A Ramsey Theory Approach to Ghostbusting

Navin Kashyap
Queen's University
Kingston, ON, K7L 3N6, Canada
nkashyap@mast.queensu.ca

Paul H. Siegel
University of California San Diego
La Jolla, CA 92093-0407, USA
psiegel@ece.ucsd.edu

Alexander Vardy
University of California San Diego
La Jolla, CA 92093-0407, USA
vardy@kilimanjaro.ucsd.edu

Abstract —**We study bi-infinite sequences $\mathbf{x} = (x_k)_{k\in\mathbb{Z}}$ over the alphabet $\{0, 1, \ldots, q-1\}$, for an arbitrary $q \geqslant 2$, that satisfy the following q-ary ghost pulse (qGP) constraint: for all $k, l, m \in \mathbb{Z}$ such that x_k, x_l, x_m are nonzero and equal, x_{k+l-m} is also nonzero. This constraint arises in the context of coding to combat the formation of spurious "ghost" pulses in high data-rate communication over an optical fiber. We show using techniques from Ramsey theory that if x satisfies the qGP constraint, then the support of x is a disjoint union of cosets of a subgroup $k\mathbb{Z}$ of $\mathbb{Z}$ and a set of zero density.**

I. Introduction

In optical communication, a train of light pulses corresponding to a sequence of data bits is sent across an optical fiber. At high data rates (~40Gbps), a nonlinear effect known as four-wave mixing causes a transfer of energy from pulses in the kth, l-th and m-th time slots (k, l, m need not all be distinct) into the $(k + l - m)$-th time slot [1]. If this slot did not originally contain a pulse, the energy transfer creates a spurious pulse called a *ghost pulse*, which causes the original '0' in that slot to be changed to a '1'. Ghost pulse formation is phase-sensitive, so it can be mitigated by changing the phases of some of the pulses. However, an optical receiver cannot detect the phase of a pulse, so phase cannot be used to encode information.

To counter the ghost pulse effect, we consider a constrained coding scheme based on a class of "ghost pulse constraints."

II. Constrained Codes for Ghostbusting

For $q \geqslant 2$, let $\mathcal{A}_q = \{0, 1 \ldots, q-1\}$. For $\mathbf{x} = (x_k)_{k\in\mathbb{Z}} \in \mathcal{A}_q^{\mathbb{Z}}$, let $\operatorname{supp}(\mathbf{x}) = \{k \in \mathbb{Z} : x_k \neq 0\}$. A sequence $\mathbf{x} \in \mathcal{A}_q^{\mathbb{Z}}$ is said to satisfy the *q-ary ghost pulse constraint* if for all $k, l, m \in \operatorname{supp}(\mathbf{x})$ (k, l, m not necessarily distinct) such that $x_k = x_l = x_m$, we also have $k + l - m \in \operatorname{supp}(\mathbf{x})$. Let $\mathcal{T}_q$ be the set of all $\mathbf{x} \in \mathcal{A}_q^{\mathbb{Z}}$ that satisfy the qGP constraint, and let $\mathcal{S}_q$ denote the set of all *binary* sequences $\mathbf{y}$ such that there exists an $\mathbf{x} \in \mathcal{T}_q$ with $\operatorname{supp}(\mathbf{x}) = \operatorname{supp}(\mathbf{y})$. The object of this paper is to study the sequences in $\mathcal{S}_q$, particularly in the cases when q is 2 or 3.

To transmit a binary data sequence $a_0a_1 \ldots a_{M-1}$, we first encode it as a subblock $b_0b_1 \ldots b_{N-1}$ of a sequence in $\mathcal{S}_q$, which is then converted to a subblock $c_0c_1 \ldots c_{N-1}$ of some sequence in $\mathcal{T}_q$. The q-ary sequence $c_0c_1 \ldots c_{N-1}$ corresponds to a train of N light pulses, with the phases of the nonzero pulses being determined by a one-to-one mapping from $\{1, 2, \ldots, q-1\}$ to $[0, 2\pi]$. Under the simplifying assumption that only pulse triples with the same phase can interact to create ghost pulses, the sequence $c_0c_1 \ldots c_{N-1}$ can be transmitted without error across an optical fiber. This is because the qGP constraint ensures that the positions where ghost pulses could potentially be created already contain nonzero pulses.

*This work was supported in part by the Applied Micro Circuits Corporation, and by the UC Discovery Grant Program.

The efficiency of such a coding scheme is limited by the *capacity* $h(\mathcal{S}_q)$ of $\mathcal{S}_q$, which is defined as

$$h(\mathcal{S}_q) \stackrel{\text{def}}{=} \lim_{n\to\infty} \frac{\log_2 |\mathcal{B}_{q,n}|}{n} \tag{1}$$

where $\mathcal{B}_{q,n}$ is the set of all length-n subblocks of sequences in $\mathcal{S}_q$. The closer $h(\mathcal{S}_q)$ is to 1, the more efficient are the qGP constrained coding schemes. Herein, we analyze the structure of sequences in $\mathcal{S}_q$ with a view towards determining $h(\mathcal{S}_q)$.

III. Results

The case $q = 2$ is easily analyzed to obtain the following simple characterization of sequences in $\mathcal{S}_2$.

Theorem 1. *A binary sequence* $\mathbf{x}$ *is in* $\mathcal{S}_2$ *iff* $\operatorname{supp}(\mathbf{x}) = \varnothing$ *or* $\operatorname{supp}(\mathbf{x}) = a + k\mathbb{Z}$ *for some* $a, k \in \mathbb{Z}$.

It follows from Theorem 1 that $|\mathcal{B}_{2,n}| = O(n^2)$, which implies that $h(\mathcal{S}_2) = 0$.

The analysis for $q > 2$ is considerably more difficult. Using results from the branch of mathematics known as Ramsey theory (in particular, the theorems of Schur and Szemerédi [2]), we can prove the following result. To state this result, we need the following definition: the *upper density* of a subset $I \subset \mathbb{Z}$ is defined as $\overline{d}(I) = \limsup_{n\to\infty} \frac{|I\cap[-n,n]|}{2n+1}$.

Theorem 2. *For* $q > 2$, *if* $\mathbf{y} \in \mathcal{S}_q$ *then there exist an integer* $k \geqslant 0$ *and a set* $I \subset [0, k-1]$, *both depending on* $\mathbf{y}$, *such that*

$$\bigcup_{i\in I}(k\mathbb{Z} + i) \subset \operatorname{supp}(\mathbf{y}) \quad \textit{and} \quad \overline{d}\left(\operatorname{supp}(\mathbf{y}) \setminus \bigcup_{i\in I}(k\mathbb{Z} + i)\right) = 0.$$

This shows that any sequence $\mathbf{y} \in \mathcal{S}_q$ is "almost periodic," in the sense that it can be transformed into a periodic sequence by changing a relatively sparse subset of the 1's to 0's. However, this result needs to be strengthened considerably before we can determine $h(\mathcal{S}_q)$.

For $q = 3$, we can prove a stronger result which asserts that any $\mathbf{y} \in \mathcal{S}_3$ can be made periodic by changing at most two 1's to 0's. In fact, we have a simple and complete description of the aperiodic sequences in $\mathcal{S}_3$. However, the problem of fully classifying the periodic sequences in $\mathcal{S}_3$ remains largely open. We derive a characterization of such sequences, which can be used to completely describe the sequences of prime period in $\mathcal{S}_3$. Based on these results and numerical evidence, we conjecture that $h(\mathcal{S}_3) = 0$. For detailed descriptions of these results and their proofs, we refer the reader to our full paper [3].

References

[1] M.J. Ablowitz and T. Hirooka, "Intrachannel pulse interactions in dispersion-managed transmission systems: energy transfer," *Optical Lett.*, vol. 27, no. 3, pp. 203–205, February 2002.

[2] R.L. Graham, B.L. Rothschild, and J.H. Spencer, *Ramsey Theory*, New York: Wiley-Interscience, 1980.

[3] N. Kashyap, P.H. Siegel, and A. Vardy, "An application of Ramsey theory to coding for the optical channel," *SIAM J. Discrete Math*, submitted for publication (preprint available on request).