Faster-than-Nyquist Sinyalleşme için Öz-Dikkat Transformer Tabanlı Dedektör
Self-Attention Transformer-Based Detector for Faster-than-Nyquist Signaling

Nurettin Şafak, Osman Tokluoglu, Enver Cavus
Elektrik Elektronik Mühendisliği Bölümü, Ankara Yıldırım Beyazıt Üniversitesi, Ankara, Türkiye
nsafaked@gmail.com, otokluoglu@aybu.edu.tr, ecavus@aybu.edu.tr

***Özetçe*—Bu çalışmada, $\tau = 0.8$ sıkıştırma faktörüyle kasıtlı semboller arası girişim (ISI) oluşturan Nyquist-ötesi (FTN) sinyalleşme kanalları üzerinden iletilen BPSK sinyalleri için özgün bir encoder-only Transformer tabanlı alıcı mimarisi sunuldu. BPSK modülasyonu, RRC darbe şekillendirme ve eşleştirilmiş filtrelemeden kaynaklanan ISI katsayılarını kapsayan uçtan uca tam bir haberleşme zinciri kurularak değerlendirildi. Önerilen Transformer alıcısı, 0–8 dB E_b/N_0 aralığında optimal BCJR dedektörü ile karşılaştırıldı. BCJR ile BER farkını sistematik biçimde kapatmak amacıyla çok-SNR ön eğitim ve SNR başına müfredat ince ayarını birleştiren iki aşamalı bir eğitim stratejisi geliştirildi. Transformer alıcısının hesaplama karmaşıklığı ve çıkarım süresi GRU ile karşılaştırmalı olarak analiz edildi. Attention haritası görselleştirmeleri, Transformer'ın herhangi bir kanal bilgisi gerektirmeksizin FTN kaynaklı ISI bellek yapısını özerk olarak tanımladığını; SNR arttıkça dikkat ağırlıklarının merkez token ve yakın komşularında belirgin biçimde yoğunlaştığını ortaya koydu.**

Anahtar Kelimeler*—*Transformer, BCJR, BPSK, FTN sinyalleşme, ISI kanalı, paralel çıkarım, attention görselleştirme, müfredat öğrenmesi

***Abstract*—In this study, a novel encoder-only Transformer-based receiver architecture is presented for BPSK signals transmitted over Faster-than-Nyquist (FTN) signaling channels that introduce intentional inter-symbol interference (ISI) with a compression factor of τ=0.8. A complete end-to-end communication chain encompassing BPSK modulation, RRC pulse shaping, and the ISI coefficients arising from matched filtering was constructed and evaluated. The proposed Transformer receiver was benchmarked against the optimal BCJR detector over an E_b/N_0 range of 0–8 dB. To systematically close the BER gap to the BCJR, a two-stage training strategy combining multi-SNR pretraining and per-SNR curriculum fine-tuning was developed. The computational complexity and inference latency of the Transformer receiver were analyzed in comparison with a GRU based receiver. Attention map visualizations revealed that the Transformer autonomously identifies the FTN-induced ISI memory structure without requiring any prior channel knowledge; as the SNR increases, the attention weights become significantly concentrated around the center token and its nearest neighbors.**

Keywords*—*Transformer, BCJR, BPSK, FTN signaling, ISI channel, parallel inference, attention visualization, curriculum learning

979-8-3195-1046-4/26/$31.00 ©2026 IEEE

I. GİRİŞ

Yüksek veri hızlarına olan artan talep ve sınırlı spektral kaynaklar, spektral verimliliği artırmaya yönelik tekniklerin geliştirilmesini zorunlu kılmaktadır. Özellikle kablosuz haberleşme sistemlerinde artan kullanıcı yoğunluğu ve bant genişliği kısıtları, mevcut kaynakların daha verimli kullanılmasını kritik bir gereksinim haline getirmiştir. Bu tekniklerden biri, ISI'yi kasıtlı olarak ortaya çıkaracak biçimde Nyquist kriterinin altında sembol aralığıyla iletim gerçekleştiren Nyquist-ötesi (FTN) sinyalleşmedir [1], [2]. FTN sinyalleşmesi bant genişliği sabit tutulurken iletim hızını artırarak spektral verimlilik kazanımı sağlar; $\tau = 0.8$ için bu kazanım veri oranı gerektiren uygulamalar için FTN'i cazip bir alternatif haline getirmektedir. ISI'yi gidermek için optimal çözüm olan BCJR algoritması [3] kanal modelinin tam olarak bilinmesini gerektirmekte ve karmaşıklık modülasyon mertebesine göre katlanarak artmaktadır. Bu durum, kanal bilgisi gerektirmeyen ve düşük karmaşıklıklı alıcı tasarımlarına duyulan ihtiyacı ön plana çıkarmaktadır. Ayrıca, pratik sistemlerde kanalın zamanla değişkenlik göstermesi, bu tür model-bağımlı yaklaşımların uygulanabilirliğini daha da sınırlamaktadır. Bu durum, uyarlanabilir ve model-bağımsız alıcı tasarımlarına olan ihtiyacı giderek daha kritik kılmaktadır.

Hesaplama yükünü azaltmak amacıyla önerilen MBCJR algoritması [4] karmaşıklığı belirli ölçüde düşürmekle birlikte üstel artışı engelleyememekte ve kestirim doğruluğu ile karmaşıklık arasındaki ödünleşim belirgin bir sınırlayıcı faktör olmaya devam etmektedir. Tüm bu yaklaşımların ortak kısıtı, kanal modelinin önceden bilinmesini zorunlu kılması ve değişken kanal koşullarına uyum sağlayamamasıdır. Klasik yöntemlerin bu sınırlılıkları, FTN tespitinde derin öğrenme tabanlı çözümlere olan ilgiyi artırmıştır. CNN tabanlı alıcılar [8], [9] paralel özellik çıkarma avantajıyla umut verici sonuçlar ortaya koymuş; ancak sabit alıcı yapıları ISI bellek dinamiklerini modellemede yetersiz kalmıştır [8]. GRU tabanlı alıcı [7] $\tau = 0.8$ koşulunda BCJR'a yakın BER performansı sergilemiş; fakat sıralı işleme yapısı paralel çıkarım imkânını engellemiş ve gecikmeye duyarlı uygulamalar açısından dezavantaj oluşturmuştur. Öte yandan Transformer mimarisi [5], FTN sinyalleşme kanalları için [7]–[9]'da incelenen CNN ve GRU tabanlı yaklaşımların ötesinde henüz araştırılmamıştır.

Transformer mimarisi, öz-dikkat mekanizması aracılığıyla giriş dizisindeki tüm tokenlar arasındaki ilişkileri eş zamanlı

modelleyerek GRU'nun sıralı işleme kısıtını ortadan kaldırmakta; uzun menzilli bağımlılıkları öğrenme kapasitesi sayesinde FTN kanallarındaki karmaşık ISI yapılarının modellenmesinde belirgin avantaj sunmaktadır. FTN bağlamında ISI yalnızca komşu sembollerle değil pencere içindeki tüm geçmiş sembollerle etkileşim halinde olduğundan, klasik yerel işlemeye dayalı yöntemler yetersiz kalmaktadır. Bu boşluktan yola çıkarak FTN sinyalleşme kanalları için encoder-only Transformer tabanlı, model-bağımsız bir alıcı mimarisi önerilmektedir. Çalışmanın başlıca katkıları: (1) FTN kanalları için ilk Transformer tabanlı model-bağımsız alıcı tasarımı ve tam implementasyonu, (2) çok-SNR ön eğitim ve SNR başına müfredat ince ayarını [6] birleştiren iki aşamalı eğitim stratejisi, (3) Transformer alıcısının hesaplama karmaşıklığı ve çıkarım süresinin [7] ile karşılaştırmalı analizi, (4) attention haritaları aracılığıyla FTN kaynaklı ISI bellek yapısının yorumlanabilirlik analizi. Ayrıca önerilen yaklaşımın farklı kanal koşullarında genellenebilirliği değerlendirilerek, pratik sistemlerde uygulanabilirliği tartışılmıştır.

II. Sistem Modeli

A. FTN Sinyalleşme

Bant genişliği $(1+\beta)/2T$ olan bir taban bant darbesi $g(t)$ için iletilen sinyal şu şekilde ifade edilmektedir:

$$s(t) = \sum_k a_k\, g(t - k\tau T) \tag{1}$$

Burada a_k k-inci BPSK sembolünü, T Nyquist sembol aralığını ve $\tau \in (0,1)$ sıkıştırma faktörünü göstermektedir. $\beta = 0.35$ roll-off faktörlü RRC darbesi kullanılmış olup darbe enerjisi normalize edilmiştir:

$$\int_{-\infty}^{+\infty} |g(t)|^2\, dt = 1 \tag{2}$$

B. Kanal Modeli

İletim AWGN kanalı üzerinden gerçekleştirilmektedir. Eşleştirilmiş filtre (RRC) çıkışındaki örnekler:

$$y(n\tau T) = \sum_k a_k\, x\big((n-k)\tau T\big) + w(n\tau T) \tag{3}$$

Burada $x(t) = g(t) * g(-t)$, $w(n\tau T)$ ise örnekleme anındaki renkli gürültüdür. Renkli gürültü yapısı RRC filtresinin bant sınırlayıcı etkisinden kaynaklanmakta olup BCJR bu yapıyı kanal modeline dahil ederek optimal kararı üretmektedir [3]. Önerilen Transformer ise bu korelasyon yapısını herhangi bir ön bilgi gerektirmeksizin öz-dikkat mekanizması aracılığıyla örtük biçimde öğrenmektedir. $\tau = 0.8$ için ISI katsayıları Tablo I'de verilmektedir; $\sum_k |x_k|^2 = 1$ normalizasyon koşulu sağlanmaktadır.

Tablo I. $\tau = 0.8$ için ISI katsayıları

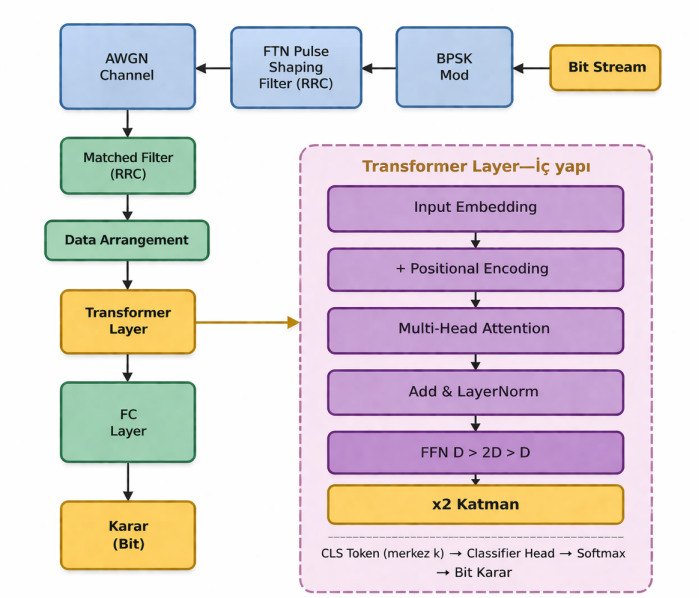
x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
.999	.222	−.152	.076	−.024	.006	.003	−.002	.001

C. Alıcı Girişi: Pencereli Gözlem

Transformer alıcısı, zaman indisi k etrafında $W = 8$ yarıçaplı bir gözlem penceresi kullanmaktadır:

$$\mathbf{r}_k = \big[y((k-W)\tau T), \ldots, y(k\tau T), \ldots, y((k+W)\tau T)\big] \in \mathbb{R}^S \tag{4}$$

Burada $S = 2W+1 = 17$ token sayısını ifade etmektedir. Pencere genişliği $W = 8$, $\tau = 0.8$ için tüm anlamlı ISI katsayılarını (x_0–x_8) kapsayacak şekilde belirlenmiştir. Sistem zincirinin tamamı Şekil 1'de gösterilmektedir.



Şekil 1. Önerilen uçtan uca FTN haberleşme sistemi zinciri ($\tau = 0.8$, $\beta = 0.35$ RRC).

III. Transformer Alıcı Tekniği

A. Genel Mimari ve Giriş Katmanı

Önerilen alıcı encoder-only Transformer mimarisine [5] dayanmaktadır. k sembolüne ilişkin bit kararı sabit uzunluklu bir bağlam penceresine dayalı sınıflandırma görevi olduğundan kod çözücü bileşeni gereksiz kalmakta; kodlayıcı yapısı çift yönlü dikkat mekanizması aracılığıyla hem geçmiş hem gelecek komşu sembollerle eş zamanlı etkileşim kurma imkânı sunmaktadır. Her skaler sinyal örneği $D = 20$ boyutlu gömme uzayına taşınmakta, öğrenilebilir konumsal kodlama eklenmektedir:

$$\mathbf{e}_i = y\big((k+i-W)\tau T\big)\cdot \mathbf{W}_e + \mathbf{b}_e, \quad \mathbf{H} = \mathbf{E} + \mathbf{PE} \in \mathbb{R}^{S\times D} \tag{5}$$

B. Çok Başlı Öz-Dikkat ve Kodlayıcı Bloğu

$N_H = 4$ başlı öz-dikkat mekanizması ve $N_L = 2$ kodlayıcı bloğu kullanılmaktadır. Her baş h için sorgu, anahtar ve değer matrisleri hesaplanarak ölçeklendirilmiş nokta çarpımı ile dikkat çıktısı elde edilmektedir:

$$\mathbf{A}_h = \mathrm{softmax}\left(\frac{\mathbf{Q}_h\mathbf{K}_h^\top}{\sqrt{D_H}}\right)\mathbf{V}_h, \quad D_H = D/N_H = 5 \tag{6}$$

$$\mathbf{H}' = \mathrm{LN}(\mathbf{H} + \mathrm{MHA}(\mathbf{H})), \quad \mathbf{H}_f = \mathrm{LN}(\mathbf{H}' + \mathrm{FFN}(\mathbf{H}')) \quad (7)$$

FFN gizli boyutu $D_F = 2D = 40$'tır. Merkez token temsili $\mathbf{cls} = \mathbf{H}_f[W, :]$ iki katmanlı sınıflandırıcı kafasına beslenerek $\hat{b} = \arg\max_j P(\hat{b} = j)$ bit kararı üretilmektedir. Toplam parametre sayısı ≈ 8.5K olup GRU'nun [7] 10K–50K parametresine kıyasla son derece hafif bir modeldir.

C. Eğitim Stratejisi

Modelin eğitiminde iki aşamalı bir strateji benimsenmiştir. **Ön eğitimde** model SNR $\in \{0, \ldots, 10\}$ dB aralığında her epoch için rastgele seçilen bir SNR noktasında eğitilmektedir. Toplam 100 epoch boyunca 5.000 sembollik mini-batch'lerle 500.000 sembol işlenmiştir; öğrenme hızı $\eta = 3 \times 10^{-3}$, gradyan kırpma eşiği ± 0.5, dropout $p = 0.1$ (Adam, $\beta_1 = 0.9$, $\beta_2 = 0.999$). **İnce ayarda** her hedef SNR için $\eta = 8 \times 10^{-4}$, 50 epoch, batch 2K olarak sürdürülmüştür. **Müfredat öğrenmesi** [6]: eğitimin ilk %40 epoch'unda model hedef SNR'den 2 dB yüksekte eğitilmekte; kalan epoch'larda doğrudan hedef SNR üzerinde sürdürülmektedir. Bu yaklaşım düşük SNR bölgelerinde yakınsamayı hızlandırmakta ve yerel minimumlara takılma riskini azaltmaktadır. Tüm deneyler PyTorch 2.x ve NVIDIA T4 GPU üzerinde gerçekleştirilmiş olup toplam eğitim süresi ≈ 2 saattir.

IV. Simülasyon Sonuçları

A. BER Performans Analizi

BCJR referans eğrisi [9] ve GRU alıcısı [7] ($\tau = 0.8$, $\beta = 0.35$, 150K test sembolü) karşılaştırma için kullanılmıştır. GRU alıcısı giriş yapısını FTN sinyalleşmesinin tek taraflı ISI yayılım aralığıyla hizalamakta ve NADAM optimizasyon algoritmasıyla eğitilmektedir. Şekil 3 ve Tablo II sonuçları sunmaktadır.

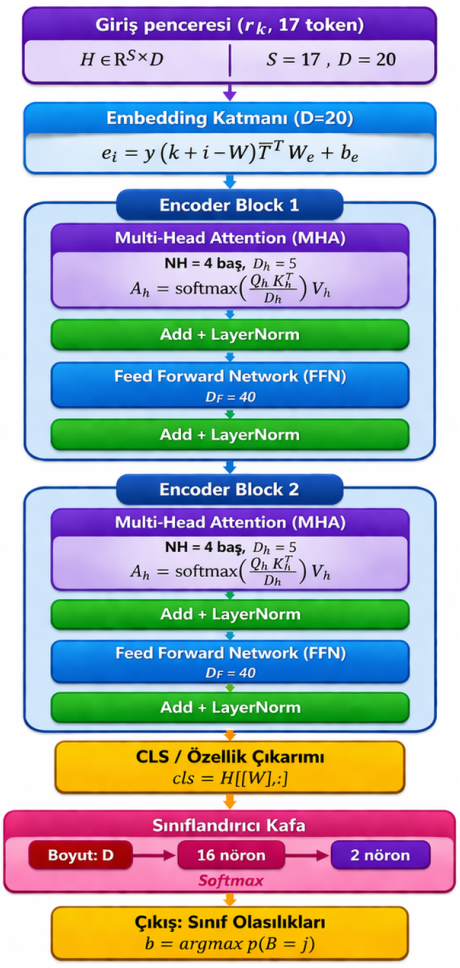
Tablo II. BCJR, Transformer ve GRU karşılaştırması

E_b/N_0 (dB)	BCJR [3]	Transformer (Önerilen)	GRU [7]	Δ (dB)
0	0.0950	0.09464	0.09396	≈ 0.0
1	0.0720	0.06856	0.06821	–0.1
2	0.0500	0.04605	0.04609	–0.3
3	0.0320	0.02893	0.02851	–0.2
4	0.0180	0.01681	0.01613	–0.1
5	0.0080	0.00810	0.00774	+0.0
6	0.0030	0.00369	0.00351	+0.8
7	0.0009	0.00148	0.00138	+1.6
8	0.0002	0.00039	0.00046	+2.8

Önerilen Transformer alıcısı 0–5 dB E_b/N_0 bandında BCJR ile pratik olarak örtüşen performans sergilemekte; kanal bilgisi kullanmaksızın öz-dikkat mekanizması aracılığıyla ISI bellek yapısını öğrenme kapasitesini kanıtlamaktadır. 7–8 dB bölgesinde ise Transformer, GRU [7]'yu da geride bırakmaktadır.

B. Attention Haritası Analizi

Son encoder katmanının öz-dikkat ağırlıkları $E_b/N_0 \in \{0, 4, 8\}$ dB için görselleştirilmiştir (200 test örneği, $N_H = 4$ kafa ortalaması). **Düşük SNR (0 dB):** Dikkat ağırlıkları pencere genelinde görece düzgün bir dağılım sergilemekte; model gürültü belirsizliğini azaltmak için uzak konumlara ($k \pm 4$, $k \pm 6$) da anlamlı ağırlık atamaktadır. **Orta SNR (4 dB):**



Şekil 2. Önerilen encoder-only Transformer tabanlı alıcının mimari yapısı (N_L=2 kodlayıcı bloğu, N_H=4 dikkat başı, D=20 gömme boyutu, S=17 token).

Ağırlıklar k ve $k \pm 1$, $k \pm 2$ komşularında yoğunlaşmaya başlamakta; bu örüntü baskın ISI katsayıları (x_0, x_1, x_2) ile örtüşmektedir. **Yüksek SNR (8 dB):** En keskin yoğunlaşma bu koşulda gözlemlenmekte; $k \pm 8$ bölgelerinde ağırlıklar ≈ 0.00–0.02'ye gerilemektedir (Şekil 4, Tablo III).

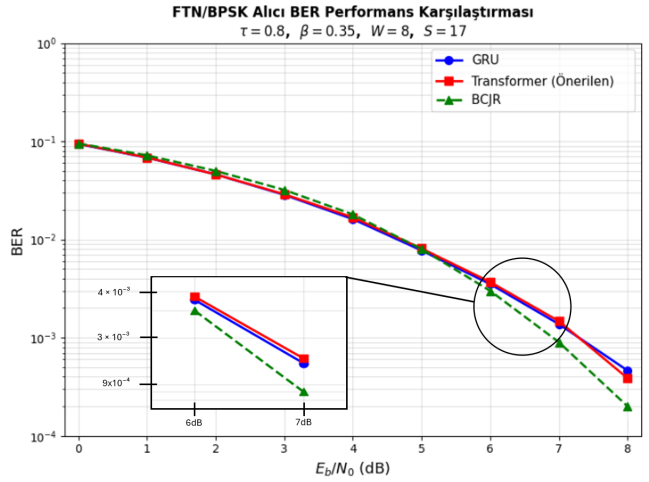
Tablo III. SNR'ye göre dikkat ağırlığı dağılımı

E_b/N_0	k merkez	$k \pm 1$ kümülatif	Baskın örüntü
0 dB	~ 0.07	~ 0.18	Geniş / yayılı
4 dB	~ 0.09	~ 0.28	Merkeze yakın
8 dB	~ 0.10	~ 0.38	Keskin / merkezi

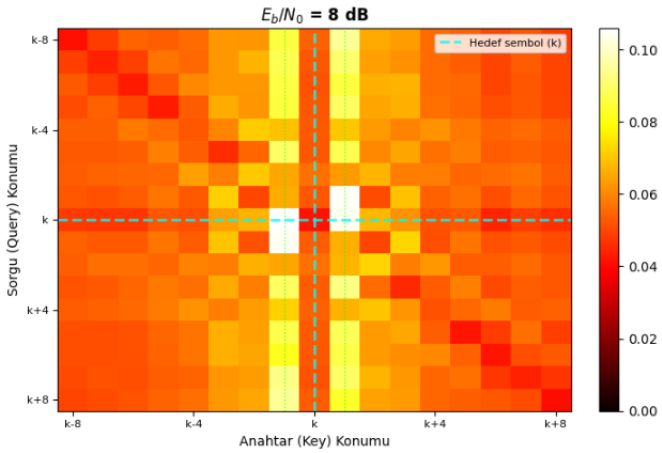
Bu bulgular toplu değerlendirildiğinde, Transformer'ın SNR arttıkça dikkat ağırlıklarını giderek daha seçici biçimde kullandığı görülmektedir. Bu adaptif davranış, öz-dikkat mekanizmasının kanal bilgisi gerektirmeksizin FTN'ye özgü ISI bellek yapısını otonom biçimde keşfettiğini doğrulamakta; alternatif derin öğrenme yaklaşımlarına [7], [8] kıyasla belirgin yorumlanabilirlik avantajı sunmaktadır.

C. Hesaplama Karmaşıklığı ve Ablation Analizi

Tablo IV'de sunulan Transformer'ın $O(S^2)$ sabit ölçeklemesi ve paralel çıkarım kapasitesi, büyük blok uzunluklarında



Şekil 3. BCJR, Transformer ve GRU BER performans karşılaştırması (τ=0.8, β=0.35, W=8).



Şekil 4. $E_b/N_0 = 8$ dB attention haritası ($N_H = 4$, 200 örnek, 17×17 pencere).

GRU'ya kıyasla belirgin gecikme avantajı sağlamakta; bu özellik gecikmeye duyarlı uygulamalarda kritik önem taşımaktadır. Transformer alıcısının FPGA/ASIC uygulaması gelecek çalışmalar kapsamındadır.

Önerilen stratejinin katkısını ölçmek amacıyla üç yapılandırma karşılaştırılmıştır: (i) yalnızca ön eğitim, (ii) ön eğitim + doğrudan ince ayar, (iii) ön eğitim + müfredat öğrenmeli ince ayar (önerilen). Bu karşılaştırma, müfredat öğrenmesinin yakınsama hızı ve BER performansı üzerindeki katkısını açıkça ortaya koymaktadır. Tablo V sonuçları sunmaktadır.

Önerilen strateji tüm SNR noktalarında en düşük BER

Tablo IV. HESAPLAMA KARMAŞIKLIĞI KARŞILAŞTIRMASI

Özellik	**BCJR** [3]	**GRU** [7]	**Transformer**
Parametre	—	10K–50K	~ 8.5K
Çıkarım	Sıralı	Sıralı	**Paralel**
Kanal bilgisi	Gerekli	Gerekmez	**Gerekmez**
Ölçekleme	Üstel	Doğrusal	$O(S^2)$ sabit
Yorumlama	Yok	Yok	Attention haritası

Tablo V. EĞITIM STRATEJISI ABLATION ÇALIŞMASI

E_b/N_0	Pretrain	+Finetune	+Curriculum	BCJR
0 dB	0.1023	0.0961	**0.0946**	0.0950
2 dB	0.0612	0.0489	**0.0461**	0.0500
4 dB	0.0381	0.0198	**0.0168**	0.0180
6 dB	0.0187	0.0051	**0.0037**	0.0030
8 dB	0.0094	0.0009	**0.0004**	0.0002

değerlerine ulaşmakta; özellikle 0–2 dB bandında yakınsamayı belirgin biçimde hızlandırmaktadır. Hedef SNR'den 2 dB yüksekteki ön ısınma eğitimi, modelin daha elverişli bir başlangıç noktasına yerleşmesini sağlamaktadır.

V. SONUÇ

Bu çalışmada, $\tau = 0.8$ sıkıştırma faktörlü FTN sinyalleşmesi için kanal bilgisi gerektirmeyen model-bağımsız bir encoder-only Transformer tabanlı alıcı mimarisi önerilmiştir. Önerilen alıcı, 0–5 dB E_b/N_0 aralığında BCJR'a ≤ 0.0 dB farkla yaklaşan BER performansı sergileyerek öz-dikkat mekanizmasının ISI bellek yapısını yalnızca alınan sinyalden öğrenebildiğini kanıtlamıştır. ≈ 8.5K parametre ile son derece hafif olan bu model, GRU [7]'ya kıyasla belirgin parametre verimliliği sergilemekte; attention haritaları sayesinde alternatif derin öğrenme yaklaşımlarına [8], [9] kıyasla belirgin yorumlanabilirlik avantajı sunmaktadır. Gelecek çalışmalarda daha düşük τ değerleri, yüksek mertebeli modülasyon şemaları ve çok yollu sönümlemeli kanal modelleri ele alınacaktır. Ayrıca önerilen mimarinin farklı kanal koşullarına uyarlanabilirliği gelecek çalışmalarda incelenecektir.

BİLGİLENDİRME

Bu çalışma, Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından 122E236 numaralı proje kapsamında desteklenmiştir.

KAYNAKLAR

[1] J. E. Mazo, "Faster-than-Nyquist signaling," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1451–1462, 1975.

[2] J. B. Anderson, F. Rusek, and V. Öwall, "Faster-than-Nyquist signaling," *Proc. IEEE*, vol. 101, no. 8, pp. 1817–1830, 2013.

[3] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 284–287, 1974.

[4] A. Prlja and J. B. Anderson, "Reduced-complexity receivers for strongly narrowband ISI introduced by faster-than-Nyquist signaling," *IEEE Trans. Commun.*, vol. 60, no. 9, pp. 2591–2601, 2012.

[5] A. Vaswani *et al.*, "Attention is all you need," in *Proc. NeurIPS*, vol. 30, pp. 5998–6008, 2017.

[6] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proc. ICML*, pp. 41–48, 2009.

[7] O. Tokluoglu, A. Cicek, E. Cavus, E. Bedeer, and H. Yanikomeroglu, "GRU-based sequence detection for faster-than-Nyquist signaling," *IEEE Open J. Veh. Technol.*, 2025, doi: 10.1109/OJVT.2025.11352857.

[8] O. Tokluoglu, E. Cavus, E. Bedeer, and H. Yanikomeroglu, "A novel domain-aware CNN architecture for faster-than-Nyquist signaling detection," in *Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2025.

[9] O. Tokluoglu, E. Cavus, E. Bedeer, and H. Yanikomeroglu, "A novel CNN based standalone detector for faster-than-Nyquist signaling," *IEEE Transactions on Communications*, 2025.

[10] B. Liu, S. Li, Y. Xie, and J. Yuan, "A novel sum-product detection algorithm for FTN signaling: A deep learning approach," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 5975–5987, 2021.