

# MATLAB Paralel Hesaplama Araç Kutusu ile Shannon Entropi Hesaplanması

\*<sup>1</sup>Sezgin Kaçar, <sup>2</sup>Ziya Ekşi, <sup>3</sup>Akif Akgül, <sup>4</sup>Fahrettin Horasan

\*<sup>1,3</sup>Elektrik – Elektronik Mühendisliği Bölümü, Teknoloji Fakültesi, Sakarya Üniversitesi, Türkiye

\*<sup>2,4</sup>Bilgisayar Mühendisliği Bölümü, Teknoloji Fakültesi, Sakarya Üniversitesi, Türkiye

## Özet:

Bu çalışmada MATLAB paralel hesaplama araç kutusu tanıtılmış ve örnek uygulama olarak en basit şekliyle paralel-for (parfor) döngüsü ile Shannon Entropi hesaplama algoritması paralelleştirilmiştir. Seri ve paralel hesaplamalar yapılarak sonuçlar elde edilmiştir. Elde edilen sonuçlar kullanılarak hızlanma ve verimlilik kriterleri temelinde paralel hesaplamanın performansı değerlendirilmiştir.

**Anahtar Kelimeler:** MATLAB, Paralel Hesaplama, Entropi

## Computation of Shannon Entropy with MATLAB Parallel Computing Toolbox

\*<sup>1</sup>Sezgin Kaçar, <sup>2</sup>Ziya Ekşi, <sup>3</sup>Akif Akgül, <sup>4</sup>Fahrettin Horasan

\*<sup>1</sup>Department of Electrical – Electronics Engineering, Technology Faculty, Sakarya University, Turkey

\*<sup>1</sup>Department of Computer Engineering, Technology Faculty, Sakarya University, Turkey

## Abstract:

In this study, MATLAB parallel computing toolbox has been introduced and for a sample application, Shannon Entropy algorithm has been simply parallelized by using parallel-for (parfor) loop. Sequential and parallel computations have been performed and the results have been obtained. By using the results, the performance of parallel computing has been evaluated based on speedup and efficiency criteria.

**Keywords:** MATLAB, Parallel Computing, Entropi

## 1. Giriş

Bugün masaüstü veya dizüstü kişisel bilgisayarlardan akıllı telefonlara kadar bir çok cihazda yüksek performans ihtiyacının karşılanması için paralel işlem yapabilen çok çekirdekli işlemciler kullanılmaktadır. Buna karşın ülkemizde paralel hesaplama teknikleri ve araçlarının kullanımı yeteri kadar yaygınlaşmamıştır. Bu durum kullanıcıların ellerindeki yüksek işlem kapasitesinin tam olarak kullanılmadığını göstermektedir. Bu sebeple paralel hesaplama tekniklerinin daha yaygın olarak kullanımı için bu alandaki çalışmaların sayısı arttırılmalıdır. Örnek uygulamalar ile bilim insanlarının kendi alanlarında paralel hesaplamayı kullanmaları için yeni anlayışlar geliştirilmesine yardımcı olunmalıdır.

Bilimsel ve teknik hesaplamalar konusunda tüm dünyada çok yaygın biçimde kullanılan MATLAB programı da paralel hesaplamalar için bir araç kutusuna sahiptir. Bu araç kutusu ile MATLAB programında karmaşık ve yüklü hesaplamalar için uygun donanımlar ile yüksek performanslı paralel hesaplama gerçekleştirilebilmektedir. Bu durum göz önüne alınarak, sunulan çalışmada öncelikle MATLAB paralel hesaplama araç kutusu tanıtılmıştır. MATLAB programının bir paralel hesaplama aracı olarak performansının görülmesi için örnek bir paralel hesaplama işlemi yapılmıştır. Bunun için Shannon Entropi hesaplama metodu örnek bir uygulama olması açısından paralel-*for* döngüsü ile paralelleştirilerek sonuçlar 8 çekirdekli CPU'ya sahip birbilgisayarda elde edilmiştir. Elde edilen sonuçlar kullanılarak yapılan işlemin performansı değerlendirilmiştir.

Çalışmanın ikinci bölümünde MATLAB paralel hesaplama araç kutusu tanıtılmıştır. Üçüncü bölümde entropi işlemi ile ilgili bilgi verilmiş ve Shannon Entropi metodunun matematiksel alt yapısı sunulmuştur. Dördüncü bölümde metodun seri ve paralel hesaplama sonucunda elde edilen zaman değerleri tablo ve grafikler halinde sunulmuştur. Ayrıca sonuçlar kullanılarak hızlanma ve verimlilik değerleri elde edilmiştir. Son bölümde ise elde edilen sonuçlar ile ilgili değerlendirme ve yorumlara yer verilmiştir.

## 2. MATLAB Paralel Hesaplama Araç Kutusu

Bilindiği üzere bilimsel ve teknik hesaplamalar çoğu zaman yüksek bir işlem kapasitesi gerektirmektedir. Bu nedenle paralel hesaplama teknikleri günümüzde oldukça yaygın biçimde kullanılmaktadır. MATLAB programı çok sayıda araç kutusu ile neredeyse tüm mühendislik alanlarında kendisine kullanım alanı bulmaktadır. Bu bağlamda MATLAB programı kullanıcılarının paralel hesaplama yapabilmeleri için MATLAB paralel hesaplama araç kutusu geliştirilmiştir.

MATLAB paralel hesaplama araç kutusunu kullanarak çok çekirdekli işlemcilerle, grafik işlemci birimleri (GPU) ve bilgisayar kümeleriyle paralel hesaplama yapmak mümkündür. Bu araç kutusunun sağladığı yüksek seviyeli yapılar olan paralel-*for* (*parfor*) döngüleri, özel dizi tipleri ve paralelleştirilmiş sayısal algoritmaları kullanarak CUDA veya MPI programlamaya gerek duymadan paralel hesaplama uygulamaları geliştirilebilir[1].

MATLAB paralel hesaplama araç kutusunun temel özellikleri arasında; *parfor* ile görev-paralel algoritmaların çok çekirdekli işlemcilerde çalıştırılması, GPU programlama desteği, çok çekirdekli bilgisayarlar için on iki adet yerel işçi ile çalışabilme imkanı, bilgisayar kümeleri ile çalışma desteği, geniş veri setleri ve veri-paralel algoritmalar için ayrık diziler sayılabilir[1].

## 3. Entropi İşlemi

Entropi termodinamik ve bilgi teorisi alanlarında farklı anlamlar ifade eden bir kavramdır. Bilgi teorisinde entropi kavramı bir zaman serisinin düzensizliğini belirlemek için kullanılır. Elde edilmiş bir veri setinin entropi değeri ne kadar büyükse düzensizliğinin o oranda fazla olduğu söylenir [2,3]. Entropi değeri biyomedikal alanı başta olmak üzere farklı sinyal işleme uygulamalarında da kullanılmaktadır.

Entropi değerinin hesaplanması için kullanılan birçok yöntem mevcuttur. İlk olarak Shannon tarafından ortaya konan entropi hesaplama metodundan sonra Norm, Eşik, Logaritmik gibi entropi hesaplama metodları da geliştirilmiştir [3]. Ayrıca zaman serilerinde entropi hesaplanması için geliştirilmiş Sample entropi ve Approximate entropi gibi iki farklı entropi metodu da bulunmaktadır [4,5]. Bu çalışmada ilk ve en temel entropi hesaplama yöntemi olan Shannon metodu üzerine bir uygulama yapılmıştır. Aşağıda Shannon metodunun matematiksel alt yapısını veren formül görülmektedir [6].

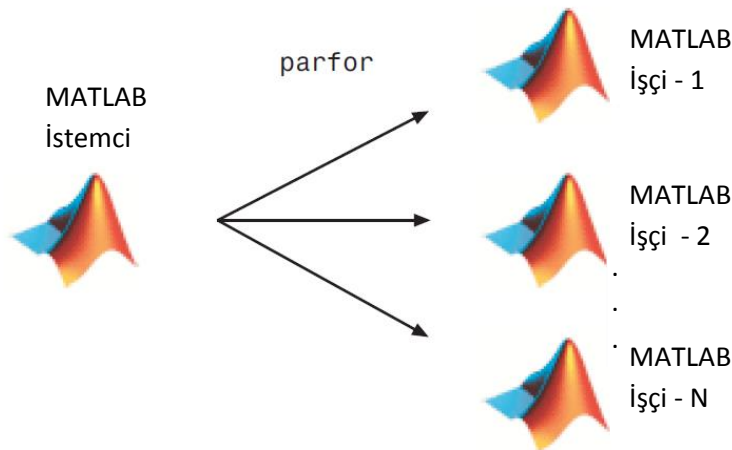
$$ShanEn(x) = - \sum_{i=1}^N (p_i(x))^2 (\log_2(p_i(x)))^2 \quad (1)$$

Eşitlik 1’de görülen  $p_i(x)$  veri setinde  $i$ . sıradaki olasılık kütle fonksiyonu değeri,  $N$  ise olasılık kütle fonksiyonu değerlerinin sayısıdır.

#### 4. Shannon Entropi Metodunun Parfor Döngüsüyle Paralleleştirilmesi

##### 4.1. MATLAB’da parfor döngüsü

MATLAB’da *parfor* (paralel-for) döngüsü, çok çekirdekli işlemciye sahip bilgisayarlarda döngü içerisine yazılan kodları her bir işlemci eşzamanlı kullanarak işlemek amacıyla oluşturulmuş paralel programlama bir yapısıdır. Kullanım olarak normal *for* döngüsünden farklı olmayan bu yapının çalışma mantığı Şekil 1’deki gibidir. *Parfor* döngüsünün içine yazılan kodların her bir çalışmasına iterasyon adı verilir. *Parfor* döngüsünün bulunduğu kodu çalıştıran MATLAB oturumu istemci (Client) olarak, iterasyonların dağıtıldığı, her biri belirli sayıda iterasyonu işleyen ve ayrı çekirdeklerde çalışan MATLAB oturumları ise işçi olarak adlandırılır[1]. Bu yapıyla 12 işçiye kadar paralel işlem yapmak mümkündür.



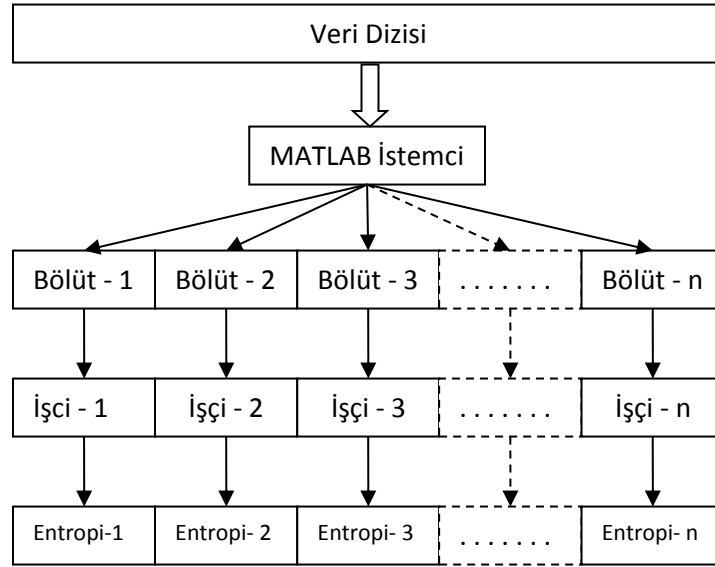
Şekil 1.Parfor ile işlemlerin paralelleştirilmesi[1]

Paralel işlem yapılırken işçiler iterasyonları sıralı olarak takip etmezler. Bu nedenle *parfor*

döngüsü ile paralel işlenecek iterasyonların birbirleriyle etkileşimli olmaması gerekir. Örneğin bir önceki iterasyonun sonucu bir sonraki iterasyonda kullanılıyorsa bu işlem *parfor* ile gerçekleştirilemez. Bu açıdan bakıldığında Eşitlik 1’de verilen Shannon entropi formülü *parfor* ile paralel işlenebilecek bir yapıda olduğu görülmektedir.

#### 4.2. Shannon entropi metodunun paralelleştirilmesi

Bilgi teorisi alanında yapılan uygulamalarda genellikle eldeki verilerin tümünün bir seferde işlenmesi yerine belirli bir parçası üzerinde işlem yapılması tercih edilir. Bu işleme ise bölütleme adı verilir. Örneğin bir ses kaydında konuşma olan ve olmayan bölümler bölütlenerek üzerlerinde işlem yapılan çalışmalar mevcuttur [7]. Entropi işlemi de bilgi teorisi alanında yaygın kullanılan bir işlem olarak verilerin bölütler halinde değerlendirilmesinde kullanılmaktadır. Bu çalışmada, entropi işleminin paralel olarak gerçekleştirilmesi esnasında bölütlere ayrılmış verilerin işlenmesi üzerine çalışma yapılmıştır (Şekil 2).



Şekil 2. Shannon entropi işleminin bölütler halinde paralelleştirilmesi

Sunulan çalışma herhangi bir amaçla elde edilmiş verilerin işlenmesinden çok eldeki verilerin bölütler halinde paralel olarak işlenmesinde performansın artırılmasını ele aldığından, çalışmada rastgele üretilmiş  $10^6$  adet veriden oluşan bir dizi kullanılmıştır. Çalışma Tablo 1’de özellikleri verilen bilgisayar kullanılarak gerçekleştirilmiştir.

Tablo 1. Bilgisayar özellikleri

İşlemci Modeli	Intel(R) Core(TM) i7 CPU 950
Çekirdek Sayısı	Gerçek 4 – Sanal 8
İşlemci Frekansı	3.07 GHz

Ram Miktarı	4.00 GB
İşletim Sistemi	Win 7 64 Bit

Paralel olarak entropi hesaplama işlemi iki farklı şekilde ele alınmıştır. Birinci uygulamada bölütlerdeki veri uzunluğu 1000 adet ile sabit tutulup tüm veri setindeki eleman sayıları ( $10^4$ ,  $10^5$ ,  $10^6$ ) değiştirilerek bölütlere ait entropi değerleri elde edilmiştir. İkinci uygulamada ise toplam veri uzunluğu  $10^6$  olarak sabit tutulmuş ve bölüt uzunlukları ( $10^2$ ,  $10^3$ ,  $10^4$ ) değiştirilmiştir. Her iki uygulama da 1'den çekirdekten 8 çekirdeğe kadar işlemcinin farklı kapasitelerde çalıştırılması ile gerçekleştirilmiştir. Her iki uygulamada gerçekleştirilen işlemler için harcanan zamanlar Tablo 2 ve 3'deki gibi kaydedilmiştir.

**Tablo 2.**Bölüt uzunluğu sabit ( $10^3$ ) iken harcanan zamanlar (Saniye)

		Kullanılan Çekirdek Sayısı							
		1 (seri)	2	3	4	5	6	7	8
Veri Uzunluğu	$10^4$	0.1150	0.0734	0.0637	0.0682	0.0735	0.0813	0.0924	0.0950
	$10^5$	0.7330	0.4022	0.2822	0.2441	0.2484	0.2613	0.2512	0.2589
	$10^6$	6.9790	3.7785	2.5311	2.0150	2.0444	2.0577	2.1030	2.1269

**Tablo 3.**Veri uzunluğu sabit ( $10^6$ ) iken harcanan zamanlar (Saniye)

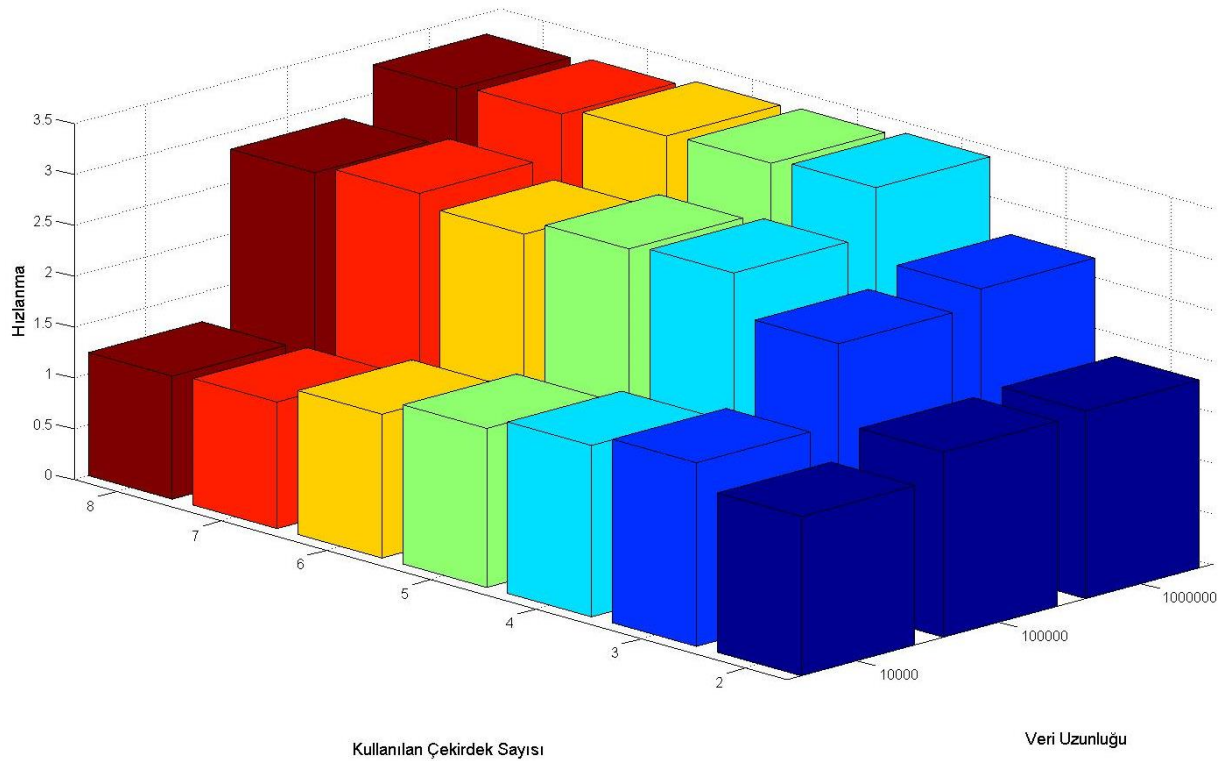
		Kullanılan Çekirdek Sayısı							
		1 (seri)	2	3	4	5	6	7	8
Veri Uzunluğu	$10^2$	3.4183	1.8256	1.3338	1.1372	1.1792	1.2081	1.2515	1.2744
	$10^3$	6.8910	3.5428	2.5232	1.9567	2.1647	1.9049	1.9215	1.9177
	$10^4$	49.0354	24.9664	17.7908	13.1236	13.1648	13.3447	13.8867	14.5092

Paralel hesaplama işlemlerinde performans değerlendirmesi genellikle hızlanma ve verimlilik olmak üzere iki kriter kullanılarak gerçekleştirilir[8]. Paralel hesaplama sonucunda elde edilen zaman değerleri Eşitlik 2 ve 3 kullanılarak gerçekleştirilen paralel hesaplamanın performans değerleri elde edilmiştir. Elde edilen performans değerleri Şekil 3, 4, 5 ve 6'da grafiksel olarak sunulmuştur.

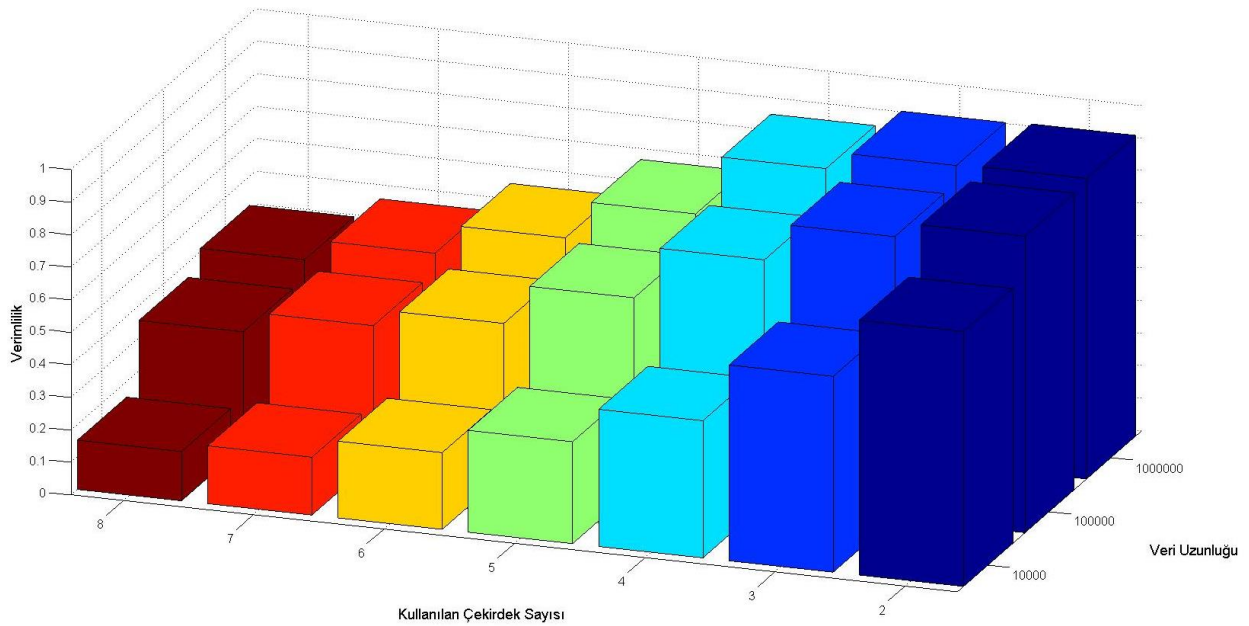
$$S_p(n) = \frac{T^*(n)}{T_p(n)} \quad (2)$$

$$E_p(n) = \frac{S_p(n)}{p} \quad (3)$$

Burada  $S_p(n)$  p adet işlemcinin kullanıldığı işlemde hızlanmayı,  $T^*(n)$  tek işlemci ile işlem süresini,  $T_p(n)$  p adet işlemcinin kullanıldığı işlemde işlem süresini,  $E_p(n)$  p adet işlemcinin kullanıldığı işlemde verimi ve n ise işlenen veri sayısını ifade etmektedir.



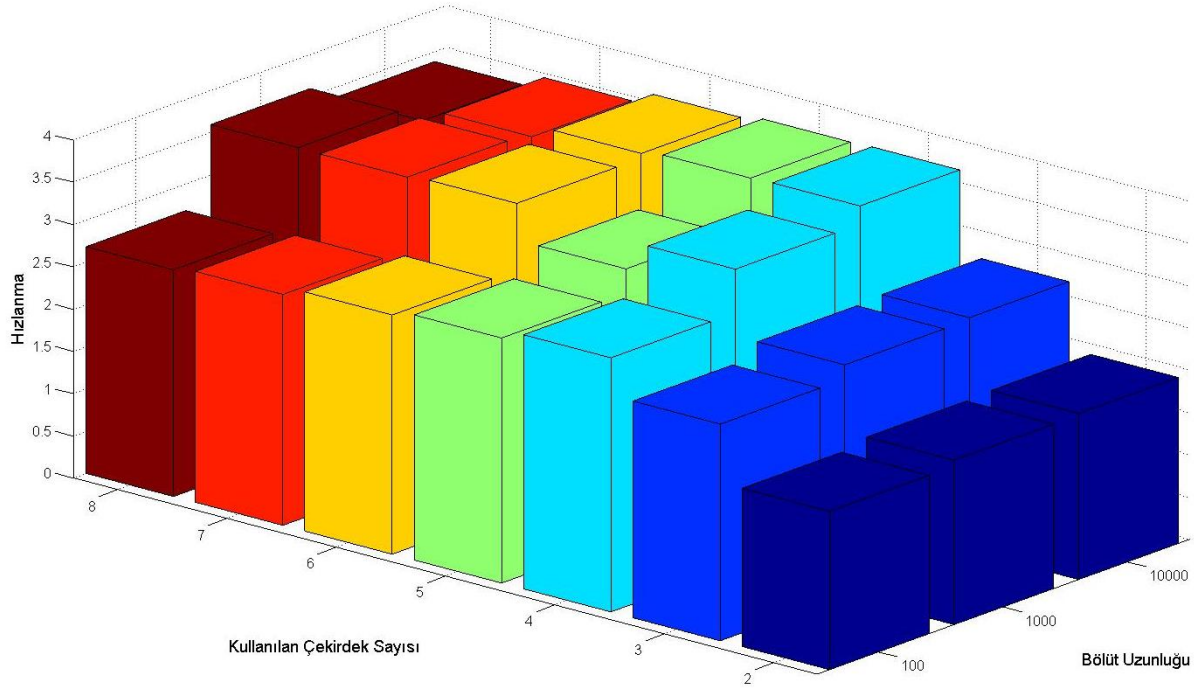
Şekil 3.Tablo 2'deki değerlerden elde edilen hızlanma grafiği



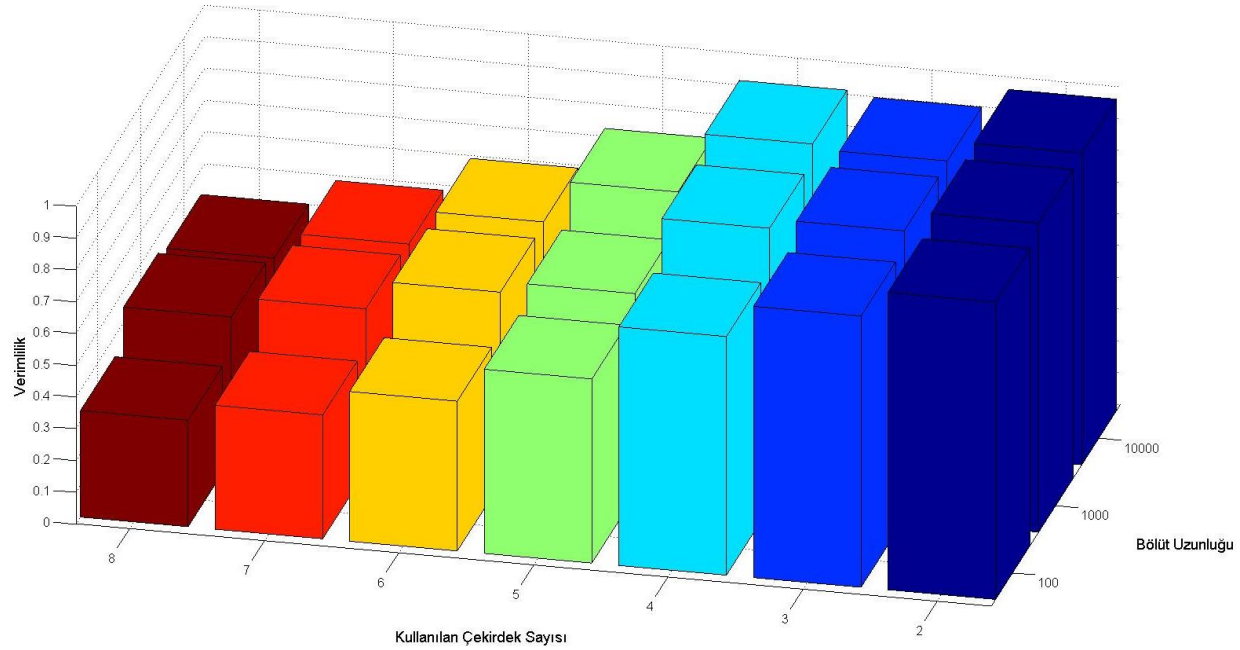
Şekil 4.Tablo 2'deki değerlerden elde edilen verimlilik grafiği

Şekil 3 ve 4'deki grafikler bölüt uzunluğu sabit veri uzunluğu değişken olduğunda elde edilmiştir. Bu grafikler incelendiğinde veri uzunluğu arttıkça hızlanma ve verimliliğin arttığı

görülmektedir. Buna karşın kullanılan çekirdek sayısına bakıldığında hızlanma grafiği dört çekirdekten sonra değişmemekte ve verimlilik değerleri kullanılan çekirdek sayısı arttıkça azalmaktadır.



Şekil 5.Tablo 3'deki değerlerden elde edilen hızlanma grafiği



Şekil 6.Tablo 3'deki değerlerden elde edilen verimlilik grafiği

Şekil 5 ve 6'daki grafikler veri uzunluğu sabit bölüt uzunlukları değişken olduğunda elde edilmiştir. Bu grafiklerde görüldüğü üzere bölüt uzunlukları değiştiğinde 2 ve 3 çekirdek için hızlanma ve verimlilik çok fazla değişmemiştir. 4 ve daha fazla çekirdek kullanıldığında ise özellikle bölüt sayısının 1000 olduğu değerler daha hızlı ve verimli olarak görülmektedir. Bununla birlikte bu sonuçlarda da çekirdek sayısı arttıkça verimliliğin düştüğü gözlemlenmektedir.

## 5. Sonuçlar

Sunulan bu çalışmada MATLAB paralel hesaplama araç kutusunun kullanıldığı bir örnek uygulama gerçekleştirilmiş ve sonuçlar elde edilmiştir. Uygulama konusu olarak özellikle biyomedikal sinyal işleme alanında yaygın olarak kullanılan bir analiz metodu olan Shannon entropi hesaplama yöntemi ele alınmıştır. Ele alınan yöntem MATLAB programında çok çekirdekli CPU'ya sahip bilgisayarlarda paralel hesaplama yapmak için kullanılan temel yapılardan bir tanesi olan *parfor* döngüsü ile paralelleştirilerek farklı çekirdek sayıları, veri ve bölüt uzunlukları için sonuçlar elde edilmiştir. Elde edilen sonuçlar hızlanma ve verimlilik grafiklerine dönüştürülerek gerçekleştirilen paralel hesaplamanın performansı değerlendirilmiştir.

Elde edilen grafiksel sonuçlar incelendiğinde Shannon entropi metodunun paralel olarak işlenmesinin uzun veri dizileri için daha iyi sonuçlar verdiği, daha fazla hızlanma ve verimlilik sağlandığı görülmüştür. Bölüt uzunluğunun artması da hızlanma ve verimlilik açısından daha iyi sonuçlar vermiştir. Bununla birlikte kullanılan çekirdek sayısının artması 4 çekirdeğe kadar oldukça iyi sonuçlar vermesine karşın daha da arttırıldığında verimliliğin önemli ölçüde azalmasına neden olmuştur. Tüm bu söylenenler ışığında Shannon entropi metodunun uzun veri dizileri ve bölütleme uzunluğu büyük olan durumlar için sunulan çalışmadaki gibi paralelleştirilmesinin daha uygun olduğu sonucuna varılabilir.

## Kaynaklar

- [1] Matlab Parallel Computing Toolbox User's Guide R2012a, Mathworks Inc.
- [2] Blesy Anu Mathew. Entropy Of Electroencephalogram (Eeg) Signals Changes With Sleep State. University of Kentucky Master's Thesis; 2006.
- [3] Yıldız A, Akın M, Poyraz O, Kirbas G. Estimation of Alertness Level by Using Wavelet Transform Method and Entropy. Signal Processing and Communications Applications Conference. p. 313–316. SIU 2009. IEEE 17th, Turkey; 2009.
- [4] Gómez C, Hornero R. Entropy and Complexity Analyses in Alzheimer's Disease: An MEG Study. The Open Biomedical Engineering Journal. p. 223-235; 2010.
- [5] Joshua S, Richman J, Randall M. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol. 278: p. H2039–H2049; 2000.



- [6] Aydın S, Saraoğlu HM, Kara S. Epilepside EEG Tabanlı Entropi Değişimleri. VI. Ulusal Tıp Bilişimi Kongresi Bildirileri. p. 270-275. ENMI Vol V No 1; 2009.
- [7] Danışman T, Alpkoçak A. Speech vs Nonspeech Segmentation of Audio Signals Using Support Vector Machines. Signal Processing and Communications Applications, 2007. SIU 2007. IEEE 15th, p. 1-4, Turkey; 2007.
- [8] Rauber T, Rünger G. Paralel Programming for Multicore and Cluster Systems. New York: Springer; 2007.