

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

$F_2 + uF_2 + vF_2 + v^2F_2$ HALKASI ÜZERİNDEKİ DEVİRLİ
KODLAR VE DNA KODLARA UYGULAMALARI

Merve BULUT YILGÖR

DOKTORA TEZİ
Matematik Anabilim Dalı
Matematik Programı

Danışman
Doç. Dr. Fatih DEMİRKALE
Eş-Danışman
Dr. Öğr. Üyesi Elif Segah ÖZTAŞ

Şubat, 2020

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

$F_2 + uF_2 + vF_2 + v^2F_2$ HALKASI ÜZERİNDEKİ DEVİRLİ KODLAR VE
DNA KODLARA UYGULAMALARI

Merve BULUT YILGÖR tarafından hazırlanan tez çalışması 17.02.2020 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Matematik Anabilim Dalı Matematik Programı **DOKTORA TEZİ** olarak kabul edilmiştir.

Doç. Dr. Fatih DEMİRKALE
Yıldız Teknik Üniversitesi
Danışman

Dr. Öğr. Üyesi Elif Segah ÖZTAŞ
Karamanoğlu Mehmetbey Üniversitesi
Eş-Danışman

Jüri Üyeleri

Doç. Dr. Fatih DEMİRKALE, Danışman
Yıldız Teknik Üniversitesi

Prof. Dr. Mehmet ÖZEN, Üye
Sakarya Üniversitesi

Doç. Dr. Gürsel YEŞİLOT, Üye
Yıldız Teknik Üniversitesi

Doç. Dr. Murat ALAN, Üye
Yıldız Teknik Üniversitesi

Dr. Öğr. Üyesi Fatma ÇALIŞKAN, Üye
İstanbul Üniversitesi

Danışmanım Doç. Dr. Fatih DEMİRKALE sorumluluğunda tarafımca hazırlanan $F_2 + uF_2 + vF_2 + v^2F_2$ Halkası Üzerindeki Devirli Kodlar ve DNA Kodlara Uygulamaları başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Merve BULUT YILGÖR



Bu çalışma, Yıldız Teknik Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimince desteklenmiştir. Proje Numarası: 2016-07-03-DOP05.

Oğlum Selim'e



TEŞEKKÜR

Hayatımın en zor dönemine denk gelen doktora eğitimi sürecimde desteğini her zaman hissettiren ve bana yol gösteren değerli hocam Doç. Dr. Fatih DEMİRKALE'ye, fikirleriyle ve varlığıyla bana her zaman destek olan kıymetli hocam Dr. Öğr. Üyesi Elif Segâh ÖZTAŞ'a, değerli yorumlarıyla tezimin gelişmesine katkıda bulunan hocalarım Prof. Dr. Mehmet ÖZEN'e, Doç. Dr. Gürsel YEŞİLOT'a, Doç. Dr. Murat ALAN ve Dr. Öğr. Üyesi Fatma ÇALIŞKAN'a ve çalışmalarım sırasında bana fikirleriyle destek olan kıymetli dostum Dr. Fatmanur Gürsoy'a teşekkürlerimi sunuyorum.

Aileme gelince, onlar benim bu hayattaki en büyük şansım. Anneme ve babama beni ben yapan değerleri verdikleri için, her zaman yanımda oldukları için ve her konudaki destekleri ile hayatımı kolaylaştırdıkları için ve her türlü desteği için biricik kardeşime çok teşekkür ediyorum. Motivasyon kaynağım olan hayat yoldaşım eşime ve canım oğlum Selim'e varlıklarıyla hayatımı anlamlandırdıkları için çok teşekkür ederim.

Merve BULUT YILGÖR

İÇİNDEKİLER

SİMGE LİSTESİ	viii
KISALTMA LİSTESİ	ix
ŞEKİL LİSTESİ	x
TABLO LİSTESİ	xi
ÖZET	xii
ABSTRACT	xiv
1 GİRİŞ	1
1.1 Literatür Özeti	1
1.2 Tezin Amacı	3
1.3 Orijinal Katkı	3
1.4 Tezin İçeriği	3
2 TEMEL BİLGİLER	5
2.1 Halkalar	5
2.2 Sonlu Cisimler	8
2.3 Modüller	9
2.4 Hata Düzeltken Kodlar	9
2.4.1 Lineer Kodlar	11
2.4.2 Devirli Kodlar	13
2.5 DNA ile İlgili Bazı Kavramlar	15
2.6 DNA Replikasyonu ile Kodlama Teorisi Arasındaki İlişki	17
2.7 DNA Hesaplama ve DNA Kod	18
3 $\mathbb{F}_2 + u\mathbb{F}_2 + v\mathbb{F}_2 + v^2\mathbb{F}_2$ HALKASI ÜZERİNDEKİ DEVİRLİ KODLAR	22
3.1 $\mathbb{F}_2 + u\mathbb{F}_2 + v\mathbb{F}_2 + v^2\mathbb{F}_2$ Halkası	22
3.2 R Halkasının İdeal Yapısı	23
3.3 R Halkası Üzerindeki Devirli Kodlar	25
3.4 Homojen Ağırlık ve R Halkası Üzerindeki Devirli Kodların İkili Görüntüleri	29

4	$\mathbb{F}_2 + u\mathbb{F}_2 + v\mathbb{F}_2 + v^2\mathbb{F}_2$ HALKASI ÜZERİNDEKİ DEVİRLİ KODLARIN DNA UYGULAMALARI	34
4.1	DNA Kodun Ters Sıralılık Problemi	34
4.2	R Halkası Üzerindeki DNA kodlar	35
5	GERÇEK DNA DİZİLERİNDE HATA DÜZELTEN KODLAR	39
5.1	Metod	39
5.2	TRAV7 Geni ile Hata Düzeltten Kodlar Arasındaki İlişki	43
5.3	Elde Edilen Sonuçlar	48
6	SONUÇ ve ÖNERİLER	50
6.1	Sonuç	50
6.2	Öneriler	51
6.2.1	GC -miktar Kısıtını Sağlayan DNA Kodlar	51
6.2.2	R Halkasından Elde Edilen Kodların CSS Quantum Kod İnşasında Kullanılması	51
A	TRAV7 VE ARABİDOPSİS THALİANA GENLERİNİN DİZİLİMLERİ	53
	KAYNAKÇA	57
	TEZDEN ÜRETİLMİŞ YAYINLAR	60

SİMGE LİSTESİ

A	Adenin
$wt(c)$	c Kodsözünün Ağırlığı
$C^\perp$	C Kodunun Duali
$d(C)$	C Kodunun Minimum Hamming Uzaklığı
$der(f(x))$	$f(x)$ Polinomunun Derecesi
$f(x)^*$	$f(x)$ Polinomunun Ters-sıralısı
G	Guanin
$g_i g_j$	g_i Polinomu g_j Polinomunu Böler
$C = \langle g(x) \rangle$	$g(x)$ Polinomu Tarafından üretilen devirli kod
C	Sitozin
$\mathbb{Z}$	Tamsayılar Kümesi
$\mathbb{Z}_n$	Tamsayıların Modülo n 'ye Göre Kalan Sınıflarının Kümesi
T	Timin
u^r	u Dizisinin Ters-sıralısı
u^c	u Dizisinin Tamlayanı
u^{rc}	u^r Dizisinin Tamlayanı
$\lfloor x \rfloor$	x 'e Eşit ya da x 'ten Küçük En Büyük Tamsayı

KISALTMA LİSTESİ

DNA	Deoksiribo Nükleik Asit
NCBI	National Center for Biotechnology Information
RNA	Ribo Nükleik Asit
WCC	Watson Crick Complement



ŞEKİL LİSTESİ

Şekil 2.1	<i>S</i> halkasının ideal latisi	7
Şekil 2.2	İletişim kanal modelinin sistematik diyagramı	10
Şekil 2.3	DNA çift sarmal görüntüsü [24]	16
Şekil 2.4	WCC kuralına göre DNA dizilimi	16
Şekil 2.5	Kodlama varsayımıyla birlikte gen replikasyonlu gürültülü kanal modeli [31]	18
Şekil 2.6	Şehirler arasındaki uçuşlar için yönlü graf	19
Şekil 2.7	Çift sarmal molekül çözümü	20
Şekil 3.1	<i>R</i> Halkasının ideal latisi	24
Şekil 5.1	TRAV7 genindeki kodlanan ve kodlanmayan bölgeler	46
Şekil 5.2	TRAV7 geni $N=7$, $n=6$, $k=2$ için hata düzeltme grafiği	46
Şekil 5.3	TRAV7 geni $N=8$, $n=7$, $k=2$ için hata düzeltme grafiği	47
Şekil 5.4	TRAV7 geni $N=9$, $n=8$, $k=2$ için hata düzeltme grafiği	47
Şekil 5.5	TRAV7 geni $N=10$, $n=9$, $k=2$ için hata düzeltme grafiği	47
Şekil 5.6	Arabidopsis Thaliana genindeki kodlanan ve kodlanmayan bölgeler	48
Şekil 5.7	Arabidopsis Thaliana geni $N=8$, $n=7$, $k=2$ için hata düzeltme grafiği	48

TABLO LİSTESİ

Tablo 2.1	Şehirler ve DNA karşılıkları	19
Tablo 3.1	R halkası üzerindeki devirli kodların Gray görüntülerinden elde edilen optimal ikili lineer kodlar ve [22]'ye göre adım sayıları	32
Tablo 3.2	Döngüsel matrislerden elde edilen optimal ikili lineer kodlar	33
Tablo 4.1	R halkasının elemanlarına karşılık gelen DNA ikilileri	35
Tablo 5.1	Kendine dual ve devirli kodlar	41
Tablo 5.2	Optimal, MDS ve devirli kodlar	42
Tablo 5.3	Optimal, kendine dik ve devirli kodlar	43
Tablo 5.4	Kendine dik ve devirli kodlar	43
Tablo 5.5	Optimal ve MDS kodlar	44
Tablo 5.6	Optimal devirli kodlar	44
Tablo 5.7	Devirli kodlar	45

$F_2 + uF_2 + vF_2 + v^2F_2$ Halkası Üzerindeki Devirli Kodlar ve DNA Kodlara Uygulamaları

Merve BULUT YILGÖR

Matematik Anabilim Dalı
Doktora Tezi

Danışman: Doç. Dr. Fatih DEMİRKALE

Eş-Danışman: Dr. Öğr. Üyesi Elif Segah ÖZTAŞ

Bu çalışmada, öncelikle yerel Frobenius olup zincir halkası olmayan $R = F_2[u, v]/\langle u^2 = v^2, uv = 0 \rangle$ bölüm halkası olarak da ifade edilen $R = F_2 + uF_2 + vF_2 + v^2F_2$ halkası üzerinde durulacaktır. R halkası üzerinde tanımlanan bir C devirli kodunu üreten ideal yapısı verilecek ve R halkası üzerinde ters-sıralı DNA kod elde etmek için tek ya da çift uzunluklu devirli kodların 3 tipi belirlenecektir. Belirlenen bu devirli kodların üreteç matrisi ve kodsöz sayısı elde edilecektir. Ayrıca dual kodunun üreteci belirlenecek ve R halkası üzerindeki lineer kodlar için kendine dik olma koşulu verilecektir. R halkası üzerindeki kodların Gray dönüşümlerinden optimal ikili lineer kodlar elde edilecektir. Ayrıca elde edilen kodların kendine dik kod oldukları ve tümünün bir-ağırlıklı kod oldukları bilgisi verilecektir. Farklı bir kod inşa metodu olan, döngüsel matrisler, yani R halkası üzerindeki temsilci vektörünü ve bu vektörün devirsel ötelemelerini satır kabul eden matrisler kullanılarak optimal kodlar elde edilecek ve bu kodların arasında iki-ağırlıklı kodların bulunmasının yanı sıra optimal iki-ağırlıklı kodların da olduğundan bahsedilecektir. R halkası üzerindeki devirli kodların DNA'ya uygulamalarına yer verilerek R halkası üzerinde tanımlanan devirli kodların cebirsel özelliklerinden faydalanılmak suretiyle DNA kodlar için önemli bir eşik olan ters sıralılık problemi çözülerek ters-sıralı DNA kod ve ters-sıralı tamlayan DNA kodlar elde edilecektir.

Son olarak ise gerçek DNA dizileri üzerine çalışmalar yapılacaktır. İnsanda bulunan TRAV7 (T cell receptor alpha variable 7) geni ve farede bulunan TRAV7-1 ile TRAV7-6

arasındaki 7 gen üzerindeki lineer kod yapısı incelenecektir. TRAV7 genlerinden kayan pencere metodu ile elde edilen dizilerin F_4 'teki karşılıklarını satır kabul eden matrislerin ürettiği kodların parametreleri elde edilecektir. Ayrıca elde edilen lineer kodlar MDS, kendine dual, kendine dik, devirli ve optimal olma özelliklerine göre sınıflandırılacaktır. Literatürde ilk olarak, TRAV7 geni ve Arabidopsis Thaliana geni için hata düzeltme grafikleri elde edilerek genlerle hata düzeltme grafikleri arasındaki ilişki araştırılacaktır. Kodlanan ve kodlanmayan bölgelere karşılık hata düzeltme grafiğindeki durumlar incelenerek hata düzeltemeyen kodların kodlanmayan bölgede yoğunlaştığı gösterilecektir.

Anahtar Kelimeler: Lineer kodlar, devirli kodlar, DNA kod, optimal kod



Cyclic Codes over the Ring $F_2 + uF_2 + vF_2 + v^2F_2$ and Their Applications to DNA Codes

Merve BULUT YILGÖR

Department of Mathematics
Doctor of Philosophy Thesis

Advisor: Assoc. Prof. Dr. Fatih DEMİRKALE
Co-advisor: Assist. Prof. Dr. Elif Segah ÖZTAŞ

In this study, we focus on $R = F_2 + uF_2 + vF_2 + v^2F_2$, $u^2 = v^2$, $uv = 0$ which is a local Frobenius but non-chain ring. The ideal structure that generates a cyclic code C defined on the ring R will be given. 3 types of single or double length cyclic codes will be determined to obtain the reversible DNA code on the ring R . In addition, the generator of the dual code will be determined. It is given the condition of being self orthogonal for linear codes on the R ring. Optimal binary linear codes will be obtained from Gray images of codes on the ring R . It will also be given that the codes obtained are self orthogonal codes and all are one-weight codes. By using circulant matrices, a different code constructing method, that is, optimal codes will be obtained from the matrix which accepts the representative vector on the ring R and the cyclical displacements of this vector, and will include two-weight codes as well as optimal two-weight codes. Using the algebraic properties of the cyclic codes defined on the ring R by applying the cyclic codes on the ring R to the DNA, the problem of reversibility, an important threshold for the DNA codes, will be solved and reversible DNA code and reversible complement DNA codes will be obtained.

Finally, real DNA sequences will be studied. The human TRAV7 (T cell receptor alpha variable 7) gene and 7 genes between TRAV7-1 and TRAV7-6 in mice will be examined on the linear code structure.

The parameters generated by the matrices that accept the lines in the F_4 of the sequences obtained by the sliding window method from TRAV7 genes will be obtained.

In addition, the linear codes obtained will be classified according to MDS, dual, vertical, cyclic and optimal. In the literature, first of all, error correction graphics for TRAV7 gene and Arabidopsis Thaliana gene will be obtained and the relationship between genes and error correction graphics will be investigated. The cases in the error correction graph will be examined for exon and intron regions and it will be shown that the non-corrected codes are concentrated in the non-coded region.

Keywords: Lineer codes, cyclic codes, DNA code, optimal code



1.1 Literatür Özeti

1997’de NASA’nın Mars’a gönderdiği Pathfinder isimli robot 16.500’den fazla görüntü ve 15’ten fazla kaya ve toprağın analizi ve rüzgarlar ile diğer hava koşulları hakkında kapsamlı bilgiler içeren 2.3 milyar bitlik bilgi gönderdi [1]. Bir radyo vericisi ile gerçekleşen bu bilgi aktarımının ortalama 225 milyon kilometre [2] uzaktan güvenilir ve tamamen bozulmamış bir şekilde gerçekleşmesi nasıl mümkün olabilir? Matematik ve mühendislik gibi farklı disiplinleri bir araya getiren *Kodlama Teorisi (Coding Theory)* ile.

İletişim teorisi, dijital dünyada bilgi iletimi ve bilginin depolanmasını konu olarak ele alır. Claude Shannon "A mathematical theory of communication" [3] başlıklı makalesi ile iletişimdeki matematiksel teoriyi geliştirdi. Bu teoriye göre uygun kodlama teknikleri kullanılırsa bilginin istenilen oranda güvenli iletiminin sağlanacağından bahsedildi.

Kodlama teorisi, gürültülü bir kanal boyunca bilgi iletimi ve bu sırada bozulan iletilerin tamir edilmesi ile ilgilenmektedir. Kodlamanın nasıl yapılacağına dair ilk temel ise Richard W. Hamming’in [4] *hata düzelten kodlar (error-correcting codes)* üzerine yaptığı çalışmalar ile atıldı.

DNA hesaplama fikri ise ilk olarak 1987 yılında Tom Head [5] tarafından ortaya atılmış olmasına rağmen bu konudaki ilk başarılı çalışma 1994 yılında Leonard Adleman tarafından gerçekleştirildi [6]. Adleman DNA hesaplama ile NP-tam problemi olan Hamilton yolu problemini sentetik DNA iplikçiklerini bir test tüpünde bir araya getirmek suretiyle birbirleri ile olan etkileşimlerinden faydalanarak çözdü. Adleman bu deneyinde DNA’nın WCC (Watson-Crick Complement) özelliğinden faydalandı. Bu özellik, DNA molekülleri olarak isimlendirilen adenin, guanin, sitozin ve timin moleküllerinin biraraya gelmesi durumunda adenin ile timinin, guanin ile sitozinin yapışmasına dayanır. Yapışan iplikçikler elektromanyetik alandan geçirilerek ayklanır

ve aranan sonuçlar elde edilir. Bu metodun bilinen yöntemlere göre problemin çözümünde daha pratik olması, farklı NP problemlerin çözümünde kullanılabileceği fikrini akıllara getirdi. Kırılması zor olan DES (Data Encryption Standard) yani veri şifreleme standardı bu metod ile kırıldı [7]. NP-tam problemi olan SAT probleminin çözümüne [8] ve kombinatoriyel optimizasyon problemlerinin çözümüne [9] DNA molekülleri yardımıyla ulaşıldı. Bilgi depolama maksadıyla da DNA'nın yapısında bulunan depolama özelliğinden faydalanılabilecek çalışmalar yapıldı [10, 11].

DNA moleküllerini hesaplama aracı olarak kullanırken başvurulacak en önemli özellik DNA hibridizasyonudur. Ancak meydana gelebilecek hatalar da yine hibridizasyondan kaynaklanır. Bu yüzden DNA hesaplamasının başarısı hata düzeltme tekniklerine bağlıdır. Cebirsel kodlama teorisi teknikleri mevcut bilgi ve iletişim teknolojisi (BIT) için temel sağlamıştır. Dolayısıyla, kodlama teorisinin DNA hesaplama süreci için temel konu olacağını beklemek doğaldır. Bu ise DNA kodların doğuşuna sebep olmuştur.

DNA hesaplama, esas olarak canlı hücrelerde bulunan kısıtlar altında mantıksal hesaplamaları gerçekleştirmeyi amaçlamaktadır. Kısıtlar hibridizasyonda meydana gelebilecek hataları en aza indirmek için yapılır. Başlıca DNA kısıtları; Hamming uzaklık kısıtı, ters-sıra (reverse) kısıtı, ters-sıra tamlayan (reverse-complement) kısıtı ve sabitlenmiş GC-miktar kısıtıdır (fixed GC-content) [12].

Günümüze kadar DNA kısıtları göz önünde bulundurularak cebirsel yapılardan olan sonlu cisimler ve sonlu halkalar üzerine bir çok DNA kod çalışması gerçekleştirildi. DNA'nın 4 molekülünün olması 4 veya 4'ün kuvveti elemanlı sonlu yapılar üzerinde çalışmayı gerektirdi. Abulraub vd., F_4 cisminin elemanları ile DNA molekülleri arasında eşleme yaparak DNA kod inşa ettiler [13]. Öztaş vd. ise F_{16} cisminin elemanları ile 2 uzunluklu DNA bazları (DNA 2-bazları) arasında ilişki kurarak bazı özel polinomlar yardımıyla DNA kod elde ettiler [14] ve bu polinomları kullanarak DNA kodları $F_{4^{2k}}$ cisminde genelleştirdiler [15]. Şiap vd. ise $F_2[u]/(u^2 - 1)$ zincir halkasının (chain ring) elemanları ile DNA molekülleri arasında ilişki kurarak ters-sıralı DNA kod elde ettiler. Yıldız vd. tarafından yapılan çalışmada $F_2[u]/(u^4 - 1)$ zincir halkası kullanılarak DNA 2-bazları ilişkilendirildi [16]. Bayram vd. tarafından ise $F_4 + vF_4$ zincir olmayan halkası (non-chain ring) üzerinde DNA kod çalışması yapıldı [17].

Matematiksel biyoloji alanında çalışan araştırmacıların, DNA dizilerinin yapısındaki hata düzelten kodların varlığına ilişkin merakı olumlu sonuç verdi. Faria vd. tarafından yapılan çalışmada [18], Trav7 (T cell receptor alpha variable 7) geninin $n = 511$ uzunluklu bir kodsöz olarak, sadece bir bit fark ile, devirli kodların (cyclic

codes) önemli bir alt sınıfı olan $GF(4)$ üzerindeki bir BCH kodun elemanına karşılık geldiği görüldü.

1.2 Tezin Amacı

Yerel Frobenius olup zincir olmayan bir halka üzerinde tanımlanan bir C devirli kodunu üreten ideal yapısının, kodun üreteç matrisinin ve kodsöz sayısının belirlenmesi ayrıca dual kodun üreticinin belirlenmesi amaçlanmaktadır. R halkası üzerindeki kodların Gray dönüşümü ile ve döngüsel matrisler ile optimal ikili lineer kodlar elde edilmesi amaçlanmaktadır. R halkasının cebirsel yapısından faydalanılarak ters-sıralı DNA kodların ve ters-sıralı tamlayan DNA kodların elde edilmesi amaçlanmaktadır. Ayrıca gerçek DNA dizileri üzerinde hata düzelten kodların gösterilmesi ve gerçek DNA dizilerinde bulunan özel bölgeler ile hata düzelten kodlar arasında bir ilişki kurulması amaçlanmaktadır.

1.3 Orijinal Katkı

Bu çalışmada yerel Frobenius olup zincir olmayan bir halka üzerinde tanımlanan bir C devirli kodunu üreten ideal yapısı belirlenmekte olup kodun üreteç matrisi ve kodsöz sayısı verilmektedir. Ayrıca dual kodun üretici gösterilmektedir. R halkası üzerindeki kodların Gray dönüşümü ile ve döngüsel matrisler ile optimal ikili lineer kodlar elde edilmektedir. Ters sıralı DNA kod elde etmek için öncelikle ters sıralılık probleminin çözülmesi gerekir. Bu çalışmada ters sıralılık problemini çözmek için daha önce bulunan yöntemlerin dışında R halkasının cebirsel yapısından faydalanılarak, ters sıralılık problemi çözülmekte ve böylece ters-sıralı DNA kodlar ve ters-sıralı tamlayan DNA kodlar belirlenmektedir. Ayrıca insanda ve farede bulunan 7 farklı Trav7 geni ve moleküler biyoloji üzerine yapılan çalışmalarda oldukça sık kullanılan Arabidopsis Thaliana bitkisinden alınan gen (NRPB5 Eukaryotic rpb5 RNA polymerase subunit family protein [Arabidopsis Thaliana (thale cress)]) için kodlanmayan bölge (intron) ve kodlanan bölge (exon) gibi protein sentezinde aktif rol oynayan özel bölgeler ile hata düzelten kodlar arasında bir ilişki kurulmaktadır. Kodlanmayan bölgeyi belli bir yüzde ile tahmin edebilen bir algoritma geliştirilmiştir.

1.4 Tezin İçeriği

Bölüm 2'de öncelikle halkalar, sonlu cisimler ve modüller ile ilgili tezde bahsi geçen genel tanım ve teoremler yer alacaktır. Ayrıca hata düzelten kodlar, lineer kodlar ve devirli kodlardan bahsedilecektir. Son olarak ise DNA ile ilgili kavramlara

değınılecektır.

Bölüm 3'te $\mathbb{F}_2 + u\mathbb{F}_2 + v\mathbb{F}_2 + v^2\mathbb{F}_2$ halkasının yapısı ıncelenecek ve bu halka üzerindeki devırlı kodlar belırlenecektır. Belırlenen kodlar ıçın üretet matrısları ve kodsöz sayısı elde edılecektır. Ayrıca dual kodun üretet belırlenecek ve R halkası üzerindeki lineer kodlar ıçın kendıne dık olma koşulu verılecektır. R halkası üzerindeki kodların Gray dönüşümlerından optımal ikılı lineer kodlar elde edılecektır. Elde edılen kodların kendıne dık kod oldukları ve tümünün bır-ağırlıklı kod oldukları gösterılecektır. Farklı bır kod inşa metodu olarak döngüsel matrıslar kullanılarak optımal kodlar elde edılecektır. Bu şekilde elde edılen kodların ıkı-ağırlıklı optımal kod oldukları hatta ıçlerinde optımal ıkı-ağırlıklı kodların da olduđu gösterılecektır.

Bölüm 4'te $\mathbb{F}_2 + u\mathbb{F}_2 + v\mathbb{F}_2 + v^2\mathbb{F}_2$ halkası üzerindeki devırlı kodlar üzerinden DNA kodlar belırlenecektır. R halkası üzerinde tanımlanan devırlı kodların cebırsel yapısından faydalananarak DNA kodlar ıçın büyük önem taşıyan ters sıralılık problemları çözülerek ters-sıralı DNA kod ve ters-sıralı tamlayan DNA kodlar belırlenecektır.

Bölüm 5'te gerçet DNA dizıları üzerinde hata düzelten kodlar ıncelenecektır. İnsanda bulunan TRAV7 (T cell receptor alpha variable 7) genı ve farede bulunan TRAV7-1 ile TRAV7-6 arasındaki 7 gen üzerindeki lineer kod yapısı ıncelenecektır. Bu genlerden kayan pencete metodu ile elde edılen vektörler satır kabul eden matrısların ürettiđi kodlar belırlenerek özelliklerine göre sınıflandırılacaklardır. Bu bölümde son olarak TRAV7 genı ve Arabidopsis Thaliana genı ıçın hata düzeltme grafiklerı belırlenerek hata düzeltemeyen kodların genler üzerindeki kodlanmayan bölgelerde yoğunlaştıkları gösterılecektır.

2 TEMEL BİLGİLER

Bu bölümde öncelikle halkalarla ilgili bazı kavramlar açıklanacaktır. Ardından hata düzelten kodlar için büyük öneme sahip olan lineer kodlarla ilgili önemli tanım ve teoremler verilecektir ve devirli kodlar açıklanacaktır. Son olarak ise DNA kavramına dair bilgilerden ve DNA koddan bahsedilecektir.

2.1 Halkalar

Bu kısımda halkalar ile ilgili kavramlardan bahsedilecektir. Tanım ve teoremler Thomas W. Hungerford'un, Algebra [19] isimli eseri esas alınarak yazılmıştır.

Tanım 2.1. R boş kümeden farklı bir küme olmak üzere toplama ve çarpma ikili işlemli $(R, +, \cdot)$ cebirsel yapısı verilsin. Aşağıdaki üç koşulun sağlanması durumunda bu yapıya *halka* denir.

1. $(R, +)$ bir değişmeli gruptur.
2. Her $a, b, c \in R$ için $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
3. Her $a, b, c \in R$ için $a \cdot (b + c) = a \cdot b + a \cdot c$ ve $(a + b) \cdot c = a \cdot c + b \cdot c$.

Bundan sonra a ile b 'nin çarpımı için $a \cdot b$ yerine ab kullanılacaktır.

Tanım 2.2. Boş kümeden farklı R kümesi için $(R, +, \cdot)$ cebirsel yapısı verilsin. Buna göre aşağıdaki tanımlar mevcuttur.

1. R halkasında çarpma işlemine göre değişme özelliği varsa, yani her $a, b \in R$ için, $ab = ba$ ise R 'ye bir *değişmeli halka* denir.
2. R halkasında çarpma işleminin birim elemanı varsa, yani her $a \in R$ için, $1_R a = a 1_R = a$ olacak şekilde R halkasında bir 1_R elemanı varsa, R 'ye *birimli halka* denir.

3. R halkasındaki bir u elemanının R 'de tersi varsa, yani $uv = vu = 1_R$ şartını sağlayan bir $v \in R$ varsa u 'ya R 'nin *birimsel elemanı* denir.
4. R birimli halkası için 0_R dışındaki tüm elemanlarının tersi varsa bu halkaya *bölenler halkası* denir.

Bu bölümde geçen tanım ve teoremlerde R halkası birimli ve değişmeli bir halka olarak kabul edilecektir.

Yukarıda bahsedilen tanımlamalara göre $\mathbb{Z}$ ve $\mathbb{Z}_n$ (mod n kalan sınıflarının kümesi) toplama ve çarpma işlemlerine göre değişmeli ve birimli birer halkadır. Çift sayılardan oluşan $2\mathbb{Z}$ halkası ise değişmeli ve birimli olmayan bir halkadır.

Tanım 2.3. R bir halka ve her $r \in R$ için $nr = 0_R$ koşulunu sağlayan pozitif n tamsayılarının en küçüğü R halkasının *karakteristiği*dir. Bu şartı sağlayan n sayısı olmaması durumunda halkanın karakteristiği sıfırdır.

Karakteristiğin tanımına göre $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ ve $\mathbb{C}$ halkalarının karakteristiği 0'dır. $\mathbb{Z}_n$ halkasının karakteristiği ise n 'dir.

Tanım 2.4. R bir halka, S kümesi R halkasının elemanlarından oluşan kümenin boş kümeden farklı bir alt kümesi olsun. S kümesi R halkasındaki işlemlere göre halka olma koşullarını sağlıyorsa S kümesine R halkasının bir alt halkası denir.

Tanım 2.5. R bir halka, I , R halkasının bir alt halkası olsun.

1. Her $r \in R$ ve $b \in I$ için $rb \in I$ ise I bir sol idealdir.
2. Her $r \in R$ ve $b \in I$ için $br \in I$ ise I bir sağ idealdir.
3. I hem sağ hem de sol ideal ise I 'ya R halkasının bir *ideali* denir.

İdealin tanımına göre $k \in \mathbb{Z}$ olmak üzere, $k\mathbb{Z}$ alt halkası $\mathbb{Z}$ halkasının bir idealidir.

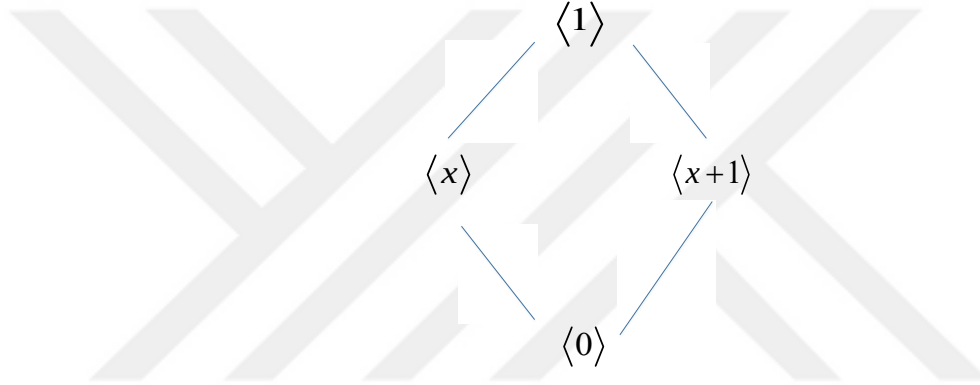
Bir $X \subset R$ kümesi tarafından üretilen ideal $\langle X \rangle$ şeklinde gösterilir. X kümesi sonlu sayıda eleman içeriyorsa, $\langle X \rangle$ idealine *sonlu üretilmiş ideal* denir. Eğer $\langle X \rangle$ ideali, $\langle X \rangle = \langle x \rangle$ gibi tek bir eleman tarafından üretiliyorsa, *temel ideal* olarak adlandırılır. Halkanın tüm idealleri temel ideal ise bu halkaya *temel ideal halkası* denir.

Tanım 2.6. Bir halkanın $\langle 0 \rangle$ ve $\langle 1 \rangle$ ideallerine *aşık ideal*ler, bunların dışındaki ideallerine ise *aşık olmayan ideal*ler denir.

Tanım 2.7. R bir halka olsun. R halkasının tüm idealleri tam sıralı ise R halkasına *zincir halkası* (*chain ring*) denir. Aksi takdirde, bu halkaya *zincir olmayan halka* (*non-chain ring*) denir.

Tanım 2.8. R bir halka olsun. R 'nin aşikar olmayan ideallerinin en büyüğüne *maksimal ideal* denir. Sadece bir maksimal ideale sahip olan halkaya ise *yerel halka* (*local ring*) denir.

Örnek 2.1. $\mathbb{Z}_{16}$ halkasının idealleri $\langle 0 \rangle \subset \langle 8 \rangle \subset \langle 4 \rangle \subset \langle 2 \rangle \subset \langle 1 \rangle = \mathbb{Z}_{16}$ şeklinde zincir oluşturur. Dolayısıyla $\mathbb{Z}_{16}$ bir zincir halkasıdır. Bu halkanın maksimal ideali ise $\langle 2 \rangle$ 'dir. Ancak $S = \mathbb{Z}_2[x]/\langle x + x^2 \rangle$ halkasının idealleri Şekil 2.1'den görüleceği üzere tek bir zincir oluşturmadığı için bir zincir olmayan halkadır.



Şekil 2.1 S halkasının ideal latisi

Tanım 2.9. R bir halka ve S boş kümeden farklı bir alt halkası olsun. $\text{Ann}(S) = \{a \in R : \text{her } x \in S \text{ için } ax = 0_R\}$ kümesi R 'nin bir sol idealidir. Eğer S , R halkasının bir sol ideali ise $\text{Ann}(S)$, R halkasının bir idealidir. $\text{Ann}(S)$ ye R 'nin *sıfırlayıcısı* (*anhilatörü*) denir.

Teorem 2.1. R bir halka, I ise bir ideali olsun. Her $r_1 + I, r_2 + I \in R/I$ için aşağıdaki gibi tanımlanan toplama ve çarpma işlemlerine göre R/I bir halkadır.

- $(r_1 + I) + (r_2 + I) = (r_1 + r_2) + I.$
- $(r_1 + I)(r_2 + I) = r_1 r_2 + I.$

Ayrıca, R halkası değişmeli halka ise R/I halkası da değişmeli halkadır. R halkası birimli halka ise R/I halkası da birimli halkadır.

Yukarıdaki teoremden bahsedilen R/I halkası R halkasının I idealine göre *bölüm halkasıdır*.

Tanım 2.10. $(R, +, \cdot)$ ve $(S, \oplus, \otimes)$ iki halka ve $f : R \rightarrow S$ bir fonksiyon olsun. Her $a, b \in R$ için aşağıdaki işlemler sağlanıyorsa, f fonksiyonuna R halkasından S halkasına bir *halka homomorfizması* denir.

- $f(a + b) = f(a) \oplus f(b)$.
- $f(ab) = f(a) \otimes f(b)$.

Birebir örten halka homomorfizmasına *izomorfizma* denir. Ayrıca R ve S halkaları arasında bir izomorfizma varsa bu iki halka birbirine izomorftur.

Örnek 2.2. n pozitif tamsayı olmak üzere $\mathbb{Z}/\langle n \rangle$ değişmeli ve birimli bir bölüm halkasıdır. Bu halka aynı zamanda $\mathbb{Z}_n$ halkasına izomorftur.

R bir değişmeli halka ve $a_0, a_1, \dots, a_n \in R$ olmak üzere $R[x] = \{a_0 + a_1x + \dots + a_nx^n : n \geq 0\}$ kümesine toplama ve çarpma işlemlerine göre R halkası üzerindeki *polinom halkası* denir.

$R[x]$ polinom halkasının elemanlarına *polinom* denir. Ayrıca $a_n \neq 0$ olmak üzere $f(x) = \sum_{i=0}^n a_i x^i$ polinomunun derecesi n 'dir ve $\deg(f(x))$ olarak gösterilir. $a_n = 1$ ise polinoma özel olarak *monik polinom* denir.

2.2 Sonlu Cisimler

Bu bölümde [19] nolu eserden faydalanılarak sonlu cisimler konusuna kısaca değinilecektir.

Tanım 2.11. Boş kümeden farklı bir R kümesi için $(R, +, \cdot)$ cebirsel yapısı verilsin. Eğer R değişmeli bölenler halkası ise R 'ye *cisim* denir. Cisim, genel olarak F veya $\mathbb{F}$ ile gösterilir.

Tanım 2.12. F bir cisim olsun. $p1_F = 0_F$ koşulunu sağlayan, eğer varsa, pozitif p tamsayılarının en küçüğüne F cisminin *karakteristiği* denir. Bu şartı sağlayan p sayısı olmaması durumunda cismin karakteristiği sıfırdır.

Teorem 2.2. Cismin karakteristiği ya sıfırdır ya da bir asal sayıdır.

Teorem 2.3. Karakteristiği p olan F sonlu cisminin eleman sayısı $q = p^n$ ($n \in \mathbb{Z}^+$)'dir ve F_q ile gösterilir.

2.3 Modüller

Bu bölümde modüllerle ilgili bazı tanım ve teoremlere yer verilecek olup bu bilgiler için [19] nolu kaynaktan faydalanılmıştır. R halkası birimli ve değişmeli bir halkadır.

Tanım 2.13. R bir halka olsun. M toplama işlemine göre bir değişmeli grup olsun. M 'deki elemanların, R 'deki elemanlarla skaler çarpımı, $R \times M \rightarrow M$ fonksiyonu aşağıdaki koşulları sağlıyorsa, M 'ye R üzerinde bir sol modül veya kısaca sol R -modül denir.

1. Her $r \in R$ ve her $m, m' \in M$ için, $r(m + m') = rm + rm'$.
2. Her $r, r' \in R$ ve her $m \in M$ için, $(r + r')m = rm + r'm$.
3. Her $r, r' \in R$ ve her $m \in M$ için, $(rr')m = r(r'm)$.

Tanım 2.14. R bir halka, M bir R -modül ve $N \subseteq M$ boştan farklı bir alt küme olsun. N kümesi bir R -modül ise N kümesine M 'nin bir alt modülü veya R -alt modülü denir.

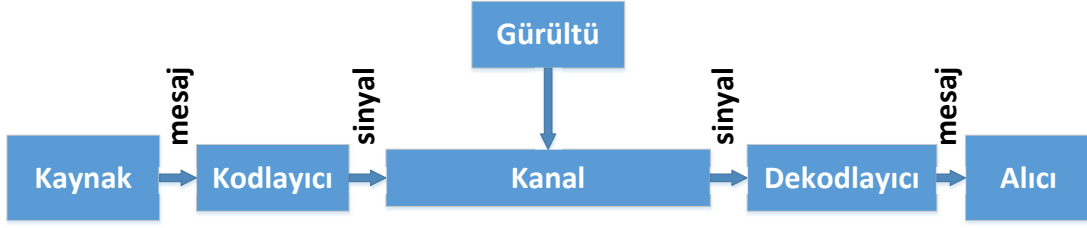
Önerme 2.1. R -modül M 'nin boştan farklı bir $N \subseteq M$ kümesinin alt modül olması için gerek ve yeter şart her $r, r' \in R$ ve her $m, m' \in N$ için $rm + r'm' \in N$ olmasıdır.

Tanım 2.15. M bir R -modül ve $m \in M$ olsun. m elemanı tarafından üretilen sol alt modül $\langle m \rangle = Rm = \{rm : r \in R\}$ olarak tanımlanır. $M = \langle m \rangle$ olacak şekilde bir $m \in M$ bulunabiliyorsa M 'ye devirli modül denir.

Tanım 2.16. R bir halka, M bir R -modül ve $S = \{y_j\}_{j \in I}$ de M 'nin bir üreteç kümesi olsun. Her $m \in M$ elemanı, $r_j \in R$ ve $y_j \in S$ olmak üzere $m = \sum_{j \in I} r_j y_j$ şeklinde yazılabiliyor ve bu yazılış tek türlü oluyorsa, $S = \{y_j\}_{j \in I}$ kümesine M için bir taban denir ve M modülüne serbest modül denir.

2.4 Hata Düzeltken Kodlar

İletişim teorisi, dijital dünyada bilgi iletimi ve bilginin depolanmasını konu olarak ele alır. Bell telefon laboratuvarlarında matematikçi olarak görev yapan Claude Shannon, 1948 yılında yayınladığı makalesinde [3], iletişim kanal modelinin sistematik yapısını sunarak iletişimdeki matematiksel teoriyi geliştirdi. Bu teoriye göre bilgi transferinde ilk olarak kaynaktan çıkan mesaj, kanala uygun hale gelmesi için kodlayıcı (encoder) tarafından kodlanır. Kodsöz (codeword) olarak isimlendirilen kodlanan mesaj kanaldan geçerken gürültüye yani hataya (error) maruz kalır. Kanaldan çıkan mesaj ise dekodlayıcı tarafından dönüştürülür ve tahmini mesaj alıcıya iletilir. Bilgi transferinin basit bir şeması Şekil 2.2'de yer almaktadır.



Şekil 2.2 İletişim kanal modelinin sistematik diyagramı

Shannon'un [3] nolu çalışmasında, eğer uygun kodlama ve kod çözme metodu kullanılırsa bilginin istenilen oranda güvenli iletiminin sağlanacağından bahsedildi. Kodlamanın nasıl yapılacağı sorusu ise kodlama teorisinin ortaya çıkmasını sağlamıştır. Hamming'in [4] nolu çalışması ise hata düzelten kodların temelini oluşturmaktadır.

Bilgi aktarımı esnasında çevresel faktörlerden ya da kaynaktan dolayı bilginin gürültüye maruz kalması kaçınılmazdır. Gürültü ise iletmek istenen bilginin hatalı bir şekilde alıcıya ulaşmasına sebep olabilir. Bir kanal boyunca bilginin iletilmesi ve varsa hatanın tespit edilmesi ve hatta mümkünse düzeltilmesi hata düzelten kodlar teorisinin konusudur.

Bu bölümde hata düzelten kodlar, lineer kodlar ve devirli kodlar ile ilgili bazı tanım ve teoremlere yer verilecektir. Bu tanım ve teoremler [20] nolu eser temel alınarak yazılmıştır.

Tanım 2.17. Alfabe olarak tanımlanan $A = \{a_1, a_2, \dots, a_q\}$ eleman sayısı q olan bir küme olsun. Aşağıdaki tanımlamalar mevcuttur.

1. $w = w_1w_2\dots w_n$ şeklindeki diziye n uzunluklu q -lu söz denir. w sözü $(w_1, w_2, \dots, w_n)$ şeklinde vektör olarak da gösterilir.
2. A kümesinin elemanlarından oluşan n uzunluklu vektörlerin oluşturduğu boştan farklı bir $C \subseteq A^n$ kümesine *kod* denir. C kodunda bulunan vektörlere ise C kodunun *kodsözleri* denir.
3. $A = \{0, 1\} = \mathbb{Z}_2$ ise C koduna *ikili kod* denir.
4. C kodunun kodsöz sayısı $|C|$ ile gösterilir.

Bir kodun kodsözleri arasındaki uzaklık hata tespitinde ve mümkünse düzeltilmesinde büyük önem taşır. Kodsözler arasındaki uzaklık tanımı ve uzaklığa bağlı bazı kavramlar aşağıda yer almaktadır.

Tanım 2.18. $u = (u_1, u_2, \dots, u_n)$ ve $v = (v_1, v_2, \dots, v_n)$ vektörleri A^n kümesindeki birbirinden farklı iki söz olsun. Sözler arasındaki Hamming uzaklık, u ve v sözlerinin koordinatlarının farklı olduğu yerlerin sayısıdır ve $d(u, v) = |\{i : u_i \neq v_i, 1 \leq i \leq n\}|$ şeklinde hesaplanır.

Örnek 2.3. $A = \{0, 1, 2, 3\}$ ise A^5 kümesinden seçilen $u = 12022$ ve $v = 10121$ sözlerinin arasındaki Hamming uzaklık $d(u, v) = 3$ olur.

Tanım 2.19. C , en az iki kodsöze sahip bir kod olsun. C kodunun *minimum Hamming uzaklığı* birbirinden farklı kodsözlerin aralarındaki uzaklıkların en küçüğüdür ve $d(C) = \min\{d(u, v) | u, v \in C, u \neq v\}$ şeklinde ifade edilir.

Tanım 2.20. C , q elemanlı bir alfabede tanımlı, uzunluğu n , eleman sayısı M ve minimum uzaklığı d olan bir kod olsun. n, M ve d sayılarına C kodunun *parametreleri* denir. C kodu ise parametreler cinsinden $(n, M, d)_q$ -kod olarak isimlendirilir.

Örnek 2.4. $C = \{1000, 1100, 1011\}$ ikili kod olsun. C kodunun kodsözleri arasındaki (Hamming) uzaklıklar $d(1000, 1100) = 1$, $d(1000, 1011) = 2$ ve $d(1100, 1011) = 3$ olduğundan $d(C) = 1$ dir ve C kodu ikili $(4, 3, 1)_2$ -koddur.

Bir kodun en küçük uzaklığı, kodun tespit edebileceği ve düzeltebileceği hata sayısı hakkında bilgi verir. Bu bilgiye ilişkin teorem aşağıda yer almaktadır.

Teorem 2.4. C bir kod ve d kodun minimum uzaklığı olsun. C kodu olarak $d - 1$ hata tespit eden koddur.

Teorem 2.5. C bir kod ve d kodun minimum uzaklığı olsun. C kodu olarak $\lfloor (d - 1)/2 \rfloor$ hata düzelten koddur.

Burada $\lfloor x \rfloor$ ile ifade edilmek istenen x 'e eşit ya da x 'ten küçük en büyük tamsayıdır.

2.4.1 Lineer Kodlar

Bu bölümde lineer kodlardan bahsedilecek olup önemli tanım ve teoremlere değinilecektir. Bölümün hazırlanmasında [21] nolu kaynaktan faydalanılmıştır.

Tanım 2.21. F_q^n vektör uzayının bir alt uzayına F_q sonlu cismi üzerinde uzunluğu n olan bir *lineer kod* denir.

Tanım 2.22. C , n uzunluklu bir lineer kod, $u = (u_1, u_2, \dots, u_n)$ C 'nin bir kodsözü olsun. u kodsözünün Hamming ağırlığı sıfırdan farklı bileşenlerinin sayısıdır ve $wt(u)$ şeklinde gösterilir.

u ve v , F_q^n vektör uzayındaki iki söz olsun. İki söz arasındaki uzaklık ile sözler arasındaki fark vektörünün ağırlığı birbirine eşittir, yani $d(u, v) = wt(u - v)$ olur.

Tanım 2.23. C bir lineer kod olsun. C kodunun *minimum Hamming ağırlığı*, kodsözlerinin Hamming ağırlıklarının sıfırdan farklı en küçük olanıdır ve $wt(C)$ şeklinde gösterilir.

Teorem 2.6. C bir lineer kod olsun. C kodunun minimum Hamming ağırlığı ile minimum Hamming uzaklığı birbirine eşittir, yani $wt(C) = d(C)$ olur.

C , F_q^n vektör uzayında bir lineer kod olsun. Uzunluğu n , boyutu k ve minimum Hamming ağırlığı d olan bir C lineer kodu parametreler cinsinden $[n, k, d]_q$ -kod şeklinde ifade edilir. C kodunun kodsöz sayısı $|C| = q^k$ 'dir. Boyutun sıfır olması durumunda ise C koduna *aşık kod* denir ve bu kod sadece sıfır vektörünü içerir.

Örnek 2.5. $C = \{00000, 10000, 10011, 00011\}$ ikili lineer kodunun kodsözlerinin ağırlıkları sırasıyla $wt(00000) = 0$, $wt(10000) = 1$, $wt(10011) = 3$ ve $wt(00011) = 2$ 'dir. C lineer kodunun uzunluğu 5, minimum Hamming uzaklığı 1 ve boyutu 2 olduğundan kısaca $[5, 2, 1]_2$ -kod olarak ifade edilir.

Tanım 2.24. $q > 1$, n ve d pozitif tamsayıları $1 \leq d \leq n$ şartını sağlasın. *Singleton sınırı* (*singleton bound*), $|C| \leq q^{n-d+1}$ olacak şekilde C kodunun kodsöz sayısı için bir üst sınır verir.

Tanım 2.25. Singleton sınırını sağlayan yani parametreleri $d = n - k + 1$ şartını sağlayan $[n, k, d]$ -koda *MDS* (*maximum distance separable*) kod denir [21].

Kodlama teorisinin temel problemi n , k ve d parametrelerinden herhangi ikisi sabit tutulduğunda diğerinin en uygun hale getirilmesidir. Bu optimizasyonu gerçekleyen kodlara *optimal kodlar* denir. Lineer kodlar için uzunluğun ($1 \leq n \leq 256$) ve boyutun ($1 \leq k \leq 256$) sabit tutulması durumunda mümkün olan en büyük Hamming uzaklık [22] nolu kaynakta yer almaktadır.

C bir lineer kod, $\mathbf{u} = (u_1, u_2, \dots, u_n)$ ve $\mathbf{v} = (v_1, v_2, \dots, v_n)$ vektörleri R^n vektör uzayında iki vektör olsun. $\mathbf{u}$ ve $\mathbf{v}$ vektörlerinin *Öklit iç çarpımı*, işlemler R halkasında olacak şekilde $[\mathbf{u}, \mathbf{v}] = \sum_{i=1}^n u_i v_i$ olarak hesaplanır.

R halkası üzerindeki n uzunluklu C lineer kodu için, C kodunun *dual kodu* $C^\perp = \{\mathbf{u} \in R^n : \text{her } \mathbf{v} \in C, [\mathbf{u}, \mathbf{v}] = 0\}$ olarak tanımlanır.

Tanım 2.26. C bir lineer kod olsun. $C \subset C^\perp$ ise C koduna *kendine dik* (*self-orthogonal*) kod denir. $C = C^\perp$ ise C koduna *kendine dual* (*self-dual*) kod denir.

Lineer kodlarda baz vektörlerinin bilinmesi kodsözleri tanımlamada kolaylık sağlar. Kodlama teorisinde lineer kodların bazıları matris olarak gösterilir. Bir lineer kodun baz vektörlerinin oluşturduğu matrise kodun *üreteç matrisi* denir. Üreteç matrisi G ile gösterilir. Lineer kodun dual kodunun üreteç matrisine ise *parite kontrol matrisi* denir. Parite-kontrol matrisi H ile gösterilir.

Lineer kodların lineer olmayan kodlara göre birçok avantajı vardır. Bu avantajlar aşağıdaki gibi sıralanabilir.

1. Lineer kodlar, vektör uzayı oldukları için bir taban kullanılarak daha kolay ifade edilirler.
2. Lineer kodun minimum uzaklığı, kodun sıfırdan farklı kodsözlerinin ağırlıklarının en küçüğüne eşittir.
3. Lineer kodlar için kullanılan kodlama ve dekodlama algoritmaları daha hızlı çalışır [20].

Kodlama teorisi üzerine yapılan çalışmalar genellikle lineer kodlar üzerinedir. Bunun sebebi lineer kodların bir cebirsel yapıya sahip olmasıdır [20]. Örneğin, lineer bir kod üreteç veya kontrol matrisi ile ifade edilir. Kodun parite kontrol matrisinden, minimum Hamming uzaklığına ulaşılabilir. Minimum Hamming uzaklığı bilinen kodun kaç hata tespit edebileceği veya düzeltebileceği bilgisi de kolayca elde edilir.

2.4.2 Devirli Kodlar

Bir lineer koddaki her bir kodsözün dairesel döndürülmesi ile yine bir kodsöz elde ediliyorsa bu tip kodlara *devirli kod* (*cyclic code*) denir. Bu fikir ilk olarak Prange tarafından 1957 yılında ortaya atılmıştır [23]. Devirli kodların yapısı incelendiğinde ise n uzunluklu, k boyutlu devirli kodun, derecesi $n - k$ olan bir polinom tarafından ifade edilebildiği görülmüştür. En önemli kod ailelerinden olan Hamming kod, Golay kod ve BCH kod devirli kodlardır [20].

Günümüze kadar devirli kodlarla ilgili farklı yapılar üzerine birçok çalışma yapılmıştır. Bu bölümde devirli kodların cebirsel yapısından bahsedilecektir. Tanım ve teoremler için [20] nolu kaynaktan faydalanılmıştır.

Tanım 2.27. C , n uzunluğunda bir lineer kod ve her $(c_0, c_1, \dots, c_{n-1}) \in C$ için $(c_{n-1}, c_0, \dots, c_{n-2}) \in C$ ise C 'ye devirli kod denir. $\sigma(c_0, c_1, \dots, c_{n-1}) = (c_{n-1}, c_0, \dots, c_{n-2})$ devirsel hareketini sağlayan σ dönüşümüne ise devirsel öteleme (*cyclic shift*) denir [34].

Örnek 2.6. $C = \{0000, 1100, 0110, 0011, 1001\}$ bir devirli koddur.

Devirli kodların kombinatoriyel yapısı polinomlar yardımıyla cebirsel yapıya dönüştürülür. $\Phi : F_q^n \longrightarrow F_q[x]/\langle x^n - 1 \rangle$ lineer dönüşümü $c = (c_0, c_1, \dots, c_{n-1})$ kodsözünü polinom halkasındaki $c(x) = c_0 + c_1x + \dots + c_{n-1}x^{n-1}$ elemanına eşler. Bu durumda c elemanının devirsel ötelemesi polinom halkasında $xc(x) \bmod(x^n - 1)$ polinomuna karşılık gelir. Ayrıca Φ lineer dönüşümü yardımıyla devirli kodlar ideallerle ilişkilendirilebilir.

Teorem 2.7. $C \subseteq F_q^n$ lineer kodunun devirli kod olması için gerek ve yeter koşul $\Phi(C)$ kümesinin $F_q[x]/\langle x^n - 1 \rangle$ polinom halkasının bir ideali olmasıdır.

Teorem 2.8. I ideali $F_q[x]/\langle x^n - 1 \rangle$ halkasının sıfırdan farklı bir ideali ve $g(x)$, I idealinin sıfırdan farklı en küçük dereceli monik polinomu olsun. $g(x)$, I idealinin üreteç polinomudur ve $x^n - 1$ polinomunu böler.

Teorem 2.9. $F_q[x]/\langle x^n - 1 \rangle$ halkasının sıfırdan farklı her ideali için en küçük dereceli monik polinom vardır.

Tanım 2.28. $F_q[x]/\langle x^n - 1 \rangle$ halkasının bir I idealinin sıfırdan farklı en küçük dereceli monik polinomuna I 'nin üreteç polinomu denir. C devirli kodu için ise $\Phi(C)$ 'nin üreteç polinomu aynı zamanda C kodunun üreteç polinomudur.

Yukarıdaki tanım ve teoremlerin uygulaması niteliğindeki örnek aşağıda yer almaktadır.

Örnek 2.7. $C = \{0000, 1100, 0110, 0011, 1001\}$ devirli kodunu ele alalım. $\Phi(C) = \{0, 1 + x, x + x^2, x^2 + x^3, 1 + x^3\}$ kümesi $F_2[x]/\langle x^4 - 1 \rangle$ halkasının $1 + x$ elemanı tarafından üretilen idealidir. $1 + x$ elemanı aynı zamanda C devirli kodunun üreteç polinomudur.

Teorem 2.10. $x^n - 1 \in F_q[x]$ polinomunu bölen her bir monik polinom, F_q^n 'de bir devirli kodun üreteğine tekabül eder.

Teorem 2.9 ve 2.10'den aşağıdaki sonuç elde edilir.

Sonuç 2.10.1. $x^n - 1 \in F_q[x]$ polinomunun monik bölenleri ile F_q^n 'deki devirli kodlar arasında birebir eşleme vardır.

Teorem 2.11. $g(x)$, $F_q[x]/\langle x^n - 1 \rangle$ polinom halkasının bir idealinin üreteç polinomu olsun. $g(x)$ polinomunun derecesi $n - k$ ise ideale karşılık gelen devirli kodun boyutu k 'dir.

Örnek 2.8. $x^6 - 1$ polinomunun $F_2[x]$ 'de çarpanlara ayrılışı $x^6 - 1 = (1 + x)^2(1 + x + x^2)^2$ şeklindedir. $\langle (1 + x + x^2)^2 \rangle$ idealine karşılık gelen devirli kod bir $[6, 2, 3]$ -koddur ve kodsözleri $\{000000, 101010, 010101, 111111\}$ şeklindedir.

Teorem 2.12. $x^n - 1 = g(x)h(x)$ ve C kodu üreteç polinomu $g(x) = g_0 + g_1x + \dots + g_{n-k}x^{n-k}$ olan bir devirli kod olsun. C devirli kodunun üreteç matrisi aşağıdaki gibidir.

$$G = \begin{pmatrix} g(x) \\ xg(x) \\ \vdots \\ x^{k-1}g(x) \end{pmatrix} = \begin{pmatrix} g_0 & g_1 & \cdots & g_{n-k} & 0 & 0 & 0 & \cdots & 0 \\ 0 & g_0 & g_1 & \cdots & g_{n-k} & 0 & 0 & \cdots & 0 \\ \vdots & & & & & & & & \vdots \\ 0 & 0 & 0 & g_0 & g_1 & \cdots & \cdots & \cdots & g_{n-k} \end{pmatrix}. \quad (2.1)$$

C devirli kodunun parite kontrol matrisi ise $h(x) = h_0 + h_1x + \dots + h_kx^k$ olmak üzere aşağıdaki gibi elde edilir.

$$H = \begin{pmatrix} h^*(x) \\ xh^*(x) \\ \vdots \\ x^{n-k-1}h^*(x) \end{pmatrix} = \begin{pmatrix} h_k & h_{k-1} & \cdots & h_0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & h_k & h_{k-1} & \cdots & h_0 & 0 & 0 & \cdots & 0 \\ \vdots & & & & & & & & \vdots \\ 0 & 0 & 0 & h_k & h_{k-1} & \cdots & \cdots & \cdots & h_0 \end{pmatrix}. \quad (2.2)$$

2.5 DNA ile İlgili Bazı Kavramlar

DNA; hücrelerin nasıl davranması gerektiği bilgisini ve biyolojik gelişmelerle ilgili genetik talimatları içinde barındıran bir moleküldür. Adenin (A), guanin (G), sitozin (C) ve timin (T) bazlarından (nükleotid) oluşur. DNA molekülünün şeker ve fosfattan oluşan yapısına tutunan bazlar DNA molekülünün bir dizisini oluştururlar. İki DNA dizisinin, birbirini tamamlayan bazların karşı karşıya gelmesiyle birbirine bağlanarak DNA çift sarmal (double helis) yapısı oluşturmaya ise *hibridizasyon* denir. Bu bağlanma WCC (Watson Crick complement) özelliği gereği birbirinin tamamlayıcı bazları olan A ile T ve G ile C baz eşleşmesi ile gerçekleşir. Tamamlayıcı bazlar arasındaki ilişki $A^c = T$ ve $G^c = C$ şeklinde gösterilir. Bağlanma esnasında tamamlayıcı bazlardan A ile T arasında iki zayıf hidrojen bağ, G ile C arasında ise üç zayıf hidrojen bağ oluşur. Bazlar yapısal olarak ise pürin ve pirimidin olarak ikiye ayrılırlar. Pürinler adenin ve guanin, pirimidinler ise timin ve sitozinden oluşur.

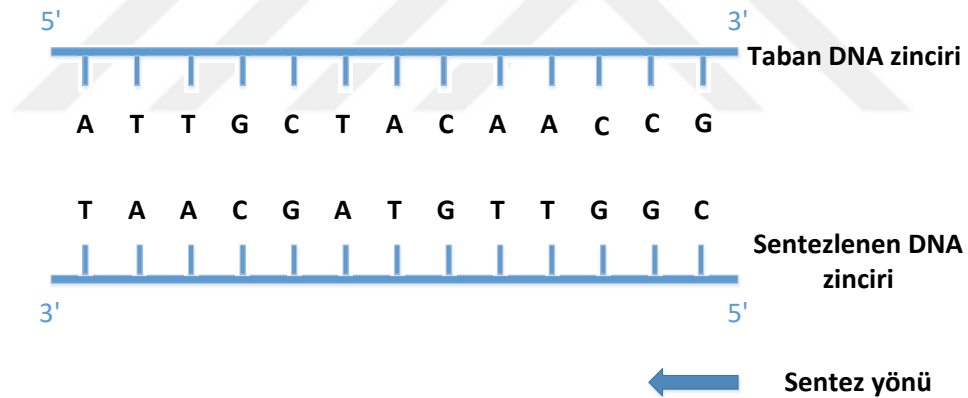
DNA dizisinin uç noktaları yapısal farklılıklardan dolayı 3' ucu ve 5' ucu olarak isimlendirilir. DNA ve RNA dizileri 5' ucundan 3' ucuna doğru sentezlenir. Hibridizasyon esnasında ise iki zincir birbirlerine antiparalel olarak yapışır. Bu ise bir taban DNA dizisi aldığımızda bu diziden ters-sıralı tamlayan bir DNA dizi



Şekil 2.3 DNA çift sarmal görüntüsü [24]

sentezlendiğini gösterir.

Şekil 2.4’de taban DNA dizisi $5' - ATTGCTACAACCG - 3'$ olsun. Bu zincirden sentezlenen DNA dizisi $5' - CGGTTGTAGCAAT - 3'$ olur. Buna göre, $5' - ATTGCTACAACCG - 3'$ ve $5' - CGGTTGTAGCAAT - 3'$ zincirleri WCC kuralına göre birbirlerine yapışarak DNA’nın çift sarmal yapısını oluştururlar.



Şekil 2.4 WCC kuralına göre DNA dizilimi

Bu tezde geçen DNA ile ilgili bazı kavramlar kısaca aşağıdaki gibidir:

- Replikasyon; Hücre bölünmesi öncesinde DNA molekülünün kendini eşlemesine denir.
- Yazılma (transkripsiyon); DNA dizisindeki genetik bilginin kopyalanıp aynısının mRNA'ya (messenger RNA) aktarılmasıdır.
- Okunma (translasyon) ise mRNA üzerindeki kodonlara (mRNA'da bulunan 3 nükleotidden oluşan diziler) uygun amino asitler ile ribozom organelinde protein sentezlenmesidir.

- Mutasyon; genetik materyalde meydana gelen deęişikliklerdir.

2.6 DNA Replikasyonu ile Kodlama Teorisi Arasındaki İlişki

Biyolojide DNA veya protein dizisinin bir bilgisayar kullanılarak, biyoenformatik yöntemlerle incelenmesine *dizi analizi (sequences analysis)* denir. DNA, sonlu ve sembolik bir dizidir. Dolayısıyla DNA'nın incelenmesi dizi analizinden kodlama teorisinin kullanımına genişletilebilir.

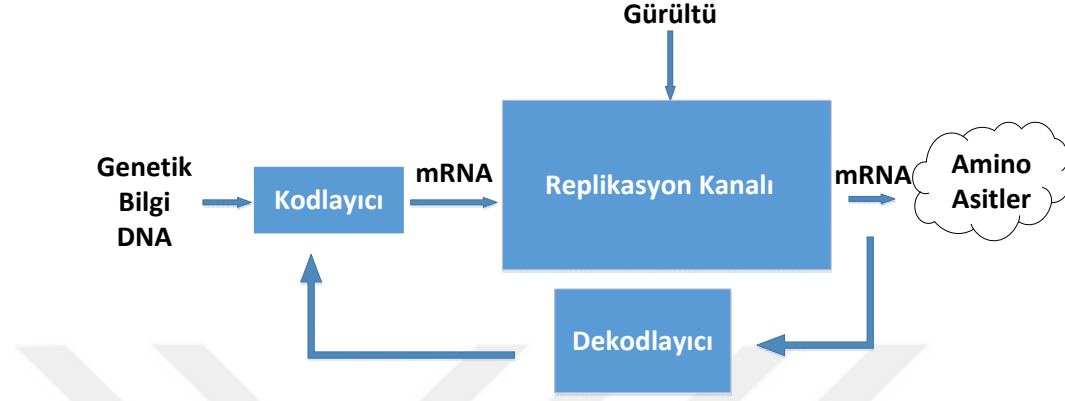
İlk olarak Battail, mutasyonla ağır şekilde deęiştirilen bir dizinin tekrardan evrimleştiğini ve bir replikasyon kod çözme çerçevesi sunduğunu belirtti [25]. May vd. ise blok ve evrişimli kodlama modellerini kullanarak E. coli çeviri başlatma dizilerini inceledi [26, 27]. Bu çalışmada mRNA gürültülü kodlanmış bir sinyal olarak görülür ve diziyi çeviren ribozom kod çözücü olarak düşünülür. Ribozomal kod çözme modelini parametreleştirmek için çeşitli biyolojik ve kimyasal faktörler kullanılır. Blok kod modeli, ribozomal bağlanma sahasının (binding site) [28] tanınmasında etkilidir ve evrişimli model, çevrilmiş ve çevrilmemiş diziler arasında kolayca ayırt edilir. May vd. [26, 27] bu çalışmalar ile kodlama modellerinin sinyal tanımada etkili olduğunu ve DNA'da doğal bir kodlama yapısı olup olmadığı sorusunu akıllara getirdi.

Replikasyon süreci tek başına ele alındığında 10^{-3} ile 10^{-5} arasında bir hata oranı vardır [29]. Ancak DNA kendi içinde hata düzeltmeyi amaçlayan "tekrar okuma (proofreading)" mekanizmasına sahiptir. Replikasyon işlemi sırasında, ilk önce sarmal yapı fermuar gibi açılmaya başlar ve açılan kısım iki ayrı zincire ayrılır. Zincirin karşısına sırasıyla uygun tamamlayıcı bazlar kendilerini bağlarlar ve birbirinin aynısı olan iki merdiven oluşur. Replikasyon sonrası esas DNA molekülü ile nükleotid dizilimi aynı olan yeni DNA molekülleri oluşur.

Replikasyon esnasında meydana gelen yerine koyma hatası, tamamlayıcı bazın yerleşmesi esnasında bağlanması gereken pürin yerine farklı bir pürin yerleşmesidir. Benzer şekilde bir primidin, tamamlanma aşamasında bağlanması beklenen primidinin yerine geçebilir. Örneğin, guanin ile sitozin primidininin bağlanması gerekirken başka bir primidin olan timin bağlanır. Meydana gelen bu uyumsuzluk DNA'da bir kıvrılma (kink) oluşmasına neden olur ve doğru baz bağlanana kadar sentez durur. Bu basit tekrar okuma, hata oranını yaklaşık 10^{-10} 'a kadar düşürür [29].

Bilim insanları iletişim kanalı modeli ile DNA işlemleri arasında paralellik olduğunu düşündüler. Bir doktrinde kanal, kodonların proteinelere dönüştüğü yer olarak görüldü [30]. May vd. ise [27], kanalın DNA için çoğalma işleminin gerçekleştiği ve genetik bilginin kuşaktan kuşağa aktarıldığı ortam olarak kabul etti. Bu doktrin

DNA'nın transkripsiyonunun ve kopyalanmasının gürültülü bir işlem olmasından dolayı mutasyon modellemesi için iyi olduğunu söyledi [27]. Yukarıda bahsettiğimiz tekrar okuma mekanizmasının etkinliği engellendiğinde, replikasyon sürecinde hata oranı artar. Bu da bizi replikasyon gürültüsüne karşı koruyacak bir blok kodlama yapısı olduğu hipotezine götürür.



Şekil 2.5 Kodlama varsayımıyla birlikte gen replikasyonlu gürültülü kanal modeli [31]

Şekil 2.2 ve Şekil 2.5 incelendiğinde DNA'nın mRNA'ya kodlanması ve mRNA'nın proteinlere dekodlanması süreci, iletişim kanal modeli ile benzerlik göstermektedir.

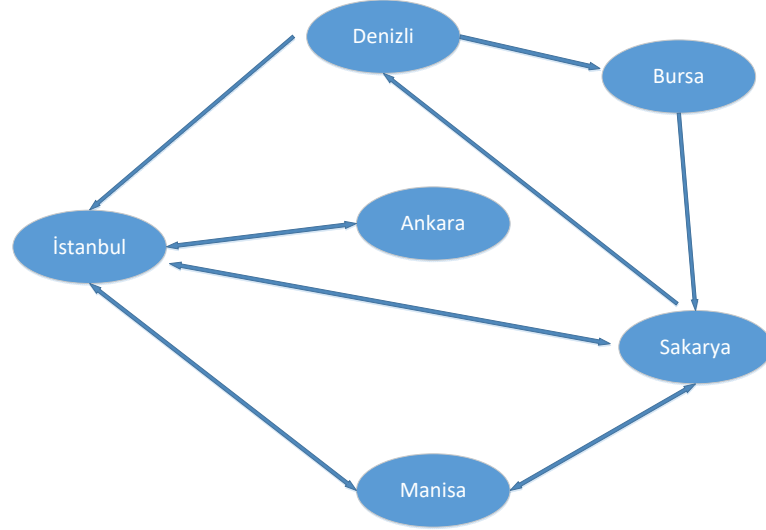
2.7 DNA Hesaplama ve DNA Kod

Adleman [6], 1994 yılında yayınlanan "Molecular Computation of Solutions to Combinatorial Problems" isimli makalesinde, sentetik DNA dizilerinin WCC kuralına göre birbirleri ile olan etkileşimlerinden faydalanarak 7 düğüm için Hamilton yolu problemini çözdü. Bu çalışma, DNA dizilerinin manipüle edilerek matematiksel problemlerin çözümünde hesaplama aracı olarak kullanılabileceğini gösterdi.

Hamilton yolu probleminin çözümü için, yönlü veya yönsüz bir graf verildiğinde her düğümden yalnızca bir kez geçmek şartıyla tüm düğümlerden geçen bir yol olup olmadığı sorusuna cevap aranır. Bu problemi 6 düğüm için modelleyelim.

Örnek 2.9. Şehirler arasındaki uçuşlar Şekil 2.6'te yönlü graf ile gösterilmiştir. Buna göre Manisa ve Sakarya arasında gidiş-dönüş uçuşu vardır. Sakarya'dan Denizli'ye ise sadece gidiş uçuşu vardır, dönüş uçuşu yoktur.

Denizli'den başlayıp tüm şehirlere bir kez uğrayarak Ankara'ya giden bir yol aradığımızı kabul edelim. Bunun için öncelikle her şehir 6 uzunluklu DNA dizileri ile kodlanır. Şehirler arasındaki tüm olası uçuşlar, şehirler için yapılan kodlamalara göre, tamamlayıcı bazlar kullanılarak kodlanır. Yollar için yapılan kodlama kalkış şehrine



Şekil 2.6 Şehirler arasındaki uçuşlar için yönlü graf

karşılık gelen DNA dizisinin son 3-bazının tamlayanı ile varış şehrinin ilk 3-bazının tamlayanının yanyana yazılması ile olur. Örneğin; Denizli şehrinin DNA dizi karşılığı *CCGATG*, Bursa şehrinin DNA dizi karşılığı *CGGTGC* olarak verildiğinden Denizli'den Bursa'ya uçuş *TACGCC* dizisi ile ifade edilir.

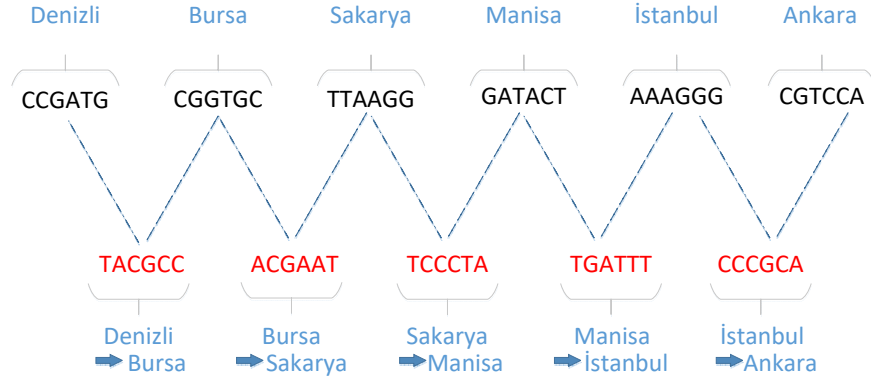
Şehir	DNA Dizisi
Denizli	CCGATG
Bursa	CGGTGC
Sakarya	TTAAGG
Manisa	GATACT
İstanbul	AAAGGG
Ankara	CGTCCA

Tablo 2.1 Şehirler ve DNA karşılıkları

Problemin çözümü için ise, şehirlere karşılık gelen DNA dizileri çoğaltılarak bir tüpe, yollara karşılık gelen DNA dizileri ise çoğaltılarak başka bir tüpe konur. İki tüp karıştırılır. Hibridizasyonu sağlayan enzim eklenerek dizilerin uygun çift sarmalları oluşturmaları sağlanır. Farklı uzunluklu çift sarmallar elektromanyetik alandan geçirilerek 6 uzunluklu zincirler ayıklanır. Son olarak da ayıklama yöntemleri kullanılarak, Denizli ile başlayıp Ankara ile biten zincirler elde edilir. Kalan DNA dizileri Şekil 2.7'deki gibi okunarak problemin çözümüne ulaşılır.

Tanım 2.29. C , n uzunluklu vektörlerden oluşan bir kod olsun. C kodunun $c = \{c_1, c_2, \dots, c_n\}$ kodsözlerinin bileşenleri $c_i \in \{A, G, C, T\}$ ise C 'ye DNA kod denir.

Tanım 2.30. C bir DNA kod ve u bir kodsöz olsun. u kodsözünün ters-sıralısı (reverse) u^r , ters-sıralı tamlayanı (reverse-complement) u^{rc} şeklinde gösterilir.



Şekil 2.7 Çift sarmal molekül çözümü

Örnek 2.10. $u = GTACAGCTT$ vektörünün ters-sıralısı $u^r = TTCGACATG$, ters-sıralı tamlayanı ise $u^{rc} = AAGCTGTAC$ vektörüdür.

Tanım 2.29'e göre yukarıdaki örnekte şehirleri temsil eden 6 uzunluklu vektörlerin oluşturduğu küme bir DNA koddur.

Adleman'ın deneyinden de görüleceği gibi problemin çözümünde DNA dizilerinin seçimi oldukça önemlidir. Hibridizasyon sürecinde dizilerin kendi üzerlerine yapışmaları ya da meydana gelebilecek istenmeyen tamamlamalar problemin çözümüne ulaşmayı engeller. Bu tip olumsuzlukların önüne geçmek maksadıyla DNA kodların kodsözlerinin bazı kombinatöriyel kısıtları sağlamaları gerekir. C bir DNA kod ve kodsözler arasındaki uzaklık d olmak üzere kodsözler için bazı kısıtlar aşağıdaki gibidir.

1. Hamming uzaklığı kısıtı: her $u, v \in C$ ($u \neq v$) için $d(u, v) \geq d$.
2. Ters-sıra kısıtı: her $u, v \in C$ için $d(u^r, v) \geq d$.
3. Ters-sıra tamlayan kısıtı: her $u, v \in C$ için $d(u^{rc}, v) \geq d$.
4. Sabitlenmiş GC-miktar kısıtı: C kodunun her bir kodsözü için toplam G ve C sayısının eşit olmasıdır.

Tanım 2.31. D bir DNA kod olsun. Her u kodsözü için u^r vektörü de D 'nin bir kodsözü oluyorsa D koduna *ters-sıralı DNA kod* denir.

Tanım 2.32. D bir DNA kod olsun. Her u kodsözü için u^{rc} vektörü de kodsöz oluyorsa D koduna *ters-sıralı tamlayan DNA kod* denir.

Ters-sıralı tamlayan DNA kod elde etmek için ters-sıralı DNA kodlar önemli bir basamaktır. Ters-sıralı DNA kodun vektörlerinin ve her vektörünün tamlayanının oluşturduğu küme bir ters-sıralı tamlayan DNA koddur.

Örnek 2.11. $D = \{AAAAA, ACTTG, GTTCA\}$ bir ters-sıralı DNA koddur. Ayrıca $D' = \{AAAAA, TTTT, ACTTG, GTTCA, TGAAC, CAAGT\}$ bir ters-sıralı tamlayan DNA koddur.



$\mathbb{F}_2 + u\mathbb{F}_2 + v\mathbb{F}_2 + v^2\mathbb{F}_2$ HALKASI ÜZERİNDEKİ DEVİRLİ KODLAR

Bu bölümde öncelikle $R = \mathbb{F}_2 + u\mathbb{F}_2 + v\mathbb{F}_2 + v^2\mathbb{F}_2$ halkası incelenecektir. Kısım 3.2’de R halkasının ideal yapısı açıklanacaktır. Kısım 3.3’te ise R halkası üzerinde tanımlı devirli kodlar belirlenecektir. Son olarak, Kısım 3.4’te homojen ağırlık tanımlanarak R halkası üzerindeki devirli kodların homojen ağırlığa göre ikili görüntüleri incelenecektir.

Kodlama teorisinde genel olarak cisimler üzerine çalışmalar yapılmıştır. Ancak 1994 yılında Hammons vd. tarafından $\mathbb{Z}_4$ halkası üzerinde kod çalışması yapıldı [32]. Bu çalışmadan sonra pek çok araştırmacı tarafından bazı özel halkalar üzerindeki kod yapıları incelendi [33–35].

Bölüm 2’de cisimler üzerinde tanımlanan lineer kodlardan bahsedildi. Ancak lineer kodlar halkalar üzerinde de tanımlanabilirler.

Tanım 3.1. R bir halka ve C kodu, R halkası üzerindeki uzunluğu n olan bir kod olsun. C kodu R^n ’in bir R -alt modülü ise C ’ye *lineer kod* denir.

Cisimler birer vektör uzayı olduklarından, cisim üzerinde tanımlanan bir lineer kodun baz vektörleri kullanılarak üreteç matrisi elde edilir. Fakat bir halka üzerindeki lineer kod için bazdan bahsedilemez. Bunun yerine en küçük üreteç kümesi kullanılır. Bu üreteç kümesinden faydalanılarak kodun üreteç matrisi elde edilir.

Tanım 3.2. C , R halkası üzerinde tanımlı bir kod olsun. Bir S kümesinin elemanları diğer elemanlarının lineer kombinasyonundan oluşmuyor ve öz alt kümeleri (proper subsets) C ’yi germiyorsa S ’ye C ’yi üreten *en küçük üreteç kümesi* denir.

3.1 $\mathbb{F}_2 + u\mathbb{F}_2 + v\mathbb{F}_2 + v^2\mathbb{F}_2$ Halkası

$\mathbb{F}_2 + u\mathbb{F}_2 + v\mathbb{F}_2 + v^2\mathbb{F}_2$ halkası karakteristiği 2 olan değişmeli bir halka olmasının yanı sıra $R = \mathbb{F}_2[u, v]/\langle u^2 = v^2, uv = 0 \rangle$ bölüm halkası olarak da ifade edilir. R halkası 16

elemanlıdır ve bu elemanlar; a, b, c, d elemanları $\mathbb{F}_2$ cisminden olmak üzere $a + bu + cv + dv^2$ şeklindedir. Elemanların tamamı aşağıdaki gibidir:

$$R = \{0, u, v, u + v, v^2, u + v^2, v + v^2, u + v + v^2, 1, 1 + u, 1 + v, 1 + u + v, 1 + v^2, 1 + u + v^2, 1 + v + v^2, 1 + u + v + v^2\}. \quad (3.1)$$

R halkasının iki elemanı $\alpha_1 = a_1 + b_1u + c_1v + d_1v^2$ ve $\alpha_2 = a_2 + b_2u + c_2v + d_2v^2$ olmak üzere R halkası üzerinde toplama işlemi ve çarpma işlemi aşağıdaki gibidir:

$$\begin{aligned} \alpha_1 + \alpha_2 &= (a_1 + a_2) + (b_1 + b_2)u + (c_1 + c_2)v + (d_1 + d_2)v^2, \\ \alpha_1 \alpha_2 &= (a_1 a_2) + (b_1 a_2 + a_1 b_2)u + (a_2 c_1 + a_1 c_2)v + (b_1 b_2 + c_1 c_2 + a_2 d_1 + a_1 d_2)v^2. \end{aligned} \quad (3.2)$$

R halkasının birimsel elemanları ise aşağıdaki gibidir:

$$U(R) = \{1, 1 + u, 1 + v, 1 + u + v, 1 + v^2, 1 + u + v^2, 1 + v + v^2, 1 + u + v + v^2\}. \quad (3.3)$$

Bundan sonraki bölümlerde $\mathbb{F}_2 + u\mathbb{F}_2 + v\mathbb{F}_2 + v^2\mathbb{F}_2$ halkası kısaca R halkası olarak gösterilecektir.

3.2 R Halkasının İdeal Yapısı

R halkasının ideal yapısı incelendiğinde ideallerinin birbirini içeren tek bir zincir oluşturmadığı görülür:

$$\{0\} = I_0 \subseteq I_{v^2} \subseteq I_u, I_v, I_{u+v} \subseteq I_{u,v} \subseteq I_1 = R. \quad (3.4)$$

R halkasının aşık ar olmayan idealleri aşağıdaki gibidir:

$$I_{v^2} = \langle v^2 \rangle = \{0, v^2\}, \quad (3.5)$$

$$I_u = \langle u \rangle = \{0, u, u + v^2, v^2\}, \quad (3.6)$$

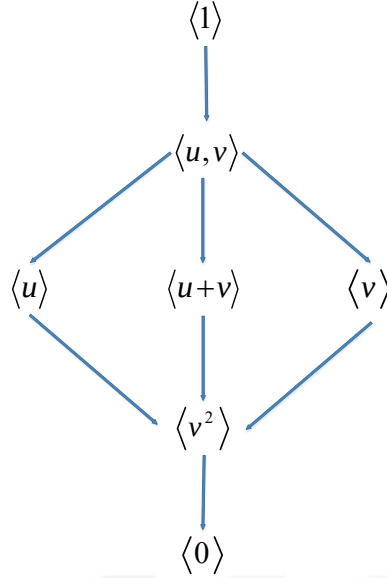
$$I_v = \langle v \rangle = \{0, v, v + v^2, v^2\}, \quad (3.7)$$

$$I_{u+v} = \langle u + v \rangle = \{0, u + v, v^2, u + v + v^2\}, \quad (3.8)$$

$$I_{u,v} = \langle u, v \rangle = \{0, u, v, u + v, v^2, u + v^2, v + v^2, u + v + v^2\}. \quad (3.9)$$

R halkasının yukarıda verilen ideallerinin lat is gösterimi Şekil 3.1'de yer almaktadır.

R halkasının aşık ar olmayan ideallerinden $I_{u,v}$ ideali, R halkasının tek maksimal idealidir. Ayrıca $I_{u,v}$ idealinin elemanları R halkasının sıfır bölenleridir. R halkasının birimsel elemanları ise a elemanı $I_{u,v}$ idealinin bir elemanı olmak üzere $1 + a$ şeklindedir.



Şekil 3.1 R Halkasının ideal latisi

Şekil 3.1 ve Tanım 2.8'den R halkası için aşağıdaki sonuçlara varılır.

1. R bir yerel halkadır, çünkü $I_{u,v}$ tek maksimal idealdir.
2. R zincir olmayan halkadır.
3. R temel ideal halkası değildir, çünkü $I_{u,v} = \langle u, v \rangle$ temel ideal değildir.

Teorem 3.1. R halkası bir Frobenius halkadır.

İspat. Teoremin ispatı için [36] nolu kaynaktaki sonuç kullanılır. Buna göre, R , maksimal ideali M ve kalan cismi (residue field) K olan bir sonlu yerel halka ise aşağıdaki koşullar birbirine denktir.

1. R bir Frobenius halkadır.
2. M maksimal idealinin sıfırlayıcısının K kalan cismi üzerindeki boyutu 1'dir.

Şekil 3.1'e göre R halkası, maksimal ideali $I_{u,v}$ olan sonlu bir yerel halkadır ve kalan cismi $\mathbb{F}_2$ 'dir. $I_{u,v}$ idealinin sıfırlayıcısı $\text{Ann}(I_{u,v}) = \{0, v^2\}$ olduğu için $I_{u,v}$ idealinin sıfırlayıcısının $\mathbb{F}_2$ kalan cismi üzerindeki boyutu 1 olur. O halde R halkası bir Frobenius halkadır. ■

Kodlama teorisi alanında kullanmaya uygun en geniş aile sonlu Frobenius halkalarıdır. Çünkü MacWilliams'ın iki önemli teoremi olan genişleme teoremi (extension theorem)

ve MacWilliams özdeşlikleri (MacWilliams identities) sonlu Frobenius halkalarda sağlanır [36].

3.3 R Halkası Üzerindeki Devirli Kodlar

Homojen ağırlık ilk olarak Contantinescu ve Heise tarafından 1997 yılında bulundu [37]. Daha sonraki yıllarda Greferath ve Schimidt sonlu zincir halkaları üzerinde homojen ağırlığa göre Gray dönüşümü tanımladılar [38]. Son zamanlarda yapılan çalışmalardan [34] ve [35]'da ise sonlu zincir halkası olmayan halkalar üzerinde homojen ağırlığa göre Gray dönüşüm tanımlandı. Bu bölümde, bir sonlu zincir halkası olmayan halka üzerinde devirli kodlar incelenecektir. Ayrıca bu kodların homojen ağırlığa göre Gray dönüşümleri üzerinden optimal kodlar araştırılacaktır.

Dougherty vd.'nin [33] nolu çalışmasında 16 elemanlı yerel Frobenius halkalar üzerinde tanımlı uzunluğu tek sayı olan devirli kodların yapısı karakterize edilmiştir. Bu çalışmaya göre R halkası üzerinde tanımlı devirli kodların üreteçleri Teorem 3.2'de verilmiştir. Bu bölümde g polinomu, R_n polinom halkası üzerindeki bir polinom olmak üzere $I_{u,v}g = \langle u, v \rangle g = \{ag | a \in I_{u,v} = \langle u, v \rangle\}$ elemanları kısaca Zg ile gösterilecektir.

Teorem 3.2. [33] n bir pozitif tek tamsayı ve C kodu, R halkasının üzerinde tanımlı n uzunluklu devirli bir kod olsun. $R[x]$ halkasının üzerinde $x^n - 1$ polinomu, $i \in \{0, \dots, 6\}$ olmak üzere F_i monik polinomlarının çarpımı olarak $x^n - 1 = F_0 F_1 \dots F_6$ şeklinde yazılsın. C kodu, $\widehat{F}_i = (x^n - 1)/F_i$ polinomları tarafından üretilen $\langle \widehat{F}_1, Z\widehat{F}_2, u\widehat{F}_3, v\widehat{F}_4, (u+v)\widehat{F}_5, v^2\widehat{F}_6 \rangle$ idealidir.

R halkası üzerinde tanımlı n uzunluklu devirli kodların üreteçleri, birbirini bölen polinomlar ile de ifade edilebilirler. Bu bölümde, g_i polinomunun g_j polinomunu kalansız böldüğü $g_i | g_j$ ile ifade edilecektir.

Teorem 3.3. n bir pozitif tek tamsayı ve C kodu, R halkasının üzerinde tanımlı uzunluğu n olan devirli bir kod olsun. g_i polinomu, $R[x]$ halkasında bir polinom ve $i \in \{1, \dots, 6\}$ olsun. Polinomlar arasında $j \in \{3, 4, 5\}$ için $g_6 | g_j | g_2 | g_1 | (x^n - 1)$ ilişkisi olmak üzere C kodu, üreteçleri g_i polinomları olan $\langle g_1, Zg_2, ug_3, vg_4, (u+v)g_5, v^2g_6 \rangle$ idealidir.

İspat. C kodu, $x^n - 1 = F_0 F_1 \dots F_6$ için Teorem 3.2'deki gibi $C = \langle \widehat{F}_1, Z\widehat{F}_2, u\widehat{F}_3, v\widehat{F}_4, (u+v)\widehat{F}_5, v^2\widehat{F}_6 \rangle$ olarak tanımlansın. n tek tamsayı olduğu için $i \neq j$ olması durumunda F_i ve F_j polinomları aralarında asaldır. Polinom halkası $R[x]$ 'de bulunan g_i polinomları $i \in \{1, \dots, 6\}$ ve $j \in \{3, 4, 5\}$ için $g_6 | g_j | g_2 | g_1$ koşulunu sağlayacak şekilde aşağıdaki

gibi tanımlansın.

$$g_1 = F_0 F_2 F_3 F_4 F_5 F_6, \quad (3.10)$$

$$g_2 = F_0 F_3 F_4 F_5 F_6, \quad (3.11)$$

$$g_3 = F_0 F_4 F_5 F_6, \quad (3.12)$$

$$g_4 = F_0 F_3 F_5 F_6, \quad (3.13)$$

$$g_5 = F_0 F_3 F_4 F_6, \quad (3.14)$$

$$g_6 = F_0. \quad (3.15)$$

Bir $A = \langle g_1, Zg_2, ug_3, (u+v)g_4, vg_5, v^2g_6 \rangle$ idealini ele alalım. Teorem 3.2'den $\widehat{F}_1 = g_1$ 'dir ve $Z\widehat{F}_2 = ZF_0F_1F_3F_4F_5F_6 = Zg_2F_1$ olup $Z\widehat{F}_2 \in A$ 'dır. Benzer şekilde $w_j\widehat{F}_j = w_jg_jF_1F_2 \in A$, $j \in \{3, 4, 5\}$ ve $w_3 = u, w_4 = v, w_5 = u+v$, ve $v^2\widehat{F}_6 = v^2g_6F_1F_2F_3F_4F_5 \in A$ olur. Dolayısıyla $C \subseteq A$ 'dır. Ters kapsama için $g_1 \in C$ olsun.

- i. F_1 ve F_2 aralarında asal polinomlar oldukları için $aF_1 + bF_2 = 1$ olacak şekilde $a, b \in R[x]$ vardır. Bu yüzden,

$$g_2 = (aF_1 + bF_2)F_0F_3F_4F_5F_6 = a\widehat{F}_2 + bg_1 \quad (3.16)$$

ve $Zg_2 = aZ\widehat{F}_2 + bZg_1 \in C$ olur.

- ii. F_1F_2 ve F_j aralarında asal polinomlar oldukları için $aF_1F_2 + bF_j = 1$ olacak şekilde $a, b \in R[x]$ vardır. Dolayısıyla,

$$g_j = (aF_1F_2 + bF_j)g_j = a\widehat{F}_j + bg_2 \quad (3.17)$$

olduğundan $w_jg_j = aw_j\widehat{F}_j + bw_jg_2 \in C$ 'dir.

- iii. Durum ii'ye göre $v^2g_3, v^2g_4, v^2g_5 \in C$ 'dir. Dolayısıyla $v^2ebob(g_3, g_4, g_5) = v^2F_0F_6 \in C$ olur. $F_1F_2F_3F_4F_5$ ve F_6 aralarında asal polinomlar olmalarından dolayı $aF_1F_2F_3F_4F_5 + bF_6 = 1$ olacak şekilde $a, b \in R[x]$ vardır. O halde,

$$g_6 = (aF_1F_2F_3F_4F_5 + bF_6)F_0 = a\widehat{F}_6 + bF_0F_6 \quad (3.18)$$

ve $v^2g_6 = av^2\widehat{F}_6 + bv^2F_0F_6 \in C$ 'dir.

Sonuç olarak, $A = C$ elde edilir. ■

Bu tezde, R halkası üzerinde ters-sıralı DNA kodları elde etmek için tek ya da çift uzunluklu devirli kodların aşağıda tanımlanan üç tipi üzerinde durulacaktır.

Tanım 3.3. n pozitif bir tamsayı, $i \in \{1, 2, 3\}$, $\gamma_1 = u$, $\gamma_2 = u + v$ ve $\gamma_3 = v$ olsun. C_i kodu, $g_4|g_3|g_2|g_1|(x^n - 1)$ şartını sağlayan $j \in \{1, 2, 3, 4\}$ olmak üzere g_j polinomları için $\langle g_1, Zg_2, \gamma_i g_3, v^2 g_4 \rangle$ şeklinde tanımlanır.

Tanım 3.3'te tanımlanan devirli kodun üreteç matrisi ve sahip olduğu kodsöz sayısı aşağıdaki teoremden yer almaktadır.

Teorem 3.4. n pozitif bir tamsayı $i \in \{1, 2, 3\}$, $\gamma_1 = u$, $\gamma_2 = u + v$ ve $\gamma_3 = v$ olsun. $C_i = \langle g_1, Zg_2, \gamma_i g_3, v^2 g_4 \rangle$ kodu $g_4|g_3|g_2|g_1|(x^n - 1)$ olmak üzere n uzunluklu devirli bir kod olsun. $j \in \{2, 3, 4\}$ için $k_1 = n - \text{der}(g_1)$, $k_j = \text{der}(g_{j-1}) - \text{der}(g_j)$ olsun. Buna göre aşağıdaki kümeler elde edilir:

$$\begin{aligned} S_1 &= \{x^t g_1 : 0 \leq t \leq k_1 - 1\}, \\ S_2 &= \{x^t z g_2 : 0 \leq t \leq k_2 - 1, z \in Z\}, \\ S_3 &= \{x^t \gamma_i g_3 : 0 \leq t \leq k_3 - 1\}, \\ S_4 &= \{x^t v^2 g_4 : 0 \leq t \leq k_4 - 1\}. \end{aligned} \quad (3.19)$$

$S = S_1 \cup S_2 \cup S_3 \cup S_4$ kümesi bir R -modül olarak C_i kodu için en küçük üreteç kümesidir. Ayrıca, C_i kodunun eleman sayısı $16^{k_1} 8^{k_2} 4^{k_3} 2^{k_4}$ olur.

İspat. C_i kodunun bir kodsözü c olsun. $a_i \in R[x]$ ve $z \in Z$ için $c = a_1 g_1 + a_2 z g_2 + a_3 \gamma_i g_3 + a_4 v^2 g_4$ olsun. Eğer $\text{der}(a_1) < n - \text{der}(g_1)$ ise $a_1 g_1 \in \text{Span}(S_1)$ olur. Aksi takdirde $\text{der}(r_1) < n - \text{der}(g_1)$ olmak üzere $a_1 = \frac{x^n - 1}{g_1} q_1 + r_1$ olacak şekilde $q_1, r_1 \in R[x]$ vardır. Dolayısıyla, $g_1 a_1 = r_1 g_1 \in \text{Span}(S_1)$ olur.

Eğer $\text{der}(a_2) < k_2 = \text{der}(g_1) - \text{der}(g_2)$ ise $a_2 z g_2 \in \text{Span}(S_2)$ 'dir. Aksi halde, $\text{der}(r_2) < k_2$ için $a_2 = \frac{g_1}{g_2} q_2 + r_2$ olacak şekilde $q_2, r_2 \in R[x]$ vardır. $r_2 z g_2 \in \text{Span}(S_2)$ olduğu için $a_2 z g_2 = q_2 z g_1 + r_2 z g_2 \in \text{Span}(S)$ 'dir. Bu işlemleri sürdürürsek $c \in \text{Span}(S)$ elde edilir.

S kümesinin elemanları diğer elemanlarının lineer kombinasyonundan oluşmadığından, C_i için en küçük üreteç kümesidir. Ayrıca $|\text{Span}(S_1)| = 16^{k_1}$, $|\text{Span}(S_2)| = 8^{k_2}$, $|\text{Span}(S_3)| = 4^{k_3}$, $|\text{Span}(S_4)| = 2^{k_4}$ ve sonuç olarak, $|C_i| = 16^{k_1} 8^{k_2} 4^{k_3} 2^{k_4}$ elde edilir. ■

Teorem 3.5. n pozitif bir tamsayı, $i \in \{1, 2, 3\}$, $\gamma_1 = u$, $\gamma_2 = u + v$ ve $\gamma_3 = v$ olsun. $g_4|g_3|g_2|g_1|(x^n - 1)$ olmak üzere $C_i = \langle g_1, Zg_2, \gamma_i g_3, v^2 g_4 \rangle$, n uzunluklu devirli bir kod olsun. $\gamma_1^\perp = v$, $\gamma_2^\perp = u + v$ and $\gamma_3^\perp = u$ olmak üzere C_i kodunun dual kodu aşağıdaki gibidir.

$$C_i^\perp = \left\langle \left(\frac{x^n - 1}{g_4} \right)^*, Z \left(\frac{x^n - 1}{g_3} \right)^*, \gamma_i^\perp \left(\frac{x^n - 1}{g_2} \right)^*, v^2 \left(\frac{x^n - 1}{g_1} \right)^* \right\rangle. \quad (3.20)$$

İspat. $\alpha_1 = 1$, $\alpha_2 = Z$, $\alpha_3 = \gamma_i$ ve $\alpha_4 = v^2$ olmak üzere $C_i = \langle \alpha_1 g_1, \alpha_2 g_2, \alpha_3 g_3, \alpha_4 g_4 \rangle$ olsun. $\beta_1 = 1, \beta_2 = Z, \beta_3 = \gamma_i^\perp, \beta_4 = v^2$ için $A = \langle \beta_1 \left(\frac{x^n-1}{g_4}\right)^*, \beta_2 \left(\frac{x^n-1}{g_3}\right)^*, \beta_3 \left(\frac{x^n-1}{g_2}\right)^*, \beta_4 \left(\frac{x^n-1}{g_1}\right)^* \rangle$ idealini alalım. $t_4|t_3|t_2|t_1|(x^n-1)$ için $t_i = \left(\frac{x^n-1}{g_i}\right)^*$ olsun. $j \in \{1, 2, 3, 4\}$ için $\alpha_i g_i \left(\beta_j \left(\frac{x^n-1}{g_{5-j}}\right)^*\right)^* = 0$ vardır. Bu yüzden $A \subseteq C_i^\perp$ dir. Diğer taraftan $m \in \{2, 3, 4\}$ için $k_m = \text{der}(g_{m-1}) - \text{der}(g_m)$ olup $|A| = 16^{\text{der}(g_4)} 8^{k_4} 4^{k_3} 2^{k_2}$ ve $|C_i| = 16^{k_1} 8^{k_2} 4^{k_3} 2^{k_4}$ vardır. $n = k_1 + k_2 + k_3 + k_4 + \text{der}(g_4)$ olduğundan $|A||C_i| = 16^n$ sonucuna varılır. O halde, $|A| = |C_i^\perp|$ dir. ■

Sonuç 3.5.1. n pozitif bir tamsayı olsun. $C_2 = \langle g_1, Zg_2, \gamma_2 g_3, v^2 g_4 \rangle$ kodu n uzunluklu ve $\gamma_2 = u + v$ ve $g_4|g_3|g_2|g_1|(x^n-1)$ olsun. g_i polinomları için $g_4|g_3|g_2|g_1|(x^n-1)$ şartı sağlansın. $g_1 = \left(\frac{x^n-1}{g_4}\right)^*$ ve $g_3 = \left(\frac{x^n-1}{g_2}\right)^*$ olması durumunda, C_2 kodu R halkası üzerinde kendine diktir.

Örnek 3.1. $C_2 = \langle g_1, Zg_2, \gamma_2 g_3, v^2 g_4 \rangle$ R halkası üzerinde 12 uzunluklu devirli bir kod olsun. $\{i = 1, 2, 3, 4\}$ için g_i polinomları aşağıdaki gibi tanımlansın:

$$\begin{aligned} g_1(x) &= (x^2 + x + 1)^4(x + 1)^3, \\ g_2(x) &= (x^2 + x + 1)^3(x + 1)^3, \\ g_3(x) &= (x^2 + x + 1)(x + 1), \\ g_4(x) &= (x + 1). \end{aligned} \tag{3.21}$$

Teorem 3.4'e göre C_2 kodunun en küçük üreteç kümesi aşağıdaki kümelerin birleşimi ile elde edilir. $z \in Z$ ve $i \in \{1, 2, 3, 4\}$ olmak üzere S_i kümeleri

$$\begin{aligned} S_1 &= \{g_1\}, \\ S_2 &= \{x^t z g_2 : 0 \leq t \leq 1\}, \\ S_3 &= \{x^t \gamma_2 g_3 : 0 \leq t \leq 5\}, \\ S_4 &= \{x^t v^2 g_4 : 0 \leq t \leq 1\}. \end{aligned} \tag{3.22}$$

olarak elde edilir. S kümesinin elemanları olan vektörler, C devirli kodunun üreteç matrisinin satırlarını oluşturacağından kodun üreteç matrisi aşağıdaki gibidir.

$$G = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ z & 0 & 0 & z & 0 & 0 & z & 0 & 0 & z & 0 & 0 \\ 0 & z & 0 & 0 & z & 0 & 0 & z & 0 & 0 & z & 0 \\ u+v & 0 & 0 & u+v & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & u+v & 0 & 0 & u+v & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & u+v & 0 & 0 & u+v & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & u+v & 0 & 0 & u+v & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & u+v & 0 & 0 & u+v & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & u+v & 0 & 0 & u+v & 0 & 0 & 0 \\ v^2 & v^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & v^2 & v^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

G üreteç matrisinin transpozu G^T olmak üzere $GG^T = 0$ olduğundan C_2 kodu kendine diktir. Ayrıca Sonuç 3.5.1'de belirtildiği gibi $g_1 = \left(\frac{x^{12}-1}{g_4}\right)^*$ ve $g_3 = \left(\frac{x^{12}-1}{g_2}\right)^*$ şartları sağlandığından kendine diktir.

3.4 Homojen Ağırlık ve R Halkası Üzerindeki Devirli Kodların İkili Görüntüleri

Steven vd. 16 elemanlı zincir olmayan Frobenius halkalar üzerindeki devirli kodları Lee ağırlığa göre incelediler [33]. Bu bölümde ise R halkası üzerindeki devirli ve parçalı devirli kodların homojen ağırlığa göre Gray dönüşümleri sonucu oluşan kodlar incelenecektir.

Bir kodun sıfır ağırlıklı kodsözü dışındaki tüm kodsözlerinin ağırlıkları aynı ise bu koda *bir-ağırlıklı (one-weight) kod* denir. Ayrıca bir kodun sıfır ağırlıklı kodsözü dışındaki kodsözlerinin ağırlıkları w_1 veya w_2 ise bu koda *iki-ağırlıklı (two-weight) kod* denir. Bir-ağırlıklı kodların GSM ağı [39] ve nano elektronik [40] alanlarında bir çok önemli uygulamaları vardır. İki-ağırlıklı kodlar ile güçlü düzenli graflar (strongly regular graphs) yakın ilişkilidir. Bu çalışmadaki inşa metodları ile bir-ağırlıklı kodlar ve iki-ağırlıklı kodlar için örnekler elde edilecektir.

Tanım 3.4. C kodu, R halkası üzerinde lineer kod ve

$$c = (c_{1,1}, \dots, c_{1,n-1}, c_{1,n}, c_{2,1}, \dots, c_{2,n-1}, c_{2,n}, \dots, c_{\ell,1}, \dots, c_{\ell,n-1}, c_{\ell,n})$$

bir kodsözü olsun. c kodsözünün parçalı devirsel ötelemesi (quasi-cyclic shift)

$$\sigma(c) = (c_{1,n}, c_{1,1}, \dots, c_{1,n-1}, c_{2,n}, c_{2,1}, \dots, c_{2,n-1}, \dots, c_{\ell,n}, c_{\ell,1}, \dots, c_{\ell,n-1}) \quad (3.23)$$

C 'nin bir kodsöz oluyorsa C koduna indeksi ℓ uzunluğu $n\ell$ olan *parçalı devirli (quasi-*

cyclic) kod denir.

Yıldız vd. tarafından, elemanları değişme özelliğini ve $u^2 = v^2 = 0$ şartını sağlayan $\mathbb{F}_2 + u\mathbb{F}_2 + v\mathbb{F}_2 + uv\mathbb{F}_2$ zincir olmayan halka üzerinde homojen ağırlık üzerine çalışma yapıldı [34]. Bu çalışmada, birinci Reed-Muller kodu $RM(1,3)$ kullanılarak lineer mesafe koruyan bir dönüşüm inşa edildi.

Tanım 3.5. [41] Sonlu elemanlı S halkasında reel değerli w fonksiyonuna aşağıdaki koşulları sağlaması durumunda (*sol*) *homojen ağırlık* denir.

- (i) $w(0) = 0$.
- (ii) Her $x, y \in S$ için $Sx = Sy$, $w(x) = w(y)$ eşitliğini ifade eder.
- (iii) Aşağıdaki koşulu sağlayan bir γ reel sayısı vardır:

$$\sum_{y \in Sx} w(y) = \gamma |Sx|. \quad (3.24)$$

Burada γ sayısı, S halkasındaki w ağırlığının ortalama değeridir ve (iii) koşulunda bahsedildiği gibi S halkasındaki sıfırdan farklı tüm temel idealler için sabittir.

R halkası üzerindeki her x elemanı için homojen ağırlık [34] nolu çalışmadaki gibi tanımlanabilir:

$$w_{\text{hom}}(x) = \begin{cases} 0 & , x = 0, \\ 8 & , x = v^2, \\ 4 & , \text{diğer.} \end{cases} \quad (3.25)$$

Homojen ağırlığın tanımlı olduğu R halkasından (R, w_{hom}) Hamming ağırlığının tanımlı olduğu $(\mathbb{F}_2^8, w_H)$ vektör uzayına mesafe koruyan *Gray dönüşümü* inşa edilir. Bu dönüşüm aşağıdaki gibidir:

$$\varphi(v^2) = (1, 1, 1, 1, 1, 1, 1, 1), \quad \varphi(u) = (1, 1, 1, 1, 0, 0, 0, 0), \quad (3.26)$$

$$\varphi(v) = (1, 1, 0, 0, 1, 1, 0, 0), \quad \varphi(1) = (1, 0, 1, 0, 1, 0, 1, 0). \quad (3.27)$$

Halkada bulunan her eleman için Gray görüntüsü aşağıdaki gibi elde edilir:

$$\varphi(a + bu + cv + dv^2) = a\varphi(1) + b\varphi(u) + c\varphi(v) + d\varphi(v^2). \quad (3.28)$$

R halkasındaki elemanların tümünün Gray görüntüsü alınarak elde edilen ikili lineer kodun parametreleri $[8, 4, 4]$ şeklindedir. Bu yol ile elde edilen $C = \varphi(R)$ ikili lineer

kodunun üreteç matrisi aşağıdaki gibidir:

$$G = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}. \quad (3.29)$$

Gray dönüşüm ile elde edilen C ikili lineer kodun, birinci mertebe Reed-Muller kod olduğu görülür. Ayrıca, $GG^T = 0$ şartı sağlandığından C kodu kendine dik koddur.

Gray dönüşüm $c = (c_0, c_1, \dots, c_{n-1})$ kodsözü için $\phi(c) = (\varphi(c_0), \varphi(c_1), \dots, \varphi(c_{n-1}))$ olarak n -koordinata genişletilebilir. Ayrıca, R halkası üzerindeki C lineer kodu n uzunluklu ise $\phi(C)$ ikili lineer kodu $8n$ uzunlukludur.

Teorem 3.6. C kodu R halkası üzerindeki bir lineer kod olsun. C kodunun Gray görüntüsü bir ikili kendine dik koddur.

İspat. $\varphi(R)$ kendine dik kod olduğundan her $a, b \in R$ için $[\varphi(a), \varphi(b)] = 0$ 'dır. C uzunluğu n olan bir lineer kod ve $c, d \in C$ olsun. Bu durumda $\phi(c) = \phi(c_0, c_1, \dots, c_{n-1}) = (\varphi(c_0), \varphi(c_1), \dots, \varphi(c_{n-1}))$ ve $\phi(d) = \phi(d_0, d_1, \dots, d_{n-1}) = (\varphi(d_0), \varphi(d_1), \dots, \varphi(d_{n-1}))$ olur ve $[\phi(c), \phi(d)] = \sum_{i=0}^{n-1} \varphi(c_i)\varphi(d_i) = 0$ elde edilir. Dolayısıyla $\phi(C)$ bir kendine dik koddur. ■

Teorem 3.7. C kodu R halkası üzerindeki n uzunluklu tüm vektörleri içeren bir lineer kod olsun. $\phi(C)$ kodu, $[8n, 4n, 4]$ parametrelerine sahip bir ikili kendine dual koddur.

İspat. $n = 1$ için $C = R$ 'dir. Bu durumda $\phi(C)$, $[8, 4, 4]$ parametrelerine sahiptir. Ayrıca, optimal kendine dual kodlar vertabanına göre $\phi(C)$ bir optimal Tip II (tüm ağırlıkları 4'ün katı olan) ikili kendine dual koddur [42].

$n = 2$ için $C = \langle (1, 0), (0, 1) \rangle$ 'dir. Bu durumda $\phi(C)$ kodu $[16, 8, 4]$ parametrelili bir optimal Tip II kendine dual koddur.

n uzunluğunun daha büyük değerleri için ise $C = R^n$ kodu $16^n = 2^{4n}$ elemanlıdır. $c = (0, 0, \dots, e_j, \dots, 0)$ bir kodsöz ve bu kodsözün e_j -nci bileşeni sıfırdan ve v^{2^j} 'den farklı olsun. Bu kodsözün ağırlığı $w_{\text{hom}}(\phi(c)) = w_H(\varphi(e_j)) = 4$ 'tür. C kodunda ağırlığı sıfırdan farklı olan daha küçük ağırlıklı kodsöz bulunamayacağından $\phi(C)$ 'nin minimum uzaklığı 4 olur. Bu yüzden $\phi(C)$ kodu $[8n, 4n, 4]$ parametrelerine sahiptir ve Teorem 3.6'dan ikili kendine dual koddur. ■

R halkası üzerindeki devirli kodların Gray görüntülerinden elde edilen optimal kodlar

için bazı örnekler aşağıdaki gibidir. Örnekler için Magma [43] cebirsel hesaplama sistemi kullanılmış olup elde edilen sonuçlar, [22] ve [44] nolu kaynaklardaki kod tabloları ile karşılaştırılmıştır.

Örnek 3.2. $g(x)$ polinomu, R halkası üzerinde derecesi $n - k$ olan ve $x^n - 1$ polinomunu bölen bir polinom olsun. $C = \langle v^2 g(x) \rangle$ kodu R halkası üzerinde devirli koddur ve $\phi(C)$ kodu $[8n, k, d]$ parametrelerine sahip bir ikili lineer koddur. Devirli kodların Gray görüntülerinden elde edilen kodlardan optimal kod olan kodların listesi Tablo 3.1’de yer almaktadır. Ayrıca R halkası üzerinde $(v^2 g(x), \dots, v^2 g(x))$ şeklinde üretilen parçalı devirli kodlardan elde edilen optimal kodlar da yine Tablo 3.1’den görülebilir.

Tablo 3.1’de, R halkası üzerindeki $\ell = 2$ ve $n\ell = 6$ uzunluklu parçalı devirli kodun üretici $(110|110)$ vektörü ile gösterilmektedir ve $(v^2 + v^2 x | v^2 + v^2 x)$ polinomuna karşılık gelir. Tablo 3.1’deki kodlar incelendiğinde tüm kodların bir-ağırlıklı kod olduğu görülür. Ayrıca [22] nolu kaynaktaki kod tablosunda, $[224, 3, 128]$ parametrelili optimal kod 68 adımdan sonra elde edilirken bu çalışmada bahsedilen inşa metodu ile parametreler 1 adımda elde edilir. Elde edilen kodların [22] nolu kaynakta yer alan adım sayıları Tablo 3.1’de görülmektedir.

Tablo 3.1 R halkası üzerindeki devirli kodların Gray görüntülerinden elde edilen optimal ikili lineer kodlar ve [22]’ye göre adım sayıları

n	$g(x)$	$\phi(C)$	Adım
3	(110)	[24, 2, 16]	1
3	(110 110)	[48, 2, 32]	1
7	(1011100)	[56, 3, 32]	20
7	(1011100 1011100)	[112, 3, 64]	36
7	(1011100 1011100 1011100)	[168, 3, 96]	52
7	(1011100 1011100 1011100 1011100)	[224, 3, 128]	68
9	(110110110)	[72, 2, 48]	1
9	(110110110 110110110)	[144, 2, 96]	1
9	(110110110 110110110 110110110)	[216, 2, 144]	1
15	(110110110110110)	[120, 2, 80]	1
15	(110110110110110 110110110110110)	[240, 2, 160]	1
15	(100110101111000)	[120, 4, 64]	22
15	(100110101111000 100110101111000)	[240, 4, 128]	38
21	(111010011101001110100)	[168, 3, 96]	52
21	(101101101101101101101)	[168, 2, 112]	1
27	(101101101101101101101101101)	[216, 2, 144]	1

Yukarıdaki inşa metodu dışında döngüsel (circulant) matrisler kullanılarak, yani temsilci vektörü ve onun devirsel ötelemelerini satır kabul eden matristen elde edilen optimal kodlar Tablo 3.2’de yer almaktadır.

Bu inşa metoduna örnek olarak bir temsilci vektör $\mathbf{v} = (u, v)$ olsun. C kodu bu vektörün döndüsel matrisi tarafından üretilsin. $\phi(C)$ kodunun 16 kodsözü vardır ve bu kodsözlerden 14 tanesi 8 ağırlıklı, bir tanesi 16 ağırlıklı ve bir tanesi de 0 ağırlıklıdır. Tablo 3.2’de ağırlıklar kısaca ağırlık sayacı polinomu ile $x^{16} + 14x^8 + 1$ örneğinde olduğu gibi ifade edilmektedir.

Tablo 3.2’de $[24, 8, 8]$ ve $[24, 9, 8]$ parametrelerine sahip kodlar dışındaki tüm kodlar iki-ağırlıklı koddur. Ayrıca, $[16, 5, 8]$ ve $[32, 6, 16]$ parametrelerine sahip kodlar [44] nolu kaynakta yer alan optimal iki-ağırlıklı kod tablosuna göre optimal iki-ağırlıklı kodlardır.

Tablo 3.2 Döngüsel matrislerden elde edilen optimal ikili lineer kodlar

n	C kodunun üretici	$\phi(C)$	$\phi(C)$ kodunun ağırlık sayacı polinomu
2	(u, v)	$[16, 4, 8]$	$x^{16} + 14x^8 + 1$
2	$(1, 1 + v)$	$[16, 5, 8]$	$x^{16} + 30x^8 + 1$
3	$(u + 1, u + 1, u + 1)$	$[24, 4, 12]$	$x^{24} + 14x^{12} + 1$
3	$(v, u + 1, u + v + 1)$	$[24, 8, 8]$	$45x^{16} + 168x^{12} + 42x^8 + 1$
3	$(v^2, 1, 1)$	$[24, 9, 8]$	$x^{24} + 87x^{16} + 336x^{12} + 87x^8 + 1$
4	(v, u, v, u)	$[32, 4, 16]$	$x^{32} + 14x^{16} + 1$
4	$(v + 1, 1, v + 1, 1)$	$[32, 5, 16]$	$x^{32} + 30x^{16} + 1$
4	$(1, v + 1, u + v + 1, u + 1)$	$[32, 6, 16]$	$x^{32} + 62x^{16} + 1$
6	$(1, 1, 1, 1, 1, 1)$	$[48, 4, 24]$	$x^{48} + 14x^{24} + 1$

Örnek 3.3. C kodu, $\mathbf{v} = (1, 1 + v)$ vektörü tarafından üretilen devirli kodu olsun. $\phi(C)$ kodu $[16, 5, 8]$ parametrelerine sahip optimal ikili koddur. $(\mathbf{v}|\mathbf{v})$ vektörü ile elde edilen indeksi 2 olan parçalı devirli kod ise $[32, 5, 16]$ parametrelili bir optimal koddur. Bu kod aynı zamanda iki-ağırlıklı koddur. İndeksi 8 olana kadar bu işlemi devam ettirirsek $(\mathbf{v}|\dots|\mathbf{v})$ vektörü tarafından elde edilen kod ise yine $[16\ell, 5, 8\ell]$ parametrelili bir optimal koddur. Bu kodların parametreleri sırasıyla $[48, 5, 24]$, $[64, 5, 32]$, $[80, 5, 40]$, $[96, 5, 48]$, $[112, 5, 56]$ ve $[128, 5, 64]$ şeklindedir. Bu kodların tümü iki-ağırlıklı koddur.

Örnek 3.4. C kodu $\mathbf{v} = (1, v + 1, u + v + 1, u + 1)$ vektörü ile üretilen bir devirli kod olsun. $\phi(C)$ kodu $[32, 6, 16]$ parametrelili bir optimal ikili lineer koddur. Önceki örnekte olduğu gibi $2 \leq \ell \leq 8$ için indeks ℓ ile elde edilen $(\mathbf{v}|\dots|\mathbf{v})$ vektörü ile üretilen parçalı devirli kodlar elde edilir. Bu kodların Gray dönüşümleri ile elde edilen kodların parametreleri sırasıyla $[64, 6, 32]$, $[96, 6, 48]$, $[128, 6, 64]$, $[160, 6, 80]$, $[192, 6, 96]$, $[224, 6, 112]$ ve $[256, 6, 128]$ şeklindedir. Bu ikili lineer kodlar optimaldir ve tümü iki-ağırlıklı koddur.

$\mathbb{F}_2 + u\mathbb{F}_2 + v\mathbb{F}_2 + v^2\mathbb{F}_2$ HALKASI ÜZERİNDEKİ DEVİRLİ KODLARIN DNA UYGULAMALARI

Bir iletişim sisteminin bant genişliği arttıkça, tek seferde daha fazla bilgi aktarımı sağlanır. Bu nedenle, geniş bir bant genişliğinde daha fazla veri taşımak için DNA'nın tekli bazları yerine DNA k -bazları kullanılabilir. Bu yüzden halkalar ve cisimler gibi bir cebirsel yapının elemanları ile DNA k -bazları arasında bir eşleme yapılması önemli bir problemidir.

Cebirsel yapılar için DNA k -bazlarının ($k \geq 2$) kullanılması, DNA kodları üretmek için aşılması gereken bir sorun olan ters sıralılık problemini ortaya çıkartır. Dört elemanlı bir cebirsel yapı ile DNA nükleotidleri, yani $\{A, T, G, C\}$, arasında bir eşleme yaparsak, bu cebirsel yapı üzerindeki ters sıralı herhangi bir kod, bu eşleme ile bir ters-sıralı DNA kodu verir. Ancak 4^i ($i > 1$) elemanlı cebirsel yapılarda ters sıralılık problemi olarak isimlendirilen sorun meydana gelir ve DNA kod elde etmek için öncelikle bu problemin çözülmesi gerekir.

Bu bölümde ters sıralılık problemi çözülerek R halkasından ters-sıralı DNA kodlar ve ters-sıralı tamlayan DNA kodlar elde edilecektir.

4.1 DNA Kodun Ters Sıralılık Problemi

Eleman sayısı 4^k olan halka ya da cisimler üzerinde ters-sıralı DNA kodlarla ilgili yapılan çalışmalarda $k \geq 2$ için DNA k -bazları ile bu yapıların 4^k elemanın birebir eşlenmesi sırasında oldukça zor ve ilginç bir problemle karşılaşılır. $k = 1$ için, Abulraub ve Ghrayeb tarafından yapılan çalışmada $F_4 = \{0, 1, \alpha, \alpha^2\}$ elemanları sırası ile $\{A, T, G, C\}$ bazları ile eşleştirilmiştir [13]. Bu çalışmada kodsözlerin ters-sıralısı kodsözleri sondan başa doğru yazarak elde edilir. Ancak k çift sayı olduğunda [14, 16, 17] nolu çalışmalarda karşılaşılan ters sıralılık problemi aşağıda yer almaktadır.

16 elemanlı F_{16} cismi kullanılarak DNA kod elde etmek için cismin elemanları ile

DNA 2-bazları arasında birebir eşleme olması gerekir. Bu eşlemeye göre cismin $\alpha_1, \alpha_2, \alpha_3$ elemanlarına karşılık gelen DNA 2-bazları sırasıyla AT, GA ve TC olsun. Bu durumda $(\alpha_1, \alpha_2, \alpha_3)$ kodsözünün DNA karşılığı $ATGATC$ olur. $(\alpha_1, \alpha_2, \alpha_3)$ kodsözünün ters-sıralısı $(\alpha_3, \alpha_2, \alpha_1)$ olup DNA karşılığı $TCGAAT$ olarak elde edilir. Ancak $ATGATC$ nin ters-sıralısı $TCGAAT$ değildir.

Ters-sıralı DNA kod için kodsözlerin ters-sıralılarının da kodsöz olması gerektiği bilinmektedir. Bu durumda ters-sıralı DNA kod elde etmek için öncelikle ters sıralılık probleminin çözülmesi gerekir.

4.2 R Halkası Üzerindeki DNA kodlar

16 elemanlı bir halka üzerinde yapılan DNA kod çalışmasında özel bir matris kullanılarak ters sıralılık problemi çözülmüştür [17]. Bu bölümde Tablo 4.1'deki gibi halkanın elemanları ile DNA ikilileri arasında ilişki kurularak ters sıralılık problemi çözülecektir. R halkası üzerindeki devirli kodlar kullanılarak ters-sıralı ve ters-sıralı tamlayan DNA kodlar üretilecektir.

R halkasının elemanları ile DNA ikilileri arasındaki ilişki τ dönüşümü ile aşağıdaki gibi ifade edilir.

$$\tau : R \longrightarrow \{A, T, G, C\}^2 \quad (4.1)$$

$$a \longmapsto \tau(a). \quad (4.2)$$

DNA ikilileri ve R halkasının elemanları arasında kurulan eşleme Tablo 4.1'de yer almaktadır.

Tablo 4.1 R halkasının elemanlarına karşılık gelen DNA ikilileri

0	AA	v	AT	1	AG	$1 + u + v$	GA
v^2	TT	$v + v^2$	TA	$1 + v^2$	TC	$1 + u + v + v^2$	CT
$u + v$	GG	u	GC	$1 + v$	GT	$1 + u + v^2$	TG
$u + v + v^2$	CC	$u + v^2$	CG	$1 + v + v^2$	CA	$1 + u$	AC

Yukarıdaki eşleme $c = (c_0, c_1, \dots, c_{n-1}) \in R^n$ kodsözü için $\tau(c) = (\tau(c_0), \tau(c_1), \dots, \tau(c_{n-1}))$ şeklinde n -koordinata genişletilebilir.

τ eşlemesi, R halkası üzerindeki devirli kodların özellikleri kullanılarak cebirsel yollardan ters sıralılık probleminin çözülmesini sağlar.

Tablo 4.1'de, $a \in R$ için $\tau(a)$ DNA ikilisinin tamlayanının $\tau(a + v^2)$ DNA ikilisi ve $\tau(a)$ DNA ikilisinin ters sıralısının $\tau((1 + u + v)a)$ DNA ikilisi olduğu görülür.

Örnek 4.1. $(u, 1, 1)$ vektörüne karşılık gelen DNA dizisi $\tau((u, 1, 1)) = GCAGAG$ şeklindedir. $GCAGAG$ dizisinin ters sıralısı ise $\tau((1 + u + v)(u, 1, 1)^r) = \tau((1 + u + v)(1, 1, u)) = \tau((1 + u + v, 1 + u + v, u + v^2)) = GAGACG$ olur.

Lemma 4.1. c vektörü R^n vektör uzayında bir vektör olsun. c vektörünün ters sıralısı $c^r = (c_0, c_1, \dots, c_{n-1})^r = (c_{n-1}, c_{n-2}, \dots, c_0)$ ve $\mathbf{1} = (1, 1, \dots, 1)$ olmak üzere aşağıdaki eşitlikler sağlanır:

- i. $\tau(c)^c = \tau(v^2 \mathbf{1} + c)$,
- ii. $\tau(c)^r = \tau((1 + u + v)c^r)$,
- iii. $\tau(c)^{rc} = \tau(v^2 \mathbf{1} + (1 + u + v)c^r)$.

Tanım 4.1. Bir kodun kodsözleri DNA dizilerinden oluşuyorsa bu koda *DNA kod* denir.

$f(x)$ polinomunun *ters-sıralı (reciprocal) polinomu* $x^{\deg(f(x))} f(x^{-1})$ polinomudur ve $f^*(x)$ ile gösterilir.

Örnek 4.2. $f(x) = x^3 + x + 1$ polinomu için $\deg(f(x)) = 3$ 'tür. $f(x)$ polinomunun ters-sıralı polinomu ise $f^*(x) = x^3 + x^2 + 1$ şeklindedir.

Bir $f(x)$ polinomu ile $f(x)$ polinomunun ters-sıralı polinomunun birbirine eşit olması durumunda $f(x)$ polinomuna *ters sıralısına eşit (self-reciprocal veya palindromic) polinom* denir.

Lemma 4.2. [45] $f(x)$ ve $g(x)$ polinomları $R[x]$ polinom halkasındaki iki polinom olsun. $\deg(f) \geq \deg(g)$ ve $\mu = \deg(f) - \deg(g)$ olsun. Aşağıdaki eşitlikler sağlanır:

- i. $[f(x)g(x)]^* = f(x)^* g(x)^*$,
- ii. $[f(x) + g(x)]^* = f(x)^* + x^\mu g(x)^*$.

Bu bölüm ile ilgili ana sonuçlar ve uygulama örnekleri aşağıda yer almaktadır.

Teorem 4.3. n pozitif bir tamsayı olsun. $C_i = \langle g_1, Zg_2, \gamma_i g_3, v^2 g_4 \rangle$, $i \in \{1, 2, 3\}$, $\gamma_1 = u$, $\gamma_2 = u + v$, $\gamma_3 = v$ olmak üzere n uzunluklu devirli kod ve g_j polinomları için $g_4 | g_3 | g_2 | g_1 | (x^n - 1)$ ilişkisi verilsin. $1 \leq j \leq 4$ için $g_j(x)$ polinomları ters sıralısına eşit polinomlar ise $\tau(C_i)$ kodu bir ters-sıralı DNA koddur. Ayrıca $v^2((x^n - 1)/(x - 1))$ polinomu C_i kodunun elemanı ise $\tau(C_i)$ kodu ters-sıralı tamlayan DNA koddur.

İspat. $c(x) \in C_i$ olsun. $1 \leq j \leq 4$ için a_j 'ler $R[x]$ 'te birer polinom olmak üzere

$$c(x) = a_1 g_1 + a_2 Z g_2 + a_3 \gamma_i g_3 + a_4 v^2 g_4 \quad (4.3)$$

olsun. Lemma 4.2'den μ_1, μ_2, μ_3 tamsayıları vardır, öyle ki;

$$\begin{aligned} c(x)^* &= [a_1 g_1 + a_2 Z g_2 + a_3 u g_3 + a_4 v^2 g_4]^* \\ &= (a_1 g_1)^* + x^{\mu_1} (a_2 Z g_2)^* + x^{\mu_2} (a_3 u g_3)^* + x^{\mu_3} (a_4 v^2 g_4)^* \\ &= (a_1)^* g_1 + x^{\mu_1} (a_2)^* Z g_2 + x^{\mu_2} (a_3)^* u g_3 + x^{\mu_3} (a_4)^* v^2 g_4. \end{aligned} \quad (4.4)$$

$c(x)^* \in C$ ve $c^r(x) = x^{n-\deg(c(x))-1} c(x)^*$ olduğundan, $c^r(x) \in C$ olur. Dolayısıyla C bir ters-sıralı koddur. $(1+u+v)c^r(x) \in C$ ve $\tau((1+u+v)c^r) = \tau(c)^r \in \tau(C_i)$ olduğu için $\tau(C_i)$ bir ters-sıralı DNA koddur.

Ayrıca, $v^2((x^n-1)/(x-1)) \in C_i$ yani $v^2(1+x+x^2+\dots+x^{n-1}) \in C_i$ olduğunu kabul edelim. $\mathbf{1} = (1, 1, \dots, 1)$ olmak üzere $v^2 \mathbf{1} \in C_i$ olur. Dolayısıyla $v^2 \mathbf{1} + (1+u+v)c^r \in C$ ve $\tau(v^2 \mathbf{1} + (1+u+v)c^r) = \tau(c)^{rc} \in \tau(C_i)$ 'dir. Bu yüzden, $\tau(C_i)$ ters-sıralı tamlayan DNA koddur. ■

Teorem 4.4. n pozitif bir tamsayı olsun. $C_i = \langle g_1, Z g_2, \gamma_i g_3, v^2 g_4 \rangle$, $i \in \{1, 2, 3\}$, $\gamma_1 = u$, $\gamma_2 = u+v$, $\gamma_3 = v$ olmak üzere n uzunluklu devirli kod ve g_j polinomları için $g_4 | g_3 | g_2 | g_1 | (x^n - 1)$ ilişkisi verilsin. $1 \leq j \leq 4$ için $g_j(x)$ polinomları ters sıralısına eşit polinomlar ise $\tau(C_i^\perp)$ kodu bir ters-sıralı DNA koddur.

İspat. Lemma 4.2 ve Teorem 3.5'ten, $C_i^\perp$ nin üreteç polinomları da ters sıralısına eşit polinomlar olduklarından $\tau(C_i^\perp)$ ters-sıralı DNA koddur. ■

Örnek 4.3. $n = 15$ olsun. $x^n - 1$ polinomunun aşağıdaki çarpanlara ayrılışı verilsin:

$$\begin{aligned} x^{15} - 1 &= f_1 f_2 f_3 f_4 f_5 \\ &= (x+1)(x^2+x+1)(x^4+x+1)(x^4+x^3+1)(x^4+x^3+x^2+x+1). \end{aligned}$$

Uzunluğu 15 olan devirli kodu ele alalım. $C = \langle g_1, u g_3, v^2 g_4 \rangle$ ve polinomlar $g_1 = f_2 f_3 f_4 f_5$, $g_3 = f_2 f_3 f_4$ ve $g_4 = f_2$ olsun. C kodunda $16^1 4^4 2^8 = 2^{20}$ kodsöz vardır. $\tau(C)$ kodu ise 30 uzunluklu ve 2^{20} kodsöze sahip bir ters-sıralı DNA koddur. Ayrıca $\tau(C)$ kodu $v^2 g_1(x) = v^2((x^n-1)/(x-1))$ polinomunu içerdiği için bir ters-sıralı tamlayan DNA koddur.

Örnek 4.4. $n = 4$ uzunluklu $C = \langle v g_3, v^2 g_4 \rangle$ kodu için $g_3 = 1+x^2$ ve $g_4 = x+1$ olsun.

C kodu uzunluğu 4 olan bir koddur ve üreteç matrisi aşağıdaki gibidir:

$$\begin{pmatrix} v & 0 & v & 0 \\ 0 & v & 0 & v \\ v^2 & v^2 & 0 & 0 \end{pmatrix}. \quad (4.5)$$

Üreteç matrisinin satırlarına karşılık gelen DNA dizilerini düşünelim. $1 \leq i \leq 3$ için r_i vektörü, G üreteç matrisinin $i - nci$ satırı olsun. Buna göre aşağıdaki eşitlikler sağlanır: g_3 ve g_4 polinomlarının her ikisi de ters sıralısına eşit polinomlar olduklarından $\tau(C)$ ters-sıralı DNA koddur.

Üreteç matrisinin satırlarına karşılık gelen DNA dizilerini düşünelim. $1 \leq i \leq 3$ için r_i vektörü, G üreteç matrisinin $i - nci$ satırı olsun. Buna göre aşağıdaki eşitlikler sağlanır:

$$\tau(r_1) = \tau((v, 0, v, 0)) = AT AAAT AA, \quad (4.6)$$

$$\tau((1 + u + v)r_2) = \tau((0, v + v^2, 0, v + v^2)) = AA TAAA TA. \quad (4.7)$$

Yukarıda görüldüğü gibi $\tau(r_1)$ DNA ikilisinin ters sıralısı $\tau((1 + u + v)r_2)$ DNA ikilisidir. Benzer şekilde $\tau(r_3)$ DNA ikilisinin ters sıralısı $\tau((1 + u + v)(vr_1 + vr_2 + r_3))$ DNA ikilisidir. Ayrıca C kodu, $v^2 \mathbf{1} = (v^2, v^2, v^2, v^2)$ kodsözünü içerdiğinden ters-sıralı tamlayan DNA koddur. $\tau(r_1)$ DNA dizisinin ters sıralı tamlayanının $\tau(v^2 \mathbf{1} + (1 + u + v)r_2) = \tau((v^2, v, v^2, v)) = TT AT TT AT$ olduğu görülür.

GERÇEK DNA DİZİLERİNDE HATA DÜZELTEN KODLAR

Matematiksel biyoloji alanında çalışan araştırmacılar gerçek DNA'larda bulunan genler ile hata düzelten kodlar arasındaki ilişkiyi araştırdılar. Faria vd. tarafından yapılan çalışmada [18], TRAV7 (T cell receptor alpha variable 7) geninin 511 uzunluklu bir kodsöz olarak, sadece bir bit fark ile, devirli kodların önemli bir alt sınıfı olan F_4 üzerindeki bir BCH kodun kodsözüne karşılık geldiği görüldü.

Bu bölümde TRAV7 (T cell receptor alpha variable 7)(insan) geni ve TRAV7-1 (fare) ile TRAV7-6 (fare) arasındaki 7 gen üzerindeki lineer kod yapısı incelenecektir.

5.1 Metod

DNA dizilerinin matematiksel gösterimini oluşturmak için mertebesi 4 olan Galois cisminin yani $F_4 = \{0, 1, \alpha, \alpha^2\}$ cisminin elemanları ile 4 DNA nükleotidi (A, T, G, C) arasında ilişki kurulması gerekir. Buna göre 24 eşleme yapmak mümkündür. Sıfırın seçimi kodun minimum ağırlığına etki edeceğinden 4 farklı seçim vardır. Bu seçimler aşağıda yer almaktadır.

$$\begin{array}{cc}
 \text{DNA no 1} & \text{DNA no 2} \\
 \left[\begin{array}{cccc} A & T & G & C \\ \alpha^2 & \alpha & 1 & 0 \end{array} \right] & \left[\begin{array}{cccc} A & T & G & C \\ \alpha^2 & \alpha & 0 & 1 \end{array} \right] \\
 \text{DNA no 3} & \text{DNA no 4} \\
 \left[\begin{array}{cccc} A & T & G & C \\ \alpha^2 & 0 & \alpha & 1 \end{array} \right] & \left[\begin{array}{cccc} A & T & G & C \\ 0 & 1 & \alpha & \alpha^2 \end{array} \right]
 \end{array} \tag{5.1}$$

$F_4 = \{0, 1, \alpha, \alpha^2\}$ cisminde toplama işlemi mod 2'de gerçekleşir. Çarpma işlemi için ise $\alpha^2 = \alpha + 1$ kuralı uygulanır.

Bu çalışmada *kayan pencere (sliding window)* [46] yöntemi kullanılmaktadır. Bu yöntemle göre TRAV7 geni uzunluğu 8'den 40'a kadar değişen parçalara bölünür. Her

parçanın bir başlangıç noktası (SP) ve bir bitiş noktası (EP) vardır. SP, orjinal gende parçanın ilk nükleotidine, EP ise son nükleotidine tekabül eder.

Herhangi bir TRAV7 geninden lineer kod elde etmek için öncelikle genden uzunluğu N olan bir ana pencere alınır. Kayan pencere metodu ile N uzunluklu parçadan uzunluğu n olan alt pencereler yani vektörler elde edilir. Bu vektörleri satır kabul eden matris $[n, k, d]$ kodu için üreteç matrisidir.

Kayan pencere metodu ile elde edilen vektörleri ve bu vektörleri satır kabul eden matristen kod elde edilmesini açıklayıcı örnek aşağıda yer almaktadır.

Örnek 5.1. TRAV7-3 geni için SP=28 ve EP=34 olan 7 uzunluklu DNA parçası $GTCTGT$ 'dir. Bu DNA parçasındaki $n = 5$ uzunluklu vektörler kayan pencere yöntemine göre belirlenirler. Belirlenen vektörler koyu renk ile aşağıda gösterilmektedir. Vektörler DNA no 1'e göre aşağıdaki gibi eşlenirler:

$$\begin{aligned} R1 : & \textbf{GTCTGT} \longrightarrow GTCT \longrightarrow (1, \alpha, 0, 0, \alpha), \\ R2 : & \textbf{GTCTGT} \longrightarrow TCTG \longrightarrow (\alpha, 0, 0, \alpha, 1), \\ R3 : & \textbf{GTCTGT} \longrightarrow CTGT \longrightarrow (0, 0, \alpha, 1, \alpha). \end{aligned} \quad (5.2)$$

Eşlemeye göre elde edilen vektörleri satır kabul eden üreteç matrisi aşağıdaki gibidir:

$$G = \begin{bmatrix} 1 & \alpha & 0 & 0 & \alpha \\ \alpha & 0 & 0 & \alpha & 1 \\ 0 & 0 & \alpha & 1 & \alpha \end{bmatrix}. \quad (5.3)$$

G matrisi tarafından üretilen C lineer kodunun parametreleri $[5, 3, 3]$ olup [22]'de bulunan F_4 üzerindeki lineer kodların en küçük uzaklıklarına göre sınırlar tablosuna göre bir optimal koddur. Ayrıca C kodu devirli MDS koddur.

TRAV7 genlerinden kayan pencere metodu ile elde edilen lineer kodların MDS, kendine dual, kendine dik, devirli ve optimal olma özelliklerine göre sınıflandırıldıkları tablolar aşağıda yer almaktadır. Hesaplamalar yapılırken Magma [43] programından faydalanılmıştır.

Tablolarda yer alan Gen ID'nin TRAV70 olması ile insan DNA'sında bulunan TRAV7 geni ifade edilmektedir. Fare DNA'sında bulunan TRAV7-1 tipindeki genler ise kısaca TRAV71 olarak gösterilmektedir.

Tablo 5.1'de kendine dual ve devirli olan lineer kodlar listelenmektedir. Sadece ana parça uzunluğunun 8 olması durumunda sonuca ulaşılmıştır. TRAV70, TRAV71, TRAV72, TRAV73 ve TRAV76 genlerinden elde edilen kendine dual ve devirli lineer

kodların tümü listedeki gibidir. TRAV74 ve TRAV75 genleri için kendine dual ve devirli olan lineer kod bulunmamaktadır. Listede yer alan $[6, 3, 3]$ lineer kodu [44]'ye göre optimal kendine dual koddur.

Tablo 5.1 Kendine dual ve devirli kodlar

Gen ID	N	SP	EP	DNA no	n	k	d
TRAV70	8	37	44	1,2,4	6	3	3
TRAV71	8	93	100	1,3,4	6	3	2
TRAV72	8	376	383	1,2,4	6	3	3
TRAV73	8	417	424	1,3,4	6	3	2
TRAV76	8	137	144	2,3,4	6	3	3

Tablo 5.2'de hem MDS hem de devirli kod olan optimal kodlar listelenmektedir. Ana parça uzunluğu 7, 8, 10 ve 12 için sonuca ulaşılmıştır. Üzerinde çalışılan tüm genlerde MDS devirli optimal kod elde edilmiştir. Farklı SP ve EP'ler için aynı DNA no'lardan aynı kodların elde edilmesi durumunda ilk karşılaşılan kod listeye eklenmiştir.

Tablo 5.3'te kendine dik ve devirli optimal kodlar listelenmektedir. Ana parça uzunluğu 7 ve 9 için sonuca ulaşılmıştır. Üzerinde çalışılan tüm genlerde alt sınırın sağlandığı kendine dik devirli kod elde edilmiştir. Farklı SP ve EP'ler için aynı DNA no'lardan aynı kodların elde edilmesi durumunda ilk karşılaşılan kod listeye eklenmiştir.

Tablo 5.4 kendine dik ve devirli kodlar yer almaktadır. Sadece ana parça uzunluğu 7 için sonuca ulaşılmıştır. Üzerinde çalışılan tüm genlerde kendine dik ve devirli kod elde edilmiştir. Farklı SP ve EP'ler için aynı DNA no'lardan aynı kodların elde edilmesi durumunda ilk karşılaşılan kod listeye eklenmiştir.

Elde edilen optimal MDS kodların sayısının oldukça fazla olması nedeniyle Tablo 5.5'te TRAV70 geni için elde edilen kodların bir kısmı listelenmektedir. Üzerinde çalışılan tüm genlerde aynı parametrelili kodlar elde edilmiştir.

Tablo 5.6'da devirli optimal kodlar yer almaktadır. Ana parça uzunluğu 9, 10, 15 ve 17 için sonuca ulaşılmıştır. Üzerinde çalışılan tüm genlerde devirli optimal kod elde edilmiştir. Farklı SP ve EP'ler için aynı DNA no'lardan aynı kodların elde edilmesi durumunda ilk karşılaşılan kod listeye eklenmiştir.

Tablo 5.7'de ise devirli kodlar yer almaktadır. Ana parça uzunluğu 7, 8, 16 ve 17 için sonuca ulaşılmıştır. Üzerinde çalışılan tüm genlerde devirli optimal kod elde edilmiştir. Farklı SP ve EP'ler için aynı DNA no'lardan aynı kodların elde edilmesi durumunda ilk karşılaşılan kod listeye eklenmiştir.

Tablo 5.2 Optimal, MDS ve devirli kodlar

Gen ID	N	SP	EP	DNA no	n	k	d
TRAV70	7	19	25	1,3,4	5	3	3
TRAV70	7	56	62	1,2,3	5	3	3
TRAV70	7	429	435	2,3,4	5	3	3
TRAV70	7	466	472	1,2,4	5	3	3
TRAV70	8	28	35	3	5	4	2
TRAV70	8	184	191	4	5	4	2
TRAV70	8	251	258	2	5	4	2
TRAV70	8	498	505	1	5	4	2
TRAV70	10	115	124	1,2,3,4	6	5	2
TRAV70	10	185	194	1,2,4	6	5	2
TRAV71	7	28	34	1,3,4	5	3	3
TRAV71	7	296	302	1,2,4	5	3	3
TRAV71	8	187	194	3	5	4	2
TRAV71	8	208	215	2	5	4	2
TRAV71	8	249	256	1	5	4	2
TRAV72	7	30	36	1,3,4	5	3	3
TRAV72	7	135	141	1,2,4	5	3	3
TRAV72	8	2	9	1	5	4	2
TRAV72	8	90	97	2	5	4	2
TRAV72	8	200	207	3	5	4	2
TRAV72	10	490	499	1,2,3,4	6	5	2
TRAV72	12	229	240	3	7	6	2
TRAV72	12	415	426	2	7	6	2
TRAV73	7	28	34	1,3,4	5	3	3
TRAV73	7	310	316	1,2,4	5	3	3
TRAV73	8	263	270	1	5	4	2
TRAV73	8	438	445	2	5	4	2
TRAV73	8	485	492	3	5	4	2
TRAV73	10	505	514	1,2,3,4	6	5	2
TRAV73	12	429	440	4	7	6	2
TRAV74	7	28	34	1,3,4	5	3	3
TRAV74	7	310	316	1,2,4	5	3	3
TRAV74	7	318	324	1,2,3	5	3	3
TRAV74	8	149	156	1	5	4	2
TRAV74	8	222	229	2	5	4	2
TRAV74	8	398	405	3	5	4	2
TRAV74	8	498	505	4	5	4	2
TRAV75	7	28	34	1,3,4	5	3	3
TRAV75	7	108	114	1,2,4	5	3	3
TRAV75	8	63	70	1	5	4	2
TRAV75	8	102	109	4	5	4	2
TRAV75	8	151	158	3	5	4	2
TRAV75	8	224	231	2	5	4	2
TRAV75	12	246	257	1	7	6	2
TRAV76	7	316	322	1,2,4	5	3	3
TRAV76	8	110	117	2	5	4	2
TRAV76	8	269	276	1	5	4	2
TRAV76	8	404	411	3	5	4	2

Tablo 5.3 Optimal, kendine dik ve devirli kodlar

Gen ID	N	SP	EP	DNA no	n	k	d
TRAV70	9	368	376	1	7	3	4
TRAV71	7	64	70	1	6	2	4
TRAV71	7	93	99	2	6	2	4
TRAV71	7	247	253	1,2,4	6	2	4
TRAV71	7	335	341	3	6	2	4
TRAV72	7	246	252	1,2,4	6	2	4
TRAV72	7	334	340	3	6	2	4
TRAV72	7	403	409	2	6	2	4
TRAV73	7	261	267	1,2,4	6	2	4
TRAV73	7	416	422	2	6	2	4
TRAV73	9	296	304	3	7	3	4
TRAV74	7	261	267	1,2,4	6	2	4
TRAV74	7	349	355	3	6	2	4
TRAV75	7	263	269	2,4	6	2	4
TRAV75	7	420	426	2	6	2	4
TRAV75	9	486	494	4	7	3	4
TRAV76	7	355	361	3	6	2	4

Tablo 5.4 Kendine dik ve devirli kodlar

Gen ID	N	SP	EP	DNA no	n	k	d
TRAV70	7	37	43	3	6	2	3
TRAV71	7	124	130	4	6	2	3
TRAV71	7	366	372	3	6	2	3
TRAV72	7	376	382	3	6	2	3
TRAV73	7	217	223	4	6	2	3
TRAV73	7	380	386	3	6	2	3
TRAV74	7	239	245	2	6	2	3
TRAV74	7	380	386	3	6	2	3
TRAV75	7	124	130	4	6	2	3
TRAV75	7	146	152	1	6	2	3
TRAV76	7	137	143	1	6	2	3

5.2 TRAV7 Geni ile Hata Düzelten Kodlar Arasındaki İlişki

DNA'dan protein ya da enzim üretmek için DNA'daki genetik bilgi RNA'ya RNA'dan da mRNA'ya aktarılır. RNA'dan mRNA'ya aktarım esnasında genin nükleotid dizilimi baştan sona okunmaz. Bir kısım dizilim okunur, bir kısım dizilim ise okunmadan atlanır ve oradan okunmaya devam edilir. Okunmadan atlanan bölgeye *kodlanmayan bölge (intron)* okunan bölgeye ise *kodlanan bölge (exon)* denir. İnsan DNA'sında kodlanmayan bölge tüm DNA'nın yaklaşık %97'lik kısmını oluşturmaktadır [47]. Gen araştırmalarında kodlanan ve kodlanmayan bölgelerin belirlenmesi önemli bir problemdir.

Tablo 5.5 Optimal ve MDS kodlar

Gen ID	N	SP	EP	DNA no	n	k	d
TRAV70	7	3	9	1,3	5	3	3
TRAV70	8	1	8	2	5	4	2
TRAV70	8	26	33	1	6	3	4
TRAV70	10	1	10	1	6	5	2
TRAV70	12	2	13	4	7	6	2
TRAV70	14	2	15	2	8	7	2
TRAV70	16	1	16	1	9	8	2
TRAV70	18	10	27	3	10	9	2
TRAV70	20	8	27	4	11	10	2
TRAV70	22	28	49	4	12	11	2
TRAV70	24	2	25	3	13	12	2
TRAV70	26	42	67	2	14	13	2
TRAV70	28	11	38	1	15	14	2
TRAV70	30	18	47	2	16	15	2
TRAV70	32	73	104	2	17	16	2
TRAV70	34	94	127	3	18	17	2
TRAV70	36	110	145	2	19	18	2
TRAV70	38	72	109	3	20	19	2
TRAV70	40	178	217	3	21	20	2

Tablo 5.6 Optimal devirli kodlar

Gen ID	N	SP	EP	DNA no	n	k	d
TRAV70	9	139	147	4	6	4	2
TRAV70	9	186	194	3	6	4	2
TRAV71	9	36	44	1,2,3,4	6	4	2
TRAV72	9	38	46	1,2,3,4	6	4	2
TRAV72	9	397	405	2	6	4	2
TRAV73	9	101	109	4	6	4	2
TRAV73	9	221	229	1,2,3,4	6	4	2
TRAV73	9	324	332	1	6	4	2
TRAV73	17	188	204	3	10	8	2
TRAV74	9	119	127	4	6	4	2
TRAV74	9	193	201	1,2,3,4	6	4	2
TRAV75	9	414	422	2	6	4	2
TRAV75	10	486	495	1,2,3	7	4	3
TRAV75	15	74	88	4	10	6	4
TRAV76	9	125	133	1,2,3,4	6	4	2
TRAV76	9	324	332	1,2,3	6	4	2
TRAV76	9	330	338	1	6	4	2

Protein ya da enzim sentezinde rol alan kodonun yapısında kodlanmayan bölge bulunmaması kodlanmayan bölgelerde hata düzeltme oranının düşük olduğunu düşündürmektedir. Bu düşünceden yola çıkarak bu bölümde TRAV7 geni ile hata düzeltme arasında bir ilişki olduğu gösterilecektir.

Tablo 5.7 Devirli kodlar

Gen ID	N	SP	EP	DNA no	n	k	d
TRAV70	8	153	160	3	6	3	3
TRAV70	8	396	403	2	6	3	3
TRAV71	7	11	17	1,2,3,4	6	2	3
TRAV71	8	134	141	4	6	3	3
TRAV72	7	291	297	1,2,3,4	6	2	3
TRAV72	8	175	182	1	6	3	3
TRAV72	17	441	457	1,2,3,4	12	6	3
TRAV73	7	121	127	1,2,3,4	6	2	3
TRAV73	8	190	197	1	6	3	3
TRAV74	7	306	312	1,2,3,4	6	2	3
TRAV75	7	308	314	1,2,3,4	6	2	3
TRAV75	8	195	202	2	6	3	3
TRAV75	16	74	89	1,2,3	10	7	2
TRAV76	7	87	93	1,2,3,4	6	2	3
TRAV76	8	31	38	2	6	3	3
TRAV76	8	324	331	4	6	3	3
TRAV76	8	397	404	1	6	3	3
TRAV76	8	412	419	1,2,3,4	6	3	3

Bu çalışmada izlenen algoritma aşağıdaki gibidir:

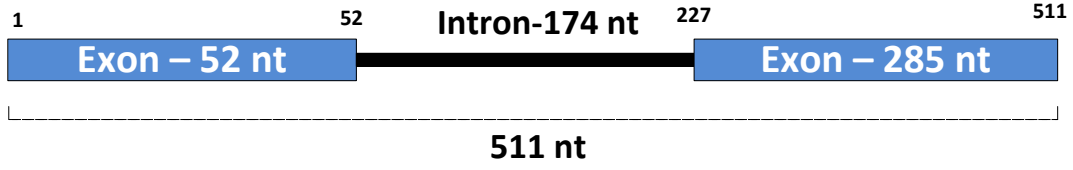
1. TRAV7 geni için öncelikle genden alınan ana pencere yani N^1 belirlenir.
2. Magma [43] programı kullanılarak kayan pencere metodu ile boyutu 2 olan kodların parametreleri belirlenir.
3. Elde edilen kodların minimum uzaklığı hesaplanarak Teorem 2.5'den kodun düzeltebileceği hata sayısı tespit edilir.
4. Hata düzeltemeyen kodların sayısı belirlenerek hangi oranda kodlanmayan bölgede bulundukları tespit edilir. Bu oran algoritma uygulayıcısına kodlanmayan bölge tahmin yüzdesini verir.

Edinilen bilgilerin grafikleri aşağıda yer almaktadır.

Şekil 5.1'de TRAV7 genindeki kodlanan ve kodlanmayan kısımlar yer almaktadır. Buna göre 511 uzunluklu gende 52 uzunluklu ve 285 uzunluklu iki kodlanan bölge ve 174 uzunluklu bir kodlanmayan bölge olduğu görülür.

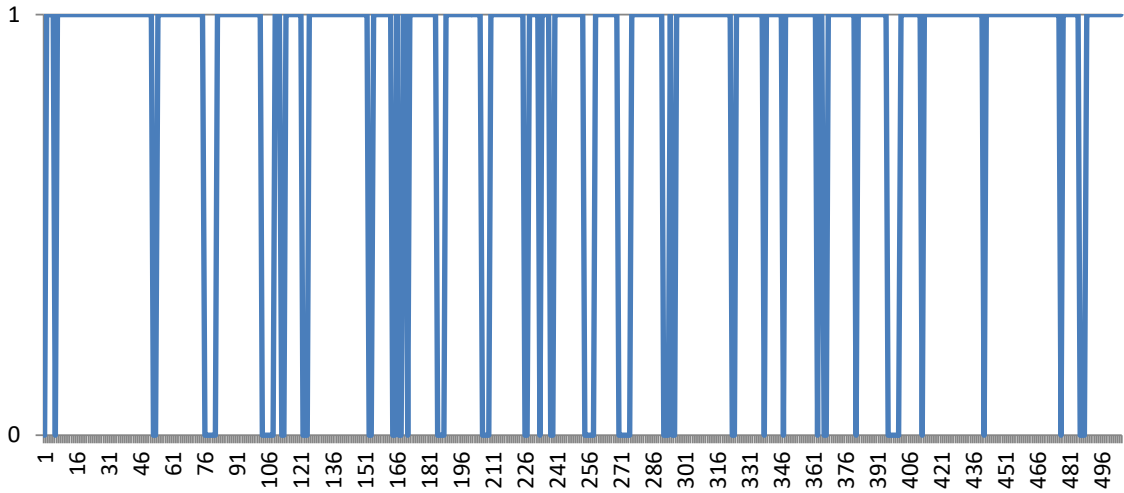
TRAV7 geninden alınan 7 uzunluklu parçalardan kayan pencere metodu ile elde edilen boyutu 2 olan kodlar incelendiğinde bu kodların minimum uzaklıklarının maksimum 2

¹Bu çalışmada ana pencere uzunluğunun 7, 8, 9 ve 10 olma durumları incelenmiştir.



Şekil 5.1 TRAV7 genindeki kodlanan ve kodlanmayan bölgeler

olduğu görülür. Şekil 5.2'deki hata düzeltme grafiği incelendiğinde hata düzeltemeyen kodların yoğunlukta olduğu görülür.



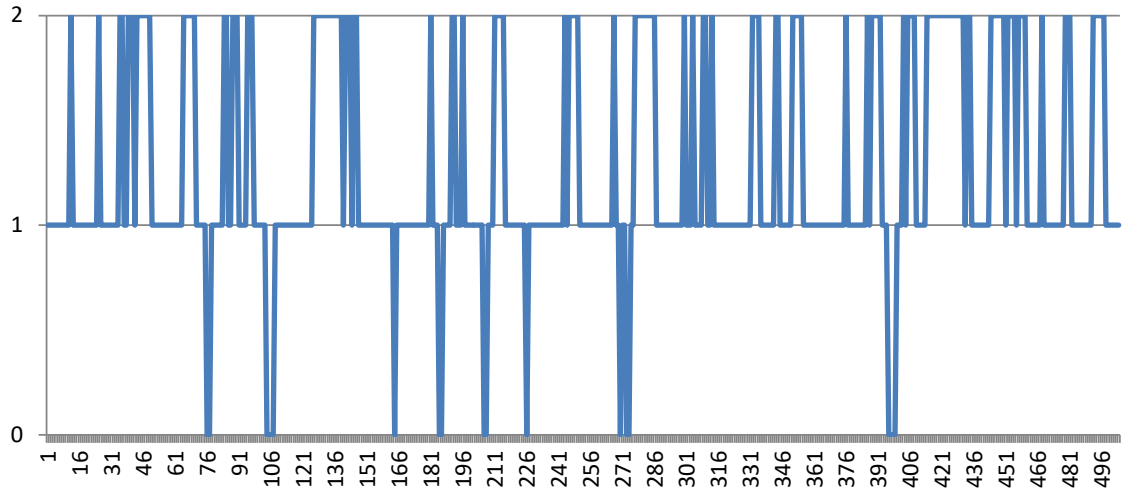
Şekil 5.2 TRAV7 geni N=7, n=6, k=2 için hata düzeltme grafiği

TRAV7 geninden alınan 8 uzunluklu parçalardan kayan pencere metodu ile elde edilen boyutu 2 olan kodlar incelendiğinde ise bu kodların minimum uzaklıklarının maksimum 5 olduğu görülür. Şekil 5.3'teki hata düzeltme grafiği incelendiğinde hata düzeltemeyen kodların yoğunlaştığı bölgelerin %63'inin kodlanmayan bölgede olduğu görülür.

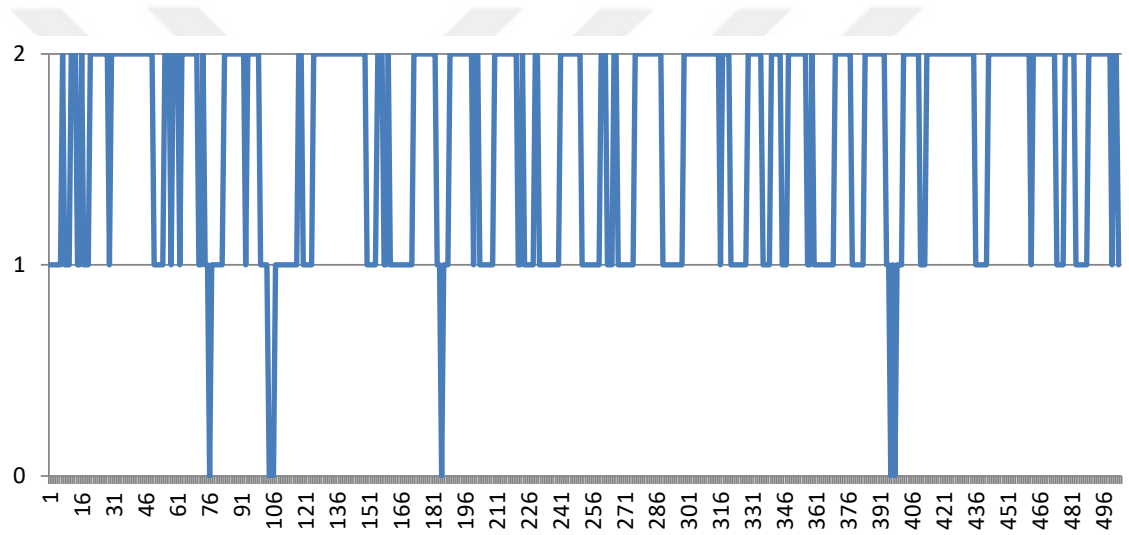
TRAV7 geninden alınan 9 uzunluklu parçalardan kayan pencere metodu ile elde edilen boyutu 2 olan kodlar incelendiğinde ise bu kodların minimum uzaklıklarının maksimum 6 olduğu görülür. Şekil 5.4'teki hata düzeltme grafiği incelendiğinde hata düzeltemeyen kodların yoğunlaştığı bölgelerin %78'inin kodlanmayan bölgede olduğu görülür.

TRAV7 geninden alınan 10 uzunluklu parçalardan kayan pencere metodu ile elde edilen boyutu 2 olan kodlar incelendiğinde ise bu kodların minimum uzaklıklarının maksimum 7 olduğu görülür. Şekil 5.5'deki hata düzeltme grafiği incelendiğinde hata düzeltemeyen kodların tamamının kodlanmayan bölgede olduğu görülür.

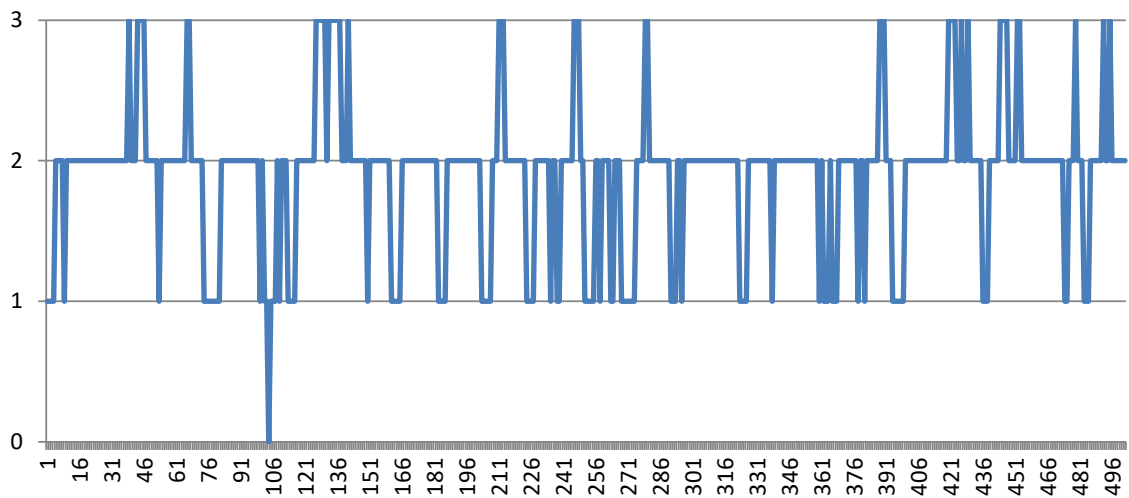
Grafiklerde DNA no 1 için elde edilen sonuçlar gösterilmektedir. Diğer seçimler için



Şekil 5.3 TRAV7 geni $N=8$, $n=7$, $k=2$ için hata düzeltme grafiği



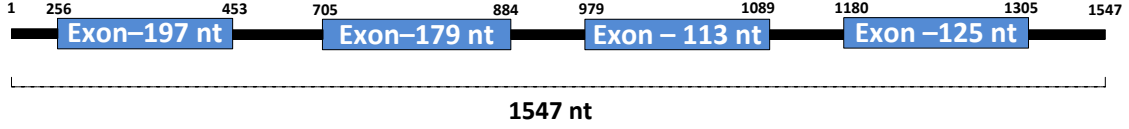
Şekil 5.4 TRAV7 geni $N=9$, $n=8$, $k=2$ için hata düzeltme grafiği



Şekil 5.5 TRAV7 geni $N=10$, $n=9$, $k=2$ için hata düzeltme grafiği

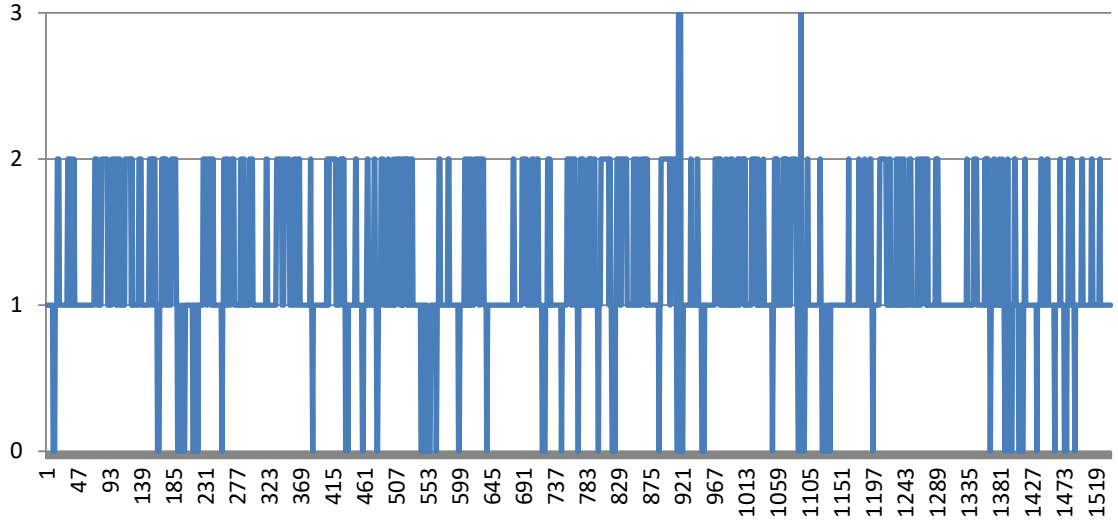
elde edilen sonuçlar DNA no 1 ile paralellik göstermektedir.

Moleküler biyoloji üzerine yapılan çalışmalarda oldukça sık kullanılan Arabidopsis Thaliana ² bitkisinden alınan gen (NRPB5 Eukaryotic rpb5 RNA polymerase subunit family protein [Arabidopsis Thaliana (thale cress)]) için kodlanan ve kodlanmayan kısımlar şekil 5.6'da yer almaktadır.



Şekil 5.6 Arabidopsis Thaliana genindeki kodlanan ve kodlanmayan bölgeler

Arabidopsis Thaliana geninden alınan ana pencere uzunluğu 8 olan parçalardan kayan pencere metodu ile elde edilen boyutu 2 olan kodlar incelendiğinde ise bu kodların minimum uzaklıklarının maksimum 7 olduğu görülür. Şekil 5.6'daki bölgeler ve Şekil 5.7'deki hata düzeltme grafiği incelendiğinde hata düzeltemeyen kodların yoğunlaştığı bölgelerin %82'sinin kodlanmayan bölgede olduğu görülür. Arabidopsis Thaliana geni için detaylı bilgi NCBI gen veritabanında [48] yer almaktadır.



Şekil 5.7 Arabidopsis Thaliana geni N=8, n=7, k=2 için hata düzeltme grafiği

5.3 Elde Edilen Sonuçlar

Proteinlerin üretimindeki anlamlı rolü nedeniyle, genler içerisindeki kodlanan ve kodlanmayan bölgeleri belirlemek önem arz etmektedir.

Bu çalışmada, hata düzelten kodlar kullanılarak oluşturulan metod ile kodlanan ve kodlanmayan bölgeleri belli bir oranda tahmin edebilen bir algoritma oluşturulmuştur.

²Bkz. <https://www.sciencedirect.com/topics/neuroscience/arabidopsis-thaliana>

Arabidopsis Thaliana bitkisi üzerinde yapılan [49] nolu çalışmada, DNA zincirlerinin ayrılması için anahtar role sahip bağlanma bölgesinin (binding site) [28] tespiti için, 8 uzunluğundaki DNA dizilerinin, 6 – 13 aralığındaki diğer uzunluktaki dizilere göre daha anlamlı olabileceği sonucu belirtilmiştir. Bu çalışmada da, kullanılan cebirsel metod ve tahmin algoritması ile [49] çalışmasına benzer olarak 8 uzunluğundaki DNA parçalarının, kodlanan ve kodlanmayan bölgelerin belirlenmesi hususunda da önem arz ettiği ortaya çıkarılmıştır.

Tanımlanan kodlanan ve kodlanmayan bölgeleri tahmin algoritması, belirli bir yüzde ile kodlanmayan bölge için doğruluk belirtmesi ve mevcut biyolojik çalışmaların sonuçları ile örtüşen verileri ortaya çıkarması, bu algoritmanın genler üzerindeki araştırma alanındaki tahmin çalışmalarında belirli bir öneme sahip olduğunu göstermektedir.



6 SONUÇ ve ÖNERİLER

Bu bölümde, yapılan çalışmalardan elde edilen sonuçlar ve yapılabilecek çalışma önerileri verilecektir.

6.1 Sonuç

Bu tezde, öncelikle yerel Frobenius olup zincir halkası olmayan $R = F_2 + uF_2 + vF_2 + v^2F_2$ halkası ve bu halkanın ideal yapısı incelendi. R halkası üzerinde tanımlanan bir C devirli kodunu üreten ideal yapısı belirlendi. Ayrıca R halkası üzerinde ters-sıralı DNA kod elde etmek için tek ya da çift uzunluklu devirli kodların 3 tipi belirlendi. Belirlenen bu devirli kodların üreteç matrisi ve kodsöz sayısı elde edildi. Ayrıca dual kodunun üretici belirlendi ve R halkası üzerindeki lineer kodlar için kendine dik olma koşulu verildi.

R halkası üzerindeki kodların Gray dönüşümlerinden optimal ikili lineer kodlar elde edildi. Elde edilen optimal kodlar, optimal lineer kodlar veritabanında [22] birden fazla adımda gerçekleşirken R halkası üzerinden Gray dönüşüm ile elde edilen aynı parametrelili kodların sadece bir adımda bulunabildiği görüldü. Ayrıca elde edilen kodların kendine dik kod oldukları anlaşıldı. Elde edilen kodlar incelendiğinde tümünün bir-ağırlıklı kod oldukları görüldü.

Farklı bir kod inşa metodu olan, döngüsel matrisler kullanılarak, yani R halkası üzerindeki temsilci vektörünü ve bu vektörün devirsel ötelemelerini satır kabul eden matristen optimal kodlar elde edildi. Bu kodların arasında iki-ağırlıklı kodların bulunmasının yanı sıra optimal iki-ağırlıklı kodların da olduğu görüldü.

Bölüm 4'te ise R halkası üzerindeki devirli kodların DNA'ya uygulamalarına yer verildi. Bu bölümde, R halkası üzerinde tanımlanan devirli kodların cebirsel özelliklerinden faydalanılmak suretiyle DNA kodlar için önemli bir eşik olan ters sıralılık problemi çözülerek ters-sıralı DNA kod ve ters-sıralı tamlayan DNA kodlar elde edildi.

Bölüm 5'in ilk kısmında, gerçek DNA dizileri üzerine çalışmalar yapıldı. İnsanda bulunan TRAV7 (T cell receptor alpha variable 7) geni ve farede bulunan TRAV7-1 ile TRAV7-6 arasındaki 7 gen üzerindeki lineer kod yapısı incelendi. TRAV7 genlerinden kayan pencere metodu ile elde edilen dizilerin F_4 'teki karşılıklarını satır kabul eden matrislerin ürettiği kodların parametreleri elde edildi. Ayrıca elde edilen lineer kodlar MDS, kendine dual, kendine dik, devirli ve optimal olma özelliklerine göre sınıflandırıldı.

Bölüm 5'in ikinci kısmında, literatürde ilk olarak, TRAV7 geni ve Arabidopsis Thaliana geni için hata düzeltme grafikleri elde edilerek genlerle hata düzeltme grafikleri arasındaki ilişki araştırıldı. Kodlanan ve kodlanmayan bölgelere karşılık hata düzeltme grafiğindeki durumlar incelendiğinde hata düzeltemeyen kodların kodlanmayan bölgede yoğunlaştığı görüldü.

6.2 Öneriler

GC-miktar kısıtını sağlayan DNA kodlar ve CSS quantum kodlar üzerine öneriler aşağıda yer almaktadır.

6.2.1 GC-miktar Kısıtını Sağlayan DNA Kodlar

Bölüm 2'de GC-miktar kısıtının tanımı verilmektedir. Tanıma göre bu tezde bahsedilen kod üretme metodu için R halkasında GC-miktar kısıtı sağlanmamaktadır.

Örneğin; (u, u, u, u) vektörünü ele alalım. (u, u, u, u) vektörünün DNA dizisi olarak karşılığı $GCGCGCGC$ 'dir. Ancak bu vektörü u ile çarptığımızda elde ettiğimiz vektör $u(u, u, u, u) = (v^2, v^2, v^2, v^2)$ olur. Bu vektörün DNA karşılığı ise $TTTTTTTT$ şeklindedir. Bu ise GC-miktar kısıtının sağlanmadığını göstermektedir.

R halkası üzerindeki kodlardan GC-miktar kısıtını sağlayan DNA kodların bulunması açık bir problemdir.

6.2.2 R Halkasından Elde Edilen Kodların CSS Quantum Kod İnşasında Kullanılması

Teorem 3.6'da R halkası üzerindeki lineer kodların Gray dönüşümleri ile oluşan kodların ikili kendine dik kodlar olduklarından bahsedilmektedir. Dolayısıyla [50] nolu çalışmada geçen Sonuç 21'e göre R halkası üzerindeki lineer kodların Gray dönüşümleri ile oluşan ikili kendine dik kodların duali ile CSS quantum kod inşası yoluyla CSS quantum kod elde edilir. Ayrıca elde edilen kodlar bazıları [22]'ye göre

bilinen en iyi quantum kodlardandır.

Örneğin, $C = (v^2 + v^2x)$ kodu R halkası üzerinde bir kod olsun. C kodunun Gray görüntüsünün parametreleri $[24, 2, 16]$ şeklindedir. Bu kodun dualinin parametreleri ise $[24, 22, 2]$ olur. Dolayısıyla, bu kod parametreleri $[[24, 22, 2]]$ olan bir ikili quantum kod elde etmek için kullanılabilir.

Yukarıda bahsedilen CSS quantum kod inşası ile iyi quantum kodlar elde edilmesi açık bir problemdir.



A

TRAV7 VE ARABİDOPSİS THALİANA GENLERİNİN DİZİLİMLERİ

Ulusal biyoteknoloji bilgi merkezi arama veritabanı [48]'den alınan TRAV7 genleri ve Arabidopsis thaliana geni ile ilgili bilgiler aşağıda yer almaktadır.

ID: 28686 T cell receptor alpha variable 7 [Homo sapiens (human)] Chromosome 14

ATGGAGAAGATGCGGAGACCTGTCCTAATTATATTTTGTCTATGTCTTGGCTGTAAGTTG
AGGGTTCTAAGAACTGGGGACCCAGGAGACATTTATTCAAGTCCTTTTGGGGAGATG
GGGATGTAGTCTGGACTTACTTGTCATTGCTTGTTTGAGATTAAAGAAATAAAATTATGAA
AGGTCTAAATTAAAATGTACATATTGTACCTGATGTCTTTCTGAATAGGGGCAAATGGAG
AAAACCAGGTGGAGCACAGCCCTCATTTTCTGGGACCCAGCAGGGAGACGTTGCCTC
CATGAGCTGCACGTACTCTGTCAAGTCGTTTAAACAATTTGCAGTGGTACAGGCAAATAC
AGGGATGGGTCCCAAACACCTATTATCCATGTATTCAGCTGGATATGAGAAGCAGAAAGG
AAGACTAAATGCTACATTACTGAAGAATGGAAGCAGCTTGATACATTACAGCCGTGCAGCC
TGAAGATTCAGCCACCTATTTCTGTGCTGTAGATG.

ID: 671652 T cell receptor alpha variable 7-1 [Mus musculus (house mouse)]
Chromosome 14

ATGAAGTCCTTGTGTGTTTCACTAGTGGTCCTGTGGCTTCAGCTACACTGTAAGTTTGG
GAATTTCTTCTGGATCCAGGTGTGATATGTAAAGTTGTTGTCTGCCCAGGCTAATGGTAG
AAACACAGATAGTTTCAATCTCTAGGAGCGGGAGGCCCATCAAAAAGCCATTCTGGAGT
GTTATGCATCTCTTCTTGATTGTCTGACCTTTGTTTTTCTCCACAGGGGTGAACAGCCAG
CAGAAGGTGCAGCAGAGCCCAGAATCCCTCATTGTTCCAGAGGGAGGCATGGCCTCTCT
CAACTGCACTTTTCAAGTATCGTAATTCTCAGTATTTCTGGTGGTACAGACAGCATTCTGG
GGAAGGCCCAAGGCACTGATGTCCATCTTCTCCAATGGTGACAAGAAGGAAGGCAGAT
TCACAGCTCACCTCAATAAGGCCAGCCTGCATGTATCCCTGCACATCAAAGACTCCCAAC
CCAGTGACTCTGCTCTCTACTTCTGTGCAGTGAGCAC.

ID: 634132 T cell receptor alpha variable 7-2 [Mus musculus (house mouse)]

Chromosome 14

GAATGAAATCCTTTAGTATTTCCCTAGTGGTCCTGTGGCTTCAGCTAAACTGTAAGTTT
GGGAATTTCTTTAGGATTGTAATATTGGTTCTAGGCTAATGGTAGGAACCCAGATGTTT
TTTAATATCTAGTTAGGGATAGGAGACCCATGGAAAAATCTTTCTAGAGTGCTAGCATC
CATATAAGTCCACTCTATTTTGTCTGATCTTTGTCTTTCTGCACAGGGGTGAACAGCCA
ACAGAAGGTGCAGCAGAGCCCAGAATCCCTCATTGTTCCAGAGGGAGGCATGGCCTC
TCTCAACTGCACTTCCAGTGATCGTAATGTTGACTACTTCTGGTGGTACAGACAGCACT
CTGGGAAAAGCCCCAAGATGCTGATGTCTATCTTCTCCAATGGTGAAAAGGAAGAAGG
CAGATTCACAGTTCACCTCAATAAAGCCAGCCTGCATACTCCCTGCACATCAGAGACT
CCCAGCCCAGTGACTCTGCTCTCTACCTCTGTGCA.

ID: 547420 T cell receptor alpha variable 7-3 [Mus musculus (house mouse)]
Chromosome 14

ATGAAATCCTTGAGTGTTTCCCTAGTGGTCCTGTGGCTCCAGTTAAACTGTAAGTTTGG
GAATTCCTTTGCGATCCAGGTGTGATATGTAAAGTTACTGTCAGCTCCAGGGGAATGGT
AGACACACAGATTTCTTAATCTCTAGTTAAGAAATGGGAAGCCCATGGAAAAATCATTCT
AGAGTGTTAGCATCCATATGCATCCAGTCTTGATTGTCTGGCCTTTGTCTTTCTGCACAG
GGGTGAACAGCCAGCAGAAGGTGCAGCAGAGCCCAGAATCCCTCATTGTCCCAGAGGG
AGCCATGACCTCTCTCAACTGCACTTTCAGCGACAGTGCTTCTCAGTATTTTGCATGGTA
CAGACAGCATTCTGGGAAAGCCCCCAAGGCACTGATGTCCATCTTCTCCAATGGTGAAA
AAGAAGAAGGCAGATTCACAATTCACCTCAATAAAGCCAGTCTGCATTTCTCCCTGCACA
TCAGAGACTCCCAGCCCAGTGACTCTGCTCTCTACCTCTGTGCAGTGAGCA.

ID: 386450 T cell receptor alpha variable 7-4 [Mus musculus (house mouse)]
Chromosome 14

ATGAAATCCTTGAGTGTTTCACTAGTGGTCCTGTGGCTCCAGTTAAACTGTAAGTTTGG
GAATTCCTTTGGGATCCAGGTGTGATATGTAAAGTTATTGTCTGCTCCAGGCTAATGGTA
CAAACACAGATGTTCTTAATCTCTAGTTAGGGGCGGGAGACCCATGGAAATGCCATTCT
GGAATGTTAGCATCCATACGCATCCAGTCTTGATTGTCTGACCTTTGTTTTTCTGTACAG
GCGTGAGGAGCCAGCAGAAGGTGCAGCAGAGCCCAGAATCCCTCAGTGTCCCAGAGG
GAGGCATGGCCTCTCTCAACTGCACTTCAAGTGATCGTAATTTTCAGTACTTCTGGTGG
TACAGACAGCATTCTGGAGAAGGCCCAAGGCACTGATGTCAATCTTCTCTGATGGTGA
CAAGAAAGAAGGCAGATTCACAGCTCACCTCAATAAGGCCAGCCTGCATGTTTCCCTGC
ACATCAGAGACTCCCAGCCCAGTGACTCCGCTCTCTACTTCTGTGCAGCTAGTGAGCA.

ID: 630061 T cell receptor alpha variable 7-5 [Mus musculus (house mouse)]
Chromosome 14

ATGAAATCCTTGAGTGTTTCACTAGTGGTCCTGTGGCTCCAGTTTAAATTGTGAGTTTGG
ACATTTCTTTTGGATATAAGGTGCTATAAGTGAGGTCACTGCCTACTCTAGGATAGTGGT
AGAAACATAGATGTTTAAATCTCTAGTTAAGGGGTGGGAGGCCCATGGAAAATTTATTC
TAGAGTGTTAGCATCCAAATGCATCCAGTCTTGATCATCTGATCTTTGTTTTTCTGCACAG
GGGTGAGCAGCCAGCAGAAGGTGCAGCAGAGCCCAGAATCCCTCACTGTCTCAGAGGG
AGCCATGGCCTCTCTCAACTGCACGTTCAGTGATGGTACTTCTAACAACCTTCAGGTGGTA
CAGACAGCATTTCTGCGAAAGGCCTTGAGGTGCTAGTGTCCATCTTCTCTGATGGTGAAA
AGGAAGAAGGCAGATTTACAGCTCACCTCAATAGAGCCAACTTGCATGTTTCCCTACACA
TCAGAGAACCACAACCCAGTGACTCTGCTGTCTACCTCTGTGCAGTGAGCA.

ID: 213480 T cell receptor alpha variable 7-6 [Mus musculus (house mouse)]
Chromosome 14

ATGCATTCCTTACATGTTTCACTAGTGTTCTCTGGCTTCAACTAGGTGGTGAGTTTGGT
CATTTCTATGGGAATATAAAAATACAATATATAAGGGTATTGCCTATTTTAGGCTAGAGGGT
AAAATGCTAATATTCTCACGCTGCAACAAAATCGGGTGGGAGGCCCGTGGAAAAGCAAT
TCTAGAATGTTAACATCCAAGTGCATCCAGTCTTGACCGTCTGACCTTTGTCTTTCTGCA
CAGGGGTGAGCAGCCAGGAGAAGGTACAGCAGAGCCCAGAATCCCTCATTGTCCCAGA
GGGAGCCATGGTCTCCCTCAACTGCAGTTTCAGCGACAGTGCTTCTCAGTCCATCTGGT
GGTACCAACAGCATCCTGGGAAAGGCCCAAAGCACTAATATCCATCTTCTCTAATGGCA
ACAAGAAAGAAGGCAGATTGACAGTTTACCTCAATAGAGCCAGCCTGCATGTTTCCCTG
CACATCAAAGACTCCCAGCCCAGTGACTCCGCTGTCTACCTCTGTGCAGTGAGCA.

ID: 821801 NRPB5 Eukaryotic rpb5 RNA polymerase subunit family protein [
Arabidopsis thaliana (thale cress)] Chromosome 3

ATAGAATTGAGAAAAGGGTTCGAGGTAAAAGGAGAAAGCTTGAAACGCCGCTTCTTCT
TCTTCTTCTTCCACTGTCGTTCTCTCTGTAAATCGCTTCTTCAATCATCTCTGTGGATTT
GTCTCTATTTCGTAAAGTTTTCGTATCTGTGTATATATTCTCAAAGAGATTGATAAACTCTTA
ATTGTTTCTGTCAATTTTTGAAAAACCCTAATTTTGATATGGTTTTGGTAAAGCTGCAGA
TAACCTTGGTGAAGAAGATGTTGACGGAAGAGGAGCTCAAAGATTGTACAGAATCCAG
AAGACGTTGATGCAGATGTTAAGGGATCGTGGATATTTTATCGCGGATTCTGAACCTTAC
GATGACTAAACAACAGTTTATTAGGAAACATGGTGATAACATGAAAAGAGAGGATCTTGT
TACTCTTAAGGCTAAGCGAAATGATAATAGTGATCAGGTTTGTAGTTGTATGTCTCTCCTT
CAATGTGATTATCAGCTTCTCTTTAGTGTTCTTTAGCTGGAGACTATCATTATTTATCTAGT
TGTTGAATGTGAATAGTAAGTAGACATGGAATGTTACAGAGAGCTAGATTTTGACAAAAG
TATAGATAAATGTCATCTCTTAAATCTCTAAGAAGCTGTGGATATGGTAGAAAGATTGTTTA
TGTGTGTTTGATTAGTCGCTGATTCATGTGCTCTTGTTGTAGCTCTATATATCTTCCCT
GATGAGGCCGAAAGTTGGTGTGAAGACGATGAAAATGTACACGAACAGGATGAAATCAGA
GAATGTGTTTCAGAGCTATTTTGGTTGTGCAACAGAACTTGACCCCGTTTGCTCGAACTT

GCATTAGTGAGATCTCTTCTAAGTTTCATCTGGAAGTTTTCCAGGTAAAACAAATCACTT
TGTGTCTTGACTCTTGATGATGATGTATCCACATCACAATCCTACTAAACCATGATTTGGT
GGTGGATGCTGCAGGAGGCGGAAATGCTAGTGAACATTAAAGAACATGTTCTTGTTCC
TGAGCATCAAGTTCTTACCACTGAAGAGAAGAAAACTTTACTGGAGAGATACACAGTG
AAGGAGACACAGGTTAGTAGTATATAGAACTTTGTTTGAATTCTTGGTGATGAAATTT
TTGTTGTGTTATTAAATGTTTGGAACTTTGTGGGTGTTGTTAGCTTCCAAGGATCCAAG
TGACTGATCCAATCGCAAGATACTTTGGACTAAAACGTGGGCAAGTGTGAAGATCATTC
GTCCGAGTGAAACGGCTGGTCGTTATGTTACCTATCGTTATGTTGTATAAGCCAAAGCAA
AAGGCTTTTGTGGATTGTAATGCGAATTGAACGCTGAGAGTTGTGAAACATAAGAGTAA
TGCTTTTGTACTAAACCAAACCTGATGACTGTGTTTTATTTTTGTTTGTCCGATAGGAGAA
ATTAAAGCTTGAAACAACTTATTTGTATTAGTTTTCTATCCAAAACTTGCAAGATCCGAT
TCCGACTTGTTTGTATTTGTTTCATCCCAACAAATGCCATATTTTGTTAATAA.



- [1] “Nasa science,” <https://mars.nasa.gov/mars-exploration/missions/pathfinder/>, 23 Ekim 2019.
- [2] “What is the distance between earth and mars?” <https://www.space.com/14729-spacekids-distance-earth-mars.html>, 23 Ekim 2019.
- [3] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [4] R. W. Hamming, “Error detecting and error correcting codes,” *Bell Labs Tech. J.*, vol. 29, no. 2, pp. 147–160, 1950.
- [5] T. Head, “Formal language theory and dna: An analysis of the generative capacity of recombinant behaviors,” *Bull. Math. Biol.*, vol. 49, pp. 737–759, 1987.
- [6] L. Adleman, “Molecular computation of solutions to combinatorial problems,” *Science*, vol. 266, no. 5187, pp. 1021–1024, 1994.
- [7] L. Adleman, “On applying molecular computation to the data encryption standard,” *J. Comput. Biol.*, vol. 6, no. 1, pp. 53–63, 1999.
- [8] X. Wang, Z. Bao, J. Hu, S. Wang, A. Zhan, “Solving the sat problem using a dna computing algorithm based on ligase chain reaction,” *BioSystems.*, vol. 91, no. 1, pp. 117–125, 2008.
- [9] M. Darehmira, “A semi-general method to solve the combinatorial optimization problems based on nanocomputing,” *Int. J. Nanosci.*, vol. 9, no. 5, pp. 391–398, 2010.
- [10] E. E. Hesketh, J. Sayir, N. Goldman, “Improving communication for interdisciplinary teams working on storage of digital information in DNA,” *F1000Research*, vol. 7, no. 39, p. 13482.1, 2018.
- [11] G. Church, Y. Gao, S. Kosuri, “Next-generation digital information storage in DNA,” *Science*, vol. 337, no. 6102, p. 1628, 2012.
- [12] P. Gaborit O. King, “Linear constructions for DNA codes,” *Theoretical Computer Science*, vol. 334, no. 1–3, pp. 99–103, 2005.
- [13] T. Abulraib, A. Ghayeb, X. Nian Zeng, “Construction of cyclic codes over $GF(4)$ for DNA computing,” *J. Franklin Inst.*, vol. 343, no. 4-5, pp. 448–457, 2006.
- [14] E. Oztas I. Siap, “Lifted polynomials over f_{16} and their applications to dna codes,” *Filomat*, vol. 27, no. 3, pp. 459–466, 2013.
- [15] E. Oztas I. Siap, “On a generalization of lifted polynomials over finite fields and their applications to DNA codes,” *Int. J. Comput. Math.*, vol. 92, no. 9, pp. 1976–1988, 2015.

- [16] B. Yildiz I. Siap, "Cyclic codes over $f_2[u]/(u^4-1)$ and applications to dna codes," *Comput. Math. Appl.*, vol. 63, no. 7, pp. 1169–1176, 2012.
- [17] A. Bayram, E. S. Oztas, I. Siap, "Codes over $F_4 + vF_4$ and some dna applications," *Des. Codes Cryptogr.*, vol. 80, no. 2, pp. 379–393, 2015.
- [18] L. C. B. Faria, A. S. L. Rocha, J. H. Kleinschmidt, M. C. Silva-Filho, E. Bim, R. H. Herai, M. E. Yamagishi, R. Palazzo, "Is a genome a codeword of an error-correcting code?" *PloS one*, vol. 7, no. 5, e36644, 2012.
- [19] T. W. Hungerford, *Algebra*. Springer-Verlag, 1974.
- [20] S. Ling C. Xing, *Coding Theory A First Course*. Cambridge University Press, 2004.
- [21] W. C. Huffman V. Pless, *Fundamentals of Error-Correcting Codes*. Cambridge University Press, 2003.
- [22] M. Grassl, "Bounds on the minimum distance of linear codes and quantum codes," <http://www.codetables.de>, 2 Nisan 2019.
- [23] E. Prange, *Cyclic Error correcting Codes in Two Symbols*. Air Force Cambridge Res. Center, 1957.
- [24] "Genetik," <https://www.scinexx.de/diaschauen/genetik>, 3 Kasım 2019.
- [25] G. Battail, "Replication decoding revisited," pp. 1–5, 2003.
- [26] E. May, M. A. Vouk, D. L. Bitzar, D. I. Rosnick, "An error-correcting code framework for genetic sequence analysis," *Journal of the Franklin Institute*, vol. 341, no. 1, pp. 89–109, 2004.
- [27] E. May, "Towards a biological coding theory discipline," *New Thesis*, vol. 1, no. 1, pp. 19–38, 2004.
- [28] S. Vajda, D. Beglov, A. Wakefield, M. Egbert, A. Whitty, "Cryptic binding sites on proteins: Definition, detection, and druggability," *Curr Opin Chem Biol.*, vol. 44, pp. 1–8, 2018.
- [29] R. H. Burdon, *Genes and the Enviroment*. Taylor and Francis online, 1999.
- [30] L. L. Gatlin, *Information Theory and the Living System*. New York: Columbia Univ. Press, 1972.
- [31] G. L. Rosen, "Examining coding structure and redundancy in DNA," *IEEE Engineering in Medicine and Biology Magazine*, vol. 25, no. 1, pp. 2126–2133, 2006.
- [32] A. R. Hammons, P. V. Kumar, A. R. Calderbank, N. J. A. Sloane, P. Solé, "The z4-linearity of kerdock, preparata, goethals, and related codes," *IEEE Trans. Information Theory*, vol. 40, pp. 301–319, 1994.
- [33] S. T. Dougherty, A. Kaya, E. Salturk, "Cyclic codes over local frobenius rings of order 16," *Adv. Math. Commun.*, vol. 11, no. 1, pp. 99–114, 2017.
- [34] B. Yildiz S. Karadeniz, "Cyclic codes over $F_2 + uF_2 + vF_2 + uvF_2$," *Des. Codes Cryptogr.*, vol. 58, no. 3, pp. 221–234, 2011.
- [35] B. Yildiz I. Kelebek, "The homogeneous weight for R_k , related gray map a new binary quasi-cyclic codes," *Filomat*, vol. 31, no. 4, pp. 885–897, 2017.
- [36] J. A. Wood, "Duality for modules over finite rings and applications to coding theory," *American Journal of Mathematics*, vol. 121, no. 3, pp. 555–575, 1999.

- [37] I. Constantinescu W. Heise, "A metric for codes over residue class rings of integers," *Problemy Peredachi Informatsii*, vol. 33, no. 3, pp. 22–28, 1997.
- [38] M. Greferath S. E. Schmidt, "Gray isometries for finite chain rings and a nonlinear ternary (36, 312, 15) code," *IEEE Trans. Inform. Theory*, vol. 45, no. 7, pp. 2522–2524, 1999.
- [39] J. N. J. Moon, L. A. Hughes, D. H. Smith, "Assignment of frequency lists in frequency hopping networks," *IEEE Trans. Veh. Technol.*, vol. 54, no. 3, pp. 1147–1159, 2005.
- [40] P. J. Kuekes, W. Robinett, R. M. Roth, G. Seroussi, G. S. Snider, R. S. Williams, "Resistor-logic demultiplexers for nano electronics based on constant-weight codes," *Nanotechnology*, vol. 17, no. 4, pp. 1052–1061, 2006.
- [41] S. E. Greferath M ve Schmidt, "Finite-ring combinatorics and macwilliams' equivalence theorem," *J. Combin. Theory Ser. A*, vol. 92, no. 1, pp. 17–28, 2000.
- [42] P. Gaborit A. Otmani, "Tables of self-dual optimal codes," http://www.unilim.fr/pages_perso/philippe.gaborit/SD/index.html, 2 Nisan 2019.
- [43] W. Bosma, J. Cannon, C. Playoust, "The Magma algebra system. I. The user language," *J. Symbolic Comput.*, vol. 24, no. 3–4, pp. 235–265, 1997.
- [44] E. Chen, "Online database of two-weight codes," <http://www.tec.hkr.se/~chen/research/2-weight-codes/search.php>, 2 Nisan 2019.
- [45] T. Abualrub R. Oehmke, "On the generators of z_4 codes of length $2e$," *IEEE Trans. Inform. Theory*, vol. 49, no. 9, pp. 2126–2133, 2003.
- [46] F. Tajima, "Error detecting and error correcting codes," *J. Mol. Evol*, vol. 33, pp. 470–473, 1991.
- [47] G. K. Wong, D. A. Passey, Y. Huang, Z. Yang, J. Yu, "Is 'junk' dna mostly intron dna?" *Genome research*, vol. 10, no. 11, pp. 1672–1678, 2000.
- [48] L. Geer, A. Marchler-Bauer, R. Geer, L. Han, J. He, S. He, C. Liu, W. Shi, S. Bryant, "National center for biotechnology information," <https://www.ncbi.nlm.nih.gov>, 10 Eylül 2019.
- [49] J. Lichtenberg, A. Yilmaz, J. D. Welch, K. Kurz, X. Liang, F. Drews, K. Ecker, S. S. Lee, M. Geisler, L. R. Grotewold Erich ve Welch, "The word landscape of the non-coding segments of the arabidopsis thaliana genome," *Bell Labs Tech. J*, vol. 10, no. 1, 2009.
- [50] A. Ketkar, A. Klappenecker, S. Kumar, P. K. Sarvepalli, "Nonbinary stabilizer codes over finite fields," *IEEE Trans. Information Theory*, vol. 52, no. 11, pp. 4892–4914, 2006.

TEZDEN ÜRETİLMİŞ YAYINLAR

İletişim Bilgileri: mbyilgor@yildiz.edu.tr

Makale

1. M. Bulut Yılğör, F. Gürsoy, E. S. Oztas ve F. Demirkale, "Cyclic codes over $F_2 + uF_2 + vF_2 + v^2F_2$ with respect to the homogeneous weight and their applications to DNA codes", AAECC, 2020. <https://doi.org/10.1007/s00200-020-00416-0>

Konferans Bildirisi

1. M. Bulut Yılğör, E. S. Oztas ve F. Demirkale", Cyclic codes over a special non-chain ring with respect to the homogeneous weight", International Conference on Mathematical Advances and Applications, Mayıs, 2019, İstanbul, Türkiye.
2. E. S. Oztas, M. Bulut Yılğör, F. Gürsoy ve F. Demirkale, "Some DNA Codes over F_2 ", International Congress on Fundamental and Applied Sciences, Haziran, 2019, Tiran, Albania.

Proje

1. Fatih DEMİRKALE, Bazı Özel Sınıflara ait Kodların Sınıflandırılması, Bilimsel Araştırma Projeleri Koordinatörlüğü (BAP), 2016-07-03-DOP05, Araştırmacı.