetilde{W}_{2b}) = 1$ if $a = b$ and 0 otherwise. \square

A set of $2k$ dipole words encoding the canonical signed $2k$ -glue function can be concatenated in pairs to yield k dipole words encoding the canonical k -glue function. For instance, the pairs W_1, W_2 and W_3, W_4 of length $6 \cdot 2 + 14 = 26$ that together attractively encode the canonical signed 4-glugue function yield

two words $W_1 \diamond^{26} W_2$ and $W_3 \diamond^{26} W_4$ encoding the canonical unsigned 2-glye function. More generally, for larger k , concatenating the pairs W_{2a-1}, W_{2a} into $W_{2a-1} \diamond^{6k+14} W_{2a}$ for all $a \in \mathbb{N}_k$ yields the canonical unsigned k -glye function.

Corollary 3.2. *For any k , there exists a length- $(18k + 42)$ dipole code that attractively encodes the canonical unsigned k -glye function.*

4 Encoding General Glys Functions

In following with the previous section, we obtain a result for bipartite bond graphs (Theorem 4.1) and then use this to prove a matching result for general bond graphs (Corollary 4.2).

Theorem 4.1. *For any k -glye function g with a bipartite bond graph, there exists a length- $(3k + 14)$ dipole code that attractively encodes g .*

Proof. Since the bond graph is bipartite, it can be written as $(V_1, V_2, E) = (\mathbb{N}_{k'}, \mathbb{N}_k \setminus \mathbb{N}_{k'}, E)$. Assume without loss of generality that $|V_1| \geq |V_2|$ and so $|V_2| \leq k/2$.

Let $\alpha : \mathbb{Z} \rightarrow \{\diamond 1, 1 \diamond\}$ with

$$\alpha(n) = \begin{cases} \diamond 1 & : n \leq 0 \\ 1 \diamond & : n > 0 \end{cases}$$

For integers $a \in V_1$ and $b \in V_2$, define $M_a = \bigcup_{b=k'+1}^{|V_1|} \alpha(g(a, b))$, i.e. M_a is the concatenation of $|V_2|$ 2-letter words encoding the neighbors of a in V_2 . Notice that $|M_a| = 2|V_2| \leq 2(k/2) = k$. The dipole word set \mathcal{W} consists of the following set of dipole words for all $a \in V_1, b \in V_2$:

- $W_a = \diamond^{k+5} 11 M_a 1 \diamond^{k+6}$
- $W_b = 1^{k+6} 0(\diamond \diamond)^{|V_2|-b} 1 \diamond (\diamond \diamond)^{b-|V_1|-1} \diamond \diamond 1^{k+5}$

The set \mathcal{W} attractively encodes g by extending the approach in Section 3 to allow half of the words, namely W_a , to have multiple ‘‘complementary’’ words. The words are shorter here due to a reduction in the number of words (from $2k$ to k).

Following the alignment portion of the proof of Theorem 3.1 establishes that every pair is aligned and only pairs W_i, W_j with $i = a \in V_1, j = b \in V_2$

can be attracted. Observe that $f(M_a, (\diamond\diamond)^{b-|V_1|-1}\mathbf{1}_{\diamond(\diamond\diamond)}^{|V_2|-b}) = 0$ if $g(a, b) > 0$ and -1 otherwise. Then since $f(\diamond^{k+5}\mathbf{1}\mathbf{1}, \mathbf{1}^{k+5}\diamond\diamond) + f(\mathbf{1}\diamond^{k+6}, \mathbf{0}\mathbf{1}^{k+6}) = 1$, $f(W_a, \widetilde{W_b}) > 0$ if and only if $g(a, b) = 1$. So a pair W_a, W_b is attracted if and only if $g(a, b) > 1$. \square

For an unsigned k -glue function g , consider the dipole code \mathcal{W} constructed by Theorem 4.1 that attractively encodes a bipartite $2k$ -glue function g' , where:

$$g'(i, j) = \begin{cases} g(i, j - k) & : i \leq k < j \\ g(i - k, j) & : i > k \geq j \\ 0 & \text{otherwise} \end{cases}$$

Concatenating signed pairs W_i, W_{i+k} for all $i \in \mathbb{N}_k$ into words $W_i \diamond^{3k+14} W_j$ attractively encodes the original unsigned k -glue function g , since the word $W_i \diamond^{3k+14} W_{i+k}$ is attracted to a word $W_j \diamond^{3k+14} W_{j+k}$ if and only if $g(i, j) > 0$.

Corollary 4.2. *For any k -glue function g , there exists a length- $(9k + 42)$ dipole code that attractively encodes g .*

As it turns out, dipole codes of $O(k)$ -length is the best possible.

Theorem 4.3. *Most k -glue functions can only be attractively encoded with dipole codes of length $\Omega(k)$.*

Proof. Observe that all k -glue functions can be partitioned into equal-size sets, where two functions g, g' are in the same set if and only if for all $i, j \in \mathbb{N}_k^2$, $g(i, j) = g'(i, j)$ or $g(i, j) = -g'(i, j) + 1$ for all i, j . That is, two functions are in the same set if they output identical or opposite strengths for all inputs. There are $2^{\binom{k}{2}+k}$ functions in each set, each specified by a distinct sequence of $\binom{k}{2} + k$ choices for whether to output a positive or non-positive value for each input pair i, j with $i \leq j$.

Thus any method of encoding k -glue functions (including as encoding dipole word sets) requires $\log(2^{\binom{k}{2}+k}) = \Omega(k^2)$ bits to specify at least half of these functions. Encoding $\Omega(k^2)$ bits in a dipole word set of k words requires that the words contain $\Omega(k^2/k) = \Omega(k)$ bits each, and so have length $\Omega(k)$. \square

5 Encoding Canonical Glue Functions

In the previous section, we adapted the construction of Section 3 to attractively encode general k -glue functions using dipole codes optimal length $\Theta(k)$. Here we do the same for the special class of canonical glue functions, improving the length of the dipole code to $O(\log k)$. This is easily seen to be optimal, as there must be at least k words in the code.

Theorem 5.1. *For any $k \in \mathbb{N}$ divisible by 4, there exists a set of length- $(20k+3)$ words that attractively encodes the canonical signed $2\binom{k}{k/2}$ -glue function.*

Proof. Let $\alpha, \beta : \{0, 1\} \rightarrow \Sigma^2$ be functions defined by

$$\alpha(n) = \left\{ \begin{array}{l} 1\diamond : n = 0 \\ \diamond 1 : n = 1 \end{array} \right\} \quad \beta(n) = \left\{ \begin{array}{l} 1\diamond : n = 0 \\ 0\diamond : n = 1 \end{array} \right\}$$

Let $\mathcal{C} = \{C_1, C_2, \dots, C_{\binom{k}{k/2}}\}$ be the set of all bit words of length k with equal numbers of 0's and 1's. Let $M_{\alpha, \mathcal{C}} = \bigcup_{i=1}^{\binom{k}{k/2}} \alpha(C_i)$ and $M_{\beta, \mathcal{C}} = \bigcup_{i=1}^{\binom{k}{k/2}} \beta(C_i)$. Define a *bit* word to be a word over the alphabet $\{0, 1\}$. We claim the following set \mathcal{W} of dipole words attractively encodes the canonical signed $2\binom{k}{k/2}$ -glue function.

- $W_{2a-1} = \diamond^{6k} 1^{k/4} \diamond^{11k/4} 11 M_{\alpha, \mathcal{C}_a} 1 \diamond^{11k/4+1} 1^{k/4-1} \diamond^{6k+1}$
- $W_{2a} = 1^{9k+1} \diamond M_{\beta, \widetilde{\mathcal{C}}_a} \diamond \diamond 1^{9k}$

These words form a “compressed” version of the dipole code in Section 3 by changing the middle word patterns from a unary-based encoding with no force to a binary encoding with large, positive force. The regions of 1 and \diamond letters at both ends of the words create a balancing set of repelling letter pairs that only fail to exceed the positive force of the middle patterns if these patterns match.

Alignment Let $W_i, W_j \in \mathcal{W}$. If i, j have the same parity, then $f(W_i, W_j) \leq -k/2$. For any $l < |W_i| = |W_j|$, $f(W_i[..l], W_j[-l..])$ is either less than 1 (if i, j are odd) or less than $-9k + (2k+3) + k/2 < 0$ (if i, j are even), and the same holds for $f(W_i, \widetilde{W}_j[-l..])$.

So consider the remaining case: $i = 2a - 1, j = 2b$ for $a, b \in \mathbb{N}_{\binom{k}{k/2}}$ (and $l < |W_i|$). For $l \leq 15k$ and $A \in \{W_{2b}, \widetilde{W}_{2b}\}$, it is the case that

$f(W_{2a-1}[\dots], A[-\dots]), f(W_{2a-1}[-\dots], A[\dots]) \leq 0$, since the 0 letters of A (that do not appear in the first or last $9k$ letters of A) coincide only with \diamond letters of $W_{2a-1}[\dots]$ or $W_{2a-1}[-\dots]$ (the first and last $6k$ letters of W_{2a-1}).

For $18k \geq l > 15k$, since the subword 1^{9k} is in A and every subword W_{2a-1} with length $9k$ excluding the last contains at least $k/4$ 1 letters, there are at least $k/4$ repelled letter pairs involving letters in A . Moreover, the portion of $M_{\beta, \widetilde{C}_b}$ in A coincides with a subword of W_{2a-1} , namely a subword of $\diamond^{6k}1^{k/4}\diamond^{11k/4}$ or $\diamond^{11k/4+1}1^{k/4-1}\diamond^{6k+1}$, that contains at most $k/4$ 1 letters. So it must be that $f(W_{2a-1}[\dots], A[-\dots]), f(W_{2a-1}[-\dots], A[\dots]) \leq -k/4 + k/4 = 0$.

For $l > 18k$, A contains non-overlapping subwords 1^{9k} and $1^{18k-9k-(2k+3)} = 1^{7k-3}$, a superword of $1^{6k+k/4}$. Then since every subword of W_{2a-1} with length $6k + k/4$ contains $k/4$ 1 letters, there are at least $2(k/4) = k/2$ repelling letter pairs involving letters in A . Moreover, W_{2a-1} and A contain a total of at most $k/2$ 0 letters. Thus $f(W_{2a-1}[\dots], A[-\dots]), f(W_{2a-1}[-\dots], A[\dots]) \leq -k/2 + k/2 = 0$.

Force Clearly $f(W_i, \widetilde{W}_j) \leq 0$ for every i, j with the same parity. So consider $f(W_{2a-1}, \widetilde{W}_{2b})$ for $a, b \in \mathbb{N}_{\binom{k}{k/2}}$. This force is equal to $k/2 - 1 + f(M_{\alpha, C_a}, \widetilde{M_{\beta, \widetilde{C}_b}})$. Inspecting α, β , it is seen that $f(M_{\alpha, C_a}, \widetilde{M_{\beta, \widetilde{C}_b}}) = k/2$ provided $a = b$, and at most $k/2 - 2$ otherwise. Thus $f(W_{2a-1}, \widetilde{W}_{2b}) = 1$ provided $a = b$ and non-positive otherwise. \square

Applying algebra and a similar transformation as done for Corollary 3.2 gives results for both signed and unsigned k -glue functions in terms of k :

Corollary 5.2. *For any k , there exists a length- $O(\log k)$ dipole code that attractively encodes the canonical unsigned k -glue or signed $2k$ -glue function.*

6 Conclusion

All of the encodings here are attractive encodings, and thus the word-word forces do not precisely match the outputs of the encoded glue functions. Removing the “attractive” modifier is of primary importance, starting with canonical glue functions.

Open Problem 6.1. *Can dipole codes encode canonical glue functions? Glue functions with no negative outputs? All glue functions?*

The existence of dipole codes that encode the glue functions

$$g(i, j) = \begin{cases} -1 & : i = j = 1 \\ 1 & : i = j > 1 \\ 0 & : \text{otherwise} \end{cases}$$

imply that systems of square components using magnet sequences embedded along their edges are capable of universal computation, following from results of Patitz, Schweller, and Summers [36].

Prior work by Bhalla et al. [14] uses geometry to prevent unaligned bonding. Such a restriction eases the difficulty of designing dipole codes, and yields an easier version of Open Problem 6.1.

Open Problem 6.2. *Can dipole codes encode all glue functions if words are not required to be aligned?*

Forthcoming work by the authors demonstrates that dipole codes can be physically implemented. The implementation consists of centimeter-sized, 3D-printed components orbitally stirred and bond with one another via sequences of magnets embedded into their faces. The components have dimensions 19 mm × 19 mm × 5 mm (see Figure 2). Neodymium disc magnets (N45, 1 mm diameter, 1 mm length) are embedded into 9 cylindrical recesses (1 mm diameter, 2 mm length) equally spaced along 19 mm × 5 mm faces.

The recess and magnet lengths leave an *air gap* between the magnet and face. The air gap and recess spacing enforce that magnets interact pairwise as in the model: magnets embedded in the same face are sufficiently far apart that no magnet can strongly interact with more than one magnet in the same face. Magnet embedding sequences are determined by dipole words: a 1 letter is a magnet with positive pole oriented outwards, a 0 letter is a magnet negative pole oriented outwards, and a \diamond letter is no magnet. The implementation uses a length-9 code that encodes the canonical signed 4-glug function: 11101 \diamond 11 \diamond , 1 \diamond 1 \diamond 1 \diamond \diamond \diamond , 1110111 \diamond \diamond , 11 \diamond \diamond 1 \diamond \diamond . Preliminary experiments demonstrate that components reliably bond irreversibly via aligned complementary dipole words, while misaligned or non-complementary bonds rarely form and are easily broken by the kinetic energy imparted by the orbital stirring.

While the codes developed in this work are asymptotically optimal with small constants, scaling physical implementations to the sizes needed to use even small instances of these codes is challenging due to increased component

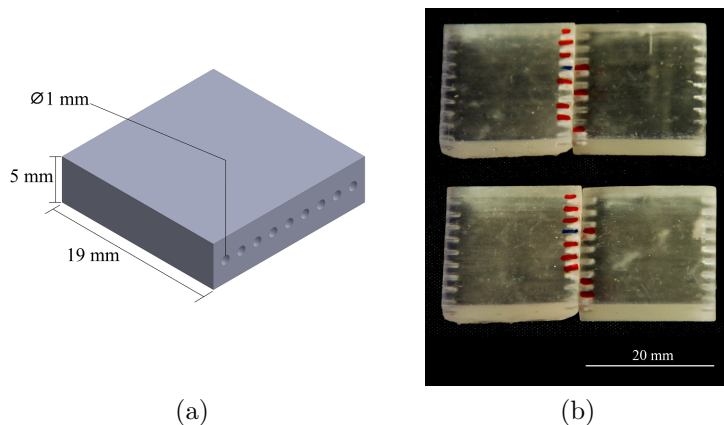


Figure 2: (a) Component dimensions. (b) Component pairs bonding along faces containing complementary sequences of embedded magnets. Blue and red marks indicate magnets embedded with north and south poles oriented outwards, respectively.

weight, agitation forces, and assembly time. Thus computing codes as close to optimal as possible is key to achieving physical implementations.

Open Problem 6.3. *Can canonical signed $2k$ -glue functions be attractively encoded by dipole codes of length $\lceil \log_3(2k) \rceil + o(\log k)$?*

Open Problem 6.4. *What is the computational complexity of the following problem: given a glue function g and integer l , does there exist a dipole code of length at most l (attractively) encoding g ?*

References

- [1] G. M. Whitesides, B. Grzybowski, Self-assembly at all scales, *Science* 295 (2002) 2418–2421.
- [2] K. Hosokawa, I. Shimoyama, H. Miura, Two-dimensional micro-self-assembly using the surface tension of water, *Sensors and Actuators A: Physical* 57 (2) (1996) 117–125.
- [3] N. Bowden, F. Arias, T. Deng, G. M. Whitesides, Self-assembly of microscale objects at a liquid/liquid interface through lateral capillary forces, *Langmuir* 17 (2001) 1757–1765.
- [4] T. D. Clark, J. Tien, D. C. Duffy, K. E. Paul, G. M. Whitesides, Self-assembly of 10- μm -sized objects into ordered three-dimensional arrays, *Journal of the American Chemical Society* 123 (2001) 7677–7682.
- [5] N. Bowden, A. Terfort, J. Carbeck, G. M. Whitesides, Self-assembly of mesoscale objects into ordered two-dimensional arrays, *Science* 276 (5310) (1997) 233–235.
- [6] N. Bowden, S. R. J. Oliver, G. M. Whitesides, Mesoscale self-assembly: capillary bonds and negative menisci, *Journal of Physical Chemistry B* 104 (12) (2000) 2714–2724.
- [7] P. W. K. Rothmund, Using lateral capillary forces to compute by self-assembly, *Proceedings of the National Academy of Sciences* 97 (3) (2000) 984–989.
- [8] M. Cheng, F. Shi, J. Li, Z. Lin, C. Jiang, M. Xiao, L. Zhang, W. Yang, T. Nishi, Macroscopic supramolecular assembly of rigid building blocks through a flexible spacing coating, *Advanced Materials* 26 (2014) 3009–3013.
- [9] A. Harada, R. Kobayashi, Y. Takashima, A. Hashidzume, H. Yamaguchi, Macroscopic self-assembly through molecular recognition, *Nature Chemistry* 3 (2010) 34–37.
- [10] M. Xiao, Y. Xian, F. Shi, Precise macroscopic supramolecular assembly by combining spontaneous locomotion driven by the marangoni effect

and molecular recognition, *Angewandte Chemie International Edition* 54 (31) (2015) 8952–8956.

- [11] J. Bishop, S. Burden, E. Klavins, R. Kreisberg, Programmable parts: a demonstration of the grammatical approach to self-organization, in: *Proceedings of International Conference on Intelligent Robots and Systems 2005*, 2005, pp. 3684–3691.
- [12] P. J. White, K. Kopanski, H. Lipson, Stochastic self-reconfigurable cellular robotics, in: *Proceedings of the International Conference on Robotics and Automation 2004*, 2004, pp. 2888–2893.
- [13] P. J. White, V. Zykov, J. Bongard, H. Lipson, Three dimensional stochastic reconfiguration of modular robots, in: *Proceedings of Robotics: Science and Systems I*, 2005, pp. 161–168.
- [14] N. Bhalla, P. J. Bentley, C. Jacob, Evolving physical self-assembling systems in two-dimensions, in: *ICES 2010*, Vol. 6274 of LNCS, Springer, 2010, pp. 381–392.
- [15] J. Breivik, Self-organization of template-replicating polymers and the spontaneous rise of genetic information, *Entropy* 3 (2001) 273–279.
- [16] K. Hosokawa, I. Shimoyama, H. Miura, Dynamics of self-assembling systems: Analogy with chemical kinetics, *Artificial Life* 1 (4) (1995) 413–427.
- [17] G. Mermoud, M. Mastrangeli, U. Upadhyay, A. Martinoli, Real-time automated modeling and control of self-assembling systems, in: *Proceedings of 2012 IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 4266–4273.
- [18] N. Bhalla, P. J. Bentley, C. Jacob, Mapping virtual self-assembly rules to physical systems, in: *Proceedings of the 6th International Conference on Unconventional Computation*, 2007, pp. 117–147.
- [19] U. Majumder, J. H. Reif, A framework for designing novel magnetic tiles capable of complex self-assemblies, in: *Unconventional Computing 2008*, Vol. 5204 of LNCS, Springer, 2008, pp. 129–145.

- [20] N. Bhalla, P. J. Bentley, P. D. Vize, C. Jacob, Programming and evolving physical self-assembling systems in three dimensions, *Natural Computing* 11 (3) (2012) 475–498.
- [21] L. Adleman, Molecular computation of solutions to combinatorial problems, *Nature* 266 (5187) (1994) 1021–1024.
- [22] C. G. Evans, Crystals that count! physical principles and experimental investigations of DNA tile self-assembly, Ph.D. thesis, Caltech (2014).
- [23] P. W. K. Rothmund, N. Padadakis, E. Winfree, Algorithmic self-assembly of DNA sierpinski triangles, *PLoS Biology* 2 (12) (2004) 2041–2053.
- [24] Y. Ke, L. Ong, W. Shih, P. Yin, Three-dimensional structures self-assembled from DNA bricks, *Science* 338 (2012) 1177–1183.
- [25] B. Wei, M. Dai, P. Yin, Complex shapes self-assembled from single-stranded DNA tiles, *Nature* 485 (2012) 623–626.
- [26] A. Brenneman, A. Condon, Stranddesign for biomolecular computation, *Theoretical Computer Science* 287 (1) (2002) 39–58.
- [27] M. H. Garzon, DNA codeword design: Theory and applications, *Parallel Processing Letters* 24 (2) (2014) 1440001.
- [28] G. Mauri, C. Ferretti, Word design for molecular computing: A survey, in: J. Chen, J. Reif (Eds.), *DNA 9*, Vol. 2943 of LNCS, Springer, 2004, pp. 37–47.
- [29] R. Montemanni, Combinatorial optimization algorithms for the design of codes: a survey, *Journal of Applied Operational Research* 7 (1) (2015) 36–41.
- [30] J. Sager, D. Stefanovic, Designing nucleotide sequences for computation: A survey of constraints, in: A. Carbone, N. A. Pierce (Eds.), *DNA 11*, Vol. 3892 of LNCS, Springer, 2006, pp. 275–289.
- [31] L. Kari, S. Konstantinidis, P. Sosik, Preventing undesirable bonds between DNA codewords, in: C. Ferretti, G. Mauri, C. Zandron (Eds.), *DNA 10*, Vol. 3384 of LNCS, Springer, 2005, pp. 182–191.

- [32] O. Milenkovic, N. Kashyap, On the design of codes for DNA computing, in: O. Ytrehus (Ed.), WCC 2005, Vol. 3969 of LNCS, Springer, 2006, pp. 100–119.
- [33] O. D. King, Bounds for DNA codes with constant GC-content, *Electronic Journal of Combinatorics* 10 (1) (2003) R33.
- [34] P. Gaborit, O. D. King, Linear constructions for DNA codes, *Theoretical Computer Science* 334 (1–3) (2005) 99–113.
- [35] R. Montemanni, D. H. Smith, N. Koul, Three metaheuristics for the construction of constant GC-content DNA codes, in: ICAOR 2014, Vol. 6 of LNMS, Springer, 2014, pp. 167–175.
- [36] M. J. Patitz, R. T. Schweller, S. M. Summers, Exact shapes and turing universality at temperature 1 with a single negative glue, in: DNA 2011, Vol. 6937 of LNCS, Springer, 2011, pp. 175–189.
- [37] J. H. Reif, S. Sahu, P. Yin, Complexity of graph self-assembly in accretive systems and self-destructible systems, *Theoretical Computer Science* 412 (2011) 1592–1605.
- [38] D. Doty, L. Kari, B. Masson, Negative interactions in irreversible self-assembly, *Algorithmica* 66 (1) (2013) 153–172.
- [39] R. T. Schweller, M. Sherman, Fuel efficient computation in passive self-assembly, in: *Proceedings of the 24th Annual Symposium on Discrete Algorithms*, 2013, pp. 1513–1525.
- [40] G. Aggarwal, Q. Cheng, M. Goldwasser, M. Kao, P. de Espanes, R. Schweller, Complexities for generalized models of self-assembly, *SIAM Journal on Computing* 34 (6) (2005) 1493–1515.