Defect tolerance: fundamental limits and examples

Jennifer Tang, Da Wang, Yury Polyanskiy, Gregory Wornell

Abstract

This paper addresses the problem of adding redundancy to a collection of physical objects so that the overall system is more robust to failures. Physical redundancy can (generally) be achieved by employing copy/substitute procedures. This is fundamentally different from information redundancy, where a single parity check simultaneously protects a large number of data bits against a single erasure. We propose a bipartite graph model of designing defect-tolerant systems where defective objects are repaired by reconnecting them to strategically placed redundant objects. The fundamental limits of this model are characterized under various asymptotic settings and both asymptotic and finite-size optimal systems are constructed.

Mathematically, we say that a k by m bipartite graph corrects t defects over an alphabet of size q if for every q-coloring of k left vertices there exists a q-coloring of m right vertices such that every left vertex is connected to at least t same-colored right vertices. We study the trade-off between redundancy m/k and the total number of edges in the graph divided by k. The question is trivial when $q \geq k$: the optimal solution is a simple t-fold replication. However, when $q < k$ non-trivial savings are possible by leveraging the inherent repetition of colors.

Index Terms

Defect-tolerant circuits, bipartite graphs, coloring, combinatorics, worst-case errors

CONTENTS

JT, YP and GW are with the Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA 02139 USA. e-mail: {jstang,yp,gww}@mit.edu. DW was with the Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, 02139 and is now with Two Sigma Investments, New York, NY, 10013. e-mail: dawang@alum.mit.edu.

The research was supported by the Center for Science of Information (CSoI), and NSF Science and Technology Center, under grant agreement CCF-09-39370. This work was also supported in part by Systems on Nanoscale Information fabriCs (SONIC), an SRC STARnet Center sponsored by MARCO and DARPA.

I. Introduction

Classical Shannon theory established principles of adding redundancy to data for combatting noise and, dually, of removing redundancy from data for more efficient storage. The central object of the classical theory is information, which unlike physical objects, can be freely copied and combined. In fact, the marvel of error-correcting codes is principally based on the counter-intuitive property that multiple unrelated information bits $X_1, \ldots, X_k$ can be simultaneously protected by adding "parity-checks" such as

$$Y = X_1 + \cdots + X_k \mod 2 . \tag{1}$$

In this example, the added parity-check Y allows the recovery of the original message even if vector

$$(X_1, X_2, \ldots, X_k, Y)$$

undergoes an erasure of an arbitrary element.

Physical objects (e.g., transistors in a chip) may also be subject to erasures (failures) and thus it is natural to ask about ways of insuring the system against failure events. Note, however, that for physical objects operations such as (1) are meaningless. If the failure event renders an object completely useless, then protecting against these failures would entail an addition of spare (redundant) elements. The required operation is to copy and then substitute. It may, therefore, seem that nothing better than a simple replication can guard against failures. This paper shows otherwise. Indeed, there exist non-trivial ways to add redundancy as long as the objects' diversity does not exceed their number. That is, if the number of types of objects is smaller than the total number of them.

The objective of this paper is to develop a study of adding redundancy to a physical system where certain objects in the system fail and can only be replaced by substitutes. This paper will explore what are good design choices in this scenario and find some relevant fundamental limits.

A. Reconfigurable defect-tolerant circuits

To facilitate defining the problem we intend to study, we will first present the application which informed the main model we developed for studying redundancy of physical objects, and that is the application of reconfigurable circuits.

Consider a chip design process, in which the chip is composed of many similar cells (e.g., standard-cell designs of ASICs). Layout of elements in each cell is dictated by the chip manufacturer. Each cell has k input/output buses and k placeholders (nodes) that can be filled in with logic realizing one of q functions. Now because of manufacturing defects, not all k primary elements will operate correctly. For this reason, each cell also contains provisions for redundant elements. In particular, there are m placeholders designated as redundant elements. The designer then selects what type of logic to instantiate into these redundant elements. Once the chip is manufactured and placed on the testbed, the testing equipment probes each cell and determines which primary elements are defective. Programmable switches are then used to reconnect input/output buses from the defective primary elements to one of the redundant elements containing the same logic. So the summary of the events happening to each cell during this process is:

1) Choose the layout of the placeholders and provisional wires
2) Choose components (from available collection of possible types) to fill in the primary elements for the reconfigurable circuit
3) Based on primary elements chosen, choose redundant components (from the same collection) to place in redundant placeholders
4) Build the circuit with these components
5) Based on where the defects occurred, reconfigure the interconnects (i.e., enable provisional wires with programmable switches) of the circuit to correct the defects

With respect to this application, our goal is to understand what wiring topologies the chip manufacturer should try to implement in order to attain optimal trade-off between the number of redundant elements, provisional wires (buses) and defect-tolerance. Notice that the two metrics, redundancy and wiring, both correspond to necessary resources. Adding redundancy requires

more silicon area and the provisional wires requires additional metal and programmable switches. [1]

B. Relation to prior work

Prior work on the subject of designing digital electronics robust to noise has been traditionally approached with the goal of combatting dynamic noise. This is epitomized in the line of work started by von Neumann [1] and contemporary variations [2]. Although significant progress has been made in understanding fundamental limits in von Neumann's model, see e.g., [3]–[11], the practical applications are limited due to a prohibitively high level of redundancy required [12].

Here, instead, we are interested in circuits robust to static manufacturing failures. As illustrated previously, this scenario has the advantage of being able to test which parts of the circuit failed and attempt to configure out (or "wire around") the defective parts. This side information enables significant savings in redundancy [13]. In fact, this method of testing the performance of a device followed by some configuration is rather popular in practice: multi-core CPUs [14], analog-to-digital converters [15], sense-amplifiers [16], self-replicating automatons [17], parallel computing [18], [19], etc.

This paper can be seen as an attempt to provide theoretical foundations for the static defect scenario. (In fact, this was our original motivation.)

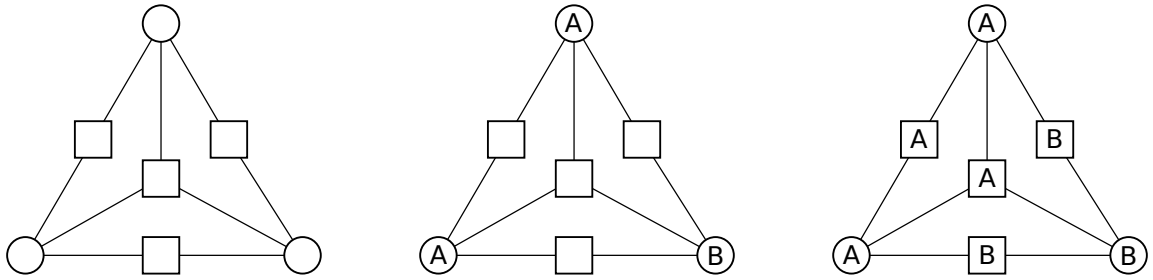
C. Problem formulation

We study the following problem formulation: Given k objects ("primary nodes"), connect each one of them to some of the available m spares ("redundant nodes") in such a way that in the event that $t \geq 1$ of the objects fail (originals or spares) the overall system can be made to function after a repair step. Such a repair step consists of replacing each failed primary node with one of the working spares that it is connected to. Each spare can only replace one failed primary node. The key assumptions are 1) the primary nodes are one of q types 2) the spares have to be programmed to one of the q types *before* the failure events are known and 3) the same connections need to repair all possible choices of types for the k primary nodes. We are interested in minimizing the redundancy m/k and the number of connections to spare nodes.

Key to our problem formulation is the idea that we want to design the interconnects before any of the node types are determined. One might argue that in some applications the interconnect could be allowed to depend on the labeling of primary nodes. Indeed, the latter will be known before the final topology for the chip is made. However, our procedure insists that the interconnection design does not depend on this labeling. The advantage of this is that in the reconfigurable circuits framework, the provisional wire-layout is usable regardless of where any element is placed, providing the same defect recovery guarantee for every possible placement. We seek a *universal* design, which is independent of element types and thus could serve as the new standard cell for all defect-tolerant circuits. We further discuss alternative design methodologies in Section VI.

We intentionally abstracted our problem to a simple model which is more fundamental and relates to other applications needing redundancy for objects and a universal design. For example, instead of parts of a reconfigurable circuit, objects can represent elements in a programmable logic device (e.g., look-up tables (LUTs) in an FPGA). As part of periodical firmware update,

[1]There are certainly other metrics (such as geometric constraints) which are relevant for circuit applications, but we leave consideration of them to future work.



(a) Design (b) primary node labeling (c) Redundant node labeling

Fig. 1. Example of a 2-defect correcting design for an alphabet $\mathcal{X} = \{A, B\}$ of size $q = 2$. Fig. 1(b) shows a labeling of primary (circle) nodes. To each such labeling, we strategically label the redundant (square) nodes, so that each primary node has $t = 2$ neighbors with matching labels (see Fig. 1(c)). Since such a choice is possible for any of the $2^3 = 8$ possible labelings of primary nodes, we conclude that this design is a $t = 2$ defect correcting in the sense of Definition 1.

a manufacturer assigns values of LUTs (both primary and redundant) without knowledge of locations of device-specific failures. Then, a built-in algorithm for each failed LUT T reconnects it to an adjacent spare LUT R, with the requirement that R and T be equivalent. This built-in local algorithm is a computationally non-demanding way to reconfigure around defective LUTs. Note that the interconnections of the LUTs need to be universal so that any update chosen by the manufacturer has the same guarantee against defects.

For $q = 2$ our problem is equivalent to finding sparsity vs. edge-size trade-off for (t, t)-colorable hypergraphs, cf. [20]. See Section VI-C. Other potential applications arise in warehouse planning, operations research, public safety etc.

Expressed mathematically, we are looking for a $k \times m$ bipartite graph with the property that for any q-coloring of the left nodes there is a q-coloring of the right nodes such that each of the k nodes is connected to at least t nodes of its color. The goal is to find bipartite which have efficient trade off in redundancy m/k vs. number of edges.

The high-level summary of our main findings is that when $q \geq k$, no strategy is better than straightforward t-fold replication. When $q < k$, there exist designs that provide savings compared to repetition. We fully or partially characterize the fundamental trade-off between redundancy m/k and the average number of edges (connections) per primary node in the following cases:

1) q, t fixed and $k, m \to \infty$;
2) q fixed and $k, m, t \to \infty$;
3) q, k fixed and $m, t \to \infty$.

Perhaps surprisingly, in this (combinatorial) problem it is possible to obtain exact analysis for asymptotics. The organization of the paper is as follows. Section II introduces the problem formally and overviews main results. Section III demonstrates small-size examples that show non-triviality of the problem. Sections IV and V address the trade-off in the regime of fixed t and $t \to \infty$ respectively. Finally, Section VI discusses implications and extensions of our results.

The notation $[n]$ will denote positive integers $1, 2, ..., n$. The notation $\mathbb{1}\{\cdot\}$ will denote the indicator function. A underlined letter (e.g, $\underline{x}$) stands for a vector quantity.

II. Problem setup and main results

A. Defect-tolerance model

This paper focuses on bipartite graph designs. The left-side nodes of the bipartite graph are called the primary nodes. These are denoted by circles and there are k of these in the bipartite graph. The right-side nodes are the redundant nodes (or secondary nodes). These are denoted by squares and there are m of these in the bipartite graph.

Let $\mathcal{X}$ be a finite alphabet where $q = |\mathcal{X}|$.

Definition 1. *Fixed an alphabet $\mathcal{X}$ with size q. A $k \times m$ bipartite graph is called a* t-defect correcting design *if for any labeling of k primary nodes by elements of $\mathcal{X}$ there exists a labeling of m redundant nodes by elements of $\mathcal{X}$ such that every primary node labeled $x \in \mathcal{X}$ has at least t neighbors labeled x. We will call such a graph a $(k, m, t, E)_q$-design, with E denoting the number of edges. (See Fig. 1 for an illustration.)*

This paper is devoted to the fundamental trade-off between the two basic parameters of t-defect correcting designs: redundancy and wiring complexity. The redundancy of a $(k, m, t, E)_q$-design is $\rho = m/(kt)$. The wiring complexity (or average degree) of a $(k, m, t, E)_q$-design is $\varepsilon = E/(kt)$. The trade-off can be encoded in a two-dimensional region:

Definition 2. *For a fixed q and $t \geq 1$ we define the region $\mathcal{R}_t$ as the closure of the set of all achievable pairs of (ε, ρ):*

$$\mathcal{R}_t \stackrel{\triangle}{=} \text{closure}\left\{\left(\frac{E}{kt}, \frac{m}{kt}\right) : \exists (k, m, t, E)_q\text{-design}\right\} \tag{2}$$

To interpret between Definition 1 and the reconfigurable circuit (and other applications), we present the following association.

Proposition 1. *An interconnect for a reconfigurable circuit can tolerate any t manufacturing defects for any choice of primary nodes if and only if the interconnect is a t-defect correcting design.*

Proof. If the interconnect corrects fewer than t defects, there is some primary node labeling where any labeling of the redundant nodes would result in some primary node with label x having fewer than t neighbors with the same label x. If this primary node and all its matching neighbors have defects, then the defect in the primary node cannot be corrected.

If the interconnects is a t-defect correcting design, with the correct redundant node labeling scheme, any primary node labeled x has t redundant neighbors with the same label x. If there are only t defects, either this primary node is working, or this primary node has an defect and at most $t - 1$ of its neighbors have an defect or are used to correct another primary node. In the latter case, there is at least one redundant node with label x available which can be used to replace this primary node. □

As noted earlier, our performance metrics, ρ and ε, correspond to the extra silicon area and wiring (and fan-out) required respectively for defect-tolerance.

Before proceeding further, we summarize some of the basic properties of regions $\mathcal{R}_t$.

Proposition 2. *(Properties of $\mathcal{R}_t$) Regions $\mathcal{R}_t$ satisfy the following:*

1) *$(\varepsilon, \rho) \in \mathcal{R}_t$ iff there exists a sequence of $(k, m, t, E)_q$-designs with $\frac{E}{kt} \to \varepsilon, \frac{m}{kt} \to \rho$ as $k, m \to \infty$;*

2) *if* $(\varepsilon, \rho) \in \mathcal{R}_t$ *and* $\varepsilon' \geq \varepsilon, \rho' \geq \rho$ *then* $(\varepsilon', \rho') \in \mathcal{R}_t$*;*
3) $\mathcal{R}_t$ *are closed convex subsets of* $\mathbb{R}_+^2$*;*
4) *We have*

$$\limsup_{t\to\infty} \mathcal{R}_t = \text{closure}\left\{\bigcup_{t\geq 1} \mathcal{R}_t\right\} \stackrel{\triangle}{=} \mathcal{R}_\infty\,. \tag{3}$$

5) *Limiting region* $\mathcal{R}_\infty$ *is also a closed convex subset of* $\mathbb{R}_+^2$ *characterized as*

$$\mathcal{R}_\infty \stackrel{\triangle}{=} \text{closure}\left\{\left(\frac{E}{kt}, \frac{m}{kt}\right) : \exists (k, m, t, E)_q - \textit{design}\right\} \tag{4}$$

See Section IV-B for the proof.

B. Preview of main results for binary alphabet

In this section, for the purpose of illustration, we give a summary of our results for the case of binary alphabet $\mathcal{X}$ (i.e., $q = 2$). The rest of the paper will present various bounds and constructions for general alphabet sizes, (i.e., arbitrarily values of q).

There are three separate results which are the main contributions of this paper. One is characterizing the region $\mathcal{R}_t$ in the regime where t is small, specifically for values where $t = 1$ and $t = 2$. The second main result is characterizing the region $\mathcal{R}_\infty$, which corresponds to the limit of regions $\mathcal{R}_t$ when t tends to infinity. The third is characterizing the result when the number of primary nodes k is finite and t tends to infinity.

The theorem for the small t case is the following:

Theorem 3. *For binary alphabet* $\mathcal{X}$*, if* $t = 1$ *or* $t = 2$*, we have*

$$\mathcal{R}_t = \{(\varepsilon, \rho) : \varepsilon \geq 1 \textit{ and } \varepsilon \geq 2 - \rho\} \tag{5}$$

This will be proved in Section IV-D. The region in Theorem 3 has two corner points. We will also discuss the designs which attain these corner points.

The theorem for the asymptotic t case is the following:

Theorem 4. *Let* $\mathcal{X}$ *be a binary alphabet. The region* $\mathcal{R}_\infty$ *defined in* (3) *is the closure of the set of points* (ε, ρ) *defined as follows. For every distribution* P_S *on* $\mathbb{Z}_+$ *with finite support, we define*

$$\varepsilon = \frac{\mathbb{E}[S]}{F(P_S)}, \quad \rho = \frac{1}{F(P_S)}\,, \tag{6a}$$

where

$$F(P_S) \stackrel{\triangle}{=} \min_{0\leq\lambda\leq 1} \max_{0\leq f(\cdot,\cdot)\leq 1} \min\left\{\mathbb{E}\left[\frac{L_0}{\lambda} f(L_0, L_1)\right], \mathbb{E}\left[\frac{L_1}{1-\lambda}(1 - f(L_0, L_1))\right]\right\} \tag{6b}$$

with the expectations being over $S \sim P_S$ *and given* S *distribution of* $L_1 \sim \textit{Bino}(S, \lambda)$ *and* $L_0 = S - L_1$.

This theorem parametrically characterizes $\mathcal{R}_\infty$ in terms of the function $F(P_S)$, which is evaluated on every P_S with finite support. Note that evaluation of the bound (6a) is non-trivial as we will discuss in Section V-E.

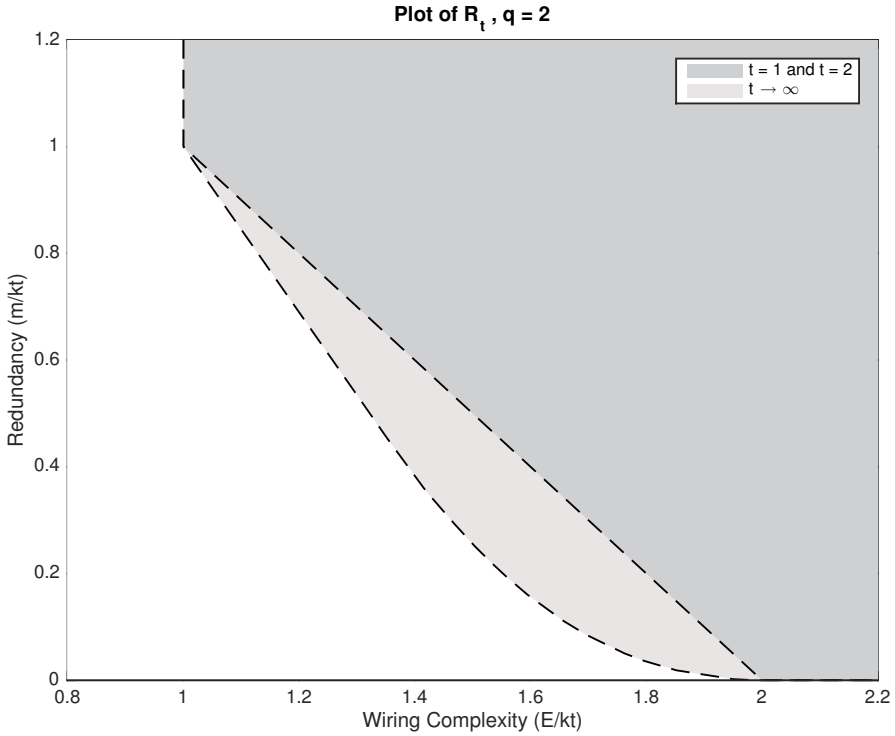


Fig. 2. Achievable regions for redundancy and wiring complexity trade-off when $q = 2$. Regions $\mathcal{R}_1$ and $\mathcal{R}_2$ are shown in darker gray. Region $\mathcal{R}_\infty$ includes lighter and darker gray areas. All other regions $\mathcal{R}_t$ lie between $\mathcal{R}_1$ and $\mathcal{R}_\infty$. The boundary of the region of $\mathcal{R}_\infty$ is calculated using the methods in Appendix E.

The generalization of Theorem 4 to larger alphabet sizes is developed in Section V. Here the designs achieving the best trade-off are more complicated then those associated with Theorem 3. We call them *subset designs* and develop them in Section III-B.

The resulting achievable regions for Theorem 3 and Theorem 4 are depicted in Fig. 2. Via these results we can determine at any fixed redundancy level, how many connections are necessary. For example, at redundancy level 10%, the figure indicates that there exists design which:

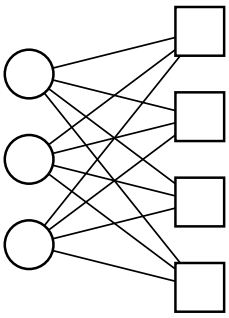
- correct 1 defect if each primary node is connected on average to about 1.9 redundant nodes
- correct 2 defects if each primary node is connected on average to about 1.9×2 redundant nodes
- correct 10^3 defects if each primary node is connected on average to about 1.7×10^3 redundant nodes

According to (4) all regions $\mathcal{R}_t$ will lie between $\mathcal{R}_1$ and $\mathcal{R}_\infty$, approaching the latter as $t \to \infty$. It is perhaps surprising that unlike most known asymptotic combinatorial problems, this one (for $t \to \infty$) admits a relatively simple solution.

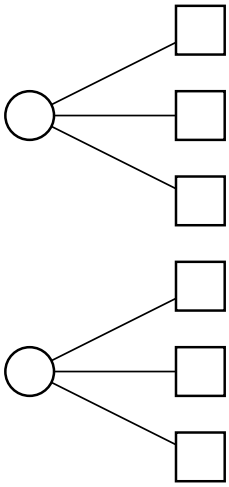
The third and the more practically useful result is the characterization of the achievable regions for asymptotic t but with finite k. This is developed in Section V-F.

III. Examples of good designs

Before developing the main results, we first develop a few basic designs and their performance. Some of the examples play major roles in subsequent developments.



(a) The complete design



(b) The repetition design

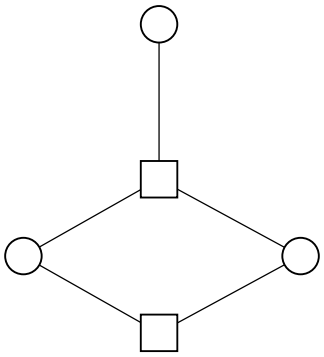
Fig. 3. Two elementary designs.

We denote by $K(k, m)$ a complete bipartite graph with k primary nodes (circles) and m redundant nodes (squares). The two most basic designs are the following:

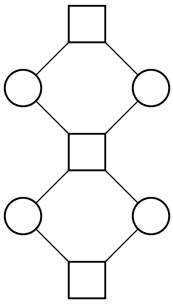
1) *Complete designs:* $K(k, qt)$ (recall that $q = |\mathcal{X}|$) is t-defect correcting (just label redundant nodes to hold t-copies of each value $\mathcal{X}$). See Fig. 3(a) for illustration.
2) *Repetition designs:* $K(1, t)$ is capable of correcting t defects over arbitrary alphabet. Taking k disjoint copies of $K(1, t)$, denoted by $kK(1, t)$, we get a repetition block achieving $\rho = \varepsilon = 1$. See Fig. 3(b) for illustration.

If we take $k \to \infty$, the complete design achieves $\varepsilon = q$ and $\rho = 0$, which is clearly the best possible trade-off given the value of ε. For finite k, however, the complete design is not the design with the minimal number of edges: it is possible to remove some of the edges and still maintain a t defect correcting property, as we will show in the next subsection.

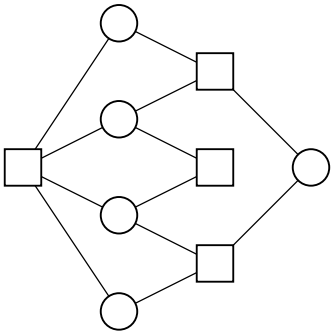
The repetition design uses the minimal number of edges, since any primary node needs at least t edges in order to be a t-defect correcting design. If all primary nodes have exactly t edges, then it is necessary for each primary node to have a distinct neighborhood, illustrating



(a) $q = 2$



(b) $q = 3$



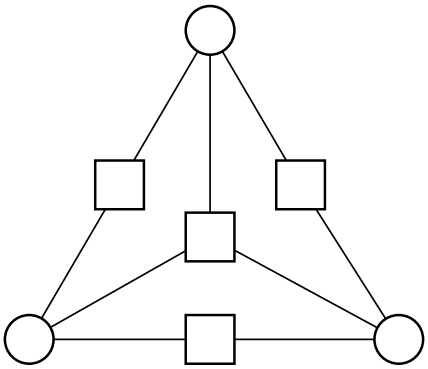
(c) $q = 4$

Fig. 4. Smallest non-trivial 1-defect correcting designs.

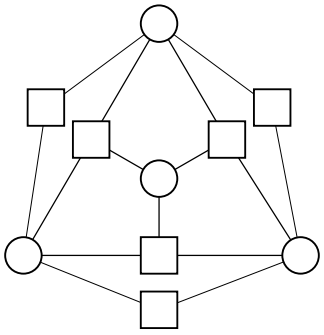
that the repetition design achieves the best trade-off at minimal wiring complexity.

A. Smallest non-trivial designs

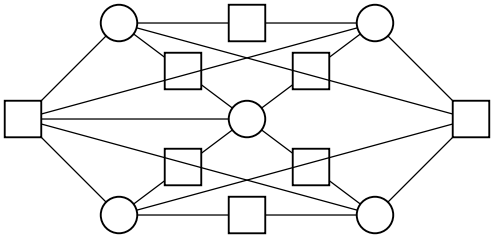
We now present designs which have the fewest number of edges given some fixed number of primary nodes k and redundant nodes m.



(a) $q = 2$



(b) $q = 3$



(c) $q = 4$

Fig. 5. Smallest non-trivial 2-defect correcting designs.

If $k \leq q$ then all primary nodes can have different values and thus one is forced to use the repetition design $kK(1,t)$ to correct t defects. For $k = q+1$ the question becomes more interesting. First, notice that the minimal possible m equals qt (this is achieved by the complete design and clearly cannot be reduced). However, some of the edges can be removed from the complete design.

The optimal designs with $k = q+1$, $m = q$ and $t = 1$ are as follows:

- Binary alphabet: $k = 3$, $m = 2$ with 5 edges. See Fig. 4(a).
- Ternary alphabet: $k = 4$, $m = 3$ with 8 edges. See Fig. 4(b). (There exist two non-isomorphic optimal designs. Fig. 4(b) shows the symmetric one.)
- Quaternary alphabet: $k = 5$, $m = 4$ with 12 edges. See Fig. 4(c). (There exist multiple non-isomorphic optimal designs. Only one is shown.)

The optimal designs with $k = q+1$, $m = 2q$ and $t = 2$ are as follows:

- Binary alphabet: $k = 3$, $m = 4$ with 9 edges, see Fig. 5(a). This design is what we call the *Hamming block*. Fig. 1 shows how it can correct 2 defects. We will discuss its optimality in Corollary 18
- Ternary alphabet: $k = 4, m = 6$ with 15 edges, see Fig. 5(b). (There exist two non-isomorphic optimal designs. Fig. 5(b) shows the symmetric one.)
- Quaternary alphabet: $k = 5, m = 8$ with 21 edges, see Fig. 5(c).

Some of these designs were found analytically and others by exhaustive search. None of these designs are at the performance boundary of any $\mathcal{R}_t$ regions. To obtain designs that are optimal in the ρ-ε trade-off setting, we need to use a larger number of primary and redundant nodes (see Proposition 9). However, a few of these designs, like the Hamming block in Fig. 5(a), achieve the best trade-off when restricted to the finite k setting (as we will develop in Section V-F).

B. Subset designs

Designs that form a key ingredient of our asymptotic (i.e., large t) constructions are subset designs. A subset design $S(k,s)$ is a bipartite graph with k primary nodes and $m = \binom{k}{s}$ redundant nodes, each connected to a distinct s-subset of $\{1,\dots,k\}$. Note that the degree of each primary node is $\binom{k-1}{s-1}$.

In general, we allow subset designs to have multiple and possibly different subset sizes. For two values s_1 and s_2, where $s_1, s_2 \in [k]$, a bipartite graph $S(k,s_1) \vee S(k,s_2)$ is defined to be the result of identifying the k primary nodes in two disjoint copies of $S(k,s_1)$ and $S(k,s_2)$. The resulting graph has k primary nodes and $m = \binom{k}{s_1} + \binom{k}{s_2}$ redundant nodes. This operation ($\vee$) we call graph merging, which we state more precisely below. We will develop the properties of merging later.

Definition 3 (Merging)**.** *For any collection of designs G_j on the same number of primary nodes k, the merging of G_j, denoted $G = \bigvee_j G_j$ is a graph formed by taking disjoint copies of G_j and identifying primary nodes.*

Definition 4 (Subset design)**.** *Given k and positive integers $s_1, s_2, \dots, s_r \in [k]$,*

$$S(k,s_1) \vee S(k,s_2) \vee \cdots \vee S(k,s_r) \tag{7}$$

is a subset design with k primary nodes and $m = \sum_{j=1}^{r} \binom{k}{s_j}$ redundant nodes.

For example, the Hamming block, Fig. 5(a), is $S(3,2) \vee S(3,3)$, the repetition design is $S(k,1) \vee \cdots \vee S(k,1)$ (t times) and complete design is $S(k,k) \vee \cdots \vee S(k,k)$ (qt times).

The subset designs are characterized by the following property:

Definition 5 (Permutation invariance)**.** *A design is called* permutation invariant *if there exists a group of bipartite-graph automorphisms (thus preserving the left/right partition) that acts as the full symmetric group S_k on primary nodes.*

Proposition 5. *Design is permutation invariant if and only if it is a subset design.*

Proof. Invariance of subset designs is clear. Conversely, given a permutation invariant design and an integer $s \geq 1$, consider the subgraph induced by all degree-s redundant nodes and their neighborhoods. By permutation invariance this subgraph must contain all k primary nodes and itself be permutation invariant (since automorphisms preserve degrees of nodes). Therefore, every s-subset of the primary nodes must appear as a neighborhood of n redundant nodes for some integer n. This degree-s subgraph corresponds to merging of n copies of $S(k, s)$ and the original graph is a merging of degree-s subgraphs. □

The number of redundant nodes used in subset designs is large and therefore it should be able to correct many defects. We will find sharp estimates for the defect-correcting properties of subset designs later (Proposition 14 below), but for now we can give a simple order-of-magnitude result:

Proposition 6. *Fix alphabet $\mathcal{X}$ and size $s \geq 1$. As $k \to \infty$ the design $S(k, s)$ corrects $t = \Theta(k^{s-1})$ defects.*

Proof. We know that $t = O(k^{s-1})$, since each primary node has at most $O(k^{s-1})$ neighbors. To show that $t = \Omega(k^{s-1})$, fix a labeling of primary nodes to elements of $\mathcal{X}$. Consider the following procedure for labeling redundant nodes. First we declare an element of $x \in \mathcal{X}$ to be *rare* if the number of primary nodes labeled x is less than $\frac{k}{cq}$ where c is a suitable constant. Now each redundant node is labeled the value $x \in \mathcal{X}$ if either all of its neighbors have label x or if x is the only rare label in their neighborhood. To see that this is an labeling that corrects $\Omega(k^{s-1})$ defects, simply notice that a non-rare primary node labeled x has $\Omega(k^{s-1})$ neighbors with all neighbors are labeled x. Similarly, each rare-labeled primary node has $\Omega(k^{s-1})$ neighboring redundant nodes that connect to only one rare label since the total number of rare-labeled primary nodes is at most k/c. □

As we will see, subset designs turn out to be optimal for achieving the boundary of $\mathcal{R}_\infty$. In other words, they can be tuned to get the optimal speed of growth for redundancy and wiring complexity as $t \to \infty$.

IV. Bounds for finite t

In this section we prove a number of basic results, which will lead to the proof of Theorem 3. We will first show how two basic operations, copying and merging, can be used to combine existing designs into a new design with certain properties. Using these operations, we then proceed to prove the claims in Proposition 2.

Then, with the convexity results from Proposition 2, we show achievability for Theorem 3. Following the achievability, we show the converse for Theorem 3 which uses a technique we call covering.

A similar result for the achievable region for ternary alphabet is stated at the end of the section.

A. Two basic operations on designs

Definition 6 (Copying). *A distinct union of a collection of designs G_j is denoted by $\sum_j G_j$.*

Proposition 7 (Copying). *Consider $(k_j, m_j, t, E_j)_q$-designs G_j. Then $\sum_j G_j$, forms a $(\sum k_j, \sum m_j, t, \sum E_j)_q$-design.*

Proof is self-evident. We note here that the value of ε and ρ for G_1+G_2 is a convex combination of those of G_j. That is

$$\rho = \frac{k_1}{(k_1+k_2)t}\rho_1 + \frac{k_2}{(k_1+k_2)t}\rho_2 \tag{8}$$

$$\varepsilon = \frac{k_1}{(k_1+k_2)t}\varepsilon_1 + \frac{k_2}{(k_1+k_2)t}\varepsilon_2\,. \tag{9}$$

where ρ_j and ε_j refer to $\frac{m}{kt}$ and $\frac{E}{kt}$ of G_j respectively and k_j is the number of primary nodes in G_j.

Proposition 8 (Merging). *Consider $(k, m_j, t_j, E_j)_q$-designs G_j and $G = \bigvee_j G_j$ (see Definition 3). G is a $(k, \sum_j m_j, \sum_j t_j, \sum_j E_j)_q$-design.*

(Note that it is possible that the merged design $G = \bigvee_j G_j$ can correct more than $\sum_j t_j$ defects.) Again, the proof is self-evident. As an example, we note that merging a design with itself, i.e., $G \vee G$, doubles all the parameters except k. However, that the wiring complexity and redundancy stays constant. This will be the basis for showing convexity of $\mathcal{R}_\infty$, cf. (4).

B. Proof of Proposition 2

With the help of the two basic operations, we can prove the convexity of $\mathcal{R}_t$ and $\mathcal{R}_\infty$, as well as other properties claimed in Proposition 2.

Proof of Proposition 2. Claim 1. From the definition of closure, $(\varepsilon, \rho) \in \mathcal{R}_t$ if and only if there is a sequence of points $\{(\varepsilon_i, \rho_i)\}_i \in \mathcal{R}_t$ approaching (ε, ρ). Each (ε_i, ρ_i) must be associated with some design G_i that is a $(k_i, m_i, t, E_i)_q$-design, where $m_i = \rho_i k_i t$ and $E_i = \varepsilon_i k_i t$. To show that $k, m, E \to \infty$, we can copy G_i with itself n_i times, where n_i is chosen so that $n_i k_i, n_i m_i, n_i E_i \to \infty$.

Claim 2. If $(\varepsilon, \rho) \in \mathcal{R}_t$, then where exists a point $(\hat{\varepsilon}, \hat{\rho}) \in \mathcal{R}_t$ in an arbitrarily small neighborhood of (ε, ρ) which corresponds to a $(k, \hat{\rho}kt, t, \hat{\varepsilon}kt)_q$-design G where k is arbitrarily large. We can always add more redundant nodes or more edges (this is possible since $\hat{\rho}kt$ can be arbitrarily large and adding a finite number of redundant nodes does not change the redundancy) to G to get a design with parameters (ε', ρ').

Claim 3. This holds using copying from Proposition 7.

If a pair of values (ε_1, ρ_1) and (ε_2, ρ_2) are in $\mathcal{R}_t$, there are sequences $(\varepsilon_{1,i}, \rho_{1,i}) \to (\varepsilon_1, \rho_1)$ and $(\varepsilon_{2,i}, \rho_{2,i}) \to (\varepsilon_2, \rho_2)$, where for each i there exists a $(k_{1,i}, \rho_{1,i}k_{1,i}t, t, \varepsilon_{1,i}k_{1,i}t)_q$-design $G_{1,i}$ and a $(k_{2,i}, \rho_{2,i}k_{2,i}t, t, \varepsilon_{2,i}k_{2,i}t)_q$-design $G_{2,i}$. For any $0 \le \alpha \le 1$, we can find a sequence of rational numbers $\alpha_i = \frac{p_i}{q_i}$ where $p_i, q_i \in \mathbb{Z}_+$ and $\alpha_i \to \alpha$. The copy $k_{2,i}p_i G_{1,i} + k_{1,i}(q_i - p_i)G_{2,i}$ achieves the point $(\varepsilon_i, \rho_i) = (\alpha_i\varepsilon_{1,i} + (1-\alpha_i)\varepsilon_{2,i}, \alpha_i\rho_{1,i} + (1-\alpha_i)\rho_{2,i})$ in $\mathcal{R}_t$ and $(\varepsilon_i, \rho_i) \to (\alpha\varepsilon_1 + (1-\alpha)\varepsilon_2, \alpha\rho_1 + (1-\alpha)\rho_2) = (\varepsilon, \rho)$.

Claim 4. Any point (ε, ρ) in $\mathcal{R}_\infty$ and any point in $\text{closure}\,\{\bigcup_{t=1}^{\infty} \mathcal{R}_t\}$ must both be the limit of some sequence of $(k_i, m_i, t_i, E_i)_q$-designs. To see that $\mathcal{R}_\infty = \limsup \mathcal{R}_t$, by merging in Proposition 8, for any t, we have $\mathcal{R}_t \subset \mathcal{R}_{2t} \subset \mathcal{R}_{4t} \subset \mathcal{R}_{8t} \ldots$.

Claim 5. This holds using merging from Proposition 8. Given two designs G_1 and G_2, where G_1 is a $(k_1, \rho_1 k_1 t_1, t_1, \varepsilon_1 k_1 t_1)_q$-design and G_2 is a $(k_2, \rho_2 k_2 t_2, t_2, \varepsilon_2 k_2 t_2)_q$-design, if we want to create a graph G with the parameter $(\alpha\varepsilon_1 + (1-\alpha)\varepsilon_2, \alpha\rho_1 + (1-\alpha)\rho_2)$ for $\alpha = \frac{p}{q}$ where $p, q \in \mathbb{Z}_+$, then we can let

$$G = pk_2 \left(\bigvee_{i=1}^{t_2} G_1 \right) \vee (q-p)k_1 \left(\bigvee_{i=1}^{t_1} G_2 \right) . \tag{10}$$

From here on, the proof proceeds similarly to the proof of Claim 3. □

C. Elementary achievability

The line of points between $(1,1)$ and $(q,0)$ is achievable by interpolating between the repetition design $K(1,t)$ and the complete design $K(k,qt)$.

Proposition 9. *The following region is achievable for any $t \geq 1$ and $q \geq 2$:*

$$\mathcal{R}_t^{(K)} \triangleq \{(\varepsilon, \rho) : \varepsilon \geq q + (1-q)\rho, \varepsilon \geq 1, \rho \geq 0\} \tag{11}$$

Furthermore, every point such that $(\varepsilon - 1)$ is a multiple of $(q-1)$ can be achieved via a design with constant degree ε primary nodes.

Proof. Note that corner points $(1,1)$ and $(q,0)$ are achieved by the repetition design and the complete design, respectively. By Proposition 2 the region $\mathcal{R}_t$ is convex and hence must contain $\mathcal{R}_t^{(K)}$. The rational points on this boundary are achieved by $r_1 K(1,t) + r_2 K(k,qt)$ for suitably chosen r_1, r_2.

In order to get primary node regularity, we can combine the repetition design and complete design by merging. Find two integers t_1, t_2 where $t_1 + t_2 = t$. The combination $kK(1,t_1) \vee K(k,qt_2)$ also achieves the boundary point at $\varepsilon = (t_1 + qt_2)/t$ as $k \to \infty$. This proves the last sentence of the Proposition 9. □

For $q = 2$ and $t = 1, 2$ the region $\mathcal{R}_t^{(K)}$ is optimal and plotted in Figure 2 as the achievable region.

D. Covering converse

Theorem 10. *Fix $q = |\mathcal{X}|$, t and suppose $(\varepsilon, \rho) \in \mathcal{R}_t$. Then there exists $\pi_t, \pi_{t+1}, \ldots, \pi_{qt} \geq 0$ satisfying*

$$\frac{1}{t} \sum_{j=t}^{qt} j\pi_j \leq \varepsilon \tag{12}$$

$$\sum_{j=t}^{qt} \pi_j = 1 \tag{13}$$

$$\sum_{j=t+1}^{qt} \pi_j \log_q \lfloor j/t \rfloor \geq 1 + (t-1)\pi_t - \rho t \, . \tag{14}$$

In other words the smallest achievable ε for a given ρ is lower bounded as

$$\varepsilon^*(\rho, t) \geq \min \left\{ \frac{1}{t} \sum_{j=t}^{qt} j\pi_j : \pi_j \geq 0 \text{ satisfy (13)-(14)} \right\} \tag{15}$$

Proof. Notice that every primary node clearly should have degree at least t. Let us define $\pi_j, j = t, t+1, ..., qt-1$ to be the fraction of primary nodes of degree j. Define π_{qt} to be the fraction of primary nodes of degree qt or larger. The fact that this satisfies (12)-(13) is obvious. We only need to show (14).

To that end, for each labeling $r^m \in \mathcal{X}^m$ of redundant nodes let $\mathcal{G}_t(r^m)$ be the set of primary node labelings for which conditions of Definition 1 are satisfied (we say that r^m covers $\mathcal{G}_t(r^m)$ of the labelings). It is clear that the design is t-defect correcting if and only if

$$\left| \bigcup_{r^m \in \mathcal{X}^m} \mathcal{G}_t(r^m) \right| = |\mathcal{X}|^k = q^k \,. \tag{16}$$

We are aiming to apply the union bound to the right-hand side to get inequality (14). Before doing so we make the following observation.

Two primary nodes of degree t should have disjoint neighborhoods (otherwise labeling them different values clearly violates Definition 1). Thus $\mathcal{G}_t(r^m)$ is empty unless each such neighborhood has a constant label. This shows that for the $tk\pi_t$ redundant nodes we are restricted to only $q^{k\pi_t}$ choices, while the rest contribute $q^{m-tk\pi_t}$ more choices.

Given any of the $q^{m-(t-1)k\pi_t}$ choices of r^m we can estimate $|\mathcal{G}_t(r^m)|$ from above by assuming that each primary node of degree d can take any of the $\lfloor d/t \rfloor$ values in $\mathcal{X}$ while still satisfying the t-wise coverage condition of Definition 1. This yields

$$|\mathcal{G}_t(r^m)| \leq \prod_{j=t}^{qt} \lfloor j/t \rfloor^{k\pi_j} \,, \tag{17}$$

and thus applying the union bound to (16), we get (14). □

Proof of Theorem 3. Achievability follows from Proposition 9. The converse is just evaluation of 15. For $t = 1$ and $t = 2$, we get that (15) simplifies to $\varepsilon \geq 2 - \rho$. □

Remark 1. *While the bound (15) is tight for $t = 1$ and $t = 2$, it is not tight in general. It however allows us to make a general conclusion: since the bound is piecewise linear, it follows that the slope of $\mathcal{R}_t$ at the point $(qt, 0)$ of minimal redundancy is non-zero. It is also the best bound known to us for values of ε near qt.*

In the next section, we will show a bound that is better for ε away from q and for large t. This converse outperforms the covering converse (Theorem 10) at certain ρ even for $q = 2$ and $t = 3$.

E. Ternary alphabet and $t = 1$

Further progress on computing regions $\mathcal{R}_t$ for values of $q > 2$ seems to require finer arguments on graph structure. We can show the following result for $q = 3$ but the proof requires significant casework.

Theorem 11. *For $q = 3$ and $t = 1$ we have*

$$\mathcal{R}_1 = \{(\varepsilon, \rho) : \varepsilon \geq 1 - 2\rho, \rho \geq 0\} \tag{18}$$

and is achievable by the simple interpolation (11)*.*

We give the proof in Appendix A.

V. Fundamental limit for $t \to \infty$

Recall that as $t \to \infty$ the fundamental limit $\mathcal{R}_\infty \stackrel{\triangle}{=} \limsup \mathcal{R}_t$ can be characterized as the set of wiring complexity-redundancy pairs, namely

$$\varepsilon \stackrel{\triangle}{=} \frac{E}{kt}, \quad \rho \stackrel{\triangle}{=} \frac{m}{kt}\,. \tag{19}$$

over all values of t. (See Proposition 2.) The goal of this section is to prove the following result, that generalizes the binary version stated earlier in Theorem 4.

Theorem 12. *Fix alphabet $|\mathcal{X}| = q$. The region $\mathcal{R}_\infty$ defined in* (3) *is the closure of the set of points (ε, ρ), parameterized by the distribution P_S on a finite support of $\mathbb{Z}_+$, and*

$$\varepsilon = \frac{\mathbb{E}\left[S\right]}{F(P_S)}, \quad \rho = \frac{1}{F(P_S)}\,, \tag{20}$$

$$F(P_S) \stackrel{\triangle}{=} \min_{P_X} \max_{P_{Y|\underline{L}}} \min_{j \in [q]} \frac{1}{P_X(j)} \mathbb{E}\left[L_j \mathbb{1}\{Y = j\}\right] \tag{21}$$

where $\mathbb{E}\left[\cdot\right]$ is computed over random variables $S \in \mathbb{Z}_+, X \in [q], \underline{L} = (L_1, \ldots, L_q) \in \mathbb{Z}_+^q, Y \in [q]$ with joint distribution

$$P_{S,\underline{L},Y}(s, \underline{\ell}, y) \stackrel{\triangle}{=} P_S(s) P_{\underline{L}|S}(\underline{\ell}|s) P_{Y|\underline{L}}(y|\underline{\ell})\,. \tag{22}$$

where[2]

$$P_{\underline{L}|S}(\underline{\ell}|s) \stackrel{\triangle}{=} \binom{s}{\ell_1, \cdots, \ell_q} \prod_{j=1}^{q} P_X(j)^{\ell_j}\,. \tag{23}$$

We start the section by proving relevant properties of $F(P_S)$. We then use these properties to prove the achievability (i.e., upper bound) of Theorem 12. Next, we present a symmetrization property which is the key idea of the converse argument of Theorem 12. Putting these elements together gives the complete proof.

Following the proof, we present a number of observations about Theorem 12. These include a section about how we compute Theorem 12 numerically and a section on our result about the achievable regions for designs where k is finite, but t and m tend to infinity. This result follows from the proof of Theorem 12. We also discuss how the Hamming block is optimal in this context.

A. Auxiliary results about $F(P_S)$

Before proceeding further, we need to describe some properties of $F(P_S)$ and related quantities:

[2] $P_{\underline{L}|S}$ is the multinomial distribution, Mult$(s, [P_X(1), \cdots, P_X(q)])$

Definition 7 (Finitary F)**.** *We define $F_{k,n}(P_S)$ and $F_k(P_S)$ as follows:*

$$F_{k,n}(P_S) \stackrel{\triangle}{=} \min_{P_X \in \frac{1}{k}\mathbb{Z}} \max_{P_{Y|\underline{L}^{(k)}} \in \frac{1}{n}\mathbb{Z}} \min_{j\in[q]} \frac{1}{P_X(j)} \mathbb{E}\left[L_j^{(k)} \mathbb{1}\{Y=j\}\right], \tag{24}$$

$$F_k(P_S) \stackrel{\triangle}{=} \min_{P_X \in \frac{1}{k}\mathbb{Z}} \max_{P_{Y|\underline{L}^{(k)}}} \min_{j\in[q]} \frac{1}{P_X(j)} \mathbb{E}\left[L_j^{(k)} \mathbb{1}\{Y=j\}\right], \tag{25}$$

where $\mathbb{E}[\cdot]$ is computed over random variables $S \in [k], X \in [q], \underline{L}^{(k)} = (L_1, \ldots, L_q) \in \mathbb{Z}_+^q, Y \in [q]$ with joint distribution

$$P_{S,\underline{L}^{(k)},Y}(s, \underline{\ell}, y) \stackrel{\triangle}{=} P_S(s) P_{\underline{L}^{(k)}|S}(\underline{\ell}|s) P_{Y|\underline{L}^{(k)}}(y|\underline{\ell}). \tag{26}$$

where[3]

$$P_{\underline{L}^{(k)}|S}(\underline{\ell}|s) \stackrel{\triangle}{=} \frac{\binom{kP_X(1)}{\ell_1} \cdots \binom{kP_X(j)}{\ell_j} \cdots \binom{kP_X(q)}{\ell_q}}{\binom{k}{s}}. \tag{27}$$

Note that the definition of $F_{k,n}$ is similar to that of $F(P_S)$, see (21), but with two changes: 1) P_X and $P_{Y|\underline{L}^{(k)}}$ (instead of $P_{Y|\underline{L}}$) are required to have denominators $\frac{1}{k}$ and $\frac{1}{n}$, respectively; and b) $P_{\underline{L}^{(k)}|S}$ is (multivariate) hypergeometric, instead of multinomial.

Proposition 13. *For any P_S with finite expectation we have*

$$F_k(P_S) - \frac{\mathbb{E}[S]}{n} \le F_{k,n}(P_S) \le F_k(P_S). \tag{28}$$

Also, there exists a sequence $\epsilon_k \to 0$ such that for any P_S on $\mathbb{Z}_+$ with finite third moment we have

$$|F_k(P_S) - F(P_S)| \le \frac{\mathbb{E}[S^3]}{2k} + \epsilon_k \tag{29}$$

See Appendix B for proofs.

B. Subset design achievability and upper bound

The next proposition give bounds on the performance of subset designs. [4]

Proposition 14. *Let $q = |\mathcal{X}|$ and fix $k \in \mathbb{Z}$. Let $G = \bigvee_{i=1}^n G'$, where G' is a subset design with $P_S(s)$ as the proportion of redundant nodes with degree s for $s \in [k]$. If G is a $(k, m, t, E)_q$-design, where $E = m\mathbb{E}[S]$ and t is the maximal number of defects G can correct, then*

$$\frac{m}{k} F_{k,n}(P_S) \le t \le \frac{m}{k} F_k(P_S) \tag{30}$$

Proof. First we show the upper bound that $t \le \frac{m}{k} F_k(P_S)$.

Consider any labeling $w^k \in \mathcal{X}^k$ of the k primary nodes of G. Let the frequency which each label occurs in the labeling have empirical distribution P_X. Given this labeling, we define the *type* of each redundant node v to be $\underline{\ell} = (\ell_1, \cdots, \ell_q)$, where ℓ_j is the number of primary nodes with label j which is a neighbor of redundant node v. (If the degree of the redundant node is

[3] $P_{\underline{L}^{(k)}|S}$ is the multivariate hypergeometric distribution, HyperGeom$(s, k, [P_X(1), \cdots, P_X(q)])$

[4] This proposition initially used random coding as an argument. The random coding as since been replaced.

s, then $\sum_{i=1}^{q} \ell_i = s$.) The proportion of degree s redundant nodes in G with type $\underline{\ell}$ is

$$P_{\underline{L}^{(k)}|S}(\underline{\ell}|s) = \frac{\binom{kP_X(1)}{\ell_1} \cdots \binom{kP_X(q)}{\ell_q}}{\binom{k}{s}}. \tag{31}$$

Now, for any choice of labeling $r^m \in \mathcal{X}^m$ of the m redundant nodes, let $P_{Y|\underline{L}^{(k)}}(j|\underline{\ell})$ represent the proportion (empirical distribution) of redundant nodes of type $\underline{\ell}$ which are labeled j. For each label j, we can count the *average* number of matching redundant node neighbors a primary node u with label j has by summing up all the edges between primary and redundant nodes both with label j, and then dividing this by the total number of primary nodes with label j. This average is

$$\tilde{t} = \frac{1}{kP_X(j)} \sum_s mP_S(s) \sum_{\underline{\ell}} P_{\underline{L}^{(k)}|S}(\underline{\ell}|s)\ell_j P_{Y|\underline{L}^{(k)}}(j|\underline{\ell}) \tag{32}$$

$$= \frac{m}{k} \frac{1}{P_X(j)} \mathbb{E}\left[L_j \mathbb{1}\{Y = j\}\right] \tag{33}$$

The label j where this average is lowest determines the upper bound on the number of defects G with labeling w^k and r^m can correct. This upper bound is given by $\min_j \tilde{t}$. We have the freedom to pick the redundant node labeling r^m with the empirical distribution $P_{Y|\underline{L}^{(k)}}$ which maximizes the average. The defect correcting number needs to hold for all possible w^k, so the empirical distribution P_X which gives the lowest value of $\max_{P_{Y|\underline{L}^{(k)}}} \min_j \tilde{t}$ determines t. This gives the upper bound on t.

We now show the lower bound $\frac{m}{k} F_{k,n}(P_S) \leq t$.

Given any labeling $w^k \in \mathcal{X}^k$ of the primary nodes with empirical distribution P_X, let

$$P_{Y_n|\underline{L}^{(k)}} = \underset{P_{Y|\underline{L}^{(k)}} \in \frac{1}{n}\mathbb{Z}}{\operatorname{argmax}} \min_{j \in [q]} \frac{1}{P_X(j)} \mathbb{E}\left[L_j \mathbb{1}\{Y = j\}\right]. \tag{34}$$

For each $\underline{\ell}$, $P_{Y_n|\underline{L}^{(k)}}(j|\underline{\ell}) = \frac{c_j}{n}$ for some $c_j \in \mathbb{Z}_+ \cup 0$, and $\sum_j c_j = n$. Because G is a merging of n copies of G', we can partition the copies of G' in G into sets of size $c_1, ..., c_q$. The jth set is a set of c_j copies of G'. Then label all redundant nodes of type $\underline{\ell}$ in the jth set the value j. We can determine that each primary node u with label j has a total of $P_S(s)m\frac{\ell_j}{P_X(j)k} P_{\underline{L}^{(k)}|S}(\underline{\ell}|s)$ redundant nodes of type $\underline{\ell}$ in its neighborhood. This redundant node labeling scheme assigns exactly $P_{Y_n|\underline{L}^{(k)}}(j|\underline{\ell})$ of these neighbors the label j.

Repeat this labeling process for each redundant node type $\underline{\ell}$. Summing over all $\underline{\ell}$ and all s will get that the total number of redundant nodes with label j in the neighborhood of primary node u is $\sum_{s,\underline{\ell}} P_S(s)m\frac{\ell_j}{P_X(j)k} P_{\underline{L}^{(k)}|S}(\underline{\ell}|s) P_{Y_n|\underline{L}^{(k)}}(j|\underline{\ell})$.

Using this scheme, G can correct at least

$$t \geq \min_{P_X \in \frac{1}{k}\mathbb{Z}} \min_{j \in [q]} \sum_{s,\underline{\ell}} P_S(s)m\frac{\ell_j}{P_X(j)k} P_{\underline{L}^{(k)}|S}(\underline{\ell}|s) P_{Y_n|\underline{L}^{(k)}}(j|\underline{\ell}) \tag{35}$$

$$= \frac{m}{k} \min_{P_X \in \frac{1}{k}\mathbb{Z}} \max_{P_{Y|\underline{L}^{(k)}} \in \frac{1}{n}\mathbb{Z}} \min_{j \in [q]} \frac{1}{P_X(j)} \mathbb{E}\left[L_j \mathbb{1}\{Y = j\}\right] \tag{36}$$

$$= \frac{m}{k} F_{k,n}(P_S) \tag{37}$$

defects. □

C. Converse and proof of Theorem 12

The converse needed to show Theorem 12 is surprisingly simple. The main idea is the following:

Proposition 15 (Symmetrization)**.** *If there exists a $(k, m, t, E)_q$-design then there exists a permutation-invariant $(k, m \cdot k!, t \cdot k!, E \cdot k!)_q$-design.*

Proof. Let G be a $(k, m, t, E)_q$-design. We will merge G exactly $k!$ number of times. The key is that each copy will be merged by identifying with a permutation of the original primary nodes.

Start with an ordering of the primary nodes in the design G. For each $\sigma \in S_k$ (the full symmetric group of k elements), let G_σ be isomorphic to the design G, with the order of its primary nodes transformed by σ. Then merge G_σ for all $\sigma \in S_k$ identifying primary nodes with the same order.

Let the result be

$$G_{\mathrm{PERM}} = \bigvee_{\sigma \in S_k} G_\sigma \,. \tag{38}$$

G_{PERM} is constructed to be permutation invariant. (For any redundant node v in G, if v has degree s, every set of s nodes in G_{PERM} needs to be connected together by a copy of v.) By Proposition 8 G_{PERM} is a $(k, m \cdot k!, t \cdot k!, E \cdot k!)_q$-design. □

In view of Propositions 5 and 15, we see that in terms of the values $\frac{E}{kt}, \frac{m}{kt}$ every design on k primary nodes is at most as good as a subset design on k primary nodes (meaning the pair of values an arbitrary design achieves has the same or worse trade-off). Performance of the latter was completely characterized by Proposition 14. Now we can combine the results to prove Theorem 12.

Proof. Proof of Theorem 12

Achievability: Fix $P_S \in \mathbb{Q}$ with finite support. For each k and n, it is always possible to construct a subset design G' on k primary nodes where the proportion of redundant nodes of degree s are given by $P_S(s)$. Let $G = \bigvee_{i=1}^{n} G'$ so that by Proposition 14, subset design G is a $(k, m, t, E)_q$-design so that $\frac{tk}{m} \geq F_{k,n}(P_S)$ and $E = m\mathbb{E}[S]$. Since $F_{k,n}(P_S) \to F(P_S)$, there must exist a sequence of subset designs G_i which are $(k_i, m_i, t_i, E_i)_q$-designs where

$$\varepsilon = \frac{E_i}{k_i t_i} \to \frac{\mathbb{E}[S]}{F(P_S)}, \rho = \frac{m_i}{k_i t_i} \to \frac{1}{F(P_S)} \,. \tag{39}$$

Thus

$$(\frac{\mathbb{E}[S]}{F(P_S)}, \frac{1}{F(P_S)}) \in \mathcal{R}_\infty \,. \tag{40}$$

Since $F(P_S)$ is continuous in P_S, (40) holds for any P_S with finite support.

Converse: For any design G which is a $(k, m, t, E)_q$-design, there exists a subset design G' which is a $(k, m \cdot k!, t \cdot k!, E \cdot k!)_q$-design by Proposition 15. Let P_S be so that $P_S(s)$ represents the proportion of redundant nodes in G' with degree s. Then $E = m\mathbb{E}[S]$. Let t' be the number of defects G' can correct. Using Proposition 14 and $F_k(P_S) \leq F(P_S)$ (cf. Lemma 23 in Appendix C),

$$t \cdot k! \leq t' \leq \frac{m \cdot k!}{k} F_k(P_S) \leq \frac{m \cdot k!}{k} F(P_S) \,. \tag{41}$$

Then for design G, $\varepsilon = \frac{E}{tk} \geq \frac{\mathbb{E}[S]}{F(P_S)}$ and $\rho = \frac{m}{tk} \geq \frac{1}{F(P_S)}$. Thus, the limit of $(\frac{E_i}{k_i t_i}, \frac{m_i}{k_i t_i})$ for any

sequence of $(k_i, m_i, t_i, E_i)_q$-designs must be in the closure of $(\frac{\mathbb{E}[S]}{F(P_S)}, \frac{1}{F(P_S)})$ for all P_S with a finite support. □

D. Observations about Theorem 12

1) Threshold solution: The optimal value for $P_{Y|\underline{L}}$ tells us what the optimal labeling of redundant nodes should be. It turns out that for most values of $\underline{\ell}$, $P_{Y|\underline{L}}(j|\underline{\ell})$ is either 0 or 1.

For an illustration of this, consider the binary alphabet case and the design $S(k, s)$. The types are $\underline{\ell} = (\ell_0, \ell_1)$. Given any empirical distribution P_X of the primary node labels, the optimal labeling of the redundant nodes must be so that redundant nodes with larger values of ℓ_1 are assigned the label 1 instead of redundant nodes with smaller values of ℓ_1. Otherwise, we can always swap the labelings and increase the number of defects corrected. In fact, even when there are multiple subsets sizes, it is possible to find an optimal solution where the value of $P_{Y|\underline{L}}$ depends only on the ratio of ℓ_0 to ℓ_1.

Proposition 16. *For $\mathcal{X} = \{0, 1\}$, the solutions $P_{Y|\underline{L}}$ which attain the maximum in (21) must have the following form*

$$P_{Y|\underline{L}}(0|\underline{\ell}) = \begin{cases} 1 & \text{if } \frac{\ell_0}{\ell_0+\ell_1} > \gamma \\ 0 & \text{if } \frac{\ell_0}{\ell_0+\ell_1} < \gamma \\ \mu(\ell_0 + \ell_1) & \text{if } \frac{\ell_0}{\ell_0+\ell_1} = \gamma. \end{cases} \tag{42}$$

where $\gamma \in [0, 1]$ and $\mu(s) \in [0, 1]$ for each $s \in \mathbb{Z}_+$. [5]

(See Appendix D for proof.) Generalizing to larger alphabet sizes, the space of all possible $\underline{\ell}$ will be partitioned into q pieces. The interior of each piece will have all types assigned to the same label, that is $P_{Y|L}(j|\underline{\ell}) = 1$ for some j. $\underline{\ell}$ on the boundary may be split between 2 or more values.

Notice that in light of Theorem 12 and Proposition 16, computing the optimal redundant node labeling for subset designs given a fixed primary node labeling is easy. For general designs, this is NP-Hard.

2) Worst-case P_X: The worst-case distribution of primary node labels which gives the result in Theorem 12 is not obvious, even in the binary alphabet case. When $\mathcal{X} = \{0, 1\}$, we can easily determine that for subset designs $S(k, s)$ with even s, the worst-case P_X is when $P_X(0) = P_X(1) = \frac{1}{2}$. However, when s is odd, this is not true. When $s = 3$, the worst-case P_X is determined by a solution to a cubic polynomial. When a merging of different subset designs are used or a larger alphabet is used, it is unclear how to find the worst-case P_X analytically. This makes finding the worst-case P_X the main difficulty in evaluating the optimization equation in Theorem 12. (The equation is non-convex in P_X.)

E. Numerical upper and lower bounds

Since the optimization presented in Theorem 4 is difficult to evaluate exactly, instead, we give an approximation for the boundary by establishing computable almost tight upper and lower bounds for when $q = 2$. The details can be found in Appendix E and a comparison is presented on Fig. 6. As can be seen, the gap between the bounds is on the order of 10^{-3} and virtually indistinguishable on the plot. The best known achievable point in $\mathcal{R}_\infty$ for selected fixed values of $\mathbb{E}[S]$ are given in Table I. These points are found by searching and using weights from the converse bound method in Appendix E

[5] There is not necessarily a unique solution for $\mu(s)$. One such solution has $\mu(s_a) = \mu(s_b)$ for all s_a, s_b.

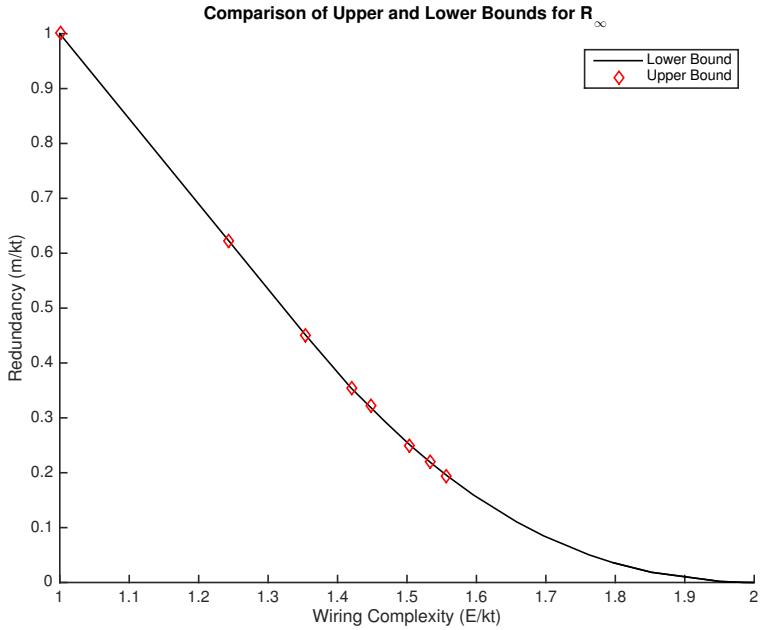


Fig. 6. Approximate converse bound compared with achievable points.

TABLE I
ACHIEVABLE POINTS

$\mathbb{E}[S]$	Support set with corresponding P_S	Point (ε, ρ) in $\mathcal{R}_\infty$
2	[1,3,4,5] , [0.62, 0.21, 0.10, 0.07]	(1.24, 0.61)
3	[1,3,4,5] , [0.24, 0.41, 0.20, 0.14]	(1.35, 0.45)
4	[3,4,5,6,7] , [0.52, 0.21, 0.13, 0.02, 0.12]	(1.42, 0.35)
5	[3,4,5,7] , [0.31, 0.23, 0.28, 0.18]	(1.40, 0.29)
6	[5,6,7,9] , [0.45, 0.31, 0.14, 0.10]	(1.47, 0.29)
7	[5,6,7,8,9] , [0.35, 0.01, 0.13, 0.32, 0.19]	(1.53, 0.22)
8	[7,8,9,11] , [0.40, 0.36,0.16,0.08]	(1.56, 0.19)

We observed the following effects about designs near the boundary of $\mathcal{R}_\infty$ while experimenting with Theorem 4:

- 4 or 5 subset sizes make up most of the design
- Odd number subset sizes are more common
- The subset sizes which make up most of the design are consecutive, possibly skipping even subset sizes

F. Results for finite k

To develop the proof for Theorem 12, we showed intermediate results on designs with k primary nodes and observed what occurs when $k \to \infty$. However, we can use the intermediate results to determine the achievable regions for designs on k primary nodes.

Definition 8. *For fixed q and $k \in \mathbb{Z}_+$, we define the region $\mathcal{R}_\infty^k$ as the closure of the set of all achievable pairs $(\frac{E}{kt}, \frac{m}{kt})$:*

$$\mathcal{R}_\infty^k \stackrel{\triangle}{=} \text{closure}\left\{\left(\frac{E}{kt}, \frac{m}{kt}\right) : \exists (k, m, t, E)_q\text{-design}\right\} \tag{43}$$

Similar to regions $\mathcal{R}_t$ and $\mathcal{R}_\infty$, the region $\mathcal{R}_\infty^k$ is convex. We can apply the proof for Claim 5 in Proposition 2 replacing the expression (10) with

$$G = p\left(\bigvee_{i=1}^{t_2} G_1\right) \vee (q-p)\left(\bigvee_{i=1}^{t_1} G_2\right) \tag{44}$$

to show this.

Claims 1 and Claim 2 of Proposition 2 also hold for $\mathcal{R}_\infty^k$.

Theorem 17. *Fix alphabet $|\mathcal{X}| = q$. The region $\mathcal{R}_\infty^k$ defined in (43) is a closure of the set of points (ε, ρ), parameterized by the distribution P_S on $[k]$, where*

$$\varepsilon = \frac{\mathbb{E}[S]}{F_k(P_S)}, \quad \rho = \frac{1}{F_k(P_S)}, \tag{45}$$

and $F_k(\cdot)$ is defined in (25).

Proof. The achievability and converse of this theorem follows from Proposition 14 (with $F_k(P_S)$ is continuous in P_S) and Proposition 15 respectively. □

Using (45), we can plot the achievable region $\mathcal{R}_\infty^3$ when $q = 2$ (see Fig. 7.) The most salient aspect of $\mathcal{R}_\infty^3$ is that the point achievable by the Hamming block (see Fig. 5(a)) is a corner point of this region. It is the only corner other than the usual corner point (t, t) achieved by the repetition design.

Corollary 18 (Hamming block corner point)**.** *The value given by the Hamming block is a corner point of $\mathcal{R}_\infty^3$ for $\mathcal{X} = \{0, 1\}$.*

The proof for Corollary 18 and the methods used to calculate $\mathcal{R}_\infty^3$ are in Appendix F. The implication of the result is that for any design on $k = 3$ primary nodes, no design has a better trade-off between redundancy $\frac{m}{kt}$ and wiring complexity $\frac{E}{kt}$ than the Hamming block, even if we allow the design to have arbitrary many edges and redundant nodes.

VI. Discussion

We conclude with a discussion of some implications of our results, some extensions and future work.

A. Implications on practical designs

The result for $\mathcal{R}_1$ and $\mathcal{R}_2$ (for $q = 2$) demonstrates that for correcting small defects, the best solution in the limit of a large number of primary nodes is a linear combination of two basic designs, the repetition design and the complete design. (Though this design is not optimal for finite k. Slight improvements can be made to this basic design.)

Theorem 12 gives a result for asymptotic t, and while practically no application is going to need to correct asymptotically many defects, the region defined by the result gives a converse

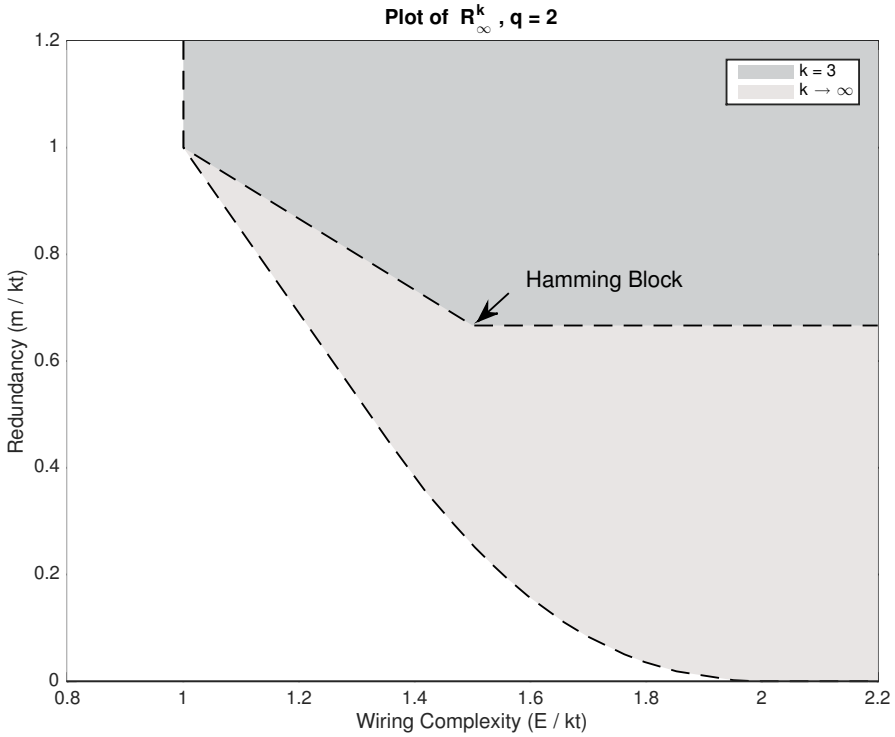


Fig. 7. Region $\mathcal{R}_\infty^3$ (see (43)) compared with $\mathcal{R}_\infty$ for $q = 2$.

bound for $\mathcal{R}_t$ for all finite t by virtue of Claim 4 from Proposition 2. All regions $\mathcal{R}_t$ must lie between $\mathcal{R}_1$ and $\mathcal{R}_\infty$, approaching the latter as $t \to \infty$. Hence Theorem 12 describes the fundamental limit for the trade-off between redundancy and wiring complexity.

The numerical results for Theorem 12 for when $q = 2$ imply that the designs which are close to optimal for large t use redundant nodes with a limited set of degrees. The best achievable points found for $\mathcal{R}_\infty$ for fixed values of $\mathbb{E}[S]$ each use redundant nodes with degrees within 2 or 3 values of $\mathbb{E}[S]$.

Results for $\mathcal{R}_\infty^k$ define what is optimal for finite k in terms of the number of defects correctable per use of redundancy and edges. We know exactly what this regions looks like for $k = 3$ and can determine that the Hamming block is in fact the optimal design. Evaluating $\mathcal{R}_\infty^k$ for larger values of k gives exactly what trade-offs are realizable.

Also note that in all the result, the optimal trade-off is obtainable by a design which has regular primary node degree. Not only that, but finding the best labeling of redundant nodes is easy to compute for these known optimal designs.

B. Comparison to other models for defect tolerance

This paper studies defect-tolerance model where steps proceed as follows:

a. bipartite graph is designed;
b. primary nodes get q-ary labeling;
c. redundant nodes are assigned q-ary labels (so that each primary node has t neighbors with matching label).

There are two natural variations where sequence of steps are interchanged:

- *adaptive graph*: b.→a.→c.
- *non-adaptive redundancy*: a.→c.→b.

In the first case, the graph is a function of the q-ary labels, while in the second case the redundant nodes are not allowed to depend on the labeling of primary nodes.

It is clear that the setting considered in this paper (a.→b.→c.) is an intermediate case. That is, any t-defect correcting design in the sense of Definition 1 is also t-defect correcting in the sense of *adaptive graph*. Similarly every design with *non-adaptive redundancy* should work in the sense of Definition 1.

The fundamental redundancy-wiring complexity trade-off is defined similarly to (2). However, for either of these cases it is rather easy to determine this trade-off for any $t \geq 1$:

- *adaptive graph*: Clearly the number of edges $E \geq kt$. This can be attained with (asymptotically) zero-redundancy by adding t redundant nodes of each label (for a total of $m = qt$) and connecting every primary node only to relevant t redundant nodes. Consequently, here
$$\mathcal{R}_t = \{(\varepsilon, \rho) : \varepsilon \geq 1, \rho \geq 0\}\,. \tag{46}$$
- *non-adaptive redundancy*: Again, clearly the number of edges $E \geq qkt$. This can be attained with (asymptotically) zero-redundancy by adding t redundant nodes of each label (for a total of $m = qt$) and connecting every primary nodes to all of qt redundant ones (i.e., using $K(k, qt)$ design). Consequently
$$\mathcal{R}_t = \{(\varepsilon, \rho) : \varepsilon \geq q, \rho \geq 0\}\,. \tag{47}$$

These observations are summarized in Fig. 8.

C. Relation to (t,t)-colorable hypergraphs

There is a purely graph-theoretic way to look at our problem. For this we recall the concept of a (t,t)-graph coloring introduced in [20]. A hypergraph is called (t,t)-colorable if for every $\{0,1\}$-coloring of hyperedges there exists a $\{0,1\}$-coloring of vertices so that each edge contains t vertices of its color. Define

$$d_t(k,m) = \min(\text{average edge-size: all } (t,t) \text{ colorable hypergraphs on } m \text{ vertices and } k \text{ hyperedges}) \tag{48}$$

It is not hard to see that our problem with binary $\mathcal{X}$ and (t,t)-coloring are one-to-one related: the vertices correspond to primary nodes and the hyperedges correspond to redundant nodes. More precisely we have

Proposition 19. *Fix binary $\mathcal{X}$. The boundary of $\mathcal{R}_t$ is given by*

$$\liminf_{k\to\infty} \frac{1}{t} d_t(k, \lceil \rho k t \rceil)\,. \tag{49}$$

The boundary of $\mathcal{R}_\infty$ is given by

$$\liminf_{t\to\infty} \liminf_{k\to\infty} \frac{1}{t} d_t(k, \lceil \rho k t \rceil) \tag{50}$$

Proof. Note that for a fixed t, if for some pair (ε, ρ) we have $\frac{1}{t} d_t(k, \lceil \rho k t \rceil) = \varepsilon$ for some k, then by copying (Proposition 7), there exists infinitely many values of $k' > k$, where $\frac{1}{t} d_t(k', \lceil \rho k' t \rceil) \leq$

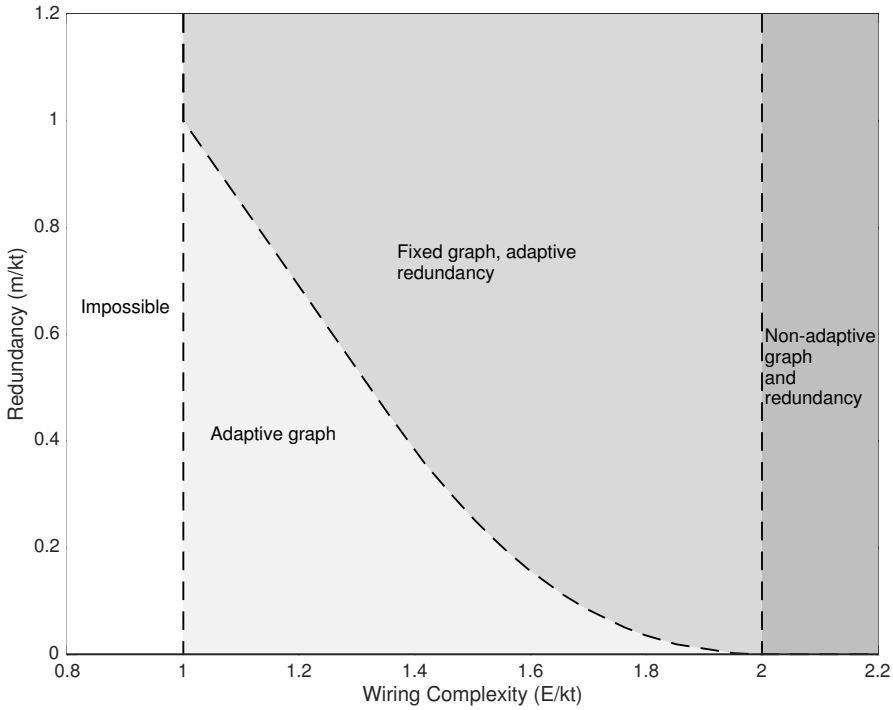


Fig. 8. Comparison of redundancy-wiring complexity trade-offs for different levels of adaptivity for $q = 2$.

ε. It follows from Proposition 2 that $\liminf_{k\to\infty} \frac{1}{t} d_t(k, \lceil \rho k t \rceil)$ must correspond to the boundary of $\mathcal{R}_t$.

Similarly, by merging (Proposition 8), any pair (ε, ρ) where there is some t such that

$$\liminf_{k\to\infty} \frac{1}{t} d_t(k, \lceil \rho k t \rceil) = \varepsilon$$

for some t, must also have infinitely many values of $t' > t$ where $\liminf_{k\to\infty} \frac{1}{t'} d_{t'}(k, \lceil \rho k t' \rceil) \leq \varepsilon$ □

D. Stochastic defects

This work considered correcting arbitrary (worst-case) defect patterns. Suppose that instead we are interested in correcting fraction α of defects (i.e., $t = \alpha(k + m)$) on k primary and m redundant nodes. In this scenario, the number of redundant nodes m would need to grow as a function of k in order to keep up with the number of defects needed to be corrected. If α is too large, it is not possible to find designs which corrects $\alpha(k + m)$ defects for arbitrarily large k.

To see this, note that correcting worst case t defects with alphabet size q requires at least qt redundant nodes.

$$m \geq qt \tag{51}$$

$$m \geq q\alpha(k + m) \tag{52}$$

$$m(1 - q\alpha) \geq q\alpha k \tag{53}$$

The quantity on the right-hand of (53) needs to be positive, so it must be that $\alpha < \frac{1}{q}$.

Additionally, the only designs which can correct fraction $\alpha < \frac{1}{q}$ of defects for growing k are designs with the same redundancy and wiring complexity as complete designs. From our results in Theorem 12, we know that there exists $(k, m, t, E)_q$-designs so that

$$\frac{m}{kt} \to c \tag{54}$$

for some constant c. When $t = \alpha(k + m)$,

$$\frac{m}{k\alpha(k+m)} > c \tag{55}$$

$$m(1 - ck\alpha) > ck^2\alpha \tag{56}$$

In order for (56) to hold, the right-hand side must be positive, so it must be that $c \to 0$ as k becomes arbitrarily large. The point in $\mathcal{R}_\infty$ where $\frac{m}{kt} \to 0$ corresponds to the complete graph design.

In light of these results, it is natural to ask what happens if instead we relaxed the requirement to correcting i.i.d. Bernoulli(α) defects in the sense of high probability (computed over distribution of defects and primary assignments). It turns out that in such probabilistic model, correcting fraction-α of defects is possible with designs possessing $O(k \log k)$ edges and $O(k)$ redundant nodes. See [21] Theorems 4.10 and 4.15 in Section 4.4 for more (pp. 63-66).

E. Future work

One direction for future work involves extensions beyond the bipartite graph. We chose to study the one-level bipartite graph model for simplicity, but experiments like Teramac [19] have demonstrated the effectiveness of multi-level hierarchical designs. This leads to the question of what are the optimal trade-offs when hierarchical models of redundancy are used. The hierarchical model would include intermediate nodes which can facilitate connections of edges. The presence of the intermediate nodes can greatly reduce the number of edges. To correct t defects, we can connect each primary node to t intermediate nodes. Regardless of the number of primary nodes, the intermediate nodes can connect to finitely many redundant nodes. This way, we are able to achieve a wiring complexity of t and redundancy of 0 (asymptotically). In such a case, we would be interested in finding the fundamental trade-offs with the number of intermediate nodes as a parameter.

F. Open Problems

Regions which are still to be determined include:

- $\mathcal{R}_t$ for $t > 2$ and $q = 2$
- $\mathcal{R}_t$ for $t > 1$ and $q \geq 3$

For $q = 2$, it is also unknown what the smallest value of t is for which $\mathcal{R}_t$ does not equal the region defined in Equation (5).

ACKNOWLEDGEMENT

YP would like to thank Prof. Jaikumar Radhakrishnan for interesting discussions at the Simons Institute for the Theory of Computing (UC Berkeley), in particular for bringing [20] to our attention.

APPENDIX

A. Proof of Theorem 11

Proof. Define $\mathcal{R}_1$ as in (18). We will show that all $(k, m, 1, E)_q$-designs must lie in $\mathcal{R}_1$ when primary nodes can have labels in $\mathcal{X} = \{0, 1, 2\}$.

Instead of saying that a given bipartite graph is 1-defect correcting for alphabet of size $q = 3$, we will say (for brevity) that a graph satisfies property (*).

(*) is the property that for any labeling of the primary nodes in $\mathcal{X}^k$, where k is the number of primary nodes, there exists a labeling of the redundant nodes so that each primary node has at least one redundant node neighbor with the same labeling.

The steps for this proof are:

1) Show that designs with primary nodes of degree 3 and greater can be disregarded.
2) Show that designs with a primary node of degree 1 cannot achieve better than $\mathcal{R}_1$.
3) Show that designs with primarys nodes all of degree 2 cannot achieve better than $\mathcal{R}_1$.
   a) Show designs containing two disjoint cycles connected by a path (see Fig. 9(a)) violate (*)
   b) Show designs containing two cycles which intersect at one point (see Fig. 9(b)) violate (*)
   c) Show designs containing two cycles which intersect at multiple points (see Fig. 9(c)) violate (*)

Step (1)

For any $(k, m, 1, E)_q$-design G in $\mathcal{R}_1$, define a new design G' with the same number of primary nodes k and the number of redundant nodes as $m' = m + qt = m + 3$. The added redundant nodes are connected to each of the primary nodes that have degree (in G) larger or equal to 3. The remaining primary nodes are connected in G' exactly as in G. It is clear that G is still satisfies (*), has the same (or smaller) number of edges and (asymptotically in m) the same redundancy ρ. This shows, that without loss of generality we can assume that there are no primary nodes of degree greater than 3 and all nodes of degree equal to 3 form a complete bipartite graph disjoint from the rest of the design which we can ignore.

Step (2)

The key to this step is to show that if the design has any primary node of degree 1, the design must be a tree.

We will call two primary nodes which share a redundant node *adjacent*.

Consider if the design has a primary node u_0 of degree 1 and no primary nodes have degree 3 or more. Label u_0 a 0. In order to satisfy (*), the redundant node neighboring u_0 must also be labeled 0. Label primary nodes adjacent to u_0 the label 1. Since the degree of each of these primary nodes is at most 2, this forces their remaining unlabeled redundant node neighbors to be labeled 1. We will call this labeling of primary nodes and neighboring redundant node the first iteration. On the nth iteration, we will label primary nodes adjacent to nodes labeled in the $n-1$ iteration, alternating 0 and 1 (meaning if iteration n labeled nodes 0, iteration $n+1$ will label the nodes 1 and vice versa.) This again forces a labeling on their redundant node neighbors. Each iteration builds a tree of labeled primary nodes and redundant nodes starting from u_0.

If during this iteration process, two primary nodes from the same iteration share a redundant node which was not labeled in the previous iteration, then label one of these two primary nodes the value 2. Their neighboring unlabeled redundant can only match the value of one of these primary node, so the design does not satisfy (*).

If there are any cycles, as the tree builds from u_0, we can label the first primary node to complete a cycle in the tree the label 2. The neighboring redundant nodes of this primary node are labeled either 0 or 1, so (*) is violated. Since the design is a tree, it must have at least the same number of redundant nodes as primary nodes, so the design lies in $\mathcal{R}_1$. We can now assume that all primary nodes are of degree 2.

Notice having $q \geq 3$ is important to avoid existence of even cycles.

Step (3)

Our goal is to prove that if all primary nodes have degree 2, then $\frac{m}{k} \geq \frac{1}{2}$. We will instead prove something stronger: For $k > 4$ we must have $m \geq k$. For $k = 4$ we must have $m \geq 3$.

Recall that it is sufficient to prove this for designs on a single component. If there is a redundant node with degree 1, we can remove it with its neighboring primary node as a separate component, so all redundant nodes must also have degree 2 or more.

We will call a labeling of primary nodes *alternating* if adjacent primary nodes have different labels.

Lemma 20. *If a design with k primary nodes, all of degree 2, and $k-1$ or fewer redundant nodes, all of degree 2 or 3, can be labeled alternatingly, then the design cannot satisfy (*).*

Proof. If there is an alternating labeling, at most each redundant node can only match the labeling of one of its neighboring primary nodes. There can only be at most $k-1$ matches, so there exists one primary node which does not have a neighboring redundant with the same label as itself. □

Suppose a design with all primary nodes of degree 2 is such that $m < k$. Then, some redundant node in the design must have degree 3 or more. Pick the separate component with this redundant node, and let A be a cycle in this separate component (if this component does not have cycles, then $m \geq k$ as in Step 2). In order for cycle A to be in this component, a redundant node with more than degree 2 must also be in A. Call this redundant node v_0. Call the neighbor of v_0 which is not in A u_0.

Let us build a path B starting at primary node u_0 as follows: The second node in path B will be the neighbor of u_0 which is not in cycle A. We can pick the next node in the path arbitrarily. The path ends when we reach a node in A or a node already in B. To show that (*) does not hold on the design, it is sufficient to show that (*) is not satisfied on $A \cup B$.

Depending on the endpoint of B, we have several cases:

Case (3a) **Endpoint of B coincides with an intermediate point of B.**
Let v_2 be the redundant node in path B where the path B ends. Rename the cycle created by path B to cycle C. The subgraph $A \cup B \cup C$ satisfies the conditions of Lemma 20 so we need only show that we can find an alternating label. v_0 and v_2 are the two redundant node with degree 3. It is clear that by starting with a alternating label of the neighbors of v_1 and v_2, we can find an alternating label for the rest of A, B and C.

Case (3b) **Endpoint of B is node v_0.**
Let cycle C be the cycle formed using path B and v_0. As long as one of cycle A and cycle C have more than 2 primary nodes, the design violates (*).
Consider when the labeling is so that the two primary nodes in the larger cycle, assume this to be cycle A, neighboring v_0 are labeled the same value, say 0. The rest of the primary nodes of A are labeled alternatingly, which is possible because cycle A has at least three primary nodes. Let the two primary nodes neighboring v_0 in cycle C be labeled 1 and 2.

v_0 must be labeled 0 in order for nodes in cycle A to each have a neighbor with the same label. Then if cycle C is labeled alternatingly, we will violate (*).
If both cycles have only 2 primary nodes, it is possible for this design to have $k = 4$ and $m = 3$ and satisfy (*). See Fig. 10.

Case (3c) **Endpoint of B is some node of A different from v_0.**
Two redundant nodes in the design have degree at least 3. Call them v_0 and v_1. Cycle A and path B make up three distinct paths which go from v_0 to v_1, which we will refer to them as E, F and G. As long as no two paths have only 1 primary node, then we can find an alternating label and use Lemma 20.
Label $u_{i,X}$ to be the primary node neighboring v_i and in path X. If all paths E, F, and G have two or more primary nodes, assign labels $0, 1, 2$ to $u_{0,E}, u_{0,F}, u_{0,G}$ and $1, 2, 0$ to $u_{1,E}, u_{1,F}, u_{1,G}$. Each path can be labeled alternatingly.
If there is one path with only one primary node, say path E, assign labels $0, 1, 2$ to $u_{0,E}, u_{0,F}, u_{0,G}$ and $2, 1$ to $u_{1,F}, u_{1,G}$. Each path can be labeled alternatingly.
If two paths both have one primary node, say E and F, as long as the third path G has at least 3 primary nodes, the design can be labeled alternatingly. We can label the two primary nodes cycle created by paths E and F the values 0 and 1. Then since G has at least 3 primary nodes, we can label $u_{0,G}$ and $u_{1,G}$ the value 2 and label the rest of G alternatingly.
If G only has 2 primary nodes, then this is a design on $k = 4$ and $m = 3$ which satisfies (*). See Fig. 11.

Note that the two exceptions with $k = 4$ are precisely the minimal non-trivial 1 defect correcting designs. One of these designs was discussed in Section III-A.

□

B. *Proof of Proposition 13*

Before we present the proof of Proposition 13, we will first show the following lemma:

Lemma 21. *Fix $P_X \in \frac{1}{k}\mathbb{Z}$ and $s \in \mathbb{Z}_+$. Let $\underline{L} \sim \mathrm{Mult}(s, P_X)$, cf. (23), and $\underline{L}^{(k)} \sim \mathrm{HyperGeom}(k, s, P_X)$, cf. (27). Then we have the following total variation estimate:*[6]

$$\mathsf{TV}(P_{\underline{L}^{(k)}}, P_{\underline{L}}) \le \frac{s^2}{2k}\,. \tag{57}$$

Similarly, if $\underline{M} \sim \mathrm{Mult}(s-1, P_X)$ and $\underline{M}^{(k-1)}$ has distribution

$$\mathbb{P}[\underline{M}^{(k-1)} = \underline{m}] = \frac{\binom{k\pi_1}{m_1}\cdots\binom{k\pi_j - 1}{m_j}\cdots\binom{k\pi_q}{m_q}}{\binom{k-1}{s-1}} \tag{58}$$

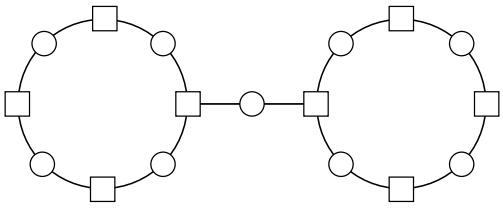
we also have

$$\mathsf{TV}(P_{\underline{M}}, P_{\underline{M}^{(k-1)}}) \le \frac{s^2}{2k}\,. \tag{59}$$

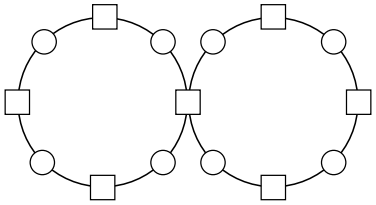
Proof. Standard estimate for total variation via coupling states that for any joint distribution $P_{\underline{L},\underline{L}^{(k)}}$:

$$\mathsf{TV}(P_{\underline{L}^{(k)}}, P_{\underline{L}}) \le \mathbb{P}[\underline{L}^{(k)} \neq \underline{L}]\,. \tag{60}$$

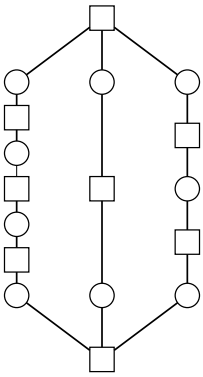
[6]Total variation distance TV for probability measures P and Q on sigma algebra $\mathcal{F}$ defined as $\mathsf{TV}(P, Q) = \sup_{A\in\mathcal{F}} |P(A) - Q(A)|$.



(a) Step (3a) case



(b) Step (3b) case



(c) Step (3c) case

Fig. 9. Different cases for designs with degree 2 primary nodes

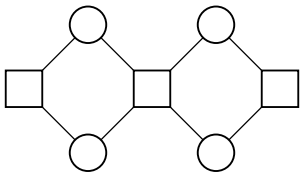


Fig. 10. Design which satisfies (*). Exception to case (3b).

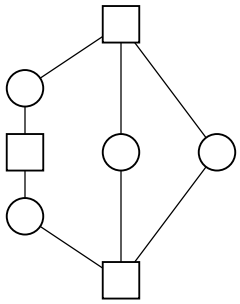


Fig. 11. Design which satisfies (*). Exception to case (3c).

Notice that $\underline{L}^{(k)}$ encodes the color distribution after sampling s balls from a collection of k colored balls (with composition given by P_X) without replacement, while $\underline{L}$ is the color distribution for sampling s balls with replacement. Let us couple these two samples as follows. Number all balls from 1 to k and define infinite string of i.i.d. uniform $X_i \in [k]$. Let our sample with replacement be the balls with indices $X_1, \dots, X_s$, while the sample without replacement be the balls $X_1, X_{i_2}, \dots, X_{i_s}$ where i_t denotes the first element of the sequence where t-th unique index appeared (i.e., for $X = (1, 2, 2, 3, \dots)$ we have $i_2 = 2$, $i_3 = 4$, etc). Now the two samples are going to be different only if $X_1, \dots, X_s$ are not distinct and this happens with probability at most

$$\sum_{i=1}^{s-1} \frac{i}{k} = \frac{s^2 - s}{2k} < \frac{s^2}{2k} \,. \tag{61}$$

This proves (57). For (59) modify distribution of X sequence by setting $X_1 = j$ and the rest are still i.i.d. uniform on $[k]$. Then $\underline{M}$ is the color composition of $X_2, \dots, X_s$ while $\underline{M}^{(k-1)}$ is the color composition of $X_{i_2}, \dots, X_{i_s}$. Again, $X_2, \dots, X_s$ are not all distinct with probability at most (61). □

Proof. Proof of (28): Simply by definition we have $F_{k,n}(P_S) \le F_k(P_S)$, so we focus on the opposite direction. First, we show that if $\underline{L}^{(k)} \sim \text{HyperGeom}(s, k, [\pi_1, \cdots, \pi_q])$, for any function $f : \underline{L}^{(k)} \to \mathbb{R}$ and any fixed $j \in [q]$ we have

$$\frac{1}{\pi_j} \mathbb{E}[L_j^{(k)} f(\underline{L}^{(k)})] = \mathbb{E}[S \cdot f(\underline{M}^{(k-1)} + \underline{e}_j)] \,, \tag{62}$$

where $\underline{e}_j$ is a vector with one in j-th position and the rest zeros, and $\underline{M}^{(k-1)}$ has hypergeometric distribution

$$\mathbb{P}[\underline{M}^{(k-1)} = \underline{m}] = \frac{\binom{k\pi_1}{m_1} \cdots \binom{k\pi_j - 1}{m_j} \cdots \binom{k\pi_q}{m_q}}{\binom{k-1}{s-1}} \,. \tag{63}$$

To that end, simply notice that

$$\frac{1}{\pi_j}\mathbb{E}[L_j^{(k)} f(\underline{L}^{(k)})] = \sum_s P_S(s) \sum_{\underline{\ell}} \frac{\ell_j}{\pi_j} \frac{\binom{k\pi_1}{\ell_1}\cdots\binom{k\pi_q}{\ell_q}}{\binom{k}{s}} f(\underline{\ell}) \tag{64}$$

$$= \sum_s P_S(s) \sum_{\underline{m}} s \frac{\binom{k\pi_1}{m_1}\cdots\binom{k\pi_j-1}{m_j}\cdots\binom{k\pi_q}{m_q}}{\binom{k-1}{s-1}} f(\underline{m}+\underline{e}_j)\,. \tag{65}$$

Now fix $P_X \in \frac{1}{k}\mathbb{Z}$ and $P_{Y^*|\underline{L}^{(k)}}$ to be the optimal distributions achieving $F_k(P_S)$ in (25). By rounding there must exist $P_{Y_n|\underline{L}^{(k)}} \in \frac{1}{n}\mathbb{Z}$ so that $|P_{Y_n|\underline{L}^{(k)}}(j|\underline{\ell}) - P_{Y^*|\underline{L}^{(k)}}(j|\underline{\ell})| \le \frac{1}{n}$ for every $\underline{\ell}$. Then for any fixed $j \in [q]$ we have in view of (62)

$$\left|\frac{1}{P_X(j)}\mathbb{E}[L_j^{(k)}\mathbb{1}\{Y^*=j\}] - \frac{1}{P_X(j)}\mathbb{E}[L_j^{(k)}\mathbb{1}\{Y_n=j\}]\right| \le \frac{\mathbb{E}[S]}{n}\,. \tag{66}$$

Taking $\min_j$ of (66) recovers the lower bound in (28).

We proceed to proving (29). Fix P_S and let

$$h(P_X, P_{Y|\underline{L}}, j) \triangleq \frac{1}{P_X(j)}\mathbb{E}\left[L_j \mathbb{1}\{Y=j\}\right], \tag{67}$$

where given $S=s$ we have $\underline{L} \sim \mathrm{Mult}(s, P_X)$, cf. (22). Similar to (62) we have

$$h(P_X, P_{Y|\underline{L}}, j) = \mathbb{E}\left[S \cdot P_{Y|\underline{L}}(j|\underline{M}+\underline{e}_j)\right], \tag{68}$$

where this time given $S=s$ we have $\underline{M} \sim \mathrm{Mult}(s-1, P_X)$. Now, for $P_X \in \frac{1}{k}\mathbb{Z}$ define also

$$h_k(P_X, P_{Y|\underline{L}}, j) \triangleq \frac{1}{P_X(j)}\mathbb{E}\left[L_j^{(k)} \mathbb{1}\{Y=j\}\right], \tag{69}$$

where given $S=s$ we have $\underline{L}^{(k)} \sim \mathrm{HyperGeom}(s, k, P_X)$. From (62), (68) and Lemma 21 (namely (59)) we have then

$$|h(P_X, P_{Y|\underline{L}}, j) - h_k(P_X, P_{Y|\underline{L}}, j)| \le \frac{\mathbb{E}\left[S^3\right]}{2k}\,. \tag{70}$$

Finally, since

$$(P_X, P_{Y|\underline{L}}) \mapsto \min_j h(P_X, P_{Y|\underline{L}}, j) \tag{71}$$

is uniformly continuous on a compact set, we also have

$$(P_X, P_{Y|\underline{L}}) \mapsto \max_{P_{Y|\underline{L}}} \min_j h(P_X, P_{Y|\underline{L}}, j) \tag{72}$$

is uniformly continuous by Proposition 22. Hence for some $\epsilon_k \to 0$ we have

$$\left|\min_{P_X \in \frac{1}{k}\mathbb{Z}} \max_{P_{Y|\underline{L}}} \min_j h(P_X, P_{Y|\underline{L}}, j) - \min_{P_X} \max_{P_{Y|\underline{L}}} \min_j h(P_X, P_{Y|\underline{L}}, j)\right| \le \epsilon_k\,. \tag{73}$$

Using (70) to replace h with h_k in the first term of the latter we get (29). □

Proposition 22. *Let $f : X \times Y \to \mathbb{R}$ where X and Y are compact and f is (uniformly) continuous. Then $\max_y f(x,y)$ is (uniformly) continuous on X.*

Proof. Let $h(x) = \max_y f(x,y)$. Because f is uniformly continuous, for every $\epsilon > 0$ there exists a δ so that if the distance between (x_1, y_1) and (x_2, y_2) is less than δ, then $|f(x_1,y_1) - f(x_2,y_2)| < \epsilon$ for all $x_1, x_2 \in X$ and $y_1, y_2 \in Y$. We want to show that for h, the same δ can be used for each ϵ. Suppose there exists values $x, x' \in X$ where $|h(x') - h(x)| > \epsilon$ and $|x - x'| < \delta$. Assume that $h(x') > h(x)$. There exists a value of y so that $f(x', y) = h(x')$. Since $|f(x', y) - f(x, y)| \leq \epsilon$ then $h(x') = f(x', y) \leq f(x, y) + \epsilon \leq h(x) + \epsilon$, which is a contradiction. □

C. Upper bound on $F_k(P_S)$

Lemma 23. *For any $P_S \in \mathbb{Q}$ with finite support,*

$$F_k(P_S) \leq F(P_S)\,. \tag{74}$$

Proof. Fix $P_S \in \mathbb{Q}$ with finite support. First, we will show that $F_k(P_S) \leq F_{2k}(P_S)$. Using Proposition 14 and (28) from Proposition 13, for each k there exists a sequence of subset designs G_i which are $(k, m_i, t_i, E_i)_q$-designs with $E_i = m_i \mathbb{E}[S]$, and $\frac{t_i k}{m_i} \to F_k(P_S)$.

For each G_i, we will construct subset design G_i' on $2k$ primary nodes by copying (see Proposition 7) two copies of G_i. G_i' is a $(2k, 2m_i, t_i, 2E_i)_q$-design. By Proposition 15, for each G_i', there exists a subset design G_i'' which is a $(2k, 2m_i \cdot (2k)!, t_i \cdot (2k)!, 2E_i \cdot (2k)!)$.

$$F_{2k}(P_S) \geq \lim_{i\to\infty} \frac{t_i \cdot (2k)! 2k}{2m_i \cdot (2k)!} = \lim_{i\to\infty} \frac{t_i k}{m_i} = F_k(P_S) \tag{75}$$

Then, $F_k(P_S) \leq F_{2k}(P_S) \leq F_{4k}(P_S) \leq F_{8k}(P_S) \leq F_{16k}(P_S) \cdots$. Since $F_{2^i k}(P_S) \to F(P_S)$ monotonically with convergence given by (29) from Proposition 13, this gives the desired result. □

D. Proof of Proposition 16

Proof. Fixed P_S with finite support and let $c = \mathbb{E}[S]$. Let $\hat{P}_S(s) = \frac{P_S(s)s}{\sum_s P_S(s)s} = \frac{1}{c} P_S(s)s$. We can substitute in $\hat{P}(s)$ and take the expectation with respect to $\hat{P}_S$ instead of P_S by adjusting (21) to

$$F(P_S) = \min_{P_X} \max_{P_{Y|L}} \min \left\{ \frac{c}{P_X(0)} \frac{L_0}{S} \mathbb{1}\{Y = 0\}, \frac{c}{P_X(1)} \frac{L_1}{S} \mathbb{1}\{Y = 1\} \right\} \tag{76}$$

Let the redundant node ratio of a redundant node with type (ℓ_0, ℓ_1) be $\nu = \frac{\ell_0}{\ell_0 + \ell_1}$. Suppose that $P_{Y|L}$ is a labeling so that

1) $P_{Y|L}(1|\underline{\ell}_a) > 0$ where $\underline{\ell}_a$ so that $\ell_0 + \ell_1 = s_a$ and has ratio ν_a
2) $P_{Y|L}(0|\underline{\ell}_b) > 0$ where $\underline{\ell}_b$ so that $\ell_0 + \ell_1 = s_b$ and has ratio ν_b
3) $\nu_a > \nu_b$

Let $P_{Y'|L}$ be equivalent to $P_{Y|L}$ except that

$$P_{Y'|L}(1|\underline{\ell}_a) = P_{Y|L}(1|\underline{\ell}_a) - \alpha \hat{P}_S(s_b) P_{\underline{L}|S}(\underline{\ell}_b|s_b) \tag{77a}$$
$$P_{Y'|L}(0|\underline{\ell}_a) = P_{Y|L}(0|\underline{\ell}_a) + \alpha \hat{P}_S(s_b) P_{\underline{L}|S}(\underline{\ell}_b|s_b) \tag{77b}$$
$$P_{Y'|L}(1|\underline{\ell}_b) = P_{Y|L}(1|\underline{\ell}_b) + \alpha \hat{P}_S(s_a) P_{\underline{L}|S}(\underline{\ell}_a|s_a) \tag{77c}$$
$$P_{Y'|L}(0|\underline{\ell}_b) = P_{Y|L}(0|\underline{\ell}_b) - \alpha \hat{P}_S(s_a) P_{\underline{L}|S}(\underline{\ell}_a|s_a) \tag{77d}$$

for an appropriate $\alpha > 0$ so that $P_{Y'|L}$ is still a valid distribution. Compared to $P_{Y|L}$, $P_{Y'|L}$ increases both quantities in the brackets in (76). So $P_{Y|L}$ cannot be optimal and any optimal $P_{Y|L}$ must have the form of(42).

For two redundant node type ratios where $\nu_a = \nu_b$, we can also see from (77a)-(77a) that there is a value of α (possibly negative unlike above) so that $P_{Y'|L}(0|\underline{\ell}_a) = P_{Y'|L}(0|\underline{\ell}_b)$ and the value of (76) is not affected by the change. Thus, there exists a solution where $P_{Y|L}$ is a function of the redundant node type ratio. □

E. Numerical results derivation

Here we develop upper and lower bounds for the expression found in Theorem 4 (the particular case when $\mathcal{X} = \{0, 1\}$).

1) Almost tight lower bound: Our lower bound for the boundary of $\mathcal{R}_\infty$ will be parametrized by c. To get this lower bound, we want to find an upper bound for

$$Z^*(c) = \max_{P_S : \mathbb{E}[S]=c} F(P_S)\,. \tag{78}$$

For notation, let $\lambda = P_X(0)$ and $1-\lambda = P_X(1)$. Let $\underline{\ell} = (\ell_0, \ell_1)$ and $f(\ell_0, \ell_1) = f(\underline{\ell}) = P_{Y|L}(0|\underline{\ell})$ where f can take any value between $[0, 1]$. For a fixed λ and s, define random variable $\underline{M} = (V, s-1-V)$ where $V \sim \text{Bino}(s-1, \lambda)$ and $\underline{e}_0 = (1, 0)$ and $\underline{e}_1 = (0, 1)$ according to Lemma 21 and the proof of Proposition 13. Fix P_S to have finite support.

First, we have that

$$\begin{aligned}
F(P_S) &= \min_{0\le\lambda\le 1} \max_{0\le f\le 1} \min\left\{\frac{1}{\lambda}\mathbb{E}[L_0 f(\underline{L})], \frac{1}{1-\lambda}\mathbb{E}[L_1(1-f(\underline{L}))]\right\} && (79)\\
&= \min_{0\le\lambda\le 1} \max_{0\le f\le 1} \min_{0\le\alpha\le 1} \alpha\mathbb{E}[S\cdot f(\underline{M}+\underline{e}_0)] + (1-\alpha)\mathbb{E}[S\cdot(1-f(\underline{M}+\underline{e}_1)] && (80)\\
&\le \min_{0\le\lambda\le 1} \frac{1}{2}\mathbb{E}\left[S\cdot \max_{0\le f\le 1}(1+f(\underline{M}+\underline{e}_0)-f(\underline{M}+\underline{e}_1))\right] && (81)\\
&= \min_{0\le\lambda\le 1} \frac{1}{2}\mathbb{E}\left[S\cdot(1+\max_{0\le\ell_0\le S-1}\mathbb{P}[\underline{M}=(\ell_0, S-1-\ell_0)])\right] && (82)\\
&\triangleq \min_{0\le\lambda\le 1} \mathbb{E}[\phi(S,\lambda)] && (83)
\end{aligned}$$

where in (80) we use (68) and convexify the minimum using α, and then (81) follows by setting $\alpha = \frac{1}{2}$. To get (82), notice that for a fixed s

$$\begin{aligned}
&\mathbb{E}[1+f(\underline{M}+\underline{e}_0)-f(\underline{M}+\underline{e}_1)] && (84)\\
&= 1+\sum_{\ell_0=0}^{s-1}\mathbb{P}[\underline{M}=(\ell_0, s-1-\ell_0)]f(\ell_0+1, s-1-\ell_0)\\
&\quad -\sum_{\ell_0=0}^{s-1}\mathbb{P}[\underline{M}=(\ell_0, s-1-\ell_0)]f(\ell_0, s-\ell_0)) && (85)\\
&= 1+\mathbb{P}[\underline{M}=(x, s-1-x)]f(x+1, s-1-x)\\
&\quad +\mathbb{P}[\underline{M}=(x+1, s-x)](1-f(x+1, s-1-x)) && (86)
\end{aligned}$$

By Proposition 16, the optimal f must have a threshold solution. We can express this threshold solution by letting x be be smallest value of ℓ_0 where $f(\ell_0, s-\ell_0)$ is non-zero. Applying the cancellations to (85), we get that only two terms remains. The value of f which obtains the

maximum must be where only the maximum value of $\mathbb{P}[\underline{M} = (x, s-1-x)]$ over all x appears in (86), and this gives (82).

We will bound

$$Z^*(c) \leq \max_{P_S:\mathbb{E}[S]=c} \min_{0\leq\lambda\leq 1} \mathbb{E}[\phi(S,\lambda)] \leq \max_{P_S:\mathbb{E}[S]=c} \min_{\lambda\in L_n} \mathbb{E}[\phi(S,\lambda)] \stackrel{\triangle}{=} Z'_n(c) \tag{87}$$

where we defined[7]

$$L_n = \left\{ \frac{\lfloor s/2 \rfloor}{s} : \text{ where } 1 < s \leq 2n \right\} .$$

Note that increasing n makes the approximation tighter. Index the elements of L_n as λ_i where $\lambda_1 = \frac{1}{2}, \lambda_2 = \frac{1}{3}, \lambda_3 = \frac{2}{5}, ..., \lambda_n = \frac{n-1}{2n-1}$, so that $\min_{\lambda\in L_n} \mathbb{E}[\phi(S,\lambda)] = \min_i \mathbb{E}[\phi(S,\lambda_i)]$.

$Z'_n(c)$ is equivalent to maximizing the value of t under the constraints that $\mathbb{E}[\phi(S,\lambda_i)] \geq t$ for all $1 \leq i \leq n$ and $\mathbb{E}[S] = c$. We can substitute

$$\phi(s,\lambda_i) = \frac{s}{2}\left(1 + \max_{1\leq\ell_0\leq s} \mathbb{P}[\underline{M} = (\ell_0, s-1-\ell_0)]\right) \stackrel{\triangle}{=} \frac{s}{2}\left(1+\psi(s,\lambda_i)\right) . \tag{88}$$

Then $\mathbb{E}[\phi(S,\lambda_i)] = \frac{1}{2}\mathbb{E}[S] + \frac{1}{2}\mathbb{E}[S\cdot\psi(S,\lambda_i)] = \frac{c}{2} + \frac{1}{2}\mathbb{E}[S\cdot\psi(S,\lambda_i)]$. Note that $\psi(S,\lambda_i) \to 0$ as $s \to \infty$ for all i.

For any value of $\pi_i \geq 0$, where $1 \leq i \leq n$, $\eta \geq 0$ and $\mu \geq 0$, we can define

$$\begin{aligned} Z''_n(c,\pi_1,...,\pi_n,\eta,\mu)] \stackrel{\triangle}{=} \max_{P_s(s)\geq 0,\forall s} & \; t + \sum_i \pi_i \left(\frac{c}{2} + \sum_{s=1}^{\infty} P_S(s)\frac{s}{2}\psi(s,\lambda_i) - t \right) \\ & - \eta\left(\sum_{s=1}^{\infty} P_S(s)s - c\right) - \mu\left(\sum_{s=1}^{\infty} P_S(s) - 1\right) \end{aligned} \tag{89}$$

Consider the set of π_i, η, μ which is the solution to

$$\textbf{minimize} \quad \frac{c}{2} + \eta c + \mu \tag{90a}$$

$$\textbf{subject to} \quad \sum_i^n \frac{1}{2}\pi_i\psi(s,\lambda_i) - \eta - \mu\frac{1}{s} \leq 0, s \in \mathbb{Z}_+ \tag{90b}$$

$$\sum_{i=1}^n \pi_i - 1 = 0 \tag{90c}$$

$$\eta \geq 0, \mu \geq 0, \pi_i \geq 0, 1 \leq i \leq n \tag{90d}$$

Such an optimization has a solution which is easy to find despite having infinitely many constraints. The constraints (90b) will hold for all s greater than some s_0 because $\psi(s,\lambda_i) \to 0$. By choosing a large enough s_0, we can solve the optimization by replacing it with an optimization where only the first s_0 constraints in (90b) are present. Set the values of π_i, η, μ in (89) to be the values which obtain the minimum for (90a)-(90d). Select a value of $s_1 \in \mathbb{Z}_+$. Then

[7] L_n is defined so that $L_n = \{\lambda \in (0, \frac{1}{2}] : \phi(s,\lambda) \leq \phi(s,\lambda')$ for some $1 < s \leq 2n$ and $\forall\lambda' \in [0,1]\}$ which is the set of all λ which minimizes $\phi(s,\lambda)$ for some $1 < s \leq 2n$.

$$Z'_n(c) \leq Z''_n(c, \pi_1, ..., \pi_n, \eta, \mu) \tag{91}$$

$$= \max_{P_s(s)\geq 0, \forall s} t + \sum_i \pi_i \left(\frac{c}{2} + \sum_{s=1}^{\infty} P_S(s) \frac{s}{2} \psi(s, \lambda_i) - t \right) \tag{92}$$

$$- \eta \left(\sum_{s=1}^{\infty} P_S(s) s - c \right) - \mu \left(\sum_{s=1}^{\infty} P_S(s) - 1 \right) \tag{93}$$

$$= \max_{P_s(s)\geq 0, \forall s} \sum_i \pi_i \left(\sum_{s=s_1}^{\infty} P_S(s) \frac{s}{2} \psi(s, \lambda_i) \right) \tag{94}$$

$$- \eta \left(\sum_{s=s_1}^{\infty} P_S(s) s \right) - \mu \left(\sum_{s=s_1}^{\infty} P_S(s) \right) + \frac{c}{2} + \eta c + \mu \tag{95}$$

$$\leq \max_{s>s_1} \max_i \frac{c}{2} \psi(s, \lambda_i) + \frac{c}{2} + \eta c + \mu \tag{96}$$

Since $\psi(s, \lambda_i) \to 0$, the optimal $\frac{c}{2} + \eta c + \mu$ given by (90a)-(90d) is an upper bound to $Z'_n(c)$ and hence also to $Z^*(c)$. This computes a lower bound on $\mathcal{R}_\infty$. In Fig. 6, we found the lower bound using $n = 10$.

2) Upper bounds: To show a point in $\mathcal{R}_\infty$ is achievable, it is sufficient to find a set of masses P_S that achieves that point. Searching all possible masses P_S is not computationally efficient. It turns out we can get decently close to the lower bound approximation by using the same masses which are solutions to $Z'_n(c)$ for each c when restricting P_S to only have finite support. While these results are close to the almost tight converse bound, they are not necessarily the best known. A few best known achievable points were found by simple search. The results are plotted in Fig. 6 and shown in Table I.

F. Proof of Corollary 18

Proof. The Hamming block achieves the point $(\frac{3}{2}, \frac{2}{3})$ in $\mathcal{R}^3_\infty$. The proof that this is a corner point amounts to computing the region $\mathcal{R}^3_\infty$.

To solve for $\mathcal{R}^3_\infty$, we will first simplify the expression for $F_3(P_S)$. For any P_S on $s \in [3]$, the labeling of primary nodes which gives the minimum value of $F_3(P_S)$ is when $P_X(0) = \frac{2}{3}$ and $P_X(1) = \frac{1}{3}$ (or these flipped). With this insight, we can simplify $F_3(P_S)$ to solve for the optimal P_S given any parameter $\mathbb{E}[S] = c$ for some $1 \leq c \leq 3$. Let $P_{Y|\underline{L}^{(3)}}(j|\ell_0, \ell_1)$ denote the

proportion of redundant nodes of type $\underline{\ell} = (\ell_0, \ell_1)$ to label j.

$$F_3(P_S) = \max_{P_{Y|\underline{L}^{(3)}}} \min_{j\in\{0,1\}} \frac{1}{P_X(j)} \mathbb{E}\left[L_j \mathbb{1}\{Y=j\}\right] \tag{97}$$

$$= \max_{P_{Y|\underline{L}^{(3)}}} \min_{j\in\{0,1\}} \left\{ \frac{1}{P_X(j)} \sum_{s=1}^{3} P_S(s) \sum_{\underline{\ell}} \ell_j P^3_{\underline{L}^{(3)}|S,P_X}(\underline{\ell}|s, P_X) P_{Y|\underline{L}^{(3)}}(j|\underline{\ell}) \right\} \tag{98}$$

$$= \max_{P_{Y|\underline{L}^{(3)}}} \min \left\{ \frac{3}{2} \left(P_S(1)\frac{2}{3} P_{Y|\underline{L}^{(3)}}(0|1,0) + \right. \right.$$
$$P_S(2)\left[\frac{2}{3} P_{Y|\underline{L}^{(3)}}(0|1,1) + 2\frac{1}{3} P_{Y|\underline{L}^{(3)}}(0|2,0)\right] + P_S(3) 2 P_{Y|\underline{L}^{(3)}}(0|2,1)\Big),$$
$$\left. \frac{3}{1}\left(P_S(1)\frac{1}{3} P_{Y|\underline{L}^{(3)}}(1|0,1) + P_S(2)\frac{2}{3} P_{Y|\underline{L}^{(3)}}(1|1,1) + P_S(3) P_{Y|\underline{L}^{(3)}}(1|2,1)\right)\right\} \tag{99}$$

We will first solve for the portion of $\mathcal{R}^3_\infty$ where $\tilde{\varepsilon} > \frac{3}{2}$.

Set $\mathbb{E}[S] = 3$. There is a unique point of the form $(\frac{3}{\eta}, \frac{1}{\eta})$ for some $\eta > 0$ which is a boundary point of the convex region $\mathcal{R}^3_\infty$. The only distribution P_S which can achieve $\mathbb{E}[S] = 3$ is when $P_S(3) = 1$ and $P_S(s) = 0$ for all other $s \neq 3$. With this P_S, we get that $F_3(P_S) = \frac{3}{2}$. Since no other P_S is possible, the point

$$\left(\frac{\mathbb{E}[S]}{F_3(P_S)}, \frac{1}{F_3(P_S)}\right) = (2, \frac{2}{3}) \tag{100}$$

must be the boundary point of the form $(\frac{3}{\eta}, \frac{1}{\eta})$ in $\mathcal{R}^3_\infty$. The line of points between this value and the value given by the Hamming block is achievable by convexity and by Claim 2 of Proposition 2 they must be optimal.

For the remaining portion of the region, we want to fixed a $1 < c < \frac{9}{4}$ (the Hamming block has $\mathbb{E}[S] = \frac{9}{4}$), and solve for $P^*_S = \operatorname{argmax}_{P_S:\mathbb{E}[S]=c} F_3(P_S)$ and determine $F_3(P^*_S)$.

Note that it is optimal to set $P_{Y|\underline{L}^{(3)}}(0|\ell_0, 0) = 1$ and $P_{Y|\underline{L}^{(3)}}(1|0, \ell_1) = 1$. Then we can simplify notation by letting $P_{Y|\underline{L}^{(3)}}(0|1,1) = x_{1,1}$ and $P_{Y|\underline{L}^{(3)}}(0|2,1) = x_{2,1}$. We can simplify (99) by applying the constraints that $\sum_{s=1}^3 P_S(s) = 1$ and $\sum_{s=1}^3 P(s)s = c$. At the maximum point, the two quantities after the minimum must be equal. Simplifying the equation with these constraints, we have

$$F_3(P^*_S) = \max_{x_{1,1}, x_{2,1}\in[0,1]} \frac{c - 1 - (3c-3)x_{1,1}}{6x_{2,1} - 6x_{1,1} - 1}(3x_{2,1} - 2x_{1,1} - 1) + (c-1)x_{1,1} + 1 \tag{101}$$

under the constraints that the variables are in $[0, 1]$.

The optimal labeling must have that either $x_{1,1} = 0$ and $x_{2,1} \in [0, 1]$ or that $x_{1,1} \in [0, 1]$ and $x_{2,1} = 1$ by Proposition 16. We try the cases $x_{1,1} = 0$ and $x_{2,1} = 1$ and take derivatives to solve for the best value of $x_{2,1}$ or $x_{1,1}$. For any value of c we pick[8], the point $(\frac{c}{F_3(P^*_S)}, \frac{1}{F_3(P^*_S)})$ lies on the line between the point achievable by the repetition design and the Hamming block. By convexity, it must be that all points on the line between the values achievable by the repetition design and the Hamming block are optimal.

[8]For example, we can pick $c = \frac{12}{7}$. We get that the maximum value occurs when $x_{1,1} = 0$ and $x_{2,1} = 1$. With this setting of variables, $P^*_S(1) = \frac{3}{7}, P^*_S(2) = \frac{3}{7}, P^*_S(3) = \frac{1}{7}$ which corresponds to the subset design $S(3,3) \vee S(3,2) \vee S(3,1)$ and achieves $(\frac{12/7}{F_3(P^*_S)}, \frac{1}{F_3(P^*_S)}) = (\frac{12}{9}, \frac{7}{9})$.

□

REFERENCES

[1] J. von Neumann, "Probabilistic logics and the synthesis of reliable organisms from unreliable components," *Automata Studies*, vol. 34, pp. 43–98, 1956.

[2] E. Moore and C. E. Shannon, "Reliable circuits using less reliable relays," *J. Franklin Inst.*, vol. 262, no. 3, pp. 191–208, 1956.

[3] R. Dobrushin and S. Ortyukov, "Lower bound for the redundancy of self-correcting arrangements of unreliable functional elements," *Problemy Peredachi Informatsii*, vol. 13, no. 1, pp. 82–89, 1977.

[4] ——, "Upper bound on the redundancy of self-correcting arrangements of unreliable functional elements," *Problemy Peredachi Informatsii*, vol. 13, no. 3, pp. 56–76, 1977.

[5] N. Pippenger, "On networks of noisy gates," in *Proc. Symp. Foundations Computer Science (FOCS)*, Oct 1985, pp. 30–38.

[6] ——, "Reliable computation by formulas in the presence of noise," *IEEE Trans. Inform. Theory*, vol. 34, no. 2, pp. 194–197, 1988.

[7] N. Pippenger, G. D. Stamoulis, and J. N. Tsitsiklis, "On a lower bound for the redundancy of reliable networks with noisy gates," *IEEE Trans. Inform. Theory*, vol. 37, no. 3, pp. 639–643, 1991.

[8] B. Hajek and T. Weller, "On the maximum tolerable noise for reliable computation by formulas," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 388–391, 1991.

[9] W. S. Evans and L. J. Schulman, "Signal propagation and noisy circuits," *IEEE Trans. Inform. Theory*, vol. 45, no. 7, pp. 2367–2373, 1999.

[10] ——, "On the maximum tolerable noise of k-input gates for reliable computation by formulas," *IEEE Trans. Inform. Theory*, vol. 49, no. 11, pp. 3094–3098, 2003.

[11] F. Unger, "Better gates can make fault-tolerant computation impossible," in *Elect. Colloq. Comp. Complexity (ECCC)*, vol. 17, 2010, p. 164.

[12] G. Norman, D. Parker, and M. Kwiatkowska, "Evaluation the reliability of defect-tolerant architectures for nanotechnology with probabilistic model checking," in *Proc. Int. Conf. VLSI Design*, 2004, pp. 907–912.

[13] K. Nikolic, A. Sadek, and M. Forshaw, "Fault-tolerant techniques for nanocomputers," *Nanotechnology*, vol. 13, no. 3, p. 357, 2002.

[14] D. Gizopoulos, M. Psarakis, S. V. Adve, P. Ramachandran, S. K. S. Hari, D. J. Sorin, A. Meixner, A. Biswas, and X. Vera, "Architectures for online error detection and recovery in multicore processors," in *Proc. Design, Automation Test in Europe Conference Exhibition (DATE)*, 2011, pp. 1–6.

[15] M. Flynn, C. Donovan, and L. Sattler, "Digital calibration incorporating redundancy of flash adcs," *IEEE Trans. Circuits and Systems II: Analog and Digital Signal Processing*, vol. 50, no. 5, pp. 205–213, May 2003.

[16] N. Verma and A. Chandrakasan, "A 256 kb 65 nm 8t subthreshold sram employing sense-amplifier redundancy," *IEEE J. Solid-State Circuits*, vol. 43, no. 1, pp. 141–149, 2008.

[17] D. Mange, M. Sipper, A. Stauffer, and G. Tempesti, "Toward robust integrated circuits: The embryonics approach," *Proc. IEEE*, vol. 88, no. 4, pp. 516–543, 2000.

[18] T. Leighton and C. Leiserson, "Wafer-scale integration of systolic arrays," *IEEE Trans. Comput.*, vol. C-34, no. 5, pp. 448–461, May 1985.

[19] J. Heath, G. Kuekes, and R. Williams, "A defect tolerant computer architecture: Opportunities for nanotechnology," *Science*, vol. 80, pp. 1716–1721, 1998.

[20] N. Alon and U. Feige, "On the power of two, three and four probes," in *Proc. ACM-SIAM Symp. Discrete Algorithms (SODA)*, 2009, pp. 346–354.

[21] D. Wang, "Computing with unreliable resources: Design, analysis and algorithms," Ph.D. dissertation, Massachusetts Institute of Technology, June 2014.