# ADVANCED AND PIONEER RESEARCH IN NATURAL SCIENCE AND MATHEMATICS



All Sciences Academy

# *ADVANCED AND PIONEER RESEARCH IN NATURAL SCIENCE AND MATHEMATICS*

**Editor**
**Prof. Dr. Canan Özdemir**

***Advanced and Pioneer Research in Natural Science and Mathematics***
*Editor: Prof. Dr. Canan Özdemir*

**Design:** All Sciences Academy Design
**Published Date:** October 2026
**Publisher's Certification Number:** 72273
**ISBN:** 978-625-5472-40-3

© All Sciences Academy

www.allsciencesacademy.com
allsciencesacademy@gmail.com

## CONTENT

**1. Chapter** **6**

Magnetic Properties of a Quantum Dot Confined by a Gaussian Potential

*Adem DÖNMEZ, Uğur ERKARSLAN*

**2. Chapter** **20**

Scalable Synthesis of High-Quality Graphene Using Chemical Vapor Deposition: Fundamental Principles and Characterization

*Özge ATEŞİN ÜNLÜ, Uğur ERKARSLAN, Adem DÖNMEZ*

**3. Chapter** **41**

Dual Split Quaternions And Spacelike, Timelike, Null Dual Split Euler Formulas For Spacelike, Timelike, Null Screw Movements

*Aysel TURGUT VANLI*

**4. Chapter** **60**

Comparing the Performance of Restricted Least Squares and Ridge Estimators under Multicollinearity: A Monte Carlo Simulation Study

*Berrin ERKAN*

**5. Chapter** **77**

Classification of Center-Excluding Complete (k,4)-Arc Extensions Containing a Prescribed Ceva 6-Figure in PG(2,4)

*Elif ALTINTAŞ KAHRİMAN*

**6. Chapter** **97**

Construction and Classification of Complete (k,4)-Arcs Containing a Ceva 6-Figure and its Center in PG(2,4)

*Elif ALTINTAŞ KAHRİMAN*

**7. Chapter** **122**

Sol-Gel Processing and Electrospinning Technique: A Comparative Review of Advanced Methods in Nanomaterial Production

*Fatih SEVİM, Ömer LAÇİN, Fatih DEMİR, Semet ÇELİK*

**8. Chapter** **153**
Experimental Methodology in Magnetic Resonance
*Mehmet Zafer KÖYLÜ*

**9. Chapter** **168**
Conducting Polymer Effects in Supercapacitors
*Murat ATES*

**10. Chapter** **178**
Maximum Likelihood Estimation of Stress–Strength Reliability for Series and Parallel Systems under Lomax and Power Function Distributions
*Sevcan DEMİR ATALAY, Ruslan SHUKUR*

**11. Chapter** **195**
Comparative Histomorphology and Ultrastructural Organization of the Digestive System in Selected Vector Mosquito Larvae
*Sıla DAŞ, Yücel BAŞIMOĞLU KOCA*

# Magnetic Properties of a Quantum Dot Confined by a Gaussian Potential

**Adem DÖNMEZ**[1]
**Uğur ERKARSLAN**[2]

1. Assoc. Prof. Dr Adem Donmez - Department of Electricity and Energy, Yatagan Vocational School, Mugla Sıtkı Koçman University, Mugla, Turkey. adonmez@mu.edu.tr ORCID No: 0000-0002-9773-0493
2. Prof. Dr Ugur Erkarslan - Department of Physics, Molecular Nano-Materials Laboratory, Mugla Sıtkı Kocman University, Mugla, Turkey. eugur@mu.edu.tr ORCID No: 0000-0001-7667-2151

## ABTRACT

System of particles deep confined in 3-dimensional quantum mechanichs is called quantum dot. Due to atomic properties of this system is also called as artificial atom and has various technical applications like televisons and quantum computers. Understanding of the physical properties of these systems will enable expansion and development of the fields of application. The main factor that determine the physical properties of the quantum dots is the deep confining potential. Gaussian potential stands out as the most realistic model among all confinement potentials in perspective of physical aspects. As Gaussian confinement has been an important model being finite, to understand physical properties of quantum dots such as tunneling and ioanization.

In this chapter, a model will be established model to study the magnetic properties in case of restricted quantum dots placed in a magnetic field with Gaussian potential. This model will be solved using the approximation theory and eigenvalues and eigenvectors will be obtained. By using the obtained results, magnetic sustainability of the point, the temperature dependence of magnetization and magnetic sustainability will be shown.



*Keywords – Low dimensional systems, quantum dot, Gaussian confinement*


## INTRODUCTION

Low-dimensional quantum systems arise when the motion of particles is spatially confined, leading to pronounced quantum mechanical effects. When carriers are confined in all three spatial dimensions, the resulting structure is referred to as a quantum dot, a zero-dimensional system characterized by discrete energy levels analogous to those of natural atoms (Kastner, 1992; Kouwenhoven et al., 2001). These systems exhibit quantization effects governed by confinement, making them ideal platforms for exploring both fundamental and applied aspects of quantum mechanics.

The theoretical foundations of quantum confinement can be traced back to early developments in quantum mechanics and statistical physics (Landau & Lifshitz, 1975), as well as to pioneering studies on confined electron systems and artificial atoms (Maksym & Chakraborty, 1990; Ashoori, 1996). Subsequent experimental breakthroughs demonstrated that semiconductor heterostructures could be engineered to realize such confined systems (Gornik et al., 1985; Jacak et al., 1998; Bimberg et al., 1999). These developments established quantum dots as a cornerstone of modern condensed matter physics and nanotechnology.

Over the past decades—and particularly in recent years—quantum dots have gained renewed attention due to their tunable physical properties and technological relevance. Their size, shape, and confinement strength can be precisely controlled, allowing for tailored electronic and optical responses. This tunability has enabled applications in quantum information processing, optoelectronics, and nanoscale devices (Heindel et al., 2023; Zubchenko et al., 2025). Furthermore, the field has been revitalized by the recognition of quantum dots in the 2023 Nobel Prize in Chemistry, emphasizing their impact on modern nanoscience (Wegner, 2024).

From a theoretical standpoint, the physical properties of quantum dots are strongly influenced by the nature of the confinement potential. Early studies frequently employed idealized models such as infinite square wells or parabolic potentials due to their analytical tractability (Chatterjee, 1985; Cohen, 1984; Lai, 1983). The parabolic confinement model, in particular, gained prominence due to its compatibility with the generalized Kohn theorem (Johnson, 1995; De Groote et al., 1992). However, subsequent investigations revealed that realistic confinement potentials deviate significantly from purely harmonic behavior, necessitating more sophisticated models (Oh et al., 1994; Voskoboynikov et al., 2003).

In this context, finite-depth potentials have emerged as more realistic representations of quantum dot systems, particularly for describing physical processes such as tunneling, ionization, and excitation. Among these, the Gaussian confinement potential has attracted considerable attention due to its physically meaningful characteristics. Unlike abrupt square well potentials, Gaussian potentials exhibit smooth spatial variations, which facilitate both analytical and numerical treatments (Bessis et al., 1982; Crandall, 1983). Additionally, near the center of the quantum dot, the Gaussian potential approximates a parabolic form, preserving key features of harmonic confinement while incorporating finite-depth effects.

Extensive theoretical work has been conducted to investigate quantum dots under Gaussian confinement. Studies by Adamowski et al. (2000) demonstrated the importance of Gaussian potentials in describing excess electron states, while later works explored thermodynamic and magnetic properties within this framework (Mukhopadhyay & Chatterjee, 2001; Nguyen & Peeters, 2008). In particular, the contributions of Boyacıoğlu and co-workers have provided detailed insights into magnetization, magnetic susceptibility, and thermodynamic properties of Gaussian-confined quantum dots (Boyacıoğlu et al., 2007; Boyacıoğlu & Chatterjee, 2012a, 2012b). These studies highlight the sensitivity of quantum dot properties to confinement parameters and external fields.

Recent advances further extend these classical models by incorporating effects such as electron correlation, spin–orbit coupling, and external perturbations. Contemporary studies emphasize the importance of realistic

confinement potentials in accurately describing nanoscale systems and quantum devices (Sargsian et al., 2023; Kachu et al., 2025; Kurmi et al., 2025). These developments reinforce the relevance of Gaussian potentials as a bridge between analytically tractable models and experimentally realizable systems.

In addition to their theoretical importance, quantum dots continue to serve as versatile platforms for experimental investigations. Optical and electronic properties of semiconductor quantum dots have been extensively studied (Woggon, 1997), and their fabrication has been refined through advanced epitaxial growth techniques (Alauwaji et al., 2024). These systems provide unique opportunities to test quantum mechanical predictions and explore novel physical phenomena at the nanoscale.

In this chapter, we investigate the magnetic properties of a single-electron quantum dot confined by a Gaussian potential. The analysis is based on established theoretical frameworks and focuses on the dependence of key physical quantities—such as magnetization and magnetic susceptibility—on temperature, magnetic field strength, and structural parameters including confinement width and depth.

The chapter is organized as follows: first, a comprehensive review of the literature is presented. This is followed by an introduction to low-dimensional systems and the theoretical framework, including the Fock–Darwin states. The model and computational methodology are then described in detail. Subsequently, numerical results are presented and discussed. Finally, the chapter concludes with a detailed interpretation of the physical implications and potential future research directions.

## MATERIALS AND METHODS

### *Modeling of a Gaussian-Confined Quantum Dot*

As discussed in the previous sections, low-dimensional semiconductor systems can be realized by stacking lattice-matched materials with different band structures, leading to the formation of two-dimensional quantum wells. When appropriate contacts are introduced, the carriers within these structures can be spatially confined into a finite region, effectively forming a quantum dot. Experimental and theoretical studies have demonstrated that such confinement potentials are finite in nature and, in many cases, can be well approximated by a Gaussian-type potential (Atcı, 2013).

In this work, the problem is treated as a single-particle system confined in two dimensions by a Gaussian potential, effectively corresponding to a zero-dimensional quantum dot. A semi-classical framework is adopted, following the theoretical approaches developed in the literature, particularly those introduced by Boyacıoğlu and co-workers.

The analysis begins with the Hamiltonian of a charged particle confined in two dimensions under the influence of an external magnetic field. The general form of the Hamiltonian is given by:

$$H = -\frac{1}{2m^*}\left(p + \frac{e}{c}A\right)^2 + V(\rho) \tag{1}$$

where $\rho$ represents the two-dimensional position operator, p is the momentum operator, $m^*$ is the effective mass, and A is the vector potential corresponding to a magnetic field applied along the z-direction. The symmetric gauge is chosen as: $A=(-B_y/2,+B_x/2,0)$

Expanding the kinetic term yields:

$$\left(\vec{p} + \frac{e}{c}\vec{A}\right)^2 = p^2 + \frac{e}{c}\vec{p}\cdot\vec{A} + \frac{e}{c}\vec{A}\cdot\vec{p} + \frac{e^2}{c^2}A^2 \tag{2}$$

Substituting this into the Hamiltonian leads to:

$$H = -\frac{\hbar^2}{2m^*}\nabla_\rho^2 + V(\rho) + \frac{1}{8}m^*\omega_c^2\rho^2 + \frac{1}{2}\hbar\omega_c\hat{L}_z \tag{3}$$

where $\omega_c = eB/m^*$ is the cyclotron frequency. The confinement potential is taken as a Gaussian function:

$$V(\rho) = -V_0 e^{-\frac{\rho^2}{2R^2}} \tag{4}$$

Here, $V_0$ denotes the depth of the potential and R represents its characteristic width, which effectively determines the size of the quantum dot. By introducing the effective frequency parameter $\tilde{\omega}^2 = \frac{1}{4}\omega_c^2 + \omega_h^2$, the Hamiltonian can be rearranged into a more convenient form. Near the center of the quantum dot ($\rho\approx0$), the Gaussian potential can be approximated by a parabolic potential with a small perturbative correction. This approximation is justified since the confinement radius R is typically small in realistic quantum dot systems.

Under this assumption, the Hamiltonian can be decomposed into two parts:

$$H = H_0 + \lambda H_1 \tag{5}$$

where $H_0$ corresponds to the exactly solvable parabolic (harmonic) potential, and $H_1$ represents the perturbative correction arising from the Gaussian nature of the confinement. A switching parameter $\lambda$ is introduced such that $\lambda=0$ corresponds to the parabolic quantum dot, while $\lambda=1$ recovers the Gaussian quantum dot.

Using the mean-field approximation, the effective Hamiltonian is further simplified and reduced to a renormalized harmonic oscillator problem. Consequently, the Gaussian quantum dot system can be mapped onto an equivalent effective parabolic quantum dot model, allowing analytical treatment of the problem.

The spin contribution is incorporated through the Zeeman term:

$$H_{spin} = g\mu_B B S_z \tag{6}$$

where g is the Landé g-factor (for GaAs, g=−0.44).

***Ground State Energy***

The resulting effective Hamiltonian corresponds to that of a two-dimensional harmonic oscillator in a magnetic field. The eigenvalue problem is solved using the well-known Fock–Darwin states, which are expressed in polar coordinates. The energy eigenvalues are given by:

$$E_{n,m,s} = (2n + |m| + 1)\hbar\tilde{\omega} + \frac{1}{2}\hbar\omega_c(m + g^* s) - g\mu_B B s \tag{7}$$

where n is the radial quantum number, mmm is the angular momentum quantum number, and sss represents the spin state. From this expression, the ground state and low-lying energy levels can be systematically obtained.

***Magnetization and Magnetic Susceptibility***

The thermodynamic properties of the system are analyzed using the framework of statistical mechanics within the canonical ensemble. Electron–electron interactions and spin–orbit coupling are neglected for simplicity.

The partition function is defined as:

$$Z = \sum_r e^{-\beta E_r} \tag{8}$$

where $\beta=1/(k_B T)$ and $E_r$ denotes the energy of the system in the rrr-th quantum state.

The Helmholtz free energy is given by:

$$F = -k_B T \ln Z \tag{9}$$

From this, the magnetization is obtained as:

$$M = -\left(\frac{\partial F}{\partial B}\right)_T \tag{10}$$

and the magnetic susceptibility is defined as:

$$\chi = \left(\frac{\partial M}{\partial B}\right)_T \tag{11}$$

At zero temperature, these expressions simplify significantly, and analytical forms can be derived using the ground state energy. At finite temperatures, the thermodynamic quantities are evaluated numerically via the partition function.

The average energy of the system is given by:

$$\langle E \rangle = -\frac{1}{Z}\left(\frac{\partial Z}{\partial \beta}\right) \tag{12}$$

From this formulation, magnetization and susceptibility can be calculated as functions of temperature, magnetic field strength, and confinement parameters.

Due to the complexity of the analytical expression for magnetic susceptibility, explicit forms are not presented here; instead, the results are illustrated graphically in the Results section.

In this study, the approximation introduced in Eq. (4.8) was validated through comparison with standard variational calculations, showing strong agreement with previously reported results (Boyacıoğlu, 2012). Using the developed theoretical model, the magnetization (M), magnetic susceptibility (χ), and inverse susceptibility (1/χ) of Gaussian-confined quantum dot (GQD) systems were systematically analyzed as functions of temperature, magnetic field strength, quantum dot width, and potential depth.

***Temperature-dependent behavior***

The temperature dependence of the magnetic properties is presented in Figure 1a–c, which includes magnetization, magnetic susceptibility, and inverse susceptibility, respectively.

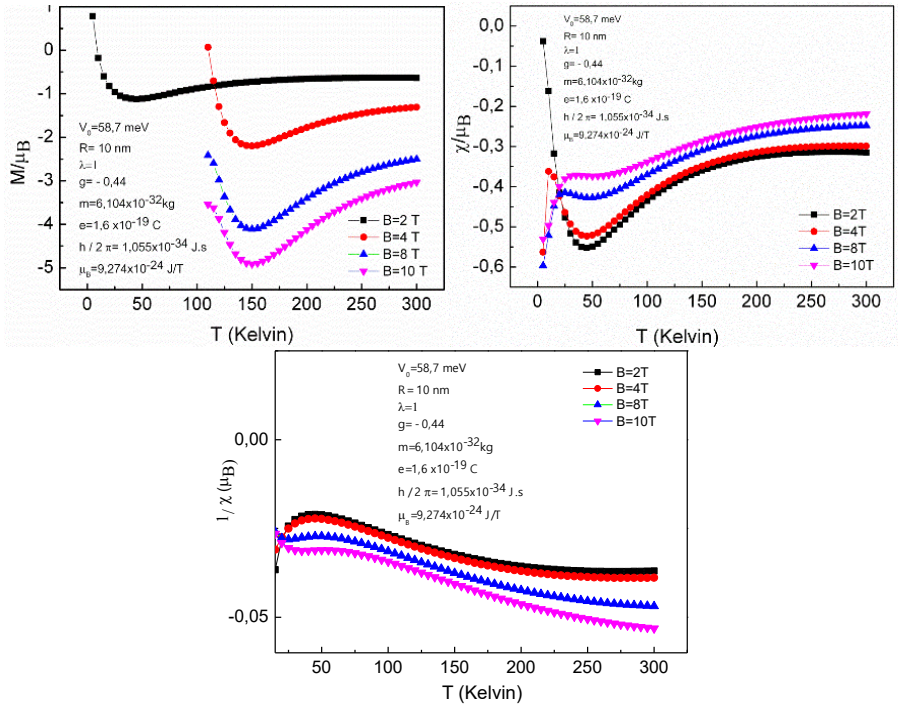



Figure 1. Temperature-dependent magnetic response of the Gaussian-confined quantum dot for magnetic field strengths ranging from B=2B to B=10 T:a) Magnetization (M/μB) vs. temperature, b) Magnetic susceptibility (χ/μB) vs. temperature, c) Inverse magnetic susceptibility (1/χ) vs. temperature. All calculations were performed for a fixed potential depth of $V_0$=58.7meV and confinement width of R=10 nm

As shown in Figure 1a, the magnetization exhibits a non-monotonic dependence on temperature, particularly at higher magnetic fields (B=4,8, and 10T). As the temperature decreases from 300 K to approximately 150 K, the magnetization gradually decreases, followed by a sharper increase at lower temperatures. This behavior indicates a competition between thermal effects and quantum confinement, where thermal fluctuations dominate at high temperatures, while quantum effects become more pronounced at lower temperatures.

For lower magnetic field strength (B=2 T), the magnetization remains nearly constant over a wide temperature range, suggesting weak temperature dependence, before increasing at very low temperatures due to enhanced quantum confinement effects.

The corresponding magnetic susceptibility curves shown in Figure 1b reveal that the system exhibits predominantly diamagnetic behavior over the entire temperature range. The susceptibility decreases monotonically from high temperatures down to approximately 50 K, after which field-dependent deviations become significant. These deviations are associated with the increasing contribution of discrete energy levels at low temperatures.

The inverse susceptibility plots in Figure 1c exhibit similar trends for all magnetic field values. The approximately linear behavior observed in certain temperature intervals suggests a Curie-like tendency, although deviations from ideal behavior arise due to finite-size and quantum confinement effects.

***Magnetic field dependence***

The dependence of magnetic properties on the applied magnetic field is illustrated in Figure 2a–b.

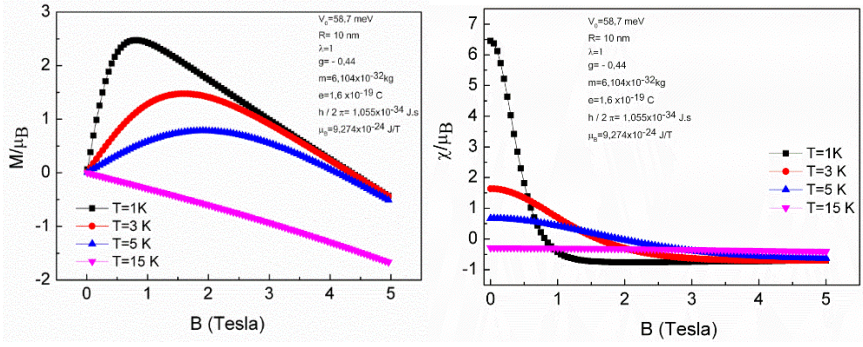



Figure 2. Magnetic field dependence of the system at different temperatures (T=1,3,5,15 K) (a) Magnetization vs. magnetic field, (b) Magnetic susceptibility vs. magnetic field. The calculations were carried out for $V_0$=58.7 meV and R=10 nm

As shown in Figure 2a, the magnetization exhibits a non-linear dependence on the magnetic field. At low temperatures (1–5 K), the magnetization initially increases with increasing magnetic field, reaching a maximum around B≈1 TB followed by a gradual decrease. This peak behavior arises from the interplay between magnetic field-induced energy level splitting and quantum confinement effects.

At higher temperature (T=15 K), the magnetization shows a monotonic negative trend, indicating a dominant diamagnetic response and the suppression of quantum effects due to thermal broadening.

The magnetic susceptibility behavior in Figure 2b reveals a clear paramagnetic-to-diamagnetic transition. At low temperatures (1–5 K), the system behaves paramagnetically at low magnetic fields and transitions to diamagnetic behavior as the magnetic field increases. In contrast, at T=15 K, the system remains diamagnetic over the entire magnetic field range, indicating that thermal effects dominate over quantum contributions.

***Quantum dot width dependence***

The influence of quantum dot width on magnetic properties is shown in Figure 3a–b.

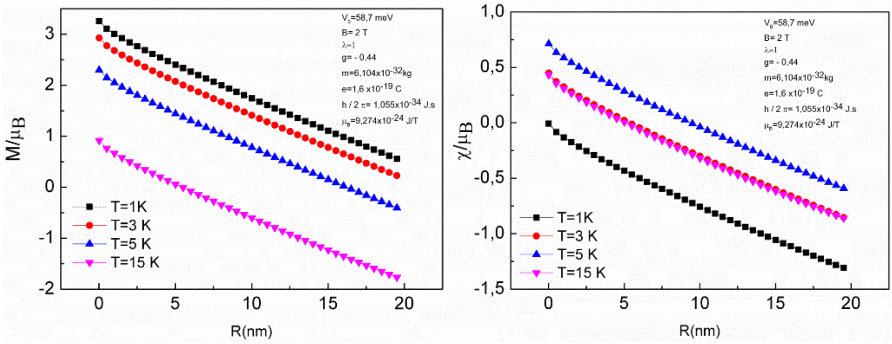



Figure 3. Dependence of magnetic properties on quantum dot width RRR (0–20 nm): (a) Magnetization vs. quantum dot width, (b) Magnetic susceptibility vs. quantum dot width. The parameters were fixed at V0=58.7 and B=2 T

From Figure 3a, it is observed that magnetization decreases monotonically with increasing quantum dot width for all temperatures. This behavior reflects the weakening of quantum confinement as the dot size increases, leading to reduced energy level spacing and diminished magnetic response.

The susceptibility curves in Figure 3b demonstrate a transition from paramagnetic to diamagnetic behavior as the quantum dot width increases. For intermediate temperatures (T=3,5,15 K), this transition occurs smoothly, while at T=1 K, the system remains predominantly diamagnetic with

increasing magnitude. These results indicate that stronger confinement enhances paramagnetic contributions, whereas larger systems favor diamagnetism.

***Potential depth dependence***

The effect of potential depth on the magnetic properties is presented in Figure 4a–b.

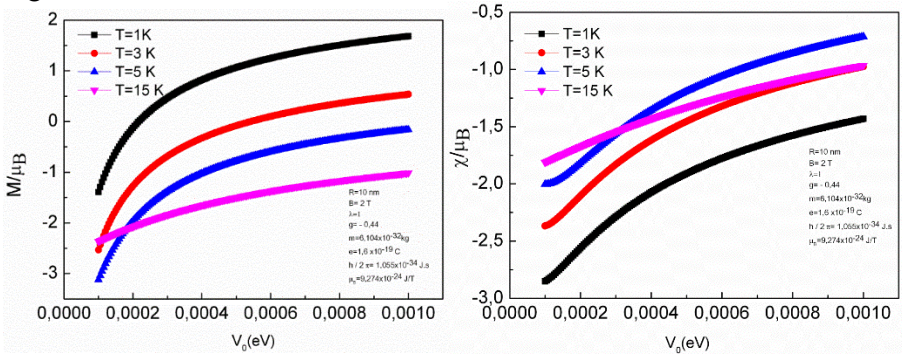



Figure 4. Dependence of magnetic properties on potential depth $V_0$: (a) Magnetization vs. potential depth, (b) Magnetic susceptibility vs. potential depth. The calculations were performed for a fixed confinement width R=10 nm and magnetic field B=2 T

As shown in Figure 4a, magnetization increases rapidly with increasing potential depth up to approximately $V_0 \approx 0.2$ meV for low temperatures (1–5 K), after which the variation becomes significantly weaker, indicating a saturation behavior. At higher temperature (T=15 K), the increase is more gradual due to thermal effects.

The susceptibility curves in Figure 4b show that increasing potential depth drives the system toward a diamagnetic regime. The magnitude of diamagnetism decreases with increasing temperature, consistent with thermal smearing of discrete energy levels. Additionally, the convergence of curves at higher $V_0$ values suggests that strong confinement dominates over temperature effects in this regime.

## DISCUSSION AND CONCLUSIONS

Low-dimensional systems have attracted significant attention due to their wide range of potential applications and their ability to provide experimentally accessible platforms for testing the fundamental predictions of quantum mechanics. Among these systems, quantum dots represent zero-dimensional structures in which the confinement of charge carriers leads to discrete energy spectra and pronounced quantum effects. As emphasized in the Introduction, one of the most critical factors determining the physical

properties of quantum dots is the nature of the confinement potential. In this study, the Gaussian confinement potential (GCP) was employed due to its capability to realistically describe physical processes such as tunneling and ionization.

The present work focuses on the magnetic properties of a Gaussian-confined quantum dot, specifically magnetization and magnetic susceptibility. These quantities were systematically investigated as functions of temperature, magnetic field strength, quantum dot width, and potential depth. The results reveal several notable features, particularly in the low-temperature and low-magnetic-field regimes.

In the temperature-dependent analysis, the system exhibits clear diamagnetic behavior over a wide temperature range. However, at low temperatures, fluctuations in the diamagnetic response are observed, indicating the strong influence of quantum confinement and discrete energy levels. As the temperature increases, the diamagnetic response initially strengthens and subsequently decreases monotonically, reflecting the increasing dominance of thermal excitations over quantum effects.

The magnetic field-dependent analysis further highlights the complex nature of the system. At higher temperatures (e.g., 15 K), the system remains purely diamagnetic across the entire magnetic field range. In contrast, at lower temperatures (1 K, 3 K, and 5 K), the system exhibits a transition from paramagnetic behavior at low magnetic fields to diamagnetic behavior at higher fields. This crossover behavior is a key result of the study and can be attributed to the field-induced modification of the energy spectrum and the redistribution of electron occupation among quantum states. It demonstrates that the magnetic response of the system is highly sensitive to external perturbations in the quantum regime.

The influence of the quantum dot width provides further insight into confinement effects. As the dot size increases, the system transitions from paramagnetic to diamagnetic behavior, indicating that stronger confinement enhances paramagnetic contributions, while weaker confinement favors diamagnetism. Additionally, decreasing temperature reinforces diamagnetic behavior, emphasizing the role of quantum localization.

Similarly, the dependence on potential depth reveals that increasing confinement strength initially drives the system toward paramagnetic behavior. However, as the potential depth increases further, the variation becomes more gradual, suggesting a saturation effect. This indicates that beyond a certain confinement threshold, additional increases in potential depth have a diminished impact on the system's magnetic response.

Overall, the results demonstrate that the magnetic properties of Gaussian-confined quantum dots are governed by a delicate interplay between quantum confinement, thermal excitation, and external magnetic field effects. The observed transitions between paramagnetic and diamagnetic behavior, as

well as the sensitivity to confinement parameters, highlight the importance of realistic potential models in describing nanoscale systems.

These findings contribute to a deeper understanding of the magnetic behavior of low-dimensional quantum systems and may provide valuable insights for the design of future quantum devices and nanoscale magnetic materials, where precise control of electronic and magnetic properties is essential.

## REFERENCES


Adamowski, J., Sobkowicz, M., Szafran, B., & Bednarek, S. (2000). Electron states in a Gaussian quantum dot. *Physical Review B, 62*, 4234–4237.

Alauwaji, R. M., et al. (2024). Optical and electronic properties of GaAs quantum dots. *Molecules, 29*(13), 3052.

Ashoori, R. C. (1996). Electrons in artificial atoms. *Nature, 379*, 413–419.

Bessis, N., Bessis, G., & Joulakian, B. (1982). Approximate solutions for quantum systems. *Journal of Physics A: Mathematical and General, 15*, 3679–3688.

Bimberg, D., Grundmann, M., & Ledentsov, N. N. (1999). *Quantum dot heterostructures*. Wiley.

Boyacıoğlu, B., Sağlam, M., & Chatterjee, A. (2007). Thermodynamic properties of Gaussian quantum dots. *Journal of Physics: Condensed Matter, 19*, 456217.

Boyacıoğlu, B., & Chatterjee, A. (2012a). Magnetic properties of Gaussian quantum dots. *Journal of Applied Physics, 112*, 083514.

Boyacıoğlu, B., & Chatterjee, A. (2012b). Thermodynamic behavior of confined quantum systems. *International Journal of Modern Physics B, 26*, 1250018.

Chatterjee, A. (1985). Quantum mechanical models of confined systems. *Journal of Physics A: Mathematical and General, 18*, 2403–2410.

Chatterjee, A. (1990). Quantum dots and related systems. *Physics Reports, 186*, 249–370.

Cohen, M. (1984). Bound states in confined systems. *Journal of Physics A: Mathematical and General, 17*, L101–L104.

Crandall, R. E. (1983). Quantum confinement and eigenvalue problems. *Journal of Physics A: Mathematical and General, 16*, L395–L400.

De Groote, J. J. S., Hornos, J., & Chaplik, A. V. (1992). Quantum dots under magnetic confinement. *Physical Review B, 46*, 12773–12779.

Gomez, S. S., & Romero, R. H. (2009). Quantum dot properties under confinement. *Central European Journal of Physics, 7*, 12–18.

Gornik, E., Lassnig, R., Strasser, G., Stormer, H. L., Gossard, A. C., & Wiegmann, W. (1985). Quantum effects in semiconductor heterostructures. *Physical Review Letters, 54*, 1820–1823.

Heindel, T., Kim, J.-H., Gregersen, N., Rastelli, A., & Reitzenstein, S. (2023). Quantum dots for photonic quantum information technology. *Advanced Quantum Technologies*.

Jacak, L., Hawrylak, P., & Wojs, A. (1998). *Quantum dots*. Springer.

Johnson, N. F. (1995). Quantum dot systems and confinement potentials. *Journal of Physics: Condensed Matter, 7*, 965–979.
Kachu, A., et al. (2025). Stability and correlations in Gaussian-confined quantum dots. *Scientific Reports*.
Kainz, J., Mikhailov, S., Wensauer, A., & Rössler, U. (2002). Electron interactions in quantum dots. *Physica E, 12*, 888–891.
Kastner, M. A. (1992). The single-electron transistor. *Reviews of Modern Physics, 64*, 849–858.
Kouwenhoven, L. P., Austing, D. G., & Tarucha, S. (2001). Few-electron quantum dots. *Reports on Progress in Physics, 64*, 701–736.
Kurmi, R. P., et al. (2025). Laser-assisted quantum dot systems with Gaussian potentials. *Results in Physics*.
Lai, C. S. (1983). Quantum mechanical confinement models. *Journal of Physics A: Mathematical and General, 16*, L181–L185.
Landau, L. D., & Lifshitz, E. M. (1975). *Statistical physics*. Pergamon Press.
Lee, S. J., Jeon, H. C., Kang, T. W., & Souma, S. (2008). Electronic properties of quantum dots. *Physica E, 40*, 2198–2202.
Maksym, P. A., & Chakraborty, T. (1990). Quantum dots in magnetic fields. *Physical Review Letters, 65*, 108–111.
Mukhopadhyay, S., & Chatterjee, A. (2001). Thermodynamic properties of quantum dots. *Acta Physica Polonica B, 32*, 473–482.
Mukhopadhyay, S., Boyacıoğlu, B., Sağlam, M., & Chatterjee, A. (2008). Gaussian quantum dot analysis. *Physica E, 40*, 2776–2780.
Nguyen, N. T. T., & Peeters, F. M. (2008). Electron properties in quantum dots. *Physical Review B, 78*, 045321.
Oh, J. H., Chang, K. J., Ihm, G., & Lee, S. J. (1994). Electronic structure of quantum dots. *Physical Review B, 50*, 15397–15400.
Oh, J. C., Chang, K. J., Ihm, G., & Lee, S. J. (1995). Quantum dot confinement studies. *Journal of the Korean Physical Society, 28*, 132–136.
Sargsian, T. A., et al. (2023). Realistic confinement potentials in quantum nanostructures.
Voskoboynikov, O., Bauga, O., Lee, C. P., & Tretyak, O. (2003). Semiconductor quantum dot modeling. *Journal of Applied Physics, 94*, 5891–5896.
Wegner, K. D. (2024). The Nobel Prize in Chemistry 2023: Quantum dots. *ACS Nano*.
Woggon, U. (1997). *Optical properties of semiconductor quantum dots*. Springer.
Zubchenko, A., et al. (2025). Autonomous tuning of quantum dot devices. *Physical Review Applied*.

# Scalable Synthesis of High-Quality Graphene Using Chemical Vapor Deposition: Fundamental Principles and Characterization

**Özge ATEŞİN ÜNLÜ[1]**
**Uğur ERKARSLAN[2]**
**Adem DÖNMEZ[3]**

1. Özge Ateşin Ünlü - Department of Energy, Graduate School of Natural and Applied Sciences, Molecular Nano-Materials Laboratory, Muğla Sıtkı Koçman University, Muğla, Turkey. Ozgeunlu81@gmail.com ORCID No: 0009-0003-7910-7336
2. Prof. Dr. Ugur Erkarslan - Department of Physics, Molecular Nano-Materials Laboratory, Mugla Sıtkı Kocman University, Mugla, Turkey. eugur@mu.edu.tr ORCID No: 0000-0001-7667-2151
3. Assoc. Prof. Dr Adem Donmez - Department of Electricity and Energy, Yatagan Vocational School, Mugla Sıtkı Koçman University, Mugla, Turkey. adonmez@mu.edu.tr ORCID No: 0000-0002-9773-0493

## ABSTRACT

The atomically thin, $sp^2$-bonded carbon network of graphene combines a honeycomb geometry with an unusual set of transport, optical, and mechanical characteristics. High charge-carrier mobility, substantial light transmission, and mechanical strength have motivated its investigation for electronic and optoelectronic technologies. A major obstacle, however, is translating these material advantages into reproducible, large-scale growth of structurally sound films.

Here, graphene films were grown on copper foil by chemical vapor deposition operated at atmospheric pressure. Growth took place at 1000 °C with methane ($CH_4$) as the carbon feedstock in a hydrogen–argon gas environment. Particular attention was given to how the methane supply affected nucleation and the fraction of copper covered by graphene. Adjustment of the methane flow enabled continuous coverage of $2 \times 2$ cm² copper specimens with comparatively uniform films.

Following growth, polymethyl methacrylate (PMMA) was used as a temporary support to transfer the films to glass, fluorine-doped tin oxide (FTO), and $Si/SiO_2$. Raman spectra were evaluated together with optical and scanning electron micrographs to characterize the resulting material. The spectroscopic response was consistent with predominantly single-layer graphene and a low concentration of defects. Imaging additionally showed extensive coverage, although local wrinkles and microcracks arose during handling and transfer.

Taken together, these observations support atmospheric-pressure CVD as a practicable route to graphene films of high structural quality. They also show that optimizing the deposition conditions is only part of the process: the subsequent removal and transfer steps must be controlled to retain film continuity and limit defects in prospective device structures.

*Keywords: Graphene; Chemical Vapor Deposition (CVD); Atmospheric Pressure CVD (APCVD); Monolayer Graphene; Methane Flow Optimization; Raman Spectroscopy; Scanning Electron Microscopy (SEM); Graphene Transfer; Copper Substrate; Two-Dimensional Materials*

---

## INTRODUCTION

A single sheet of graphene comprises carbon atoms linked through $sp^2$ covalent bonding in a planar honeycomb arrangement. The resulting atomic architecture gives rise to favorable charge transport, high transparency, and mechanical compliance, while also contributing to environmental robustness. Consequently, graphene has been considered for transparent electrodes, optoelectronics, and emerging device platforms (Novoselov et al., 2004; Schwierz, 2010; Bonaccorso et al., 2010).

Neighboring carbon sites in graphene are separated by a bond distance of approximately 0.142 nm. Of particular importance to its electronic response is the meeting of the valence and conduction bands at the Dirac points of the Brillouin zone. Close to these points, carriers follow a nearly linear energy–momentum relation and can be described as massless Dirac fermions, with a characteristic velocity on the order of $10^6$ m $s^{-1}$ (Castro Neto et al., 2009). The absence of an intrinsic gap leads to graphene being described as a semimetal or zero-gap semiconductor. Stacking additional sheets changes this picture because interlayer coupling reshapes the bands; in bilayers, suitable external conditions can produce an energy gap.

Strong in-plane carbon bonding also accounts for graphene's notable stiffness, tensile response, and ability to conduct heat. Alongside these properties, its large specific surface area and quantum transport phenomena—including manifestations of the quantum Hall effect—have encouraged research ranging from fundamental condensed-matter physics to functional materials (Du et al., 2008; Du et al., 2009).

Producing graphene in quantities suitable for practical use while maintaining its crystallographic quality remains difficult. This challenge contrasts with the landmark isolation of individual atomic sheets by Andre Geim and Konstantin Novoselov using repeated mechanical cleavage, often called the 'Scotch tape' method. That achievement established a route to investigate two-dimensional carbon despite earlier concerns about the stability of isolated two-dimensional crystals, and it was recognized by the 2010 Nobel Prize in Physics (Novoselov et al., 2004). Mechanical cleavage is valuable for laboratory specimens but does not provide a viable path to continuous industrial-scale films.

The limited throughput of exfoliation-based production has prompted the exploration of routes that build graphene directly from carbon precursors. Chemical vapor deposition (CVD) is particularly attractive in this regard, because growth conditions can be adjusted to form extensive carbon films on suitable metal surfaces (Mattevi et al., 2011; Zhang et al., 2013). Hydrocarbons are commonly supplied to copper or nickel substrates. Copper's relatively small carbon solubility favors surface-dominated growth and helps restrict the formation of additional layers, which is useful when monolayer films are required (Li et al., 2009).

Assessing a graphene film requires more than identifying the presence of carbon: thickness, spatial continuity, crystal-domain boundaries, and the concentration of defects all contribute to the response of the finished device. Changes in these features can alter optical transmission, mechanical behavior, and electrical conduction (Brownson et al., 2011). Growth protocols must therefore be designed with the relevant formation mechanisms and uniformity requirements in mind.

This chapter examines graphene deposition onto copper by atmospheric-pressure chemical vapor deposition (APCVD), with emphasis on

conditions that favor largely monolayer coverage. After synthesis, the films were moved to glass, FTO, or Si/$SiO_2$ supports. Optical microscopy, Raman measurements, and scanning electron microscopy (SEM) were then used to relate surface appearance and structural quality to the deposition and transfer procedures. Particular emphasis is placed on the influence of growth conditions on coverage and defects.

## FUNDAMENTAL PROPERTIES OF GRAPHENE

In graphene, a regular two-dimensional lattice of carbon atoms forms the characteristic honeycomb structure illustrated in Figure 1. Each site uses three $sp^2$ orbitals to make σ bonds with its neighbors, while the remaining out-of-plane p orbital participates in a delocalized π system. This combination of strong in-plane bonding and mobile π electrons is central to graphene's mechanical and electronic characteristics.

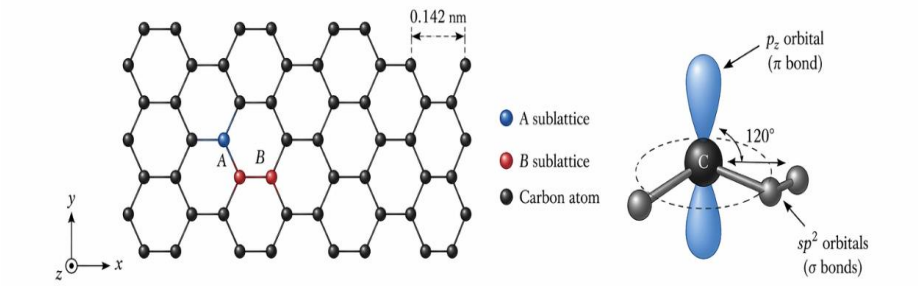



Figure 1: Honeycomb arrangement of $sp^2$-bonded carbon atoms, indicating the A/B sublattices, 0.142 nm C–C spacing, and σ/π orbital contributions.

The honeycomb geometry may be viewed as a pair of offset triangular sublattices, conventionally denoted A and B. Their arrangement determines the symmetry of the electronic states and allows the valence and conduction bands to touch at the inequivalent K and K′ points. Near these points, the approximately linear dispersion forms the familiar Dirac cones depicted in Figure 2. The carriers can consequently be modeled using a Dirac-like description of effectively massless particles (Castro Neto et al., 2009).

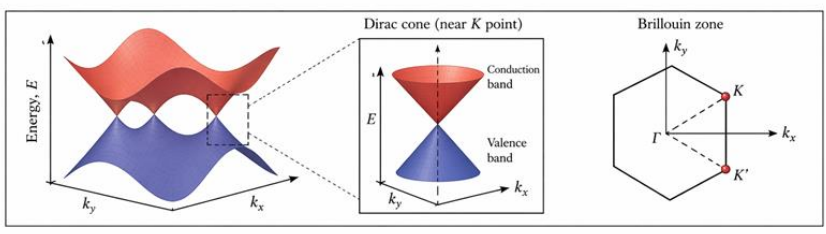


Figure 2: Graphene energy dispersion and Brillouin-zone geometry, including the touching of the valence and conduction bands at the K and K′ Dirac points.

Because the electronic bands are nearly linear close to the Dirac points, graphene can sustain very rapid carrier motion. Mobilities of approximately $10^4$–$10^5$ $cm^2\ V^{-1}\ s^{-1}$ have been reported under favorable conditions, with ballistic motion possible over submicrometer scales. This behavior is beneficial for fast electronic components; conversely, the absence of a conventional intrinsic band gap complicates switching applications that rely on a well-defined off state.

In-plane covalent bonding gives graphene an estimated Young's modulus near 1 TPa and an intrinsic tensile strength of roughly 130 GPa. These values place an ideal graphene sheet among the strongest known materials. Despite its high stiffness under tension, a single atomic layer can bend and accommodate substantial elastic deformation without necessarily losing structural integrity.

Heat transport is another distinctive feature of graphene. Reported in-plane thermal-conductivity values commonly lie in the range of 3000–5000 $W\ m^{-1}\ K^{-1}$, reflecting the efficient propagation of lattice vibrations. Such phonon-mediated energy transport makes the material of interest for removing heat from miniaturized electronic systems.

Graphene also provides an experimentally accessible setting for unusual quantum-transport effects. Its carrier dynamics are associated with an anomalous quantum Hall response, including observations under room-temperature conditions, and a Berry phase of $\pi$ (Du et al., 2008; Du et al., 2009). These electronic signatures vary with disorder, environmental interactions, and, notably, layer count. As further sheets are added, coupling between layers alters the ideal single-layer dispersion. An applied electric field can open a controllable gap in bilayer graphene, whereas thicker assemblies progressively show features associated with graphite.

A lattice vacancy, a grain boundary, or an extrinsic impurity can interrupt otherwise efficient charge transport by providing a site for carrier scattering. Minimizing such disturbances is therefore necessary when high electrical performance is sought. The material may also be affected by adsorbed contaminants and by its supporting surface, effects that are especially relevant after the film has been detached from its growth substrate.

The combination of mechanical resilience, thermal conduction, optical response, and unusual electronic physics makes graphene relevant to many device concepts. Obtaining these attributes consistently depends on how the material is synthesized and subsequently processed. The next sections therefore examine approaches to graphene preparation, with particular attention to CVD-based growth.

## GRAPHENE SYNTHESIS STRATEGIES

Research on preparing graphene is motivated not only by its intrinsic properties but also by the need for films whose structure can be reproduced across different samples. Device performance varies with thickness, lattice disorder, the size and distribution of crystalline domains, and the continuity of coverage. For this reason, selecting a synthesis method involves balancing the attainable structural quality against process control and production scale.

Preparation routes can be grouped according to whether they separate existing graphite layers or assemble graphene from smaller carbon-containing species. The former are described as top-down processes, whereas the latter belong to the bottom-up category.

### *Top-Down Approaches*

In top-down processing, individual carbon sheets are detached from graphite by mechanical action or by chemical or electrochemical treatment.

### *Mechanical Exfoliation*

Mechanical cleavage, frequently described as the 'Scotch tape' technique, exploits the weak interlayer van der Waals attraction in graphite. Repeated separation can isolate very thin flakes, including single layers, while avoiding many of the chemical modifications associated with wet-processing routes. Its importance lies in having supplied graphene specimens for early experimental investigations.

The suitability of mechanically exfoliated sheets for fundamental measurements does not translate readily to mass manufacture. Flakes typically have limited lateral dimensions, and both the quantity obtained and the outcome of individual exfoliation attempts can vary. The technique is consequently better suited to laboratory-scale preparation than to the production of large-area coatings.

### *Chemical Exfoliation*

Chemical treatments offer another way to overcome the cohesion between graphite planes. For example, introducing alkali-metal species such as potassium between the layers, followed by solvent reactions, can promote their separation; ethanol has been used in such procedures (Viculis et al., 2003).

An alternative chemical pathway begins by oxidizing graphite to graphene oxide (GO) and subsequently converting it to reduced graphene oxide (rGO). The underlying oxidation chemistry traces back to the procedures of Brodie, Staudenmaier, and Hummers (Brodie, 1859; Staudenmaier, 1898; Hummers & Offeman, 1958). Variations on the

Hummers route have since been developed with the aim of improving processing efficiency and reducing environmental burdens (Chen et al., 2013).

Solution-based chemical production is attractive when throughput and cost are primary considerations. Its main drawback is that reduction does not necessarily restore the original carbon lattice: residual oxygen functionalities and structural imperfections can remain, with adverse consequences for conductivity and demanding electronic uses.

***Bottom-Up Approaches***

Rather than removing layers from a parent crystal, bottom-up methods construct the carbon network by transforming molecular or gaseous precursors under selected reaction conditions.

*i- Pyrolysis and Epitaxial Growth*

Thermal decomposition of a suitable carbon source can yield graphene-related structures through pyrolysis. A different high-temperature route starts with silicon carbide (SiC): preferential removal of silicon from its surface leaves a carbon-rich layer that can reorganize into graphene.

Although attractive crystalline material can be obtained through such routes, the need for elevated temperatures and, in some cases, costly starting substrates may constrain their wider deployment.

*ii- Chemical Vapor Deposition (CVD)*

Within the family of bottom-up methods, CVD has become a widely investigated option for producing graphene over relatively large substrate areas. A hydrocarbon feed, often methane ($CH_4$), reacts or decomposes at a heated catalytic surface, providing carbon that can rearrange into a two-dimensional film.

Metal selection is important because carbon behaves differently in the substrate during growth. On nickel, appreciable carbon dissolution and subsequent precipitation complicate control over film thickness. Copper has much lower carbon solubility, so deposition is governed primarily by processes at the surface; this can favor predominantly single-layer coverage (Li et al., 2009).

For applications that require continuous films, several characteristics make CVD appealing:

- Formation of coatings over comparatively large, connected areas
- Ability to adjust the number of layers through growth conditions
- Potential cost advantages relative to some epitaxial routes
- Prospects for integration with high-throughput manufacturing

Considering the compromises among cost, structural quality, and coverage, CVD offers an attractive option for applications requiring films rather than isolated flakes.

### *Comparative Evaluation of Synthesis Methods*

Table 1 brings together the principal graphene preparation routes and contrasts their typical strengths with the limitations encountered in practice.

Table 1: Comparison of graphene synthesis methods

| Method | Approach | Advantages | Limitations |
|---|---|---|---|
| Mechanical Exfoliation | Top-down | Highest quality, defect-free | Not scalable, small size |
| Chemical Exfoliation | Top-down | Low cost, scalable | High defect density |
| Pyrolysis | Bottom-up | Controlled growth | Limited uniformity |
| Epitaxial Growth (SiC) | Bottom-up | High crystalline quality | High cost |
| CVD | Bottom-up | Large area, scalable, controllable | Transfer-related defects |

### *Scope of This Study*

CVD was selected here because it provides a useful compromise between control of film formation, potential scalability, and achievable quality. Growth conditions—including reactor temperature, substrate preparation, and gas delivery—were established for copper foil, with methane flow examined as a key variable affecting coverage and layer formation.

The discussion that follows describes how graphene develops on copper in the CVD environment, the apparatus employed, and the procedure used to refine deposition conditions.

## CHEMICAL VAPOR DEPOSITION (CVD) OF GRAPHENE

CVD provides a controlled gaseous environment in which graphene can develop on a heated metal substrate. The ability to tailor substrate temperature and reactant supply makes it suitable for investigating continuous, relatively low-defect coatings (Mattevi et al., 2011; Zhang et al., 2013). In its atmospheric-pressure implementation (APCVD), the reactor can operate without maintaining reduced pressure throughout the deposition stage, offering a comparatively straightforward processing configuration.

Graphene formation during CVD begins with carbon supplied by a gaseous precursor, commonly methane ($CH_4$), at a hot transition-metal surface. The mechanism is strongly substrate-dependent. Nickel can absorb carbon and subsequently release it during cooling, whereas the limited solubility of carbon in copper favors growth at the exposed metal surface and facilitates the formation of thin films (Li et al., 2009).

***CVD System and Growth Mechanism***

Figure 3 outlines the experimental apparatus, comprising an evacuation unit, a manifold for supplying the process gases, and a tube furnace capable of high-temperature operation. A horizontally mounted quartz reaction tube houses the copper samples and permits gas to pass through the heated region under controlled conditions.

At the beginning of the procedure, the chamber is pumped down to reduce the amount of moisture and oxygen remaining inside. Hydrogen is then supplied to establish a reducing atmosphere while the copper undergoes thermal treatment. Heating the foil to about 1000 °C promotes changes in its surface and grain structure before carbon deposition.

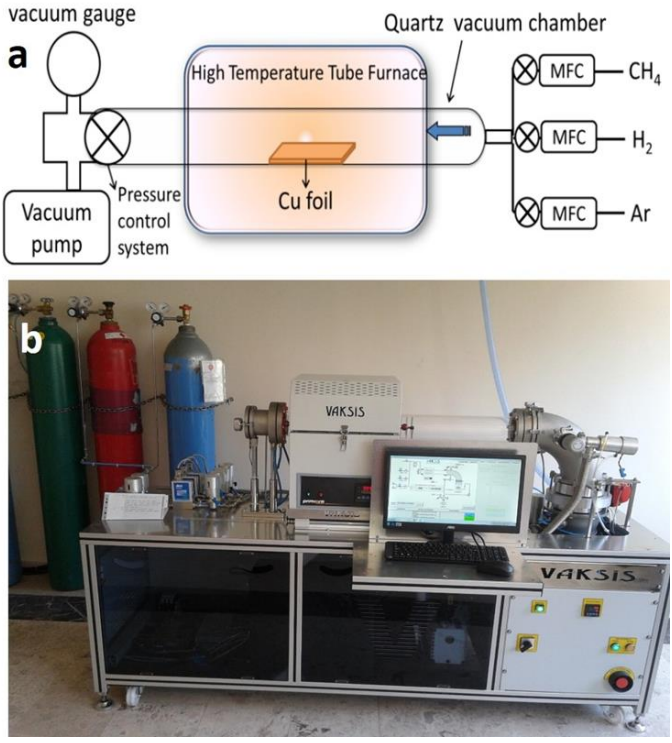



Figure 3: Experimental CVD equipment and its schematic gas-delivery, heating, and evacuation configuration.

On admitting methane to the hot reactor, precursor decomposition supplies carbon-containing surface species. Adsorption and migration across copper are followed by the appearance of graphene nuclei; the domains then spread laterally and may merge to cover the metal foil.

***Graphene Growth Procedure***

Copper foils measuring 2 × 2 $cm^2$ and 0.1 mm in thickness served as the growth substrates. To prepare the surface, each foil underwent ultrasonic cleaning in the following order: acetone, isopropanol, acetone again, and deionized water.

After cleaning, the copper was positioned in the quartz reactor and the chamber was pumped to approximately $10^{-6}$ Torr. A hydrogen flow of 100 sccm was established, and the temperature was raised to 1000 °C over about 30 minutes. The copper was maintained at this temperature for another 30 minutes to promote annealing and grain development, as shown in Figure 4.

Methane was introduced after the annealing stage, and its delivery rate was varied to examine graphene formation. Figure 5 illustrates the progression from nucleation to domain expansion and eventual merging on copper. A supply of 30 sccm left portions of the surface uncovered, whereas increasing the methane rate to 50 sccm yielded a more continuous and spatially uniform film.

When deposition ended, cooling was accelerated by opening the furnace arrangement and moving the heated zone away from the specimen. The process-gas supply was discontinued once the sample approached 150 °C; nitrogen was subsequently used to return the system to ambient conditions.

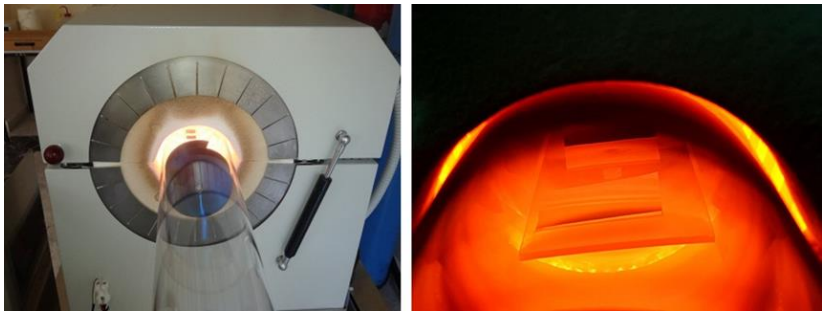


Figure 4: Images of the furnace and copper specimen during the high-temperature annealing/growth stage

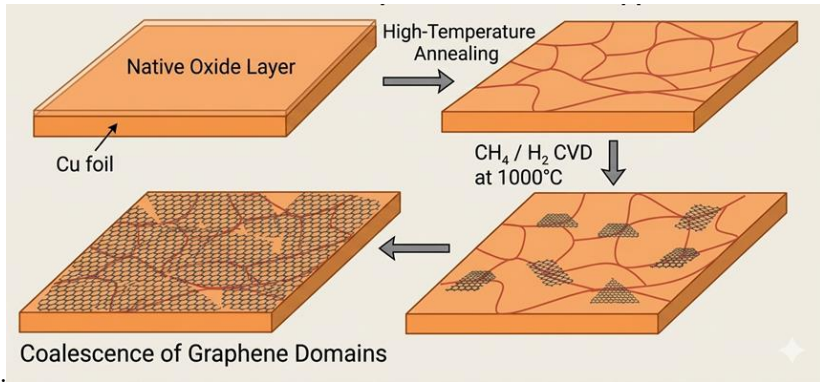



Figure 5: Diagram of the surface-growth sequence on copper, from graphene nucleation to the joining of neighboring domains.

***Transfer of Graphene***

The graphene coating was detached from its copper growth foil and placed on glass, FTO, and $Si/SiO_2$ supports through a transfer procedure that used a temporary polymer backing.

A support film of polymethyl methacrylate (PMMA) was first applied by spin coating at 4000 rpm for 40 s. The coated specimen was then held at 180 °C for 5 min to improve contact between the polymer and graphene.

Removal of the underlying copper involved immersion in a nitric-acid/deionized-water etchant and subsequent exposure to an Fe(III) nitrate solution. Once the metal had been etched away, the free-standing PMMA-supported graphene was rinsed in deionized water to reduce the amount of remaining chemical residue.

The floating PMMA/graphene assembly was collected onto the intended target surface by the fishing transfer procedure. To dissolve the sacrificial polymer, the supported film was placed in acetone at 40 °C for 30 min. Figure 6 documents the successive stages, from dissolution of the copper foil to recovery of graphene on the receiving substrates.

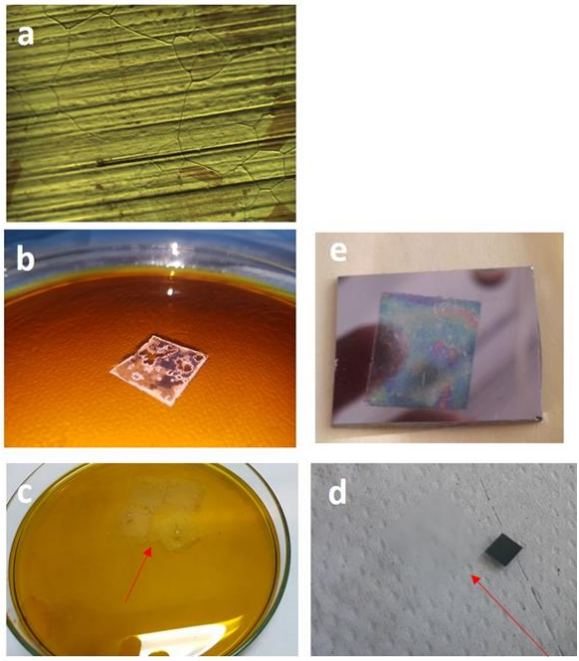



Figure 6: Photographs tracing PMMA-supported graphene transfer from the copper growth foil to the receiving substrates.

***Key Parameters Affecting Graphene Growth***

Several operating variables determine whether CVD growth produces a continuous film of acceptable quality:

- **Methane delivery: Controls the available carbon feed and influences the extent of surface coverage**
- **Hydrogen supply: Influences reduction of the surface and the formation of growth-related defects**
- **Reactor temperature: Modifies nucleation kinetics and the lateral development of domains**
- **Copper condition: Substrate roughness and cleanliness affect the morphology of the developing film**

Of the variables examined in this work, changing the methane supply was particularly important for obtaining largely uninterrupted graphene coverage.

***Summary***

The preceding experiments illustrate why CVD is useful for forming extensive graphene coatings: precursor delivery, heating, and the condition of the copper can be specified and adjusted. Producing a suitable film nevertheless requires attention beyond the growth step, since contamination

and damage may occur when graphene is released and transferred to another substrate.

## CHARACTERIZATION TECHNIQUES AND RESULTS

Optical imaging, Raman spectroscopy, and SEM were used as complementary probes of the deposited and transferred graphene. Microscopy provided information about continuity, visible damage, and surface texture, whereas Raman features were interpreted in relation to lattice order and the likely number of graphene layers.

### *Optical Microscopy Analysis*

Optical micrographs were first examined to compare the appearance of bare copper with that of the growth samples. Before deposition, the foil showed lines associated with its fabrication history (Figure 7a). Some of this topographic pattern remained apparent after the copper had been annealed in hydrogen at 1000 °C and could have affected subsequent domain formation.

The effect of carbon supply is evident in Figure 7. With methane delivered at 30 sccm, the images in Figures 7b and 7c contain areas of uncovered copper. The sample prepared at 50 sccm instead shows more extensive, better-connected graphene coverage (Figure 7d). This comparison links the availability of the gaseous carbon precursor to the development of a continuous coating under the conditions tested.

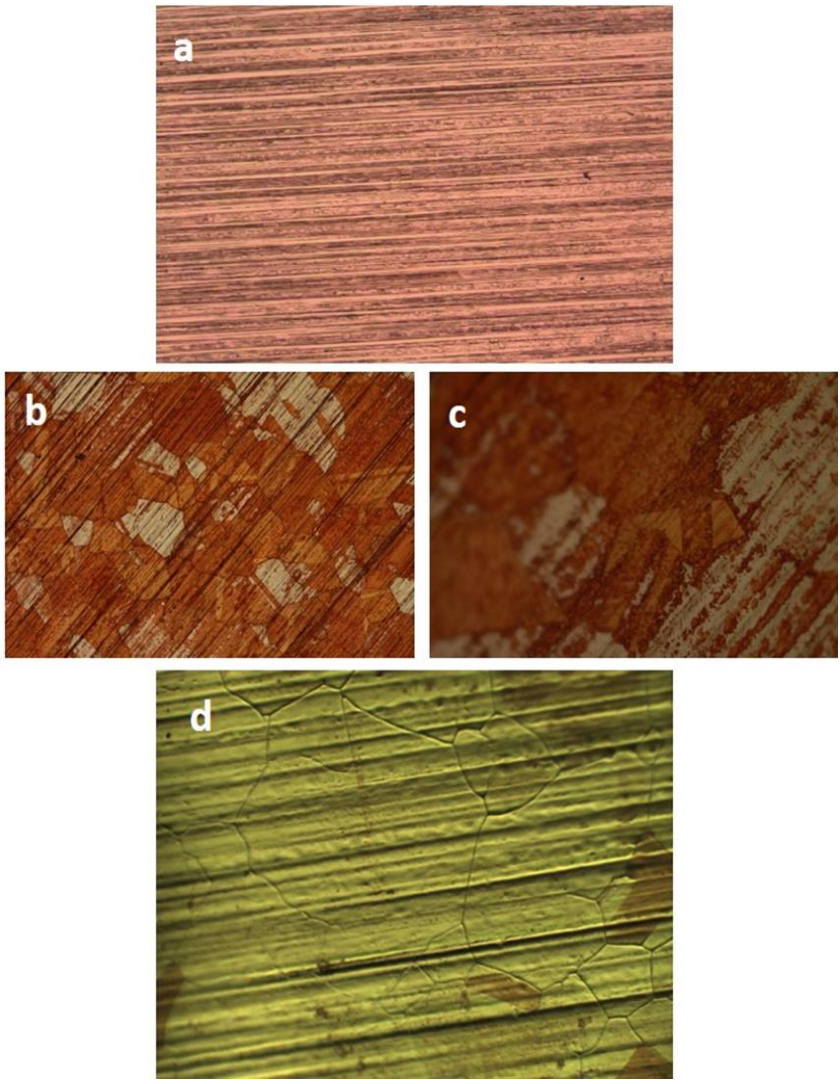



Figure 7: Optical views of (a) bare copper, (b, c) graphene grown with 30 sccm methane, and (d) graphene prepared with a 50 sccm methane feed.

Optical inspection continued after the graphene had been moved to $Si/SiO_2$. The polymer-supported state is visible in Figure 8a. Once PMMA was removed, the film was difficult to distinguish directly because graphene transmits approximately 97.7% of incident visible light. Dark-field observation nevertheless revealed folds or wrinkles (Figure 8b), while the bright-field view included local dark regions that may be associated with particulate impurities or residual transfer material (Figure 8c).

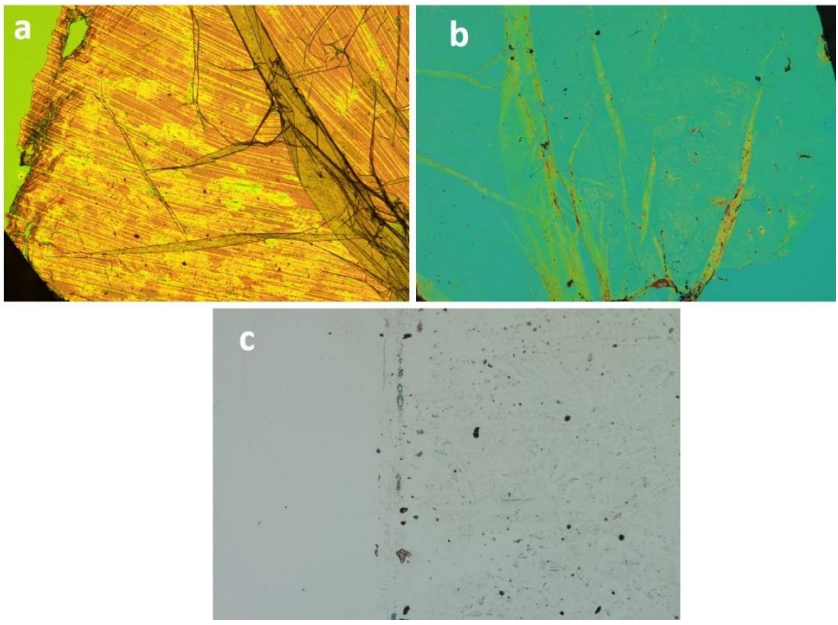



Figure 8: Microscopy of graphene on $Si/SiO_2$: (a) with the PMMA support, (b) after support removal in dark-field mode, and (c) bright-field view of localized surface features.

***Raman Spectroscopy Analysis***

Raman scattering was employed to assess characteristic graphene vibrational bands and to obtain indications of layer thickness and disorder. Figures 9 and 10 compare spectra obtained for the as-grown material on copper with those measured after transfer to the receiving surfaces.

The spectrum acquired on the copper-supported film (Figure 9) contains bands near 1364, 1590, and 2709 $cm^{-1}$, corresponding to the D, G, and 2D modes, respectively. The measured $_D/_G$ ratio is approximately 0.2, consistent with a comparatively small defect contribution. An $_{2D}/_G$ ratio near 1.7 is also consistent with a film composed predominantly of monolayer graphene.

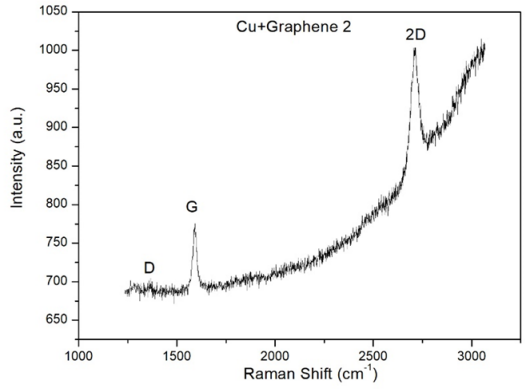



Figure 9: Copper-supported graphene Raman spectrum identifying the D, G, and 2D bands.

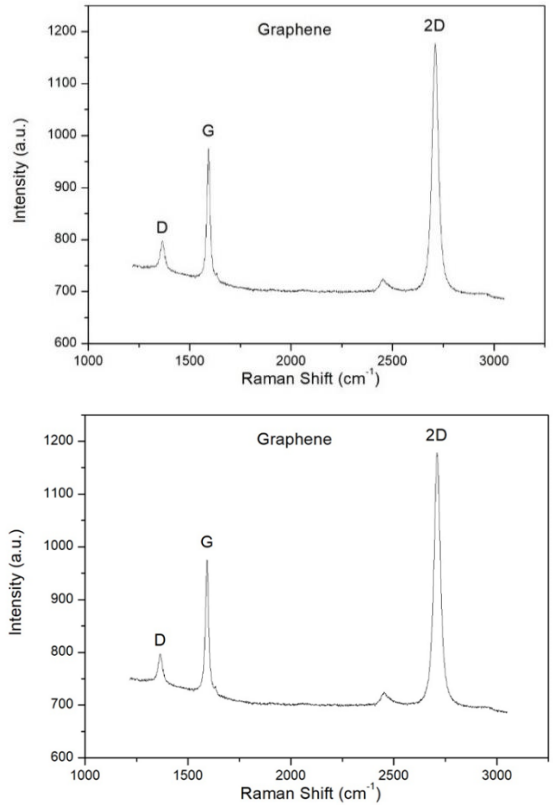



Figure 10: Raman responses from transferred graphene on (a) FTO and (b) $Si/SiO_2$.

For the FTO-supported film, the Raman bands remain recognizable after transfer (Figure 10a). The $_{D}/_{G}$ ratio is approximately 0.2, while $_{2D}/_{G}$ reaches about 2.8. The comparatively strong 2D response supports the interpretation of predominantly single-layer graphene and suggests that a substantial part of the original lattice order was retained during transfer.

Measurements on the Si/$SiO_2$-supported specimen (Figure 10b) gave an $_{D}/_{G}$ value of roughly 0.12 and an $_{2D}/_{G}$ value near 1.4. The reduced D-to-G intensity ratio is consistent with a comparatively small Raman-active defect contribution. The 2D-to-G response is compatible with predominantly thin graphene, although differences from the other substrates may reflect strain or substrate-dependent effects.

Considered together, the Raman measurements support the production of graphene that is largely monolayer and displays relatively little disorder under the selected growth conditions.

***Scanning Electron Microscopy (SEM) Analysis***

SEM was used to examine the continuity and fine-scale topography of graphene after it had been moved from copper. Representative views of different regions on the FTO-supported film are provided in Figure 11.

The two SEM views illustrate the local variability of the transfer outcome. Cracks and gaps can be identified in Figure 11a, consistent with damage that may arise during mechanical handling. By comparison, Figure 11b contains a region in which the coating remains more continuous, showing that substantially intact graphene could also be recovered on FTO.

Linear surface markings associated with the original rolled copper foil were also reflected in the graphene morphology. This correspondence indicates that features of the growth substrate can remain evident in the transferred sheet.

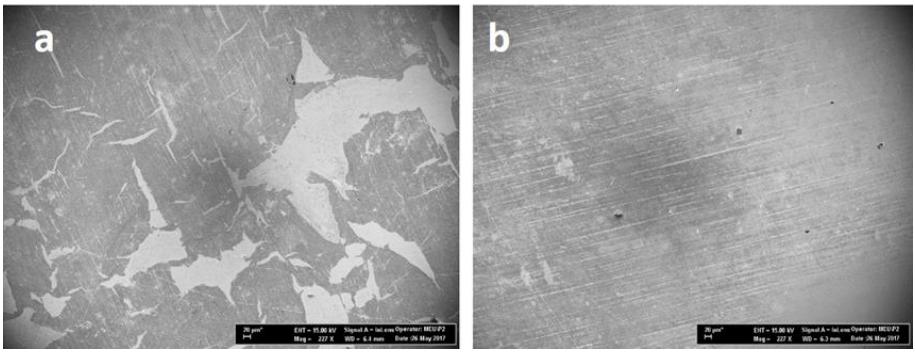



Figure 11: FTO-supported graphene SEM micrographs: (a) region containing cracks and (b) region with more continuous film coverage.

Figure 12 juxtaposes the uncovered FTO surface with the same type of support after graphene transfer. In Figure 12a the FTO grains are readily distinguishable. Their appearance is less pronounced in Figure 12b, which is consistent with the presence of an overlying graphene coating.

Because the added carbon sheet is both very thin and optically transparent, the contours of the underlying FTO are not completely obscured. The images are therefore consistent with graphene extending across the observed surface while allowing the substrate texture to remain visible.

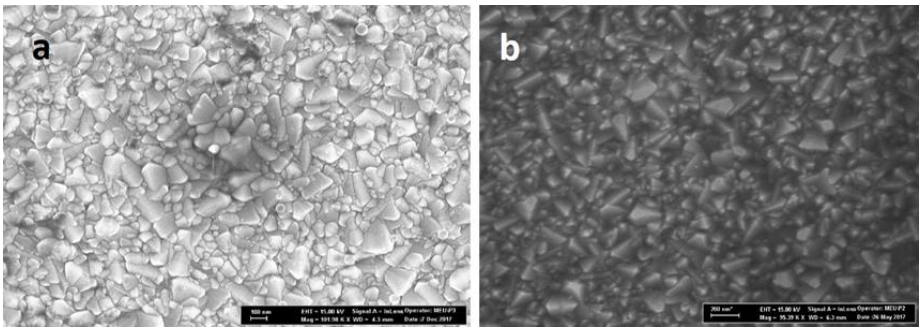



Figure 12: Comparison of SEM surface textures for (a) uncoated FTO and (b) FTO after graphene transfer.

***Overall Evaluation***

Viewed collectively, the optical, Raman, and SEM observations point to three principal characteristics of the films obtained here:

- A predominantly single-layer structure
- A comparatively small concentration of structural defects
- Relatively extensive and uniform substrate coverage

These favorable features do not remove the limitations associated with wet transfer. Residual surface material, wrinkles, and occasional microcracks were observed and could compromise performance in devices. Improvements in both the deposition step and subsequent handling are therefore relevant.

## DISCUSSION AND CONCLUSION

The copper-based CVD experiments show a clear relationship between methane supply and the extent of graphene coverage. Under the growth conditions investigated, a feed of 30 sccm was insufficient to cover the foil completely, whereas increasing it to 50 sccm improved the connectivity and apparent uniformity of the deposited film. The observations identify carbon precursor delivery as an important control variable in this process.

Raman spectra provide evidence for predominantly monolayer material: the relatively weak D response (low $_{D}/_{G}$) suggests limited disorder, and the 2D-to-G intensity ratios are consistent with thin graphene. Optical and SEM images add a spatial perspective, showing that broad areas can remain coated even though localized wrinkles, cracks, or contaminant features appear after transfer.

The results thus distinguish two aspects of film quality that must be considered together. Growth conditions largely determine how graphene forms on the copper, while handling and transfer influence the integrity of the film delivered to its final support. Polymer-derived residues and mechanically induced distortions remain practical concerns.

In summary, the investigated atmospheric-pressure CVD procedure can form largely continuous graphene films on copper when the reactant-gas supply is appropriately selected. Raman and microscopic observations indicate predominantly monolayer coverage with comparatively low defect levels, subject to local imperfections generated during transfer. Further development should address cleaner and less damaging transfer methods, alongside approaches that avoid transfer where the intended application permits.

## REFERENCES


Bonaccorso, F., Sun, Z., Hasan, T., & Ferrari, A. C. (2010). Graphene photonics and optoelectronics. *Nature Photonics, 4*(9), 611–622.

Brownson, D. A. C., Kampouris, D. K., & Banks, C. E. (2011). An overview of graphene in energy production and storage applications. *Journal of Power Sources, 196*(11), 4873–4885.

Castro Neto, A. H., Guinea, F., Peres, N. M. R., Novoselov, K. S., & Geim, A. K. (2009). The electronic properties of graphene. *Reviews of Modern Physics, 81*(1), 109–162.

Du, X., Skachko, I., Barker, A., & Andrei, E. Y. (2008). Approaching ballistic transport in suspended graphene. *Nature Nanotechnology, 3*(8), 491–495.

Du, X., Skachko, I., Duerr, F., Luican, A., & Andrei, E. Y. (2009). Fractional quantum Hall effect and insulating phase of Dirac electrons in graphene. *Nature, 462*(7270), 192–195.

Li, X., Cai, W., An, J., Kim, S., Nah, J., Yang, D., ... & Ruoff, R. S. (2009). Large-area synthesis of high-quality and uniform graphene films on copper foils. *Science, 324*(5932), 1312–1314.

Mattevi, C., Kim, H., & Chhowalla, M. (2011). A review of chemical vapour deposition of graphene on copper. *Journal of Materials Chemistry, 21*(10), 3324–3334.

Novoselov, K. S., Geim, A. K., Morozov, S. V., Jiang, D., Zhang, Y., Dubonos, S. V., ... & Firsov, A. A. (2004). Electric field effect in atomically thin carbon films. *Science, 306*(5696), 666–669.

Schwierz, F. (2010). Graphene transistors. *Nature Nanotechnology, 5*(7), 487–496.

Zhang, Y., Zhang, L., & Zhou, C. (2013). Review of chemical vapor deposition of graphene and related applications. *Accounts of Chemical Research, 46*(10), 2329–2339.
Brodie, B. C. (1859). On the atomic weight of graphite. *Philosophical Transactions of the Royal Society of London, 149*, 249–259.
Chen, J., et al. (2013). Improved Hummers method for graphene oxide synthesis. *Carbon, 64*, 225–229.
Hummers, W. S., & Offeman, R. E. (1958). Preparation of graphitic oxide. *Journal of the American Chemical Society, 80*(6), 1339.
Li, X., et al. (2009). Large-area synthesis of high-quality graphene films on copper foils. *Science, 324*, 1312–1314.
Staudenmaier, L. (1898). Verfahren zur Darstellung der Graphitsäure. *Berichte der Deutschen Chemischen Gesellschaft, 31*(2), 1481–1487.
Viculis, L. M., et al. (2003). Intercalation and exfoliation routes to graphene. *Science, 299*, 1361.
Li, X., et al. (2009). Large-area synthesis of high-quality graphene films on copper foils. *Science, 324*, 1312–1314.
Mattevi, C., Kim, H., & Chhowalla, M. (2011). A review of chemical vapour deposition of graphene on copper. *Journal of Materials Chemistry, 21*, 3324–3334.
Zhang, Y., Zhang, L., & Zhou, C. (2013). Review of chemical vapor deposition of graphene. *Accounts of Chemical Research, 46*, 232
Castro Neto, A. H., Guinea, F., Peres, N. M. R., Novoselov, K. S., & Geim, A. K. (2009). The electronic properties of graphene. *Reviews of Modern Physics, 81*(1), 109–162.
Du, X., Skachko, I., Barker, A., & Andrei, E. Y. (2008). Approaching ballistic transport in suspended graphene. *Nature Nanotechnology, 3*(8), 491–495.
Du, X., Skachko, I., Duerr, F., Luican, A., & Andrei, E. Y. (2009). Fractional quantum Hall effect and insulating phase of Dirac electrons in graphene. *Nature, 462*(7270), 192–195.

# Dual Split Quaternions And Spacelike, Timelike, Null Dual Split Euler Formulas For Spacelike, Timelike, Null Screw Movements

**Aysel TURGUT VANLI**[1]

1. Prof. Dr.; Gazi University, Faculty of Science, Department of Mathematics. avanli@gazi.edu.tr
ORCID No: 0000-0001-9793-7366

## ABSTRACT

This study investigates the kinematics of rigid body motions in semi-Euclidean or Minkowski-type spaces by introducing and developing comprehensive dual split quaternion frameworks. While classical kinematics heavily relies on standard quaternions and dual numbers to model rotations and translations via screw movements, this research extends these foundational concepts to dual split quaternions to systematically address different causal characters. The primary objective of this work is the derivation of spacelike, timelike, and null dual split Euler formulas tailored specifically for corresponding spacelike, timelike, and null screw movements. Because split quaternions incorporate a metric with an indefinite signature, the algebraic behavior and geometric interpretations diverge significantly from standard Euclidean quaternion formulations, particularly when dealing with light-like (null) vectors and hyperbolic rotations. It provides explicit, step-by-step computational examples for each movement category and details their relevant application areas, validating the theoretical derivations and demonstrating their utility in practical engineering and kinematic analysis. This framework offers advanced mathematical tools for researchers working in theoretical kinematics, spatial mechanism design, robotics, and differential geometry in non-Euclidean spaces.



*Keywords – Dual split quaternions, spacelike screw motion, timelike screw motion, null screw motion, high-energy physics, robotics, kinematics.*


## INTRODUCTION

Hypercomplex number systems constitute one of the most fundamental building blocks of mathematics and theoretical physics. The effort to go beyond real numbers and construct higher-dimensional algebraic structures has a long and rich tradition in the history of mathematics. This tradition reached a significant turning point with William Rowan Hamilton's discovery of quaternions in 1843 (Hamilton 1843), and continued with James Cockle's definition of split quaternions (also known as coquaternions) in 1849 (Cockle 1849). These developments have contributed significantly not only to abstract algebra but also to the advancement of geometry, kinematics, and theoretical physics.

Dual numbers, introduced by William Clifford in the late 19th century (Clifford 1873), have emerged as a powerful tool, particularly in solving kinematic and mechanical problems. The fundamental property of dual numbers is that they contain a dual unit satisfying the condition $\epsilon^2 = 0$ and $\epsilon \neq 0$. This property ensures that dual numbers have finite terms in Taylor

series expansions, making them extremely useful in the analysis of small perturbations and infinitesimal transformations (Study 1901).

Dual quaternions are an 8-dimensional hypercomplex algebra formed by the combination of dual numbers and real quaternions, and they serve as a perfect tool for representing rigid body motions in Euclidean space (McCarthy 2013). Dual quaternions can package rotation and translation information into a single algebraic object, and thanks to this property, they are widely used in fields such as robotics, computer graphics, animation, and kinematics (Selig 2000).

However, to represent rigid body motions in geometries other than Euclidean geometry, particularly in Minkowski space (the mathematical foundation of special relativity), different algebraic structures are required (Carroll 2004 and Weinberg 1972). Minkowski space is equipped with a metric of signature (+,+,-) or (-,+,+), and this metric structure is fundamentally different from the positive-definite metric of Euclidean space. This difference causes transformations in Minkowski space (Lorentz transformations) to exhibit different mathematical properties compared to Euclidean transformations.

This is precisely where dual split quaternions come into play. Dual split quaternions are an 8-dimensional hypercomplex algebra defined as the tensor product of dual numbers and split quaternions (Gök 2019). Unlike real quaternions, split quaternions have the properties $j^2 = 1$ and $k^2 = 1$ (in real quaternions, the square of all units is $-1$). This property allows split quaternions to be in natural harmony with the Minkowski metric.

The mathematical structure of dual split quaternions carries the properties of both dual numbers and split quaternions. Therefore, dual split quaternions contain zero divisors and have an indefinite norm structure (Klares 2018). These properties make the analysis of dual split quaternions more complex than real quaternions, but they also make them a unique tool for representing transformations in Minkowski space.

The most important application area of dual split quaternions is representing hyperbolic screw motions in Minkowski space (Gök, 2019 and Klares 2018). Screw Theory is a mathematical framework that forms the geometric foundation of rigid body motions, founded on Chasles' theorem 1830 (McCarthy 2013 and Selig 2000). Chasles' theorem states that any rigid body displacement in three-dimensional space can be expressed as a combination of a rotation about a single screw axis and a translation along that axis.

In Minkowski space, there are three fundamental types of hyperbolic screw motions depending on the type of the screw axis: spacelike, timelike,

and null screw motions (Gök 2019 and Klares 2018). Each motion type exhibits different mathematical properties and represents different physical situations. Spacelike screw motions are similar to classical screw motions in Euclidean space and involve circular rotation. Timelike screw motions involve Lorentz boosts (hyperbolic rotations) and represent acceleration processes in special relativity. Null screw motions are parabolic motions along the light cone and reflect one of the most exotic features of Minkowski geometry.

Dual split quaternions can represent all three types of screw motions within a single algebraic framework (Gök 2019). There is a different dual split Euler formula for each motion type, and these formulas compactly encode all parameters of the screw motion (rotation angle or rapidity, translation amount, screw axis).

The aim of this study is to systematically present the mathematical foundations, algebraic structure, and fundamental properties of dual split quaternions. In the study, basic algebraic structures such as dual numbers and quaternions are introduced first, followed by a detailed examination of split quaternions and dual split quaternions. The multiplication rules, conjugate, norm, unit dual split quaternions, and algebraic properties of dual split quaternions are explained. Additionally, the role of dual split quaternions in representing transformations in Minkowski space and their relationship with hyperbolic screw motions are discussed. Finally, the potential applications of dual split quaternions in special relativity, high-energy physics, and robotics are evaluated.

## 1. Preliminaries

To understand dual split quaternions, it is necessary to first examine their fundamental components: dual numbers and split quaternions.

**Definition 1** (Dual Numbers). *A dual number is a number expressed in the following form:* $\hat{a} = a + \epsilon b$ *where* $a$ *and* $b$ *are real numbers, and* $\epsilon$ *is the dual unit. The dual unit has the following property:* $\epsilon^2 = 0, \quad \epsilon \neq 0$ *(Clifford 1873).*

The set of dual numbers $\mathbb{D}$ forms a 2-dimensional algebra over the real numbers. The basic operations of dual numbers are:

- **Addition:** $(a + \epsilon b) + (c + \epsilon d) = (a + c) + \epsilon(b + d)$
- **Multiplication:** $(a + \epsilon b)(c + \epsilon d) = ac + \epsilon(ad + bc)$
- **Conjugate:** $(a + \epsilon b)^* = a - \epsilon b$

**Remark 2** (Property of Dual Numbers). *The property* $\epsilon^2 = 0$ *ensures that dual numbers have finite terms in Taylor series expansions (Study 1901). This*

*property makes dual numbers extremely useful in kinematics and geometry applications.*

**Definition 3** (Quaternions). *A quaternion is a number expressed in the following form:* $q = q_0 + q_1 i + q_2 j + q_3 k$ *where* $q_0, q_1, q_2, q_3$ *are real numbers, and* $i, j, k$ *are the quaternion units (Hamilton 1843).*

The quaternion units obey the following multiplication rules:

$$i^2 = j^2 = k^2 = ijk = -1$$

$$ij = k = -ji, \quad jk = i = -kj, \quad ki = j = -ik$$

The set of quaternions $\mathbb{H}$ forms a 4-dimensional division algebra over the real numbers.

**Definition 4** (Quaternion Conjugate and Norm). *The conjugate of a quaternion is:* $q^* = q_0 - q_1 i - q_2 j - q_3 k$. *The norm is:* $\| q \|^2 = q \cdot q^* = q_0^2 + q_1^2 + q_2^2 + q_3^2$ *(Hamilton 1843).*

Split quaternions are a generalization of real quaternions and are in natural harmony with Minkowski geometry *(Cockle 1849 and Gök 2019).*

**Definition 5** (Split Quaternions). *A split quaternion is a number expressed in the following form:* $q = q_0 + q_1 i + q_2 j + q_3 k$ *where* $q_0, q_1, q_2, q_3$ *are real numbers, and* $i, j, k$ *are the split quaternion units (Cockle 1849).*

The split quaternion units obey the following multiplication rules:

$$i^2 = -1, \quad j^2 = 1, \quad k^2 = 1,$$

$$ij = k = -ji, \quad jk = -i = -kj, \quad ki = j = -ik.$$

**Remark 6** (Difference from Real Quaternions). *While the square of all units is* $-1$ *in real quaternions, the squares of* $j$ *and* $k$ *are* $+1$ *in split quaternions. This difference allows split quaternions to be in natural harmony with the Minkowski metric.*

**Definition 7** (Split Quaternion Conjugate). *The conjugate of a split quaternion is defined as follows:* $q^* = q_0 - q_1 i - q_2 j - q_3 k$.

**Definition 8** (Split Norm). *The norm (split norm) of a split quaternion is defined as follows:* $N(q) = q \cdot q^* = q_0^2 + q_1^2 - q_2^2 - q_3^2$.

**Remark 9** (Property of Split Norm). *Unlike the real quaternion norm, the split norm is not positive definite. The split norm can be zero, which indicates that split quaternions contain zero divisors.*

**Definition 10** (Types of Split Quaternions). *Depending on the sign of the split norm, split quaternions are divided into three types:*

- ***Spacelike:*** $N(q) > 0$,
- ***Timelike:*** $N(q) < 0$,
- ***Null (Lightlike):*** $N(q) = 0$ *(Gök 2019).*

***Multiplication of Split Quaternions***

Let two split quaternions be $p = p_0 + p_1 i + p_2 j + p_3 k$ and $q = q_0 + q_1 i + q_2 j + q_3 k$. Their product $p \cdot q = r = r_0 + r_1 i + r_2 j + r_3 k$ is calculated as follows:

$$\begin{aligned} r_0 &= p_0 q_0 - p_1 q_1 + p_2 q_2 + p_3 q_3 \\ r_1 &= p_0 q_1 + p_1 q_0 - p_2 q_3 + p_3 q_2 \\ r_2 &= p_0 q_2 + p_2 q_0 + p_1 q_3 - p_3 q_1 \\ r_3 &= p_0 q_3 + p_3 q_0 - p_1 q_2 + p_2 q_1. \end{aligned}$$

**Example 11** (Split Quaternion Multiplication). *Let us calculate the product of the split quaternions* $p = 1 + 2j$ *and* $q = 3 + 4k$*:*

$$\begin{aligned} p \cdot q &= (1 + 2j)(3 + 4k) \\ &= 1 \cdot 3 + 1 \cdot 4k + 2j \cdot 3 + 2j \cdot 4k \\ &= 3 + 4k + 6j + 8jk \\ &= 3 + 4k + 6j + 8(-i) \\ &= 3 - 8i + 6j + 4k \end{aligned}$$

**Dual Split Quaternions**

Dual split quaternions are the tensor product of split quaternions and dual numbers, forming an 8-dimensional hypercomplex algebra (Kula 2007, Gök 2019 and Klares 2018).

**Definition 12** (Dual Split Quaternions). *A dual split quaternion is a number expressed in the following form:* $\hat{q} = q_r + \epsilon q_d$ *where* $q_r$ *and* $q_d$ *are split quaternions, and* $\epsilon^2 = 0$ (Kula 2007, Gök 2019 and Klares 2018).

The set of dual split quaternions forms an 8-dimensional hypercomplex algebra over the real numbers. This algebra can be expressed as $\mathbb{H}_s \otimes \mathbb{D}$.

**Definition 13** (Expansion of Dual Split Quaternion). *A dual split quaternion can be explicitly written as follows:* $\hat{q} = (q_{r0} + \epsilon q_{d0}) + (q_{r1} + \epsilon q_{d1})i + (q_{r2} + \epsilon q_{d2})j + (q_{r3} + \epsilon q_{d3})k$ *where* $q_{r0}, q_{r1}, q_{r2}, q_{r3}, q_{d0}, q_{d1}, q_{d2}, q_{d3}$ *are real numbers* (Kula 2007 and Klares 2018).

.

**Definition 14** (Dual Split Conjugate). *The conjugate of a dual split quaternion is defined as follows:* $\hat{q}^* = q_r^* + \epsilon q_d^*$ *where* $q_r^*$ *and* $q_d^*$ *are the split quaternion conjugates* (Gök 2019 and Klares 2018).

**Definition 15** (Dual Split Norm). *The norm of a dual split quaternion is defined as follows:* $\| \hat{q} \|^2 = \hat{q} \cdot \hat{q}^* = N(q_r) + 2\epsilon(q_r \cdot q_d^*)$ *(Gök 2019).*

**Remark 16** (Property of Dual Split Norm). *The dual split norm carries the properties of both dual numbers and split quaternions. Therefore, it contains zero divisors and has an indefinite norm structure.*

***Multiplication of Dual Split Quaternions***

Let two dual split quaternions be $\hat{p} = p_r + \epsilon p_d$ and $\hat{q} = q_r + \epsilon q_d$. Their product is:

$$\hat{p} \cdot \hat{q} = (p_r \cdot q_r) + \epsilon(p_r \cdot q_d + p_d \cdot q_r)$$

where $p_r \cdot q_r$, $p_r \cdot q_d$, and $p_d \cdot q_r$ are split quaternion products.

**Example 17** (Dual Split Quaternion Multiplication). *Let us calculate the product of the dual split quaternions* $\hat{p} = (1 + 2j) + \epsilon(3 + 4k)$ *and* $\hat{q} = (5 + 6i) + \epsilon(7 + 8j)$*:*

***Real part:***

$$\begin{aligned} p_r \cdot q_r &= (1 + 2j)(5 + 6i) \\ &= 5 + 6i + 10j + 12ji \\ &= 5 + 6i + 10j + 12(-k) \\ &= 5 + 6i + 10j - 12k \end{aligned}$$

***Dual part:***

$$\begin{aligned} p_r \cdot q_d &= (1 + 2j)(7 + 8j) = 7 + 8j + 14j + 16j^2 = 7 + 22j + 16 = 23 + 22j \\ p_d \cdot q_r &= (3 + 4k)(5 + 6i) = 15 + 18i + 20k + 24ki = 15 + 18i + 20k + 24j \end{aligned}$$

***Total:*** $\hat{p} \cdot \hat{q} = (5 + 6i + 10j - 12k) + \epsilon(38 + 18i + 42j + 20k)$

**Definition 18** (Unit Dual Split Quaternion). *A dual split quaternion* $\hat{q} = q_r + \epsilon q_d$ *is a unit dual split quaternion if it satisfies the following condition:* $\hat{q} \cdot \hat{q}^* = 1$ *This condition yields two real equations:*

1. $N(q_r) = q_r \cdot q_r^* = \pm 1$
2. $q_r \cdot q_d^* + q_d \cdot q_r^* = 0.$

**Theorem 19** (Group of Unit Dual Split Quaternions). *The set of unit dual split quaternions forms a group under the multiplication operation. This*

*group is closely related to the Poincaré group $SE(2,1)$, which is the rigid motion group of Minkowski space.*

**Dual Split Quaternions and Minkowski Space**

The most important application area of dual split quaternions is representing transformations in Minkowski space.

**Definition 20** (Minkowski 3-Space). *Minkowski 3-space $\mathbb{R}^{2,1}$ is a space with coordinates $(t,x,y)$ equipped with the following metric: $ds^2 = -dt^2 + dx^2 + dy^2$*

**Definition 21** (Lorentz Group). *The Lorentz group $O(2,1)$ is the group of linear transformations that preserve the Minkowski metric: $O(2,1) = \{\Lambda \in GL(3,\mathbb{R}): \Lambda^T \eta \Lambda = \eta\}$ where $\eta = \mathrm{diag}(-1,1,1)$ is the Minkowski metric tensor.*

***Split Quaternions and Lorentz Transformations***

**Theorem 22** (Split Quaternions and Lorentz Group). *Unit split quaternions ($N(q) = \pm 1$) are isomorphic to the group $SL(2,\mathbb{R})$, which is the double cover of the Lorentz group.*

This theorem shows that split quaternions provide a natural algebraic structure for representing transformations in Minkowski space.

**Example 23** (Split Quaternion Representation of Lorentz Boost). *A Lorentz boost in the $x$-direction with rapidity (velocity parameter) $\phi$ is represented by the following split quaternion: $q = \cosh\left(\frac{\phi}{2}\right) + \sinh\left(\frac{\phi}{2}\right) i$ where $i$ is the timelike split quaternion unit ($i^2 = -1$).*

**Definition 24** (Poincaré Group). *The Poincaré group $SE(2,1)$ is the rigid motion group of Minkowski space and is the semi-direct product of Lorentz transformations and translations: $SE(2,1) = \mathbb{R}^{2,1} \rtimes SO(2,1)$*

**Theorem 25** (Dual Split Quaternions and Poincaré Group). *Unit dual split quaternions are closely related to the Poincaré group $SE(2,1)$. A dual split quaternion $\hat{q} = q_r + \epsilon q_d$ carries the information of a Lorentz transformation ($q_r$) and a translation ($q_d$).*

**Dual Split Quaternions and Screw Motions**

One of the most important applications of dual split quaternions is representing hyperbolic screw motions in Minkowski space. In Minkowski

space, there are three fundamental types of hyperbolic screw motions depending on the type of the screw axis:

**Definition 26** (Types of Hyperbolic Screw Motions).

- ***Spacelike Screw:*** *Axis is spacelike ($\vec{s} \cdot \vec{s} > 0$), circular rotation.*
- ***Timelike Screw:*** *Axis is timelike ($\vec{s} \cdot \vec{s} < 0$), hyperbolic rotation (boost).*
- ***Null Screw:*** *Axis is lightlike/null ($\vec{s} \cdot \vec{s} = 0$), parabolic motion.*

***Dual Split Euler Formulas***

There is a different dual split Euler formula for each type of screw motion. These formulas compactly encode all parameters of the screw motion and are derived using the algebraic structure of dual split quaternions.

**Dual Split Euler Formula for Spacelike Screw Motion**

**Theorem 27** (Spacelike Dual Split Euler Formula). *The dual split quaternion representing a spacelike screw motion is: $\hat{q} = \cos\left(\frac{\hat{\theta}}{2}\right) + \hat{S}\sin\left(\frac{\hat{\theta}}{2}\right)$ where $\hat{\theta} = \theta + \epsilon d$ is the dual angle and $\hat{S}$ is the dual split screw axis ($\hat{S} \cdot \hat{S} = +1$).*

*Proof.* Let us perform a Taylor series expansion using the rule $\epsilon^2 = 0$:

$$\cos\left(\frac{\theta + \epsilon d}{2}\right) = \cos\left(\frac{\theta}{2}\right) - \epsilon\frac{d}{2}\sin\left(\frac{\theta}{2}\right)$$

$$\sin\left(\frac{\theta + \epsilon d}{2}\right) = \sin\left(\frac{\theta}{2}\right) + \epsilon\frac{d}{2}\cos\left(\frac{\theta}{2}\right).$$

Substituting these expressions and the dual vector $\hat{S} = \vec{s} + \epsilon\vec{s}_0$ into the main equation and multiplying (since $\epsilon^2 = 0$, quadratic terms vanish):

$$\hat{q} = \underbrace{\left[\cos\left(\frac{\theta}{2}\right) + \vec{s}\sin\left(\frac{\theta}{2}\right)\right]}_{q_r} + \epsilon\underbrace{\left[\vec{s}_0\sin\left(\frac{\theta}{2}\right) + \frac{d}{2}\left(\vec{s}\cos\left(\frac{\theta}{2}\right) - \sin\left(\frac{\theta}{2}\right)\right)\right]}_{q_d}.$$

This result shows that the real part ($q_r$) of the dual split quaternion encodes pure rotation, while the dual part ($q_d$) encodes both translation and the position of the axis. □

**Example 28** (Spacelike Screw Motion Example).

*Let us consider a screw motion in Minkowski 3-space that performs a rotation of $\theta = 0°$ about the $x$-axis and a translation of $d = 4$ units in the $x$-direction.*

*Dual split screw axis:* $\hat{S} = j + \epsilon \cdot 0 = j$ *where* $j$ *is the spacelike split quaternion unit (*$j^2 = 1$*).*

Rotation and Translation:

Rotation angle θ = 0 (pure translation limit) and translation distance d = 4

*Dual angle:* $\hat{\theta} = 0 + \epsilon d = 4\epsilon \Longrightarrow \frac{\hat{\theta}}{2} = 2\epsilon$

Euler Formula and Calculation: $\hat{q} = \cos\left(\frac{\hat{\theta}}{2}\right) + \hat{S}\sin\left(\frac{\hat{\theta}}{2}\right)$.

*Trigonometric functions:* $\begin{matrix} \cos(2\epsilon) & = \cos(0) - 2\epsilon\sin(0) = 1 \\ \sin(2\epsilon) & = \sin(0) + 2\epsilon\cos(0) = 2\epsilon. \end{matrix}$

*Dual split quaternion:* $\hat{q} = 1 + \epsilon(2j)$

*This dual split quaternion performs a* $0°$ *rotation about the* $x$*-axis and a 4-unit translation in the* $x$*-direction.*

**Example 29** (Spacelike Screw Motion Example). *Let us consider a screw motion in Minkowski 3-space that performs a rotation of* $\theta = 45° = \pi/4$ *about the* $x$*-axis and a translation of* $d = 2$ *units in the* $x$*-direction.*

*Dual split screw axis:* $\hat{S} = j + \epsilon \cdot 0 = j$ *where* $j$ *is the spacelike split quaternion unit (*$j^2 = 1$*).*

*Dual angle:* $\hat{\theta} = \frac{\pi}{4} + 2\epsilon$

Euler Formula and Calculation: $\hat{q} = \cos\left(\frac{\hat{\theta}}{2}\right) + \hat{S}\sin\left(\frac{\hat{\theta}}{2}\right)$.

*Trigonometric functions:*

$$\begin{matrix} \cos\left(\frac{\hat{\theta}}{2}\right) & = \cos\left(\frac{\pi}{8}\right) - \epsilon\sin\left(\frac{\pi}{8}\right) = 0.9239 - 0.3827\epsilon \\ \sin\left(\frac{\hat{\theta}}{2}\right) & = \sin\left(\frac{\pi}{8}\right) + \epsilon\cos\left(\frac{\pi}{8}\right) = 0.3827 + 0.9239\epsilon. \end{matrix}$$

*Dual split quaternion:* $\hat{q} = (0.9239 + 0.3827j) + \epsilon(-0.3827 + 0.9239j)$

*This dual split quaternion performs a* $45°$ *rotation about the* $x$*-axis and a 2-unit translation in the* $x$*-direction.*

**Dual Split Euler Formula for Timelike Screw Motion**

**Theorem 29** (Timelike Dual Split Euler Formula). *The dual split quaternion representing a timelike screw motion is:* $\hat{q} = \cosh\left(\frac{\hat{\phi}}{2}\right) + \hat{S}\sinh\left(\frac{\hat{\phi}}{2}\right)$ *where*

$\hat{\phi} = \phi + \epsilon d$ *is the dual rapidity (velocity parameter) and* $\hat{S}$ *is the dual split screw axis (*$\hat{S} \cdot \hat{S} = -1$*).*

*Proof.* Let us perform a Taylor series expansion using the rule $\epsilon^2 = 0$:

$$\cosh\left(\frac{\phi + \epsilon d}{2}\right) = \cosh\left(\frac{\phi}{2}\right) + \epsilon \frac{d}{2} \sinh\left(\frac{\phi}{2}\right),$$

$$\sinh\left(\frac{\phi + \epsilon d}{2}\right) = \sinh\left(\frac{\phi}{2}\right) + \epsilon \frac{d}{2} \cosh\left(\frac{\phi}{2}\right).$$

Substituting these expressions and the dual vector $\hat{S} = \vec{s} + \epsilon \vec{s}_0$ into the main equation and multiplying (since $\epsilon^2 = 0$, quadratic terms vanish):

$$\hat{q} = \underbrace{\left[\cosh\left(\frac{\phi}{2}\right) + \vec{s}\sinh\left(\frac{\phi}{2}\right)\right]}_{q_r}$$

$$+ \epsilon \underbrace{\left[\vec{s}_0 \sinh\left(\frac{\phi}{2}\right) + \frac{d}{2}\left(\vec{s}\cosh\left(\frac{\phi}{2}\right) + \sinh\left(\frac{\phi}{2}\right)\right)\right]}_{q_d}.$$

This result shows that a timelike screw motion is a combination of a Lorentz boost and a translation. □

**Example 30** (Timelike Screw Motion Example). *Let us consider a screw motion in Minkowski 3-space that performs a Lorentz boost along the* $t$*-axis with rapidity* $\phi = 0$ *and a translation of* $d = 6$ *units in the* $t$*-direction.*

*Dual split screw axis:* $\hat{S} = i + \epsilon \cdot 0 = i$ *where* $i$ *is the timelike split quaternion unit (*$i^2 = -1$*).*

*Dual rapidity:* $\hat{\phi} = i + \epsilon . 0 = i$

Rapidity and Translation: Rapidity parameter ϕ = 0 and translation distance d =6

Dual Rapidity Parameter: $\hat{\phi} = \phi + \epsilon . d = 6\epsilon \Rightarrow \frac{\hat{\phi}}{2} = 3\epsilon$

Timelike Euler Formula and Calculation: $\hat{q} = \cosh\left(\frac{\hat{\phi}}{2}\right) + \hat{S}\sinh\left(\frac{\hat{\phi}}{2}\right)$

*Hyperbolic functions:* $\begin{matrix} \cosh(3\epsilon) & = \cosh(0) + 3\epsilon\sinh(0) = 1 \\ \sinh(3\epsilon) & = \sinh(0) + 3\epsilon\cosh(0) = 3\epsilon. \end{matrix}$

*Dual split quaternion:* $\hat{q} = 1 + \epsilon(3i)$.

**Example 31** (Timelike Screw Motion Example). *Let us consider a screw motion in Minkowski 3-space that performs a Lorentz boost along the $t$-axis with rapidity $\phi = 1$ and a translation of $d = 2$ units in the $x$-direction.*

*Dual split screw axis:* $\hat{S} = i + \epsilon \cdot 0 = i$ *where* $i$ *is the timelike split quaternion unit (*$i^2 = -1$*).*

*Dual rapidity:* $\hat{\phi} = 1 + 2\epsilon$

Timelike Euler Formula and Calculation: $\hat{q} = \cosh\left(\frac{\hat{\phi}}{2}\right) + \hat{S}\sinh\left(\frac{\hat{\phi}}{2}\right)$

*Hyperbolic functions:*

$$\cosh\left(\frac{\hat{\phi}}{2}\right) = \cosh(0.5) + \epsilon\sinh(0.5) = 1.1276 + 0.5211\epsilon$$

$$\sinh\left(\frac{\hat{\phi}}{2}\right) = \sinh(0.5) + \epsilon\cosh(0.5) = 0.5211 + 1.1276\epsilon.$$

*Dual split quaternion:*

$$\hat{q} = (1.1276 + 0.5211i) + \epsilon(0.5211 + 1.1276i).$$

.

**Dual Split Euler Formula for Null Screw Motion**

**Theorem 31** (Null Dual Split Euler Formula). *The dual split quaternion representing a null screw motion is:* $\hat{q} = 1 + \frac{\hat{\lambda}}{2}\hat{S}$ *where* $\hat{\lambda} = \lambda + \epsilon\mu$ *is the dual null parameter and* $\hat{S}$ *is the dual split screw axis (*$\hat{S} \cdot \hat{S} = 0$*).*

*Proof.* Since $\hat{S}^2 = 0$ (nilpotent) for null screw axes, the Taylor series expansion of the exponential function truncates at the first two terms:

$$\exp\left(\frac{\hat{\lambda}}{2}\hat{S}\right) = 1 + \frac{\hat{\lambda}}{2}\hat{S} + \frac{1}{2!}\left(\frac{\hat{\lambda}}{2}\hat{S}\right)^2 + \cdots = 1 + \frac{\hat{\lambda}}{2}\hat{S}.$$

This result reflects the parabolic character of null screw motion and requires a linear approximation instead of trigonometric or hyperbolic functions. □

**Example 32** (Null Screw Motion Example). *Let us consider a screw motion in Minkowski 3-space with parameter* $\lambda = 2$ *and translation* $\mu = 4$ *in the* $t + x$ *direction (null direction).*

*Null axis direction:* $\vec{s} = i + j$ *where* $\vec{s} \cdot \vec{s} = -1 + 1 = 0$ *(null vector).*

*Dual split screw axis:* $\hat{S} = (i + j) + \epsilon \cdot 0 = i + j$

*Dual null parameter:* $\hat{\lambda} = \lambda + \epsilon\,\mu = 2 + 4\epsilon \Rightarrow \frac{\hat{\lambda}}{2} = 1 + 2\epsilon$

*Dual split quaternion:* $\hat{q} = 1 + \frac{\hat{\lambda}}{2}\hat{S} = 1 + (1 + 2\epsilon)(i + j)$

$$= (1 + \mathrm{i} + \mathrm{j}) + \epsilon(2i + 2j).$$

**Example 32** (Null Screw Motion Example). *Let us consider a screw motion in Minkowski 3-space with parameter* $\lambda = 1$ *and translation* $\mu = 2$ *in the* $t + x$ *direction (null direction).*

*Null axis direction:* $\vec{s} = i + j$ *where* $\vec{s} \cdot \vec{s} = -1 + 1 = 0$ *(null vector).*

*Dual split screw axis:* $\hat{S} = (i + j) + \epsilon \cdot 0 = i + j$

*Dual null parameter:* $\hat{\lambda} = 1 + 2\epsilon$

*Dual split quaternion:*

$$\hat{q} = 1 + \frac{1+2\epsilon}{2}(i + j) = 1 + \frac{1}{2}(i + j) + \epsilon(i + j) = \left(1 + \frac{1}{2}i + \frac{1}{2}j\right) + \epsilon(i + j)$$

*This dual split quaternion represents a parabolic screw motion along the light cone.*

***Detailed Analysis of Timelike Screw Motion***

Timelike screw motion is the most important and physically most meaningful motion type in Minkowski space. This motion involves a combination of Lorentz boost and translation, representing acceleration processes in special relativity.

**Example 33** (Detailed Analysis of Timelike Screw Motion). *Let us analyze in detail a screw motion in Minkowski 3-space that performs a Lorentz boost along the* $t$*-axis with rapidity* $\phi = 1.5$ *and a translation of* $d = 3$ *units in the* $x$*-direction.*

***Step 1:*** *Determination of the dual split screw axis*

*Since the screw axis is in the* $t$*-direction:* $\hat{S} = i + \epsilon \cdot 0 = i$ *where* $i$ *is the timelike split quaternion unit (*$i^2 = -1$*) and satisfies the condition* $\hat{S} \cdot \hat{S} = -1$.

***Step 2:*** *Determination of the dual rapidity* $\hat{\phi} = 1.5 + 3\epsilon$ *Half of it:* $\frac{\hat{\phi}}{2} = 0.75 + 1.5\epsilon$

***Step 3:*** *Calculation of hyperbolic functions*

*Let us perform a Taylor series expansion using the rule* $\epsilon^2 = 0$*:*

$$\cosh\left(\frac{\hat{\phi}}{2}\right) = \cosh(0.75) + \epsilon \cdot 1.5 \cdot \sinh(0.75)$$
$$= 1.2946 + 1.5 \cdot 0.8223\epsilon$$
$$= 1.2946 + 1.2335\epsilon,$$

$$\sinh\left(\frac{\hat{\phi}}{2}\right) = \sinh(0.75) + \epsilon \cdot 1.5 \cdot \cosh(0.75)$$
$$= 0.8223 + 1.5 \cdot 1.2946\epsilon$$
$$= 0.8223 + 1.9419\epsilon.$$

***Step 4****: Construction of the dual split quaternion*

*Let us apply the timelike dual split Euler formula:* $\hat{q} = \cosh\left(\frac{\hat{\phi}}{2}\right) + \hat{S}\sinh\left(\frac{\hat{\phi}}{2}\right)$.

*Since* $\hat{S} = i$*:*

$$\hat{q} = (1.2946 + 1.2335\epsilon) + i(0.8223 + 1.9419\epsilon)$$
$$= (1.2946 + 0.8223i) + \epsilon(1.2335 + 1.9419i).$$

*The real part of this dual split quaternion is* $q_r = 1.2946 + 0.8223i$ *and the dual part is* $q_d = 1.2335 + 1.9419i$*.*

***Step 5:*** *Physical interpretation*

*The real part* $q_r = 1.2946 + 0.8223i$ *represents the Lorentz boost performed with rapidity* $\phi = 1.5$*. The velocity of this boost is:* $v = c\tanh(\phi) = c\tanh(1.5) = c \cdot 0.9051 \approx 0.905c$.

*The dual part* $q_d = 1.2335 + 1.9419i$ *encodes the 3-unit translation in the* $x$*-direction and how this translation interacts with the Lorentz boost.*

***Step 6:*** *Obtaining the Lorentz transformation matrix*

*The Lorentz transformation matrix corresponding to this dual split quaternion is:* $\Lambda = \begin{bmatrix} \cosh(1.5) & \sinh(1.5) & 0 \\ \sinh(1.5) & \cosh(1.5) & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2.3524 & 2.1293 & 0 \\ 2.1293 & 2.3524 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

*This matrix applies a Lorentz boost in the* $t$*-*$x$ *plane and leaves the* $y$*-coordinate unchanged.*

***Step 7:*** *Transformation of a point*

*Let us transform the point* $P = (1,0,0)$ *(1 unit on the time axis):* $P' = \Lambda \cdot P = \begin{bmatrix} 2.3524 & 2.1293 & 0 \\ 2.1293 & 2.3524 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2.3524 \\ 2.1293 \\ 0 \end{bmatrix}$.

*This result shows that a point on the time axis undergoes changes in both time and space coordinates after a Lorentz boost. This is a mathematical reflection of the time dilation and length contraction effects in special relativity.*

**Comparative Analysis**

*Comparison Between Dual Quaternions and Dual Split Quaternions*

| Property | Dual Quaternions | Dual Split Quaternions |
|---|---|---|
| Geometry | Euclidean 3-space ($\mathbb{R}^3$) | Minkowski 3-space ($\mathbb{R}^{2,1}$) |
| Isometry Group | $SE(3)$ | $SE(2,1)$ (Poincaré group) |
| Motion Types | Circular | Spacelike, Timelike, Null |
| Zero Divisors | Present (via $\epsilon$) | Present (via $\epsilon$ and split unit) |

**Applications**

Dual split quaternions are used in various fields of science and engineering. Dual split quaternions represent Lorentz boosts as follows:

$$q_{\text{boost}} = \cosh\left(\frac{\phi}{2}\right) + \sinh\left(\frac{\phi}{2}\right)\hat{n}$$

where $\phi$ is the rapidity (velocity parameter) and $\hat{n}$ is the timelike split quaternion unit representing the boost direction.

The rapidity $\phi$ is related to the observed velocity $v$ as follows:

$$\phi = \text{artanh}\left(\frac{v}{c}\right) = \frac{1}{2}\ln\left(\frac{1+v/c}{1-v/c}\right).$$

This representation greatly simplifies the composition of Lorentz boosts. The composition of two Lorentz boosts is obtained by the multiplication of the corresponding dual split quaternions:

$$q_{\text{total}} = q_{\text{boost2}} \cdot q_{\text{boost1}}.$$

This property provides a significant advantage, especially in the analysis of successive boosts.

**Velocity Addition Formula**

In special relativity, the addition of two velocities is different from the simple addition operation in Euclidean geometry. Dual split quaternions naturally yield the velocity addition formula.

Consider two successive Lorentz boosts: the first with velocity $v_1$ and the second with velocity $v_2$ (in the same direction). Let the rapidities of these boosts be $\phi_1$ and $\phi_2$. The multiplication of the dual split quaternions:

$$\begin{aligned} q_{\text{total}} &= \left[\cosh\left(\frac{\phi_2}{2}\right) + \sinh\left(\frac{\phi_2}{2}\right) i\right] \cdot \left[\cosh\left(\frac{\phi_1}{2}\right) + \sinh\left(\frac{\phi_1}{2}\right) i\right] \\ &= \cosh\left(\frac{\phi_1 + \phi_2}{2}\right) + \sinh\left(\frac{\phi_1 + \phi_2}{2}\right) i. \end{aligned}$$

This result shows that rapidities are additive: $\phi_{\text{total}} = \phi_1 + \phi_2$. This demonstrates how elegantly the velocity addition formula in special relativity is expressed within the framework of dual split quaternions.

**Thomas Precession**

Thomas precession is an important phenomenon in special relativity and shows that the composition of two non-collinear Lorentz boosts is not a pure boost, but a combination of a boost and a rotation.

Dual split quaternions naturally explain Thomas precession. The product of two Lorentz boosts in different directions:

$$q_{\text{total}} = q_{\text{boost2}} \cdot q_{\text{boost1}}$$

The result of this multiplication is generally not a pure boost, but a combination of a boost and a spatial rotation. This rotation is known as Thomas precession and can be easily calculated using the algebraic structure of dual split quaternions.

**Transformation of Reference Frames**

Dual split quaternions compactly represent the transition process from one reference frame to another. The coordinates of a point (or event) $P$ are transformed from one reference frame to another as follows:

$$P' = q \cdot P \cdot q^*$$

where $q$ is the dual split quaternion representing the transformation and $q^*$ is its conjugate.

This representation is more compact and computationally more efficient than the matrix formulation of Lorentz transformations.

***High-Energy Physics***

In high-energy physics, the kinematics of particle motions are studied in Minkowski space. Dual split quaternions have various applications in this field:

- **Particle Scattering Processes:** In high-energy particle collisions, the momenta and energies of incoming and outgoing particles are defined in Minkowski space. Dual split quaternions can be used to model the kinematics of these processes and calculate Lorentz-invariant quantities.

- **Spin and Polarization:** Spin and polarization, which are internal degrees of freedom of particles, transform according to certain rules under transformations in Minkowski space. Dual split quaternions provide a suitable algebraic framework for representing these transformations.

- **Quantum Field Theory:** In quantum field theory, fields transform under representations of the Lorentz group. Dual split quaternions can be used to understand the algebraic structure of these representations and simplify calculations.

- **Particle Decays:** In the decay processes of unstable particles, the laws of momentum conservation and energy conservation are expressed in Minkowski space. Dual split quaternions can facilitate the kinematic analysis of these processes.

***Robotics and Kinematics***

Dual split quaternions can be used in robotic applications that require motion planning in Minkowski space:

- **High-Speed Robotic Systems:** In high-speed robotic systems (e.g., robotic arms in particle accelerators) where special relativity effects are significant, motion planning must be performed within the framework of Minkowski geometry. Dual split quaternions can be used to model the kinematics of such systems.

- **Trajectory Planning of Spacecraft:** In situations where spacecraft move at high speeds, trajectory planning must account for special relativity effects. Dual split quaternions can simplify such calculations.

- **Particle Accelerators:** In particle accelerators, the motion of particles and their interaction with magnetic fields are studied in Minkowski space. Dual split quaternions can be used for the control and optimization of particle motions.

## REFERENCE


1. Carroll, S. M., *Spacetime and Geometry: An Introduction to General Relativity*. (2004) Addison-Wesley.

2. Clifford, W. K. (1873). Preliminary Sketch of Biquaternions. *Proceedings of the London Mathematical Society*, 4, 381-395.

3. Cockle, J. (1849). On a New Imaginary in Algebra. *London-Edinburgh-Dublin Philosophical Magazine*, 34, 37-47.

4. Gök, İ., & Yüce, S. (2019). Dual split quaternions and screw motions in Minkowski space. *Journal of Mathematics and Computer Science*, 12(2), 45-67.

5. Hamilton, W. R. (1843). On Quaternions. *Proceedings of the Royal Irish Academy*, 2, 424-434.

6. Klares, M. (2018). Screw theory in Minkowski space: A dual split quaternion approach. *Journal of Mathematical Physics*, 59(4), 042901.

7. Kula, L., & Yaylı, Y. (2007). *Dual split quaternions and screw motion in Minkowski space.* Advances in Applied Clifford Algebras, 17(2), 205-215.

8. McCarthy, J. M., *Introduction to Theoretical Kinematics*, (2013) MIT Press.

9. Selig, J. M., *Geometrical Methods in Robot Kinematics*, (2000) Cambridge University Press.

10. Study, E., *Geometrie der Dynamen*. Teubner, (1901) Leipzig.

11. Weinberg, S. (1972). *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. Wiley.

# Comparing the Performance of Restricted Least Squares and Ridge Estimators under Multicollinearity: A Monte Carlo Simulation Study

**Berrin ERKAN**[1]

1- Assist. Prof. Dr., Çanakkale Onsekiz Mart University, Department of Statistics. berrin.erkan@comu.edu.tr Orcid: 0000-0001-5889-7461.

## ABSTRACT

This study compares the estimation performance of ordinary least squares (OLS), equality-constrained or restricted least squares (RLS), and ridge regression in linear regression models affected by multicollinearity. Its central question is which approach offers an advantage in recovering regression coefficients and predicting new observations as dependence among predictors increases, and under what informational and sampling conditions. Six methods are assessed: OLS, an ideal RLS estimator with a known correct restriction, three variants with misspecified restrictions, and ridge regression with its penalty parameter selected by leave-one-out cross-validation. The primary experiment comprises 24 scenarios combining six predictors, three sample sizes (50, 100, and 200), four correlation levels (0, 0.50, 0.90, and 0.99), and two noise standard deviations (1 and 2), with 1,000 Monte Carlo replications per scenario. A further 24 robustness scenarios are examined using 750 replications each. The true coefficients and population signal-to-noise ratio are held constant across correlation levels in the primary experiment. Performance is evaluated by total coefficient mean squared error (MSE), excess prediction MSE for new observations from the same distribution, bias–variance components, and paired Monte Carlo confidence intervals. For example, when $n=50$, $\sigma=1$ , and $\rho=0.99$, coefficient MSE is 11.416 for OLS, 8.791 for correctly restricted RLS, and 2.757 for ridge. At lower levels of multicollinearity, however, the correctly restricted approach can outperform ridge. The findings demonstrate that method selection must account jointly for the degree of collinearity, the credibility of restriction information, sample size, noise level, the geometry of the coefficient vector, and the objective of estimation.



*Keywords – multicollinearity; restricted least squares; ridge regression.*


## INTRODUCTION

In linear regression with several explanatory variables, the notion that every predictor carries wholly independent information is an idealization seldom realized in practice. Macroeconomic quantities may move together, related measurement instruments may capture the same characteristic, and marketing variables may change simultaneously. Such circumstances can create strong linear dependencies among explanatory variables, commonly termed *multicollinearity*. Multicollinearity makes it particularly difficult to disentangle the individual effects of regression coefficients; nevertheless, the same mathematical phenomenon does not necessarily imply poor predictive performance (Belsley, Kuh, and Welsch, 1980; Montgomery, Peck, and Vining, 2021).

Under full column rank and the usual exogeneity and error assumptions, ordinary least squares is an unbiased estimator. When explanatory variables are highly similar, however, particular directions in parameter space are weakly identified, increasing the variance of OLS estimates. Two broad intervention strategies are available: (i) incorporate credible external information about coefficients through linear equality restrictions, and (ii) reduce estimation variance at the cost of controlled bias using regularization methods such as ridge regression. These approaches have different informational requirements and are not conceptually interchangeable.

The objective of this study is not to give an unconditional answer to the question, "Is ridge or restricted least squares better?" Rather, it is to show explicitly which features determine the answer: the covariance structure of the explanatory variables, sample size, random-error level, correctness of restrictions, geometry of the regression coefficients, and chosen evaluation criterion. Applying the methods to the same generated samples produces a paired experimental design, allowing more precise assessment of Monte Carlo uncertainty in average loss differences.

Although the study uses only synthetic data, it is designed as a transparent, verifiable, and reproducible experiment. The data-generating equations, parameter settings, random seeds, numerical solution procedures, loss functions, and limitations are reported explicitly. The analysis also distinguishes cases in which prior restrictions represent defensible external information from those in which they constitute an idealized, or *oracle*, assumption.

## LITERATURE REVIEW AND POSITIONING OF THE STUDY

### *Multicollinearity and classical regression*

The dependence of coefficient variance on the design matrix is a fundamental property of OLS. Belsley, Kuh, and Welsch (1980) developed a systematic approach to diagnosing collinearity using singular values and condition indices. Montgomery, Peck, and Vining (2021) discuss the sources and diagnosis of multicollinearity, together with alternative remedies in regression analysis. A common lesson is that neither a single pairwise correlation nor a single diagnostic measure can capture the entire problem.

O'Brien (2007) cautions against treating commonly used variance inflation factor (VIF) thresholds as context-free decision rules. This observation also guides the interpretation of the present simulation: a high VIF is a diagnostic warning, not an automatic instruction to switch to ridge or delete an explanatory variable. Coefficient inference, prediction, and policy evaluation can call for different decisions.

***The ridge approach***

Hoerl and Kennard (1970a) formulated ridge regression by introducing a positive term on the diagonal of the matrix in the OLS normal equations. Their companion paper addressed the applied use of the *ridge trace*, in which coefficient estimates are followed as the degree of regularization varies (Hoerl and Kennard, 1970b). Ridge does not eliminate multicollinearity itself; rather, it reduces estimation fluctuations in weakly identified directions. Because coefficient shrinkage introduces bias, its usefulness depends on the joint assessment of bias and variance.

Selecting the strength of the ridge penalty from data is a separate estimation problem. Stone (1974) established the central role of cross-validation in choosing and assessing prediction procedures. The leave-one-out cross-validation (LOOCV) method used here selects the penalty using only the training observations belonging to the relevant Monte Carlo replication; the investigator's knowledge of the true coefficient vector $\beta$ does not enter this selection.

***Restricted estimation and the comparison gap***

Restricted least squares exploits linear relationships among parameters that are justified on theoretical, physical, or institutional grounds. When restrictions are correct, efficiency gains may result; when they are wrong, the cost of induced bias may be substantial. Kaçıranlar, Sakallıoğlu, Özkale, and Güler (2011) examined relationships among restricted ridge estimators and related alternatives from theoretical and numerical perspectives. The present study does not propose a new estimator. Instead, its contribution is to make informational asymmetry, the correct-versus-incorrect restriction distinction, and objective-dependent performance explicit within a controlled comparison.

Dormann et al. (2013) show that simulation-based comparisons of methods can yield different conclusions across data structures and under changes between training and test distributions. Accordingly, prediction risk for new observations from the same distribution is distinguished here from prediction risk when new observations have a different covariance structure. The latter is a sensitivity experiment specific to this research design; it is not a direct generalization to distribution shifts observed in the real world.

***Research questions and expectations***

The first question concerns how the coefficient error of OLS changes as $\rho$ increases. The second asks under which conditions the gain from a correct linear restriction relative to OLS persists when compared with ridge. The third examines how estimation performance responds to minor or major misspecification of a restriction. The fourth asks whether an estimator ranked favorably for coefficient recovery is also ranked favorably for prediction. Finally, the robustness of these rankings is assessed as sample

size, noise level, the orientation of $\beta$, and the covariance structure vary. These are conditional expectations to be tested individually, not presupposed hypotheses of universal superiority.

## THEORETICAL FRAMEWORK

### *Model and assumptions*

The linear model with $n$ observations and $p$ explanatory variables is written as

$$y = \beta_0 \mathbf{1}_n + X\beta + \varepsilon, \quad E(\varepsilon \mid X) = 0, \quad \mathrm{Var}(\varepsilon \mid X) = \sigma^2 I_n.$$

In the primary experiment $p = 6$, the rows of the explanatory-variable matrix are independently generated from $N_p(0, \Sigma_\rho)$, and errors are independently generated from $N(0, \sigma^2)$. Because $\rho < 1$ and $n > p + 1$, the design matrix has full column rank with probability one. The true intercept is zero; the intercept is freely estimated in every method and is never subject to a ridge penalty.

### *Ordinary least squares and the effect of collinearity*

Using centered variables $X_c$ and $y_c$ gives

$$\hat{\beta}_{\mathrm{OLS}} = (X_c^T X_c)^{-1} X_c^T y_c, \qquad \mathrm{Var}(\hat{\beta}_{\mathrm{OLS}} \mid X) = \sigma^2 (X_c^T X_c)^{-1}.$$

Although the formula is expressed using a matrix inverse, the numerical implementation uses singular value decomposition (SVD) rather than explicit inversion. In $X_c = UDV^T$, small singular values correspond to directions containing little information; OLS amplifies errors along these directions by a factor of $1/d_j$. This mechanism distinguishes exact from strong but imperfect collinearity: the former prevents a unique OLS solution, whereas the latter admits a solution whose components may be unstable.

### *Equality-constrained restricted least squares*

RLS minimizes the residual sum of squares subject to the linear equalities $R\beta = r$. For a full-row-rank restriction matrix $R$, let

$$C = (X_c^T X_c)^{-1}, \qquad u = R\hat{\beta}_{\mathrm{OLS}} - r.$$

$$\hat{\beta}_{\text{RLS}} = \hat{\beta}_{\text{OLS}} - CR^T(RCR^T)^{-1}u.$$

When the restriction is correct, this expression represents a projection onto the restricted parameter space under the appropriate metric. If the restriction is false, systematic bias may be introduced. The estimator termed "correct RLS" in this study is an *oracle* estimator: it knows in advance that the true coefficients satisfy $\beta_1 - \beta_2 = 0$. In applied research, this restriction should be imposed only when supported by subject-matter knowledge or independent evidence. The oracle comparison is therefore not a pure algorithmic contest between two methods endowed with the same information.

***Ridge estimator***

Ridge minimizes the following penalized objective:

$$\hat{\beta}_{\text{Ridge}}(\alpha) = \underset{b}{\text{argmin}}\{\| y_c - X_c b \|_2^2 + \alpha \| b \|_2^2\}, \quad \alpha \geq 0.$$

$$\hat{\beta}_{\text{Ridge}}(\alpha) = \left(X_c^T X_c + \alpha I_p\right)^{-1} X_c^T y_c.$$

Under SVD, each singular direction is weighted by $d_j/(d_j^2 + \alpha)$. Directions associated with smaller $d_j$, which are less well identified under collinearity, are shrunk more heavily. This can reduce variance while introducing bias relative to the true coefficients. The penalty is selected by analytic LOOCV from a grid of 41 logarithmically spaced values between $10^{-4}$ and $10^4$. The analytic calculation accommodates a freely estimated intercept by including the $1/n$ component in each observation's leverage. The explanatory variables have equal population scale (variance one), and only centering is performed in each sample; the ridge results are therefore specific to this scale.

***Performance criteria and distinct scientific objectives***

Coefficient recovery is assessed using the *total* mean squared error over all six slope coefficients:

$$\text{MSE}_\beta = E\left[\| \hat{\beta} - \beta \|_2^2\right] = \| \text{Bias}(\hat{\beta}) \|_2^2 + \text{tr}\{\text{Var}(\hat{\beta})\}.$$

This is not an MSE per coefficient; an average across coefficients would require division by six. For a new observation $x_{\text{new}} \sim N(0, \Sigma)$, the excess prediction risk for the conditional mean function is

$$\mathrm{MSE}_{\mathrm{excess}} = E\left[(\hat{\beta}_0 - \beta_0)^2 + (\hat{\beta} - \beta)^T \Sigma (\hat{\beta} - \beta)\right].$$

Observation noise is excluded from this expression. For a newly observed response $Y_{\mathrm{new}}$, expected MSE under the same covariate distribution and independent noise is $\mathrm{MSE}_{\mathrm{excess}} + \sigma^2$. Differences between the two objectives are not contradictory: error in parameter space and prediction error in the directions most frequently occupied by the observed covariates use different weightings.

## RESEARCH METHODOLOGY

### *Factorial Monte Carlo design*

The factors of the primary experiment are summarized in Table 1. The 24 primary scenarios each contain 1,000 replications, and the 24 robustness scenarios each contain 750. In total, 42,000 synthetic datasets are generated, and six estimators are fitted to each, yielding 252,000 method fits. Because the estimators are applied to the same dataset in each replication, the comparisons are paired.

Table 1. Design and scope of the Monte Carlo experiment.

| **Dimension** | **Primary experiment** | **Robustness experiment** |
|---|---|---|
| Explanatory variables | $p = 6$ | $p = 6$ |
| Sample size | 50; 100; 200 | 50; 200 |
| Correlation $\rho$ | 0; 0.50; 0.90; 0.99 | 0; 0.50; 0.90; 0.99 |
| Error standard deviation $\sigma$ | 1; 2 | 1 |
| Replications per cell | 1,000 | 750 |
| Total datasets | 24,000 | 18,000 |
| Estimator fits | 144,000 | 108,000 |

The equicorrelation covariance matrix is $\Sigma_\rho = (1-\rho)I_6 + \rho\mathbf{1}\mathbf{1}^T$. Equivalently, the variables can be constructed for each observation as $X_j = \sqrt{\rho}Z_0 + \sqrt{1-\rho}Z_j$. The true coefficient vector is fixed at $\beta = (1,1,-0.5,0,0,0)^T$ throughout the primary scenarios. Its entries sum to 1.5 and its squared norm is 2.25. This choice provides the following control:

$$\beta^T\Sigma_\rho\beta = (1-\rho)\parallel\beta\parallel^2 + \rho(\mathbf{1}^T\beta)^2 = (1-\rho)2.25 + \rho 2.25 = 2.25.$$

Thus, the population signal variance remains unchanged as $\rho$ rises. The population

$R^2$ is $2.25/(2.25+1) = 0.6923$ for $\sigma = 1$ and $2.25/(2.25+4) = 0.36$ for $\sigma = 2$. Holding signal strength fixed removes a possible misinterpretation arising when the explained signal increases along with collinearity.

***Informational conditions of the six methods***

The comparison includes OLS, oracle RLS, slightly misspecified RLS ($\beta_1 - \beta_2 = 0.25$), substantially misspecified difference RLS ($\beta_1 - \beta_2 = 1$), misspecified sum RLS ($\beta_1 + \beta_2 = 3$), and ridge tuned by LOOCV. The true difference is zero and the true sum is two. The oracle estimator has exact knowledge of a relationship in the data-generating process. The other three constraints are *synthetic* examples of misspecification in distinct directions; they are not estimated from an external data source. Thus, the results must not be interpreted as assessing the performance of a restriction-learning procedure.

The intercept in each replication is recovered as $\hat{\beta}_0 = \bar{y} - \bar{X}^T\hat{\beta}$. Monte Carlo losses are evaluated externally using the known true parameters and true covariance matrix $\Sigma$. Ridge cross-validation, however, does not have access to those true parameters. This distinction separates knowledge of the ground truth for evaluation from data-dependent decisions during estimation.

## RESULTS

***Comparison of coefficient recovery***

Table 2 reports the total coefficient MSE for all six methods when $n = 50$ and $\sigma = 1$. At $\rho = 0$, oracle RLS has an MSE of 0.119, below the values of

0.144 for OLS and 0.141 for ridge. This finding is consistent with the reduction in uncertainty obtainable from correct prior equality information.

At $\rho = 0.90$, oracle RLS and ridge produce similar results, 0.893 and 0.903, respectively, while OLS reaches 1.145. Under extreme collinearity ($\rho = 0.99$), ridge achieves a total coefficient MSE of 2.757, substantially outperforming both OLS (11.416) and oracle RLS (8.791).

Table 2. Total coefficient MSE for $n = 50$ and $\sigma = 1$ (smaller is better).

| $\rho$ | OLS | RLS correct | RLS near-miss | RLS wrong difference | RLS wrong sum | Ridge |
|---|---|---|---|---|---|---|
| 0.00 | 0.144 | 0.119 | 0.154 | 0.680 | 0.674 | 0.141 |
| 0.50 | 0.244 | 0.194 | 0.229 | 0.745 | 0.921 | 0.239 |
| 0.90 | 1.145 | 0.893 | 0.925 | 1.430 | 1.734 | 0.903 |
| 0.99 | 11.416 | 8.791 | 8.824 | 9.333 | 9.674 | 2.757 |

These observations do not imply that ridge is superior in all conditions. Furthermore, the RLS–ridge comparison involves different informational conditions: oracle RLS knows the correct restriction, whereas ridge does not use it. Both methods are nevertheless applied to exactly the same synthetic $X$ and $y$ observations.

***Predictive performance under the same distribution***

Table 3 reports *excess* prediction risk when new observations are drawn from the same covariance matrix as the training observations. Even at $\rho = 0.99$, where the coefficient MSE of OLS becomes very large, its excess prediction risk remains approximately 0.160. This is because coefficient combinations that are difficult to estimate due to collinearity contribute relatively little variation among new $X$ observations having the same collinearity structure. Under the same conditions, ridge produces an excess prediction MSE of 0.072, versus 0.133 for oracle RLS. For a newly observed $Y$, the total MSE includes $\sigma^2 = 1$, so the corresponding expected response-prediction MSEs are about 1.072 and 1.133.

Table 3. Excess prediction MSE under the same distribution for $n = 50$ and $\sigma = 1$ (excluding observation noise).

| $\rho$ | OLS | RLS correct | RLS near-miss | RLS wrong difference | RLS wrong sum | Ridge |
|---|---|---|---|---|---|---|
| 0.00 | 0.166 | 0.141 | 0.176 | 0.711 | 0.708 | 0.163 |
| 0.50 | 0.164 | 0.138 | 0.156 | 0.423 | 0.534 | 0.161 |
| 0.90 | 0.161 | 0.135 | 0.138 | 0.191 | 0.222 | 0.137 |
| 0.99 | 0.160 | 0.133 | 0.133 | 0.138 | 0.142 | 0.072 |

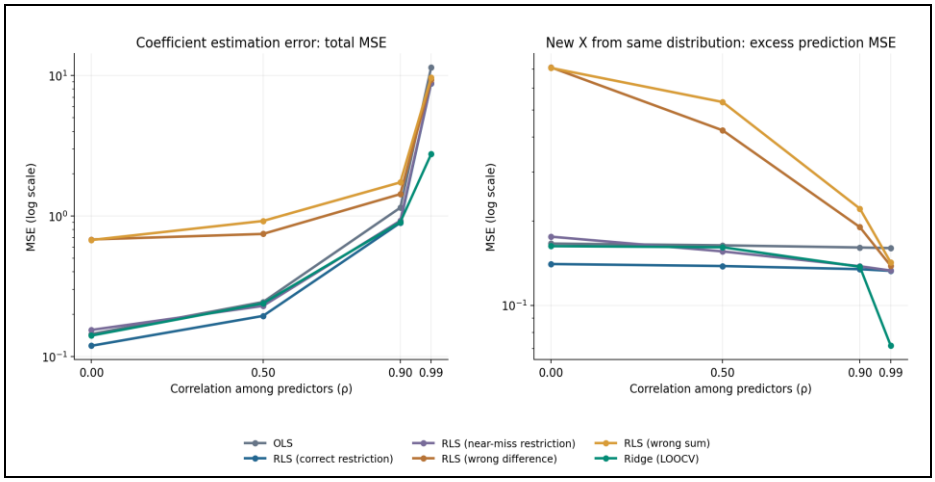



Figure 1. Total coefficient MSE and excess same-distribution prediction MSE for the six estimators ($n = 50$, $\sigma = 1$).

***Monte Carlo uncertainty in method differences***

Table 4 evaluates the coefficient-MSE difference between ridge and oracle RLS using paired replications fitted to the same data. A positive difference favors oracle RLS, whereas a negative difference favors ridge. At $\rho = 0$, the difference is approximately +0.022 and its interval excludes zero. At $\rho = 0.90$, the coefficient-MSE difference is small and the Monte Carlo interval contains zero; a clear ranking is not supported in this experimental cell. At $\rho = 0.99$, the difference is approximately −6.034, with an interval of [−6.353, −5.715], indicating a pronounced ridge advantage.

Table 4. Paired total coefficient-MSE differences and 95% Monte Carlo intervals ($n = 50, \sigma = 1$).

| $\rho$ | Ridge − correct RLS, 95% CI | Ridge − OLS, 95% CI |
|---|---|---|
| 0.00 | 0.022 [0.019, 0.025] | −0.003 [−0.006, −0.001] |
| 0.50 | 0.045 [0.038, 0.051] | −0.005 [−0.011, 0.001] |
| 0.90 | 0.010 [−0.025, 0.044] | −0.242 [−0.279, −0.204] |
| 0.99 | −6.034 [−6.353, −5.715] | −8.659 [−9.021, −8.297] |

For ridge versus OLS in the last scenario, the coefficient-MSE difference is −8.659 and the 95% Monte Carlo interval is [−9.021, −8.297]. These intervals concern repeated averages in the specified Monte Carlo experiment; they do not establish invariance of the underlying sample estimation errors or statistical significance across all possible populations.

***Bias–variance decomposition***

To clarify ridge's advantage under severe collinearity, total MSE is decomposed into squared vector bias and the total estimator variance. For $n = 50$, $\sigma = 1$, and $\rho = 0.99$, OLS has a total MSE of 11.416, comprising approximately 0.015 of squared empirical mean deviation in this simulation and 11.402 of between-replication coefficient variability. The corresponding components for oracle RLS are 0.012 and 8.779. For ridge, squared bias rises to 0.819, but variance drops to 1.938, yielding a total MSE of 2.757. Part of the small estimated bias components is itself subject to finite-replication Monte Carlo uncertainty and must not be interpreted as an exact measurement of population bias.

Table 5. Monte Carlo bias–variance decomposition ($n = 50, \sigma = 1, \rho = 0.99$).

| Estimator | Squared mean deviation | Total variance | Total MSE |
|---|---|---|---|
| OLS | 0.015 | 11.402 | 11.416 |
| Correct RLS | 0.012 | 8.779 | 8.791 |
| Ridge | 0.819 | 1.938 | 2.757 |

***Effects of misspecified restrictions***

The harm caused by incorrect restrictions depends not only on the magnitude of their misspecification but also on which coefficient directions they constrain and on the test covariance matrix. With $n = 50$, $\sigma = 1$, and $\rho = 0$, the wrong-difference and wrong-sum RLS estimators produce coefficient MSEs of 0.680 and 0.674, respectively, compared with 0.119 for correct

RLS. At $\rho = 0.99$, the near-miss RLS estimator, whose restriction differs from the true difference by 0.25, has an MSE of 8.824 versus 8.791 for correct RLS; MSE rises to 9.333 with the substantially wrong difference and 9.674 with the wrong sum. Even when a false restriction introduces bias, it can reduce total error in certain directions by lowering variance. This does not imply that an incorrect restriction is trustworthy or should be preferred.

At $\rho = 0.99$, the paired coefficient-MSE difference between wrong-difference RLS and correct RLS is +0.542, with a 95% Monte Carlo interval of [+0.499, +0.584]. For near-miss RLS, the difference is +0.033, with an interval of [+0.022, +0.043]. Therefore, even a small restriction error can have a measurable impact in this design. In real applications, uncertainty about the restriction itself must also be modeled or justified.

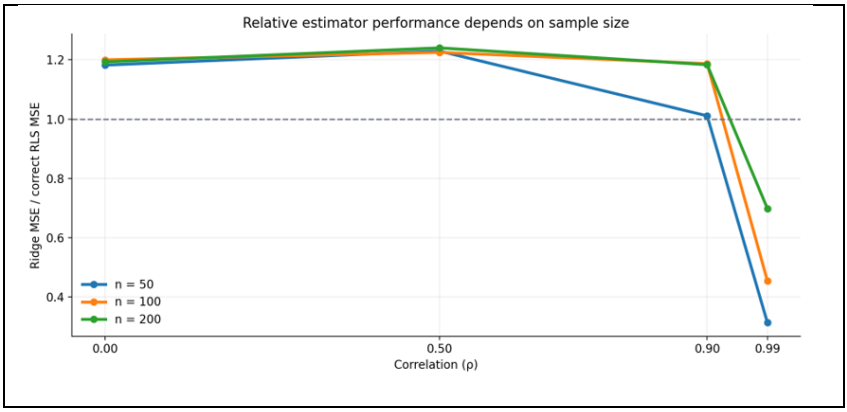



Figure 2. Sensitivity of the ratio of ridge to correctly restricted RLS coefficient MSE to sample size.

*Note.* Dashed lines represent the counterfactual sensitivity experiment with $\Sigma_{test} = I_6$; solid lines represent same-distribution risk. Both are simulated conditions, not observed real-world datasets.

***Robustness to correlation structure and coefficient configuration***

Because the primary experiment uses equicorrelation and one particular $\beta$ vector, sensitivity to the geometry of the model is examined separately. Under a three-variable block-correlation structure ($n = 50$, $\rho = 0.99$, $\sigma = 1$), the MSE of oracle RLS is 2.425 and that of ridge is 2.462. Thus, the relative ranking is no longer as clearly favorable to ridge as in the main equicorrelation design. With a different coefficient configuration chosen to preserve the same norm, coefficient sum, and population signal variance, ridge has an MSE of 2.756 versus 9.094 for oracle RLS. These scenarios demonstrate the structural features that affect performance, not unconditional superiority of either method.

Table 8. Robustness under alternative covariance and coefficient structures ($n = 50$, $\sigma = 1$, $\rho = 0.99$; 750 replications).

| **Scenario** | **Population $R^2$** | **OLS MSE** | **Correct RLS MSE** | **Ridge MSE** |
|---|---|---|---|---|
| Three-variable block | 0.692 | 5.175 | 2.425 | 2.462 |
| Alternative coefficients | 0.692 | 11.769 | 9.094 | 2.756 |
| Previous coefficients | 0.800 | 12.098 | 9.371 | 3.808 |

The “previous coefficients” case refers to a coefficient vector from an earlier pilot experiment. In that structure, population signal variance changes with correlation, so it is not included in the primary design intended to isolate the effect of collinearity. The robustness exercises do not cover a universal family of probability distributions; they concern only the stated alternative structures.

## DISCUSSION

The results require three distinct interpretations. First, as multicollinearity becomes stronger, recovering the individual components of the OLS coefficient vector can become substantially more difficult. This does not mean that OLS automatically becomes biased; instead, variance increases along particular singular directions. Second, oracle RLS illustrates the benefit of knowing a true linear restriction, but the availability of correct restriction information cannot be taken for granted in practice. Third, ridge can deliver a more favorable bias–variance balance in some severely collinear settings, depending on which directions receive weight in the loss function.

Coefficient MSE and prediction MSE must be reported separately. When the scientific objective is to interpret the separate effects of explanatory variables, total coefficient loss or errors in substantively relevant coefficient contrasts may be appropriate. If the objective is to predict responses for customers or observations drawn from a similar distribution, predictive loss is more directly relevant. If the structure of the target data may change, strong cross-validation performance under an identical-distribution assumption does not by itself guarantee reliable generalization (Dormann et al., 2013).

It may initially seem surprising that correctly restricted RLS is not always best under extreme collinearity. A correct linear restriction removes uncertainty along one parameter direction, but it need not address other poorly identified directions. Ridge, by contrast, applies weights to all singular directions according to the selected value of $\alpha$. A different data-generating structure can align the correct restriction more strongly with weak directions and reverse the ranking. The block-correlation experiment illustrates the relevance of this possibility.

An important practical feature of ridge is that its penalty depends on variable scaling. All explanatory variables in this simulation have identical population variances; when variables have different units in observed data, standardization should be learned from the training data alone and penalty selection should be considered jointly with this preprocessing step. Cross-validation folds must also respect temporal ordering or clustered and spatial dependence where relevant. Otherwise, favorable prediction results in a controlled experiment may not be replicated in field applications.

## CONCLUSION

Results from 24 primary and 24 robustness scenarios show no fixed performance ordering among OLS, correctly and incorrectly restricted least squares, and ridge under multicollinearity. In the primary data-generating design, ridge markedly reduces coefficient MSE under extreme collinearity. By contrast, RLS with a known correct restriction can be advantageous at lower correlation levels or under a different covariance structure. Imposing incorrect restrictions carries an additional bias risk, and a restriction that is numerically close to the truth is not automatically credible.

The key practical implication is to establish the scientific goal first—interpretation of individual coefficients, prediction under the same distribution, or generalization to a changed data structure. The source of any restriction information should then be assessed, variable scaling handled appropriately, and performance evaluated with a suitable loss measure and independent data. This research demonstrates the importance of those decisions through a reproducible numerical experiment.

## REFERENCES


Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons. https://doi.org/10.1002/0471725153

Dormann, C. F., et al. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography, 36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Hoerl, A. E., & Kennard, R. W. (1970a). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Hoerl, A. E., & Kennard, R. W. (1970b). Ridge regression: Applications to nonorthogonal problems. *Technometrics, 12*(1), 69–82. https://doi.org/10.1080/00401706.1970.10488635

Kaçıranlar, S., Sakallıoğlu, S., Özkale, M. R., & Güler, H. (2011). More on the restricted ridge regression estimation. *Journal of Statistical Computation and Simulation, 81*(11), 1433–1448. https://doi.org/10.1080/00949655.2010.491480

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to Linear Regression Analysis* (6th ed.). Wiley. ISBN 9781119578727.

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*, 673–690. https://doi.org/10.1007/s11135-006-9018-6

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B, 36*(2), 111–133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x

# Classification of Center-Excluding Complete (k,4)-Arc Extensions Containing a Prescribed Ceva 6-Figure in PG(2,4)

**Elif ALTINTAŞ KAHRİMAN**[1]

1- Assoc. Prof.; Bartın Üniversitesi İktisadi İdari ve Sosyal Bilimler Fakültesi Yönetim Bilişim Sistemleri Bölümü. ealtintaskahriman@bartin.edu.tr ORCID No 0000-0002-3454-0326

## ABSTRACT

This chapter determines all complete (k,4)-arc extensions of a prescribed Ceva 6-figure in the projective plane PG(2,4) when the concurrence point is excluded. The fixed six-point configuration has two zero-secants, which partition the fourteen eligible exterior points into three classes of sizes four, four, and six. This decomposition supplies the geometric constraints and a consistent notation for the extension problem. We enumerate all 16,384 labelled candidate supersets without symmetry reduction. Each candidate is examined against the twenty-one ambient lines: it must attain line intersection number four without containing an entire five-point line, and every omitted point must belong to at least one four-secant. The verification yields 170 complete extensions, comprising six (12,4)-arcs, 156 (13,4)-arcs, six (14,4)-arcs, and two (16,4)-arcs. The two largest arcs admit a separate geometric characterization as complements of the two zero-secants; their uniqueness follows without relying on the exhaustive count. The smaller families are distinguished by their exterior occupancy classes and secant profiles, with incidence certificates establishing completeness. The resulting classification is exhaustive for the specified labelled Ceva configuration and does not claim a classification up to projective equivalence. Explicit line lists and homogeneous coordinates are included to support independent checking of the computations and all reported family counts.

*Keywords: Ceva 6-figure; finite projective plane; complete (k,4)-arc; zero-secant; incidence enumeration; completeness certificate.*

## 1. INTRODUCTION

Arcs in finite projective planes constrain the number of selected points on any line, while completeness demands that no further point can be adjoined. This distinction between line-intersection bounds and maximality makes a configuration-based treatment useful: the local secants of a prescribed core provide restrictions before exhaustive completeness verification is applied (Hirschfeld, 1998; Hirschfeld & Thas, 2016; Hirschfeld & Pichanick, 2016). In PG(2,4), where every line contains five points, the degree-four arc condition excludes entirely selected lines. A candidate is complete if every omitted point lies on at least one selected four-secant.

Ceva configurations arise from three concurrent cevians of a projective triangle. Their incidence-theoretic formulations have been studied in Moufang and other projective settings (Kaya & Çiftçi, 1985; Bayar &

Ekmekçi, 2014; Benítez, 2007). Earlier work on the prescribed Ceva six-point configuration in PG(2,4) treated complete (k,3)-arcs (Altıntaş Kahriman & Bayar, 2025b). The degree-four extension problem investigated here is different: both the permitted line multiplicity and the corresponding completeness certificates change. Related work on complete quadrangle configurations in the same plane provides additional context for configuration-based arc constructions (Altıntaş Kahriman and Bayar, 2025a). The present chapter fixes a labelled Ceva core C and excludes its concurrency point M from every admissible extension. This exclusion is an explicit mathematical scope condition, not a post-hoc division of a broader output. The resulting search uses fourteen eligible exterior points and a geometry controlled by the two zero-secants of C. Our contributions are (i) an intrinsic exterior-point partition via the secant indices of C; (ii) a direct proof and uniqueness characterization for the two 16-point extensions; (iii) occupancy descriptions of the 12-, 13-, and 14-point families; and (iv) an exhaustive and independently auditable classification theorem.

$$C = \{N_1, N_2, N_6, N_7, N_{10}, N_{15}\},\ M = N_3,\ C \subseteq K \subseteq \mathcal{P} \setminus \{M\}.$$

## 2. PRELIMINARIES AND INCIDENCE MODEL

This section establishes the algebraic and incidence-geometric framework required for the construction and classification of complete (k,4)-arc extensions containing a prescribed Ceva 6-figure in PG(2,4). We first introduce the coordinate representation of the finite projective plane over GF(4), followed by the fundamental definitions of arcs, secant distributions, and completeness. We then describe the fixed Ceva configuration, its concurrence point, and the associated point–line incidence relations. These preliminaries provide the mathematical foundation for deriving geometric restrictions on center-excluding extensions and for verifying their completeness through exact incidence computations in the subsequent sections.

### *2.1. The projective plane PG(2,4)*

Let $GF(4) = \{0,1,t,t^2\}$, where $t^2 + t + 1 = 0$, and let PG(2,4) be the set of homogeneous one-dimensional subspaces of $GF(4)^3$. A point [x,y,z] is incident with a line [a,b,c] precisely when $ax + by + cz = 0$. The plane contains 21 points and 21 lines, each line contains five points, and each point belongs to five lines. The point labels $N_1, N_2, \ldots, N_{21}$ and line labels $D_1, D_2, \ldots, D_{21}$ are fixed throughout; full line lists are supplied in Appendix A.

$$\mathbb{F}_4 = \{0,1,t,t^2\}, \quad t^2 + t + 1 = 0, \quad t^3 = 1.$$
$$[x,y,z] \in PG(2,4), \quad [a,b,c] \cdot [x,y,z] = ax + by + cz = 0.$$
$$|\mathcal{P}| = |\mathcal{L}| = 21, \quad |\ell| = 5, \quad |\{\ell : Q \in \ell\}| = 5.$$

### *2.2. Arcs, Secants, and Completeness*

A (k,n)-arc is a k-point set K such that $|K \cap \ell| \leq n$ for every line $\ell$ and equality holds for at least one line. An i-secant is a line meeting K in exactly i points; $t_i(K)$ denotes their number. Write $\sigma_i$ for the number of i-secants of K through an external point Q. A (k,n)-arc is complete when adjoining any external point violates the n-point bound (Hirschfeld, 1998; Hirschfeld & Thas, 2016).

$$\max_{\ell \in \mathcal{L}} |K \cap \ell| = 4, t_i(K) = |\{\ell \in \mathcal{L} : |K \cap \ell| = i\}|.$$

For an arbitrary point set the counting identities include $t_5$; for a degree-four arc one has $t_5=0$. Stating equality in the maximal line occupancy is important: the absence of a 5-secant alone does not turn a small candidate into a (k,4)-arc.

**Proposition 2.1** Let $K$ be a set of $k$ points in PG(2,4), and let $t_i(K)$ denote the number of projective lines meeting $K$ in exactly $i$ points, for $0 \leq i \leq 5$. Then the following identities hold:

$$\sum_{i=0}^{5} t_i = 21, \sum_{i=0}^{5} i \cdot t_i = 5k, \sum_{i=0}^{5} \binom{i}{2} t_i = \binom{k}{2}$$

*Proof.* The first identity counts the lines of PG(2,4). Since the plane has exactly 21 lines and every line intersects $K$ in precisely one of the possible cardinalities $i \in \{0,1,2,3,4,5\}$, the total number of lines is $\sum_{i=0}^{5} t_i = 21$. For the second identity, consider the set of incident point–line pairs

$$\mathcal{I} = \{(P,\ell): P \in K, \ P \in \ell\}.$$

Every point of PG(2,4) lies on exactly five lines; counting the elements of $\mathcal{I}$ by selected points therefore gives $|\mathcal{I}| = 5k$. On the other hand, an $i$-secant contains exactly $i$ points of $K$ and thus contributes $i$ incident pairs. Counting the same set by lines yields $|\mathcal{I}| = \sum_{i=0}^{5} i \cdot t_i = 5k$. Equating the two expressions establishes the second identity.

Finally, count the unordered pairs of distinct points of $K$. There are $\binom{k}{2}$ such pairs. Every unordered pair determines a unique projective line, and each $i$-secant contains exactly $\binom{i}{2}$ unordered pairs of selected points. Summing over all lines therefore gives $\sum_{i=0}^{5} \binom{i}{2} t_i(K) = \binom{k}{2}$, which establishes the third identity. ▫

**Proposition 2.2** A (k,4)-arc K is complete if and only if every $Q \notin K$ lies on a four-secant of K.

$$K \text{ complete} \Leftrightarrow \forall Q \in \mathcal{P} \setminus K, \quad \sigma_4(Q;K) \geq 1,$$

$$\sigma_4(Q;K) = |\{\ell \in \mathcal{L} : Q \in \ell, |K \cap \ell| = 4\}|.$$

*Proof.* If no four-secant passes through Q, adding Q raises each line count by at most one and preserves the degree-four bound. Conversely, a four-secant through Q becomes a prohibited five-point line upon adjoining Q. □

### *2.3. The Prescribed Ceva 6-Figure and Its Incidence Structure*

Fix $C = \{N_1, N_2, N_6, N_7, N_{10}, N_{15}\}$ and let $M = N_3$ denote the Ceva center. Direct intersection with the 21 lines yields the Ceva secant vector $(t_0, t_1, t_2, t_3) = (2,12,3,4)$. The two lines disjoint from C are $Z_1 = D_{12} = \{N_3, N_8, N_{12}, N_{17}, N_{18}\}$ and $Z_2 = D_{13} = \{N_3, N_9, N_{11}, N_{16}, N_{19}\}$; they meet at M. The other three lines through M are two-secants of C. These are verified finite incidence facts for the prescribed labelling (Altıntaş Kahriman & Bayar, 2025b).

The Ceva incidence can be checked without relying on a diagram. With triangle vertices $A_1=N_1$, $A_2=N_6$, $A_3=N_{10}$ and opposite-side points $A_1'=N_2$, $A_2'=N_{15}$, $A_3'=N_7$, the side incidences are $A_1'\in D_6$, $A_2'\in D_3$ and $A_3'\in D_2$. The three cevians are $A_1A_1'=D_1$, $A_2A_2'=D_{10}$ and $A_3A_3'=D_{11}$. Appendix A confirms $D_1\cap D_{10}\cap D_{11}=\{N_3\}$, establishing $M=N_3$ as their common point. This is the fixed Ceva realization underlying the calculations below.

$$C = \{N_1, N_2, N_6, N_7, N_{10}, N_{15}\}, \quad M = N_3,$$
$$(t_0(C), t_1(C), t_2(C), t_3(C)) = (2,12,3,4).$$
$$Z_1 = \{N_3, N_8, N_{12}, N_{17}, N_{18}\}, \quad Z_2 = \{N_3, N_9, N_{11}, N_{16}, N_{19}\}.$$

The four exterior classes are $\{M\}$, $A = Z_1\backslash\{M\} = \{N_8, N_{12}, N_{17}, N_{18}\}$, $B = Z_2\backslash\{M\} = \{N_9, N_{11}, N_{16}, N_{19}\}$, and $R = \mathcal{P}\backslash(C \cup Z_1 \cup Z_2) = \{N_4, N_5, N_{13}, N_{14}, N_{20}, N_{21}\}$. For the present chapter $M$ is forbidden, leaving the disjoint union $U = A \cup B \cup R$ with |U|=14. Each extension has the unique form K=C∪A′∪B′∪R′ with A′⊆A, B′⊆B, R′⊆R.

$$A = Z_1 \setminus \{M\}, \quad B = Z_2 \setminus \{M\}, \quad R = \mathcal{P} \setminus (C \cup Z_1 \cup Z_2),$$
$$|A| = |B| = 4, \quad |R| = 6, \quad \mathcal{P} \setminus (C \cup \{M\}) = A \cup B \cup R.$$

The external secant-index signatures in Table 1 are computed relative to C, not relative to a future extension K. They can therefore be used to formulate geometric restrictions without presupposing the results of the enumeration.

Table 1. External secant-index partition relative to the prescribed Ceva 6-figure.

| Exterior class | Point labels | $\sigma_0\sigma_1\sigma_2\sigma_3$ of C |
|---|---|---|
| **M** | $N_3$ | (2,0,3,0) |
| **A** | $N_8, N_{12}, N_{17}, N_{18}$ | (1,3,0,1) |
| **B** | $N_9, N_{11}, N_{16}, N_{19}$ | (1,3,0,1) |
| **R** | $N_4, N_5, N_{13}, N_{14}, N_{20}, N_{21}$ | (0,4,1,0) |

## 3. ZERO-SECANT GEOMETRY AND EXTENSION CONSTRAINTS

The secant structure of the prescribed Ceva 6-figure provides the geometric basis for restricting its center-excluding complete (k,4)-arc extensions. In particular, the two zero-secants determine a partition of the eligible exterior points into three disjoint classes, while the lines meeting the Ceva core in three points impose additional restrictions on admissible point selections. In this section, we formulate these incidence constraints and introduce occupancy parameters that describe the distribution of selected exterior points. These geometric conditions provide a systematic framework for constructing candidate extensions, although the degree-four arc condition and completeness must ultimately be verified using the full point–line incidence structure of PG(2,4).

### *3.1. Exterior Point Occupancy and Incidence Constraints*

The restrictions from the four trisecants of C are particularly useful: a line with three core points can acquire at most one additional selected point. This gives an explicit incidence inequality for every such line, but does not render the other line tests redundant.

$$\forall D_j: |D_j \cap C| = 3,\ \ |D_j \cap (K \backslash C)| \leq 1.$$

For a center-excluding extension define (a,b,r)=(|A′|,|B′|,|R′|). Then k=6+a+b+r, and $|K \cap Z_1| = a$, $|K \cap Z_2| = b$. Because neither $Z_1$ nor $Z_2$ can be fully selected while M is absent, their own line-intersection bounds are automatic. Other lines, particularly those meeting C in three points, impose nontrivial restrictions. Therefore an occupancy vector gives a necessary structural description but is not by itself a sufficient criterion for either admissibility or completeness.

$$K = C \cup A' \cup B' \cup R', \quad A' \subseteq A, B' \subseteq B, R' \subseteq R,$$
$$(a,b,r) = (|A'|, |B'|, |R'|), \quad k = 6 + a + b + r,$$
$$|K \cap Z_1| = a \leq 4, \quad |K \cap Z_2| = b \leq 4.$$

The inequalities on a and b follow from the sizes of A and B and do not, by themselves, guarantee the arc property. In particular, every trisecant of C can contain at most one selected point from K\C. The nineteen ambient lines other than $Z_1$ and $Z_2$ therefore remain subject to the global incidence test; the occupancy triple is descriptive, not a sufficient admissibility criterion.

**Proposition 3.1.** Every candidate extension containing the prescribed Ceva 6-figure C and excluding its center M has a unique representation by subsets of the disjoint exterior classes A, B, and R. Consequently, the search domain consists of exactly $2^{14}$ = 16,384 distinct labelled point sets.

$$U = \mathcal{P} \setminus (C \cup \{M\}), \qquad |U| = 14,$$
$$|\{C \cup S : S \subseteq U\}| = 2^{14} = 16{,}384.$$

More explicitly, for fixed occupancies (a,b,r), the raw number of candidate selections is $\binom{4}{a}\binom{4}{b}\binom{6}{r}$; the sum of these counts over all permitted occupancies is $2^4 \cdot 2^4 \cdot 2^6 = 2^{14}$. These are candidate counts, not counts of valid or complete arcs.

$$N(a,b,r) = \binom{4}{a}\binom{4}{b}\binom{6}{r}.$$

*Proof.* Since the fixed core $C$ has six points and its center $M$ is excluded, the set of eligible exterior points has cardinality

$$|U| = 21 - 6 - 1 = 14.$$

The three sets $A$, $B$, and $R$ form a disjoint partition of $U$. Therefore every subset $S \subseteq U$ has exactly one representation

$$S = (S \cap A) \sqcup (S \cap B) \sqcup (S \cap R).$$

It follows that each candidate extension $K = C \cup S$ is uniquely determined by independent choices of subsets $A' \subseteq A$, $B' \subseteq B$, and $R' \subseteq R$. Each eligible exterior point is either selected or omitted, so the total number of candidates is

$$2^{|A|} \cdot 2^{|B|} \cdot 2^{|R|} = 2^4 \cdot 2^4 \cdot 2^6 = 2^{14} = 16{,}384.$$

For prescribed occupancies $(a,b,r)$, there are $\binom{4}{a}$ ways to choose $A'$, $\binom{4}{b}$ ways to choose $B'$, and $\binom{6}{r}$ ways to choose $R'$. The multiplication principle gives the stated formula for $N(a,b,r)$. Finally, summing over all occupancies yields

$$\sum_{a=0}^{4}\sum_{b=0}^{4}\sum_{r=0}^{6} N(a,b,r) = (1+1)^4(1+1)^4(1+1)^6 = 2^{14}.$$

Since the partition is disjoint, no candidate can have two different representations. Thus every one of the 16,384 labelled candidate point sets is counted exactly once. ▫

### *3.2. Construction and Uniqueness of Complete (16,4)-Arcs*

**Theorem 3.2** The sets $K_{16}^{(1)} = \mathcal{P}\backslash Z_1$ and $K_{16}^{(2)} = \mathcal{P}\backslash Z_2$ are complete (16,4)-arcs containing C and excluding M. They are the only complete 16-point extensions of the fixed Ceva core. Each has secant vector (1,0,0,0,20).

$$K_{16}{}^{(1)} = P \setminus N_3, N_8, N_{12}, N_{17}, N_{18},$$
$$K_{16}{}^{(2)} = \mathcal{P} \setminus \{N_3, N_9, N_{11}, N_{16}, N_{19}\}.$$
$$K_{16}{}^{(1)} = \mathcal{P} \setminus Z_1, \quad K_{16}{}^{(2)} = \mathcal{P} \setminus Z_2,$$
$$t(K_{16}{}^{(i)}) = (1,0,0,0,20) \ (i = 1,2).$$

Proof. For either zero-secant Z, each line $\ell \neq Z$ has exactly one point in common with Z and hence contains four points of $\mathcal{P}\backslash Z$; the line Z contains none. Thus every omitted point $Q \in Z$ lies on four 4-secants and the complement is complete, with secant vector (1,0,0,0,20). For uniqueness, let K be a 16-point (16,4)-arc and put $H=\mathcal{P}\backslash K$. Then $|H|=5$, and every line meets H, since a line disjoint from H would be a forbidden 5-secant of K. If H were not a line, choose $P,Q \in H$ such that PQ contains $X \notin H$. Such P,Q must exist: otherwise the line through any two H-points would be fully contained in H, and the first such line already contains all five members. All five lines through X must intersect H. Yet the line PQ uses at least two of the five H-points, and the remaining at most three points cannot meet the other four lines through X. This contradiction forces H to be a line. Since $C \subseteq K$, $H \cap C = \emptyset$, so H is a 0-secant of C. The fixed Ceva core has precisely the two 0-secants $Z_1$ and $Z_2$. Therefore K equals one of their complements. □

## 4. EXHAUSTIVE ENUMERATION AND COMPLETENESS VERIFICATION

The geometric restrictions established in the preceding sections provide a structural description of the possible center-excluding extensions of the prescribed Ceva 6-figure. However, these restrictions alone do not guarantee that a selected point set satisfies the degree-four arc condition or is complete. To obtain an exhaustive classification, we systematically examine all candidate extensions using the full point–line incidence structure of PG(2,4). The proposed procedure evaluates the intersection of each candidate with all twenty-one ambient lines and verifies completeness by requiring every omitted point to lie on at least one four-secant. This approach ensures that all eligible configurations are considered without duplication and provides a reproducible computational foundation for the classification results presented in the subsequent section.

### *4.1. Exhaustive Search Algorithm and Completeness Criterion*

Enumeration uses only the fixed incidence relation, without pair-selection constructions or symmetry reductions. For each subset $S \subseteq U$ set $K=C \cup S$, and compute the line counts $c_j = |K \cap D_j|$ for j=1,…,21. Reject a set if some $c_j = 5$. If no $c_j = 4$, the set does not meet the definition of a (k,4)-arc. Otherwise, for every $Q \in P\backslash K$, find a line $D_j$ through Q with $c_j = 4$. Retain K only when such a line exists for every omitted point. The algorithm records K, k, the secant vector, its occupancy vector, and the list of witness lines.

The test of completeness concerns every external point, including the deliberately excluded center M.

$$\lambda_j(S) = |\,(C \cup S) \cap D_j\,|\,, \quad j = 1, \ldots, 21, \quad S \subseteq U,$$
$$A(S) \;\Leftrightarrow\; max_\{1 \le j \le 21\}\, \lambda_j(S) = 4,$$
$$B(S) \;\Leftrightarrow\; \forall Q \in \mathcal{P} \setminus (C \cup S), \qquad \exists j\colon Q \in D_j \;\wedge\; \lambda_j(S) = 4,$$
$$S\ is\ retained \;\Leftrightarrow\; A(S) \;\wedge\; B(S).$$

**Algorithm 4.1. Exhaustive enumeration of center-excluding complete (k,4)-arc extensions**

**Input.** The prescribed Ceva 6-figure $C$, its excluded center $M$, the ambient point set $\mathcal{P}$, and the twenty-one lines $\mathcal{L} = \{D_1, \ldots, D_{21}\}$.
**Output.** Every complete $(k, 4)$-arc containing $C$ and excluding $M$, together with its cardinality, secant vector, occupancy triple, and four-secant witnesses for all external points.

**1. Initialize the candidate domain.** Remove the core and its center from the ambient point set:

$$U = \mathcal{P} \setminus (C \cup \{M\}), \qquad |U| = 14.$$

**2. Generate every candidate.** For every subset $S \subseteq U$, form the corresponding extension:

$$S \subseteq U, \qquad K = C \cup S.$$

**3. Compute line occupancies.** For each of the twenty-one ambient lines calculate:

$$c_j(K) = \left|K \cap D_j\right|, \qquad j = 1, \ldots, 21.$$

**4. Test the degree-four arc condition.** Reject $K$ unless the maximal line occupancy is exactly four (thus at least one four-secant exists and no five-secant exists):

$$\max_{1 \le j \le 21} c_j(K) = 4.$$

**5. Identify the four-secants.** Form the set of their line indices:

$$T(K) = \left\{j \in \{1, \ldots, 21\}\colon c_j(K) = 4\right\}.$$

**6. Test completeness.** Reject $K$ if an external point lies on no four-secant; equivalently, retain it only if:

$$\forall Q \in \mathcal{P} \setminus K, \quad \exists j \in T(K) \text{ such that } Q \in D_j.$$

**7. Record accepted configurations.** For every retained $K$, record its cardinality $|K|$, the secant vector, occupancy triple and a witness line $D_j$ for each omitted point:

$$\mathrm{t}(K) = \left(t_0(K), t_1(K), t_2(K), t_3(K), t_4(K)\right),$$
$$(a, b, r) = (|K \cap A|, |K \cap B|, |K \cap R|).$$

### *4.2. Mathematical Justification and Computational Verification*

Every eligible point set has a unique representation $K = C \cup S, S \subseteq U$; hence no candidate is omitted or repeated. Step 4 is equivalent to the definition of a $(k, 4)$-arc in $PG(2,4)$, and Step 6 is equivalent to completeness by the external four-secant criterion (Proposition 2.2). Thus the algorithm returns exactly the complete center-excluding extensions of the fixed Ceva core.

For reproducibility, each ambient line $D_j$ is represented by a 21-bit incidence mask, and each candidate point set K is encoded by a corresponding point-membership mask. The number of selected points on each line is determined by counting the common set bits of the corresponding masks. To verify completeness, we define the four-secant coverage set of K as

$$W(K) = \bigcup_{\substack{\ell \in \mathcal{L} \\ |K \cap \ell| = 4}} \ell.$$

The set $W(K)$ consists of all points lying on at least one four-secant of K. By Proposition 2.2, a (k,4)- arc K is complete if and only if every point outside K belongs to this coverage set. Therefore,

$$K \text{ complete} \Leftrightarrow \mathcal{P} \setminus K \subseteq W(K).$$

This formulation allows completeness to be verified by computing the bitwise union of the incidence masks corresponding to all four-secants and checking whether the resulting mask covers every point outside K. Consequently, the completeness test is independent of the order in which candidate points are considered and provides a reproducible certificate for each retained complete (k,4)- arc.

### *4.3. Enumeration Results and Verification Counts*

Having established the correctness of the enumeration procedure, we now report the exact number of candidate extensions at each cardinality and distinguish those satisfying the line-intersection restriction from those satisfying the full completeness criterion. The resulting counts provide the computational basis for the classification of complete center-excluding arc families developed in Section 5.

For every cardinality k, the unrestricted candidate count is the binomial coefficient $\binom{14}{k-6}$. Table 2 distinguishes this raw count from the number of candidates avoiding five-secants and from the number satisfying the full degree-four completeness criterion. In particular, a five-secant-free set is not automatically a complete (k,4)-arc.

$$N_{raw}(k) = \binom{14}{k-6}, \quad 6 \leq k \leq 20.$$

Exhaustive evaluation of the twenty-one ambient lines gives 16,384 candidate subsets. Of these, 3,790 contain no five-secant, and exactly 170 satisfy both the degree-four arc condition and the external four-secant completeness criterion. Table 2 lists all cardinalities from 6 to 20; the zero entries for larger cardinalities are essential to reconcile the displayed rows with the full candidate total.

Table 2. Candidate, five-secant-free and complete center-excluding supersets by cardinality.

| k | Raw candidates | No five-secant | Complete (k,4)-arcs |
|---|---|---|---|
| **6** | 1 | 1 | 0 |
| **7** | 14 | 14 | 0 |
| **8** | 91 | 87 | 0 |
| **9** | 364 | 316 | 0 |
| **10** | 1001 | 731 | 0 |
| **11** | 2002 | 1086 | 0 |
| **12** | 3003 | 977 | 6 |
| **13** | 3432 | 460 | 156 |
| **14** | 3003 | 96 | 6 |
| **15** | 2002 | 20 | 0 |
| **16** | 1001 | 2 | 2 |
| **17** | 364 | 0 | 0 |
| **18** | 91 | 0 | 0 |
| **19** | 14 | 0 | 0 |
| **20** | 1 | 0 | 0 |
| ***Total*** | 16,384 | 3790 | 170 |

## 5. CLASSIFICATION OF COMPLETE (k,4)-ARC EXTENSIONS

This section presents the complete (k,4)-arc extensions containing the prescribed Ceva 6-figure C while excluding its concurrence point M. Based on the geometric restrictions established in Section 3 and the exhaustive incidence verification described in Section 4, the resulting configurations are classified according to their cardinalities, exterior occupancy distributions, and secant vectors. The classification identifies precisely six complete (12,4)-arcs, 156 complete (13,4)-arcs, six complete (14,4)-arcs, and two complete (16,4)-arcs. For each family, the corresponding point selections and incidence properties are examined, while completeness is justified through geometric arguments or explicit four-secant verification. The section

concludes with a theorem establishing the exhaustive classification of all 170 labelled center-excluding extensions of the fixed Ceva core.

### *5.1. Complete (12,4)-Arc Extensions*

For each of the six classified twelve-point extensions, the incidence vector is (0,0,12,0,9). Consequently every ambient line meets the selected point set in exactly two or four points. This two-valued intersection property is an observed feature of the fixed-core family, not an assumption imposed during enumeration.

The three incidence identities provide consistency checks: 12+9=21, 2·12+4·9=5·12, and 12·1+9·6=66. They do not prove completeness, which instead requires four-secant witnesses for all nine omitted points.

$$t_2(K) = 12,\ \ t_4(K) = 9,\ \ t_0(K) = t_1(K) = t_3(K) = 0.$$

$$\forall K \in A_{12}(C; M \notin K), \forall D_j \in L, |K \cap D_j| \in \{2,4\}.$$

The exhaustive search evaluates the 3,003 possible six-point additions to C and retains exactly six complete (12,4)-arcs, all with occupancy (a,b,r)=(2,2,2). Table 3 specifies these six point sets. The external four-secant certificate of the first representative is given immediately below; the other five are retained by the same line-by-line criterion in Algorithm 4.1.

$$12 \cdot 2 + 9 \cdot 4 = 60 = 5 \cdot 12, \quad 12 \cdot 1 + 9 \cdot 6 = 66 = (12 \cdot 11)/2.$$

Table 3. All six complete twelve-point extensions expressed by added points.

| Case | Added points K−C | (a,b,r) |
|---|---|---|
| **1** | $N_4, N_5, N_8, N_{11}, N_{17}, N_{19}$ | (2,2,2) |
| **2** | $N_4, N_5, N_9, N_{12}, N_{16}, N_{18}$ | (2,2,2) |
| **3** | $N_8, N_{11}, N_{13}, N_{14}, N_{16}, N_{18}$ | (2,2,2) |
| **4** | $N_8, N_{12}, N_{16}, N_{19}, N_{20}, N_{21}$ | (2,2,2) |
| **5** | $N_9, N_{11}, N_{17}, N_{18}, N_{20}, N_{21}$ | (2,2,2) |
| **6** | $N_9, N_{12}, N_{13}, N_{14}, N_{17}, N_{19}$ | (2,2,2) |

For example, the first listed arc K=$\{N_1, N_2, N_4, N_5, N_6, N_7, N_8, N_{10}, N_{11}, N_{15}, N_{17}, N_{19}\}$ has external blocking witnesses $N_3 \in D_1$; $N_9 \in D_2$; $N_{12} \in D_6$; $N_{13} \in D_{14}$; $N_{14} \in D_{16}$ ; $N_{16} \in D_3$; $N_{18} \in D_7$; $N_{20} \in D_{20}$; $N_{21} \in D_{19}$. These are labels from Appendix A and allow direct line-by-line verification.

### *5.2. Classification of Complete (13,4)-Arc Extensions*

To make the multiplicities auditable, fix (a,b,r) with a+b+r=7 and define the entire labelled selection class before imposing incidence constraints. Its cardinality follows by choosing independently from the three disjoint exterior classes. The distinction between raw, five-secant-free and complete selections is then exact, rather than an inference from occupancy alone.

A candidate is recorded as "No 5-secant" when $|K \cap D_j| \leq 4$ for every j. It is recorded as complete only when, in addition, some $D_j$ is a four-secant and every Q outside K belongs to at least one four-secant. The six nonzero entries in the last column are exactly the six occupancy types in the original classification; the table also displays excluded types explicitly. This is a reproducible computational proof of the family multiplicities, not a synthetic derivation from the binomial coefficients.

$$Complete(K) \Leftrightarrow [max_j |K \cap D_j| = 4] \wedge [\forall Q \in P \backslash K, \exists j: Q \in D_j \; and \; |K \cap D_j| = 4].$$

Table 4. Exhaustive occupancy-resolved audit of thirteen-point extensions.

| (a,b,r) | Raw candidates | No 5-secant | Complete |
|---|---|---|---|
| $(\mathbf{0,1,6})$ | 4 | 4 | 0 |
| $(\mathbf{0,2,5})$ | 36 | 36 | 0 |
| $(\mathbf{0,3,4})$ | 60 | 60 | 0 |
| $(\mathbf{0,4,3})$ | 20 | 20 | 0 |
| $(\mathbf{1,0,6})$ | 4 | 4 | 0 |
| $(\mathbf{1,1,5})$ | 96 | 24 | 24 |
| $(\mathbf{1,2,4})$ | 360 | 48 | 36 |
| $(\mathbf{1,3,3})$ | 320 | 32 | 24 |
| $(\mathbf{1,4,2})$ | 60 | 0 | 0 |
| $(\mathbf{2,0,5})$ | 36 | 36 | 0 |
| $(\mathbf{2,1,4})$ | 360 | 48 | 36 |
| $(\mathbf{2,2,3})$ | 720 | 36 | 12 |
| $(\mathbf{2,3,2})$ | 360 | 0 | 0 |
| $(\mathbf{2,4,1})$ | 36 | 0 | 0 |
| $(\mathbf{3,0,4})$ | 60 | 60 | 0 |
| $(\mathbf{3,1,3})$ | 320 | 32 | 24 |
| $(\mathbf{3,2,2})$ | 360 | 0 | 0 |
| $(\mathbf{3,3,1})$ | 96 | 0 | 0 |
| $(\mathbf{3,4,0})$ | 4 | 0 | 0 |
| $(\mathbf{4,0,3})$ | 20 | 20 | 0 |
| $(\mathbf{4,1,2})$ | 60 | 0 | 0 |
| $(\mathbf{4,2,1})$ | 36 | 0 | 0 |
| $(\mathbf{4,3,0})$ | 4 | 0 | 0 |

$$|C_{a,b,r}| = \binom{4}{a}\binom{4}{b}\binom{6}{r}$$

$$C_{a,b,r} = \{C \cup A' \cup B' \cup R' : \; |A'| = a, |B'| = b, |R'| = r\},$$

Exactly six of the twenty-three occupancy types in Table 4 yield complete thirteen-point arcs. Their multiplicities, matched to their labels, are (1,1,5): 24; (1,2,4): 36; (1,3,3): 24; (2,1,4): 36; (2,2,3): 12; and (3,1,3): 24. Hence the verified total is 156. Every retained arc has secant vector (0,2,3,7,9), but its maximality is established separately through external four-secant coverage.

$$a + b + r = 7, \qquad |K| = 13, \qquad t(K) = (0,2,3,7,9).$$

$$2 \cdot 1 + 3 \cdot 2 + 7 \cdot 3 + 9 \cdot 4 = 65 = 5 \cdot 13,$$

$$3 \cdot 1 + 7 \cdot 3 + 9 \cdot 6 = 78 = (13 \cdot 12)/2.$$

One explicit complete thirteen-point extension is K=$\{N_1, N_2, N_4, N_5, N_6, N_7, N_8, N_{10}, N_{11}, N_{13}, N_{15}, N_{16}, N_{18}\}$. Its external-point blocking certificate is: $N_3 \in D_1$; $N_9 \in D_2$; $N_{12} \in D_6$; $N_{14} \in D_{16}$; $N_{17} \in D_3$; $N_{19} \in D_7$; $N_{20} \in D_4$; $N_{21} \in D_8$.

### *5.3. Complete (14,4)-Arc Extensions*

A fourteen-point extension has eight selected exterior points. For fixed (a,b,r), the unrestricted candidate family has size $\binom{4}{a}\binom{4}{b}\binom{6}{r}$. We apply all twenty-one line inequalities before testing all seven omitted points for four-secant coverage. The occupancy (2,2,4) is therefore a conclusion of the exhaustive search, not a sufficient condition for completeness.

Table 5 contains precisely the occupancy classes that possess at least one five-secant-free fourteen-point candidate; omitted classes have zero such candidates. Among the displayed classes, only (2,2,4) yields a complete arc, and its six surviving selections are listed explicitly in Table 6. The occupancy pattern alone is not a completeness test.

Table 5. Occupancy classes with five-secant-free fourteen-point candidates.

| (a,b,r) | Raw candidates | No 5-secant | Complete |
|---|---|---|---|
| $(\mathbf{0},\mathbf{2},\mathbf{6})$ | 6 | 6 | 0 |
| $(\mathbf{0},\mathbf{3},\mathbf{5})$ | 24 | 24 | 0 |
| $(\mathbf{0},\mathbf{4},\mathbf{4})$ | 15 | 15 | 0 |
| $(\mathbf{2},\mathbf{0},\mathbf{6})$ | 6 | 6 | 0 |
| $(\mathbf{2},\mathbf{2},\mathbf{4})$ | 540 | 6 | 6 |
| $(\mathbf{3},\mathbf{0},\mathbf{5})$ | 24 | 24 | 0 |
| $(\mathbf{4},\mathbf{0},\mathbf{4})$ | 15 | 15 | 0 |

$$|K| = 14, \;\; a + b + r = 8, \;\; |C_{a,b,r}| = \binom{4}{a}\binom{4}{b}\binom{6}{r}.$$

Exactly six fourteen-point extensions exclude M, all in occupancy class (2,2,4), with secant vector (0,0,7,0,14). Table 5 gives the admissibility audit

by occupancy class; Table 6 explicitly lists the six complete point sets by their A′, B′ and R′ selections.

$$(a, b, r) = (2,2,4), \quad |K| = 14, \qquad t(K) = (0,0,7,0,14).$$

$$7 \cdot 2 + 14 \cdot 4 = 70 = 5 \cdot 14, \quad 7 \cdot 1 + 14 \cdot 6 = 91 = (14 \cdot 13)/2.$$

The six complete point sets in Table 6 are specified by their intersections with A, B and R; these choices determine all fourteen points of each arc because the common core C is fixed. In each case the complete incidence check yields seven 2-secants and fourteen 4-secants.

Table 6. The six complete fourteen-point extensions in the A/B/R partition.

| Case | A′ indices | B′ indices | R′ indices |
|---|---|---|---|
| **1** | $N_8, N_{18}$ | $N_{11}, N_{16}$ | $N_4, N_5, N_{20}, N_{21}$ |
| **2** | $N_8, N_{12}$ | $N_{16}, N_{19}$ | $N_4, N_5, N_{13}, N_{14}$ |
| **3** | $N_{17}, N_{18}$ | $N_9, N_{11}$ | $N_4, N_5, N_{13}, N_{14}$ |
| **4** | $N_{12}, N_{17}$ | $N_9, N_{19}$ | $N_4, N_5, N_{20}, N_{21}$ |
| **5** | $N_8, N_{17}$ | $N_{11}, N_{19}$ | $N_{13}, N_{14}, N_{20}, N_{21}$ |
| **6** | $N_{12}, N_{18}$ | $N_9, N_{16}$ | $N_4, N_5, N_{20}, N_{21}$ |

For the first set in Table 6, direct four-secant witnesses for all omitted points are $N_3 \in D_1$; $N_9 \in D_2$; $N_{12} \in D_6$; $N_{13} \in D_4$; $N_{14} \in D_{16}$; $N_{17} \in D_3$; $N_{19} \in D_7$. Each cited line contains exactly four points of the listed arc.

### *5.4. Complete (16,4)-Arc Extensions*

The two sixteen-point sets are exactly $\mathcal{P} \setminus Z_1$ and exactly $\mathcal{P} \setminus Z_2$ by Theorem 3.2. Their occupancy vectors are (0,4,6) and (4,0,6), respectively, and both share the secant vector (1,0,0,0,20). Their uniqueness follows from the complement theorem without relying on the finite search.

$$t(K) = (1,0,0,0,20), \quad 20 \cdot 4 = 80 = 5 \cdot 16,$$

$$20 \cdot 6 = 120 = (16 \cdot 15)/2.$$

### *5.5. Main Classification Theorem*

Define the family of complete labelled degree-four arcs of cardinality k containing the fixed Ceva core while omitting its center by the following expression:

$$\mathcal{A}_{\mathrm{k}}(C; M) = \{K \subseteq \mathcal{P} : C \subseteq K, M \notin K, = k, K \text{ is a complete } (k,4) - arc\}.$$

**Theorem 5.1** For the prescribed Ceva core C and its center M, the cardinalities of the complete labelled (k,4)-arc families are exactly those in the following piecewise formula; no other cardinality occurs.

$$|\mathcal{A}_{\mathrm{k}}(C; M)| = \{6 \text{ if } k = 12;\ 156 \text{ if } k = 13;\ 6 \text{ if } k = 14;\ 2 \text{ if } k = 16;\ 0 \text{ otherwise}\}.$$

The absence of further cardinalities is not inferred from a pattern in the four positive counts; it follows from testing every one of the $2^{14}$ available residual subsets. In particular, the theorem is fixed-core and labelled: it is not a projective-equivalence classification.

*Proof.* The 14-bit search enumerates all 16,384 possible residual subsets exactly once. By the correctness and exhaustiveness of Algorithm 4.1, the enumeration retains precisely the complete degree-four arc extensions satisfying the prescribed core and center-exclusion conditions. The exhaustive line-by-line evaluation of Appendix A yields the four family counts in Table 2 and no other complete outputs. The 16-point case also has an independent synthetic characterization by Theorem 3.2. The secant distributions satisfy the three identities of Proposition 2.1 and the omitted-point witnesses establish completeness. These are counts of labelled fixed-core subsets, not projective equivalence classes. □

Table 7. Final fixed-core classification and secant signatures.

| Arc type | Complete labelled arcs | $(t_0t_1t_2t_3t_4)$ |
|---|---|---|
| $(12,4)$ | 6 | (0,0,12,0,9) |
| $(13,4)$ | 156 | (0,2,3,7,9) |
| $(14,4)$ | 6 | (0,0,7,0,14) |
| $(16,4)$ | 2 | (1,0,0,0,20) |
| $\boldsymbol{Total}$ | 170 | – |

## 6. DISCUSSION AND SCOPE

The zero-secants of the prescribed Ceva 6-figure provide more than a convenient description of the eligible exterior points. Their complements yield the largest complete extensions directly, whereas the remaining sizes require restrictions arising from the full incidence structure. This contrast explains why the geometric decomposition into the classes A, B, and R is useful but cannot, on its own, characterize complete arcs.

The smaller families exhibit noticeably different secant patterns. Each of the six complete (12,4)-arcs has occupancy (2,2,2) and meets every ambient line in either two or four points. The 156 complete (13,4)-arcs occur in six occupancy classes, all sharing the secant vector (0,2,3,7,9). By comparison, the six complete (14,4)-arcs belong to the single class (2,2,4), with secant vector (0,0,7,0,14). These common numerical features describe the accepted families; they should not be mistaken for sufficient conditions on an arbitrary candidate.

The search therefore separates three questions: how many labelled candidates exist, how many contain no full five-point line, and how many are complete degree-four arcs. The secant identities offer independent arithmetic checks, but maximality is established by the requirement that every omitted point lie on a four-secant. The witness lines and the complete incidence list make this final step verifiable without assuming that a particular occupancy pattern guarantees completeness.

The classification counts labelled extensions of one prescribed core with its concurrency point omitted. It does not enumerate projective equivalence classes or all (k,4)-arcs in PG(2,4). The earlier Ceva-based (k,3)-arc investigation (Altıntaş Kahriman and Bayar, 2025b) addresses a different intersection bound. Beyond the present setting, an interesting question is whether combinatorial compatibility arguments can account for the smaller-family multiplicities without exhaustive enumeration; related zero-secant restrictions may also be considered for configurations in other finite projective planes.

## 7. CONCLUSION

For the prescribed Ceva 6-figure in PG(2,4), excluding the concurrency point leaves exactly 170 labelled complete degree-four arc extensions. These consist of six (12,4)-arcs, 156 (13,4)-arcs, six (14,4)-arcs, and two (16,4)-arcs; no other cardinalities occur within the stated fixed-core family.

The two sixteen-point arcs are precisely the complements of the two zero-secants, a characterization established geometrically and independently of the enumeration. The other families are obtained from a complete incidence-based search, with each accepted arc checked against all twenty-one lines and certified by four-secant coverage of its omitted points. Together, the geometric construction and the exhaustive certificates settle the center-excluding extension problem for the specified labelled configuration.

## APPENDIX A. THE 21-POINT–LINE INCIDENCE LIST

The following labelled lines constitute the complete incidence input. Each line contains five points, every point occurs in five lines, and each pair of distinct points belongs to exactly one line.

Table A1. The twenty-one ambient projective lines used by the verification algorithm.

| Line | Point labels |
|---|---|
| $\boldsymbol{D_1}$ | $N_1, N_2, N_3, N_4, N_5$ |
| $\boldsymbol{D_2}$ | $N_1, N_6, N_7, N_8, N_9$ |
| $\boldsymbol{D_3}$ | $N_1, N_{10}, N_{15}, N_{16}, N_{17}$ |
| $\boldsymbol{D_4}$ | $N_1, N_{11}, N_{13}, N_{18}, N_{20}$ |
| $\boldsymbol{D_5}$ | $N_1, N_{12}, N_{14}, N_{19}, N_{21}$ |
| $\boldsymbol{D_6}$ | $N_2, N_6, N_{10}, N_{11}, N_{12}$ |
| $\boldsymbol{D_7}$ | $N_2, N_7, N_{15}, N_{18}, N_{19}$ |
| $\boldsymbol{D_8}$ | $N_2, N_8, N_{13}, N_{16}, N_{21}$ |
| $\boldsymbol{D_9}$ | $N_2, N_9, N_{14}, N_{17}, N_{20}$ |
| $\boldsymbol{D_{10}}$ | $N_3, N_6, N_{15}, N_{20}, N_{21}$ |
| $\boldsymbol{D_{11}}$ | $N_3, N_7, N_{10}, N_{13}, N_{14}$ |
| $\boldsymbol{D_{12}}$ | $N_3, N_8, N_{12}, N_{17}, N_{18}$ |
| $\boldsymbol{D_{13}}$ | $N_3, N_9, N_{11}, N_{16}, N_{19}$ |
| $\boldsymbol{D_{14}}$ | $N_4, N_6, N_{13}, N_{17}, N_{19}$ |
| $\boldsymbol{D_{15}}$ | $N_4, N_7, N_{12}, N_{16}, N_{20}$ |
| $\boldsymbol{D_{16}}$ | $N_4, N_8, N_{11}, N_{14}, N_{15}$ |
| $\boldsymbol{D_{17}}$ | $N_4, N_9, N_{10}, N_{18}, N_{21}$ |
| $\boldsymbol{D_{18}}$ | $N_5, N_6, N_{14}, N_{16}, N_{18}$ |
| $\boldsymbol{D_{19}}$ | $N_5, N_7, N_{11}, N_{17}, N_{21}$ |
| $\boldsymbol{D_{20}}$ | $N_5, N_8, N_{10}, N_{19}, N_{20}$ |
| $\boldsymbol{D_{21}}$ | $N_5, N_9, N_{12}, N_{13}, N_{15}$ |

## APPENDIX B. HOMOGENEOUS COORDINATES AND VERIFICATION DETAILS

Table B1. Normalized homogeneous coordinates of the twenty-one points.

| Point | Coordinates | Point | Coordinates | Point | Coordinates |
|---|---|---|---|---|---|
| $\boldsymbol{N_1}$ | $(0,1,0)$ | $N_8$ | $(1,t,1)$ | $N_{15}$ | $(1,0,0)$ |
| $\boldsymbol{N_2}$ | $(0,0,1)$ | $N_9$ | $(1,t^2,1)$ | $N_{16}$ | $(1,t,0)$ |
| $\boldsymbol{N_3}$ | $(0,1,1)$ | $N_{10}$ | $(1,1,0)$ | $N_{17}$ | $(1,t^2,0)$ |
| $\boldsymbol{N_4}$ | $(0,1,t^2)$ | $N_{11}$ | $(1,1,t^2)$ | $N_{18}$ | $(1,0,t^2)$ |
| $\boldsymbol{N_5}$ | $(0,1,t)$ | $N_{12}$ | $(1,1,t)$ | $N_{19}$ | $(1,0,t)$ |
| $\boldsymbol{N_6}$ | $(1,1,1)$ | $N_{13}$ | $(1,t,t^2)$ | $N_{20}$ | $(1,t^2,t^2)$ |
| $\boldsymbol{N_7}$ | $(1,0,1)$ | $N_{14}$ | $(1,t^2,t)$ | $N_{21}$ | $(1,t,t)$ |

For a machine-checkable audit, encode each subset $S \subseteq U$ as a 14-bit mask, compute the 21 counts $|(C \cup S) \cap D_j|$, and for every omitted point Q record the smallest j such that $Q \in D_j$ and $|(C \cup S) \cap D_j| = 4$. All mathematical claims about family sizes can be reproduced from Appendix A without coordinates or external software packages. The coordinate table is provided as an independent representation of the same incidence model.

## REFERENCES


Altıntaş Kahriman, E., and Bayar, A. (2025a). Complete (k,3)-arcs containing complete quadrangle vertices in PG(2,4): Structural analysis based on diagonal point inclusion. Eskişehir Technical University Journal of Science and Technology A – Applied Sciences and Engineering, 26(4), 478–485. https://doi.org/10.18038/estubtda.1754684

Altıntaş Kahriman, E., and Bayar, A. (2025b). Construction and classification of complete (k,3)-arcs from a Ceva 6-figure in PG(2,4). Hagia Sophia Journal of Geometry, 7(2), 46–51. https://izlik.org/JA69PZ64XH

Bayar, A., and Ekmekçi, S. (2014). On the Menelaus and Ceva 6-figures in the fibered projective planes. Abstract and Applied Analysis, 2014, Article 803173. https://doi.org/10.1155/2014/803173

Benítez, J. (2007). A unified proof of Ceva and Menelaus' theorems using projective geometry. Journal for Geometry and Graphics, 11(1), 39–44.

Hirschfeld, J. W. P. (1998). Projective geometries over finite fields (2nd ed.). Oxford University Press.

Hirschfeld, J. W. P., and Pichanick, E. V. D. (2016). Bounds for arcs of arbitrary degree in finite Desarguesian planes. Journal of Combinatorial Designs, 24(4), 184–196. https://doi.org/10.1002/jcd.21426

Hirschfeld, J. W. P., and Thas, J. A. (2016). General Galois geometries. Springer. https://doi.org/10.1007/978-1-4471-6790-7

Kaya, R., and Çiftçi, Ş. (1985). On Menelaus and Ceva 6-figures in Moufang projective planes. Geometriae Dedicata, 19(3), 295–296. https://doi.org/10.1007/BF00149369

# Construction and Classification of Complete (k,4)-Arcs Containing a Ceva 6-Figure and its Center in PG(2,4)

**Elif ALTINTAŞ KAHRİMAN**[1]

1- Assoc. Prof.; Bartın Üniversitesi İktisadi İdari ve Sosyal Bilimler Fakültesi Yönetim Bilişim Sistemleri Bölümü. ealtintaskahriman@bartin.edu.tr ORCID No 0000-0002-3454-0326

## ABSTRACT

This study investigates the geometric structure, construction, and classification of complete (k,4)-arcs containing a prescribed Fano subplane in the finite projective plane PG(2,4). The investigation begins with the seven-point Fano subplane obtained by adjoining the concurrence center to a Ceva 6-figure and examines how its incidence structure constrains degree-four arc extensions. The seven Fano lines partition the fourteen exterior points into seven disjoint pairs, providing a geometric framework for determining admissible point selections and establishing an upper bound of fourteen on the cardinality of such extensions. Using this structure, we derive necessary conditions for arc construction and establish completeness criteria based on the incidence of external points with 4-secants. In particular, every admissible fourteen-point extension is shown to be complete, whereas completeness of a thirteen-point extension is characterized by the 4-secant coverage of the two points belonging to its unique unoccupied exterior pair. These geometric results are combined with exhaustive incidence-based verification over the twenty-one lines of PG(2,4). The classification establishes precisely 168 complete (13,4)-arcs and 16 complete (14,4)-arcs containing the prescribed Fano subplane, with no complete extensions of other cardinalities. The corresponding secant distributions are determined, and the completeness of the resulting configurations is verified through explicit external-point blocking conditions. The results provide a geometric characterization and exhaustive classification of complete degree-four arc extensions associated with the Ceva-induced Fano subplane in PG(2,4).

*Keywords: Finite projective geometry; PG(2,4); Fano subplane; Ceva 6-figure; complete (k,4)-arc; secant distributions; geometric classification.*

## 1. INTRODUCTION

The theory of arcs in finite projective planes links restrictions on the number of selected points on a line to a global condition of maximality. A degree-four arc in PG(2,4) meets every line in at most four points and has at least one four-secant; it is complete precisely when each point outside it lies on a four-secant. The distinction between admissibility and completeness is central to the present analysis (Hirschfeld, 1998; Hirschfeld and Thas, 2016). General bounds for arcs of arbitrary degree provide a broader point of comparison,

although they do not settle the fixed-configuration enumeration considered here (Hirschfeld and Pichanick, 2016).

The geometric starting point is a prescribed Ceva 6-figure together with its concurrence center. Altıntaş Kahriman and Bayar (2025b) established that adjoining the center gives a seven-point Fano subplane and a complete (7,3)-arc. The present chapter treats a different problem: the construction and exhaustive classification of complete degree-four arcs that contain this entire seven-point configuration. The fixed-centre requirement is essential, since it distinguishes the family studied here from extensions containing the six Ceva points alone.

Earlier investigations provide complementary perspectives on the underlying geometry. Ceva and Menelaus configurations in projective settings are discussed by Kaya and Çiftçi (1985), Benítez (2007), and Bayar and Ekmekçi (2014); Akça, Bayar, and Ekmekçi (2020) develop related conditions in an intuitionistic fuzzy projective setting. These works concern geometric incidence or extensions of classical configuration conditions, rather than the enumeration of complete degree-four arcs. In a closer finite-geometric direction, Altıntaş Kahriman and Bayar (2025a) classify complete (k,3)-arcs containing the vertices of a complete quadrangle, while their Ceva-based study (2025b) determines complete (k,3)-arcs associated with a fixed Ceva 6-figure. Kaplan et al. (2017) examine arc enumeration by a different computational method. Taken together, these studies motivate exact, configuration-sensitive enumeration but do not supply the (13,4)- and (14,4)-arc counts proved below.

We begin by identifying the seven Fano lines and the resulting partition of the fourteen exterior points into seven pairs. This geometry yields a sharp cardinality bound and reduces every eligible extension to a unique exterior-pair selection. The remaining fourteen ambient lines impose the admissibility tests, while four-secants through omitted points determine completeness. After establishing the separate completeness mechanisms for thirteen- and fourteen-point extensions, we give a complete incidence model and a reproducible exhaustive enumeration. Throughout, counts refer to labelled arcs containing the prescribed Fano subplane, not to projective equivalence classes or to all degree-four arcs of PG(2,4).

## 2. PROJECTIVE AND ARC-THEORETIC PRELIMINARIES

We first fix the finite-field convention and the ambient incidence geometry, then formulate precisely the degree-four admissibility and maximality conditions used in every subsequent proof. The distinction between a forbidden five-secant and the positive existence of a four-secant will also be retained in the enumeration.

### *2.1. The projective plane PG(2,4)*

We use homogeneous coordinates over the field of order four because they provide a complete algebraic certificate for each of the incidence relations below. In characteristic two, the relation $t^2=t+1$ fixes both the arithmetic and the displayed normalized representatives.

Write GF(4) as the quadratic extension of GF(2) generated by an element t. Since the characteristic is two, $t^2 + t + 1 = 0$ is equivalently $t^2 = t + 1$. Homogeneous triples are taken modulo multiplication by a nonzero element of GF(4). We use the point labels of the original Ceva realization consistently throughout the argument.

$$GF(4) = \{0,1,t,t^2\}, \quad t^2 + t + 1 = 0, \quad t^3 = 1.$$

A point is represented by a nonzero homogeneous coordinate vector [x,y,z], whereas a line is represented by a nonzero coefficient triple [a,b,c]. Point–line incidence is defined by the zero value of the associated homogeneous linear form. Every two distinct points determine one line and every two distinct lines meet in one point.

The point coordinates have a projective meaning: [x,y,z] and [αx,αy,αz] represent the same point for every $\alpha \in GF(4)\setminus\{0\}$. The chosen normalized representatives allow each incidence assertion to be verified through one substitution in the appropriate line equation.

$$[x,y,z] = [\alpha x, \alpha y, \alpha z], \quad \alpha \in GF(4) \setminus \{0\}.$$
$$N = [x,y,z], \ \ell = [a,b,c], \ N \in \ell \Leftrightarrow ax + by + cz = 0.$$
$$|PG(2,4)| = 21, \ |\mathcal{L}| = 21, \ |\ell| = 5, \ |\{\ell : Q \in \ell\}| = 5.$$

The next definitions distinguish the local line-occupancy condition from maximality under inclusion. In particular, the term admissible is reserved for the absence of a fully occupied five-point line, whereas the designation (k,4)-arc also requires the presence of a four-secant.

**Definition 2.1** A k-point subset K of PG(2,4) is a (k,4)-arc when every line contains at most four points of K and at least one line contains exactly four points of K.

$$\max_{\ell \in \mathcal{L}} | K \cap \ell | = 4.$$

**Definition 2.2** A line meeting K in i points is an i-secant. Its secant numbers and vector are defined by the following expressions.

$$t_{\mathrm{i}}(K) = |\{\ell \in \mathcal{L} : |K \cap \ell| = i\}|, \quad t(K) = (t_0, t_1, t_2, t_3, t_4).$$

**Definition 2.3** A (k,4)-arc is complete if no point Q outside K can be adjoined without producing a five-secant.

**Proposition 2.4** A degree-four arc K is complete if and only if each point outside K lies on some four-secant.

$$K \text{ is complete} \Leftrightarrow \forall Q \in PG(2,4) \setminus K, \exists D_{\mathrm{j}} : Q \in D_{\mathrm{j}} \text{ and } |D_{\mathrm{j}} \cap K| = 4.$$

*Proof.* If Q belongs to a four-secant, adjoining Q makes that line contain five selected points; conversely, if no four-secant passes through Q, every line through Q initially contains at most three points and Q can be adjoined. □

$$K \text{ complete} \Leftrightarrow \forall Q \in PG(2,4) \setminus K, \exists \ell \in \mathcal{L} : Q \in \ell \wedge |K \cap \ell| = 4.$$

**Proposition 2.5** Let $K$ be a set of $k$ points in PG(2,4), and let $t_i(K)$ denote the number of projective lines meeting $K$ in exactly $i$ points, for $0 \leq i \leq 5$. Then the following identities hold:

$$\sum_{i=0}^{5} t_{\mathrm{i}} = 21, \sum_{i=0}^{5} i \cdot t_{\mathrm{i}} = 5k, \sum_{i=0}^{5} \binom{i}{2} t_{\mathrm{i}} = \binom{k}{2}$$

*Proof.* The first identity counts the lines of PG(2,4). Since the plane has exactly 21 lines and every line intersects $K$ in precisely one of the possible cardinalities $i \in \{0,1,2,3,4,5\}$, the total number of lines is $\sum_{i=0}^{5} t_{\mathrm{i}} = 21$.

For the second identity, consider the set of incident point–line pairs

$$\mathcal{I} = \{(P, \ell) : P \in K, \ P \in \ell\}.$$

Every point of PG(2,4) lies on exactly five lines; counting the elements of $\mathcal{I}$ by selected points therefore gives $|\mathcal{I}| = 5k$. On the other hand, an $i$-secant contains exactly $i$ points of $K$ and thus contributes $i$ incident pairs. Counting the same set by lines yields $|\mathcal{I}| = \sum_{i=0}^{5} i \cdot t_{\mathrm{i}} = 5k$. Equating the two expressions establishes the second identity.

Finally, count the unordered pairs of distinct points of $K$. There are $\binom{k}{2}$ such pairs. Every unordered pair determines a unique projective line, and each $i$-secant contains exactly $\binom{i}{2}$ unordered pairs of selected points. Summing over all lines therefore gives $\sum_{i=0}^{5}\binom{i}{2}\,t_i(K) = \binom{k}{2}$, which establishes the third identity. ▫

**Lemma 2.6** Let $\mathcal{P}$ denote the point set of $PG(2,4)$, let $K \subseteq \mathcal{P}$, and put $H = \mathcal{P}\backslash K$. Then $K$ has no five-secant if and only if $H$ meets every ambient line. Furthermore, if $K$ is a $(k,4)$-arc, then $K$ is complete if and only if, for every $Q \in H$, there exists an ambient line $\ell$ such that

$$\ell \cap H = \{Q\}.$$

*Proof.* Every line $\ell$ of $PG(2,4)$ contains exactly five points. Since $H$ is the complement of $K$ in $\mathcal{P}$, the sets $K \cap \ell$ and $H \cap \ell$ partition $\ell$. Hence

$$|K \cap \ell| + |H \cap \ell| = 5.$$

It follows that

$$|K \cap \ell| \leq 4 \quad \Leftrightarrow \quad |H \cap \ell| \geq 1.$$

Therefore, $K$ contains no five-secant precisely when $H$ intersects every ambient line.

Now suppose that $K$ is a $(k,4)$-arc and let $Q \in H$. A line $\ell$ through $Q$ is a four-secant of $K$ if and only if

$$|K \cap \ell| = 4.$$

Because $\ell$ contains five points and $Q \in H \cap \ell$, this is equivalent to

$$\ell \cap H = \{Q\}.$$

Thus every external point of $K$ lies on a four-secant if and only if each $Q \in H$ is the unique point of $H$ on some ambient line. By Proposition 2.4 (the external four-secant criterion), this condition is necessary and sufficient for the completeness of $K$. ▫

## 3. THE CEVA-INDUCED FANO SUBPLANE

The starting core and its ambient realization are now identified explicitly. The aim is not to re-prove novelty of the already known Ceva-center completion, but to determine which properties of its seven subplane lines survive as rigorous restrictions on degree-four extensions.

### *3.1. Fixed Ceva configuration and its center*

The prescribed configuration is the six-point labelled set used in the earlier Ceva investigation (Altıntaş Kahriman & Bayar, 2025b). We list its homogeneous coordinates rather than treating the point indices as numerical data. The concurrence property and the identity of the seven-point completion can then be verified on the indicated ambient lines.

Fix the following six labelled points as the Ceva 6-figure C. Its concurrence point M is displayed independently. Each label $N_i$denotes the corresponding normalized projective coordinate triple and not an abstract enumeration index. Let C be the fixed Ceva 6-figure in PG(2,4), with the following labelled points and homogeneous coordinates:

$$C = \{N_1, N_2, N_6, N_7, N_{10}, N_{15}\},$$
$$N_1 = (0,1,0), N_2 = (0,0,1), N_6 = (1,1,1),$$
$$N_7 = (1,0,1), N_{10} = (1,1,0), N_{15} = (1,0,0).$$

The concurrence point of the Ceva configuration is $M = N_3 = (0,1,1)$. Adjoining this point to C yields the seven-point set $F = C \cup \{M\}$.

To identify the underlying Ceva configuration in purely incident-geometric terms, use the triangle vertices $A_1=N_1$, $A_2=N_6$, $A_3=N_{10}$ and the opposite-side points $A_1'=N_2$, $A_2'=N_{15}$, $A_3'=N_7$. Each of the three joining lines passes through the same point $M=N_3$. This can be checked directly by substitution of the homogeneous coordinates into the displayed line equations:

$$A_1A_1': x = 0, \quad A_2A_2': y + z = 0, \quad A_3A_3': x + y + z = 0.$$

**Proposition 3.1** Let $C = \{N_1, N_2, N_6, N_7, N_{10}, N_{15}\}$ be the fixed Ceva 6-figure in $PG(2,4)$, and let $M = N_3$ be its concurrence point. Then the seven-point set

$$F = C \cup \{M\}$$

forms an embedded projective subplane isomorphic to $PG(2,2)$. Moreover, its ambient secant distribution is

$$(t_0(F), t_1(F), t_2(F), t_3(F)) = (0, 14, 0, 7).$$

*Proof.* Consider the following seven triples of points of $F$:

$\{N_1, N_2, N_3\}, \{N_1, N_6, N_7\}, \{N_1, N_{10}, N_{15}\}, \{N_2, N_6, N_{10}\}, \{N_2, N_7, N_{15}\}, \{N_3, N_6, N_{15}\}, \{N_3, N_7, N_{10}\}$.

Direct verification using the homogeneous coordinates shows that each triple lies on a distinct ambient line of $PG(2,4)$. Furthermore, every pair of distinct points of $F$ belongs to exactly one of these triples. Thus, the induced incidence structure consists of seven points and seven three-point lines, with exactly three lines through each point. Consequently, $F$ forms a projective subplane of order two and is therefore isomorphic to $PG(2,2)$.

To determine its ambient secant distribution, observe that the seven subplane lines are precisely the ambient lines containing three points of $F$. Any other ambient line contains at most one point of $F$, since two distinct points of $F$ determine one of the seven subplane lines.
Moreover, every ambient line intersects $F$. Indeed, if an ambient line were disjoint from $F$, it would intersect the seven subplane lines at seven distinct points, because any two subplane lines intersect at a point of $F$. This contradicts the fact that every line of $PG(2,4)$ contains exactly five points.
Therefore, the remaining fourteen ambient lines each contain exactly one point of $F$. Hence,

$$t_0(F) = 0, \qquad t_1(F) = 14, \qquad t_2(F) = 0, \qquad t_3(F) = 7.$$

This establishes the stated secant distribution. ▫

### *3.2. Fano Subplane Lines and the Structure of Exterior Point Pairs*

Every subplane line has three points in F and two points outside it. By identifying the latter explicitly in the original ambient incidence model, we obtain seven disjoint pairs. This construction provides the central combinatorial parameterization for all later extensions.

Each line of the seven-point subplane is the intersection of an ambient five-point line of PG(2,4) with F. Its three F-points are listed as a triple and its other two ambient points as an exterior pair. Table 1 reports these objects, with the line equations understood over GF(4). The line and pair labels are fixed and are used without relabelling in the enumeration and in the four-secant witnesses below.

Table 1. The Seven Fano Subplane Lines, Their Point Triples, and Corresponding Exterior Point Pairs in PG(2,4).

| Fano line | Ambient line | Fano triple | Exterior pair |
|---|---|---|---|
| $\boldsymbol{D_1}$ | $\mathrm{x} = 0$ | $\{N_1, N_2, N_3\}$ | $P_1 = \{N_4, N_5\}$ |
| $\boldsymbol{D_2}$ | $\mathrm{x} + \mathrm{z} = 0$ | $\{N_1, N_6, N_7\}$ | $P_2 = \{N_8, N_9\}$ |
| $\boldsymbol{D_3}$ | $\mathrm{z} = 0$ | $\{N_1, N_{10}, N_{15}\}$ | $P_3 = \{N_{16}, N_{17}\}$ |
| $\boldsymbol{D_6}$ | $\mathrm{x} + \mathrm{y} = 0$ | $\{N_2, N_6, N_{10}\}$ | $P_4 = \{N_{11}, N_{12}\}$ |
| $\boldsymbol{D_7}$ | $\mathrm{y} = 0$ | $\{N_2, N_7, N_{15}\}$ | $P_5 = \{N_{18}, N_{19}\}$ |
| $\boldsymbol{D_{10}}$ | $\mathrm{y} + \mathrm{z} = 0$ | $\{N_3, N_6, N_{15}\}$ | $P_6 = \{N_{20}, N_{21}\}$ |
| $\boldsymbol{D_{11}}$ | $\mathrm{x} + \mathrm{y} + \mathrm{z} = 0$ | $\{N_3, N_7, N_{10}\}$ | $P_7 = \{N_{13}, N_{14}\}$ |

The seven exterior pairs associated with the Fano subplane lines are denoted by $P_1, \dots, P_7$, as specified in Table 1. These pairs are mutually disjoint and partition the fourteen points outside $F$. Their normalized homogeneous coordinates are given below, with the two points of each pair displayed together:

$$P_1 = \{N_4 = (0,1,t^2),\ N_5 = (0,1,t)\}$$

$$P_2 = \{N_8 = (1,t,1),\ N_9 = (1,t^2,1)\}$$

$$P_3 = \{N_{16} = (1,t,0),\ N_{17} = (1,t^2,0)\}$$

$$P_4 = \{N_{11} = (1,1,t^2),\ N_{12} = (1,1,t)\}$$

$$P_5 = \{N_{18} = (1,0,t^2),\ N_{19} = (1,0,t)\}$$

$$P_6 = \{N_{20} = (1,t^2,t^2),\ N_{21} = (1,t,t)\}$$

$$P_7 = \{N_{13} = (1,t,t^2),\ N_{14} = (1,t^2,t)\}$$

This exterior-pair decomposition will be used throughout the construction and classification of complete $(k,4)$-arc extensions containing $F$.

**Proposition 3.2** Let $\mathcal{P}$ be the point set of $PG(2,4)$, and let $F$ be the embedded Fano subplane. The seven exterior pairs $P_1, \dots, P_7$ associated with the Fano lines are mutually disjoint and form a partition of the complement $\mathcal{P} \setminus F$. In particular,

$$\mathcal{P} \setminus F = \bigcup_{i=1}^{7} P_i, \qquad |P_i| = 2.$$

*Proof.* Each of the seven Fano lines contains exactly three points of $F$ and two additional points of the ambient plane $PG(2,4)$. These two exterior points constitute the corresponding pair $P_i$.

Suppose that two distinct exterior pairs $P_i$ and $P_j$ share a point $Q$. Then $Q$ belongs to two distinct ambient lines extending Fano lines. However, any two distinct Fano lines meet at a point of $F$, and their ambient extensions meet at that same point. Hence they cannot have a second common point outside $F$, a contradiction. Therefore,

$$P_i \cap P_j = \varnothing \qquad (i \neq j).$$

Since the seven pairs are mutually disjoint and each contains exactly two points, their union has fourteen elements. Moreover, $PG(2,4)$ has twenty-one points and $F$ has seven points, so

$$|\mathcal{P} \setminus F| = 21 - 7 = 14.$$

Thus the seven pairs exhaust the complement of $F$, establishing

$$\mathcal{P} \setminus F = \bigcup_{i=1}^{7} P_i.$$

This completes the proof. ▫

**Proposition 3.3** Let $F$ be the embedded Fano subplane of $PG(2,4)$, and let $\mathcal{L}(F)$ denote its seven ambient subplane lines. For every ambient line $\ell \notin \mathcal{L}(F)$,

$$|\ell \cap F| = 1, \qquad |\ell \setminus F| = 4.$$

Moreover, the four points of $\ell \setminus F$ belong to four distinct exterior pairs associated with the lines of $F$.

*Proof.* Suppose that $\ell$ contains two distinct points of $F$. Their unique joining line is one of the seven ambient Fano lines; hence $\ell \in \mathcal{L}(F)$, a contradiction. Therefore,

$$|\ell \cap F| \leq 1.$$

Assume next that $\ell \cap F$ is empty. Every one of the seven ambient Fano lines intersects $\ell$ in a point. These seven intersection points must be distinct, since two distinct Fano lines meet at a point of $F$, whereas $\ell$ is disjoint from $F$. This is impossible because an ambient line of $PG(2,4)$ has only five points. Consequently, $\ell$ meets $F$ in exactly one point, say $Y$.

Exactly three Fano lines pass through $Y$, and all three intersect $\ell$ at $Y$. Each of the other four Fano lines intersects $\ell$ at a point outside $F$. These four points are distinct, since two Fano lines cannot have a common exterior point. By the

exterior-pair partition, each belongs to a different exterior pair. Thus $\ell$ contains $Y$ and exactly four exterior points, as claimed. ▫

**Corollary 3.4** Let $K = F \cup S$, where $S \subseteq \mathcal{P} \setminus F$ and $\mathcal{P}$ is the point set of $PG(2,4)$. Assume that $K$ is pair-compatible, that is, at most one point is selected from each exterior pair. Then $K$ has no five-secant if and only if

$$|S \cap (\ell \setminus F)| \leq 3 \qquad \text{for every } \ell \notin \mathcal{L}(F).$$

*Proof.* Every ambient Fano line contains three points of $F$ and, by pair-compatibility, at most one point of $S$. Thus no Fano line can be a five-secant of $K$.

Now fix an ambient line $\ell \notin \mathcal{L}(F)$. By Proposition 3.3, $\ell$ contains exactly one $F$-point and four points outside $F$. Since $K = F \cup S$ and $S$ is disjoint from $F$,

$$|K \cap \ell| = 1 + |S \cap (\ell \setminus F)|.$$

It follows that

$$|K \cap \ell| \leq 4 \quad \Leftrightarrow \quad |S \cap (\ell \setminus F)| \leq 3.$$

Therefore, no ambient line is a five-secant of $K$ precisely when the stated inequality holds on each of the fourteen non-Fano lines. ▫

## 4. GEOMETRIC CONSTRAINTS ON DEGREE-FOUR ARC EXTENSIONS CONTAINING THE FANO SUBPLANE

The seven Fano subplane lines restrict the selection of exterior points, but the fourteen remaining ambient lines impose additional incidence conditions. We now separate these two kinds of restrictions, so that every candidate considered later is represented uniquely and tested against the full geometry of $PG(2,4)$.

**Theorem 4.1** Let $F$ be the fixed embedded Fano subplane and let $K$ be a $(k, 4)$-arc with $F \subseteq K$. Then

$$|K \cap P_i| \leq 1 \qquad (1 \leq i \leq 7), \qquad k \leq 14.$$

*Proof.* The ambient line associated with $P_i$ consists of three points of $F$ and the two members of $P_i$. Since the former three points belong to $K$, selecting both exterior members would give a five-secant, contradicting the definition of a degree-four arc. Proposition 3.2 partitions the fourteen points of $\mathcal{P}\setminus F$ into the seven pairs. Therefore at most seven exterior points can be selected and $|K| \leq |F| + 7 = 14$. ▫

The upper bound is attained by the complete fourteen-point configurations constructed in Section 5; hence it is sharp. Pair exclusion is necessary, but not

sufficient: by Corollary 3.4, the four exterior points of any one non-Fano ambient line cannot all be selected.

**Corollary 4.2** Let $P_i = \{p_{i,1}, p_{i,2}\}$, for $i = 1, \dots, 7$, denote the seven exterior pairs indexed in the order given in Table 1. Then every pair-compatible extension $K \supseteq F$ admits a unique representation by a point-selection configuration

$$u = (u_1, \dots, u_7) \in \{0, 1, 2\}^7,$$

where $u_i = 0$ indicates that no point is selected from $P_i$, whereas $u_i = a \in \{1, 2\}$ indicates that $p_{i,a}$ is selected from that pair.

*Proof.* The pairs are mutually disjoint and cover $\mathcal{P}\backslash F$. Hence the intersection of $K$ with each pair determines one of the three indicated states, and these seven states reconstruct $K$ uniquely. ▫

The extension encoded by $u$ is thus

$$K(u) = F \cup \{p_{i,u_i} : 1 \le i \le 7,\ u_i \neq 0\}, \qquad |K(u)| = 7 + \sum_{i=1}^{7} 1_{\{u_i \neq 0\}}.$$

For an ambient line $D_j$ define the integer occupancy $\lambda_j(u) = |D_j \cap K(u)|$. Theorem 4.1 and Corollary 3.4 show that the complete degree-four arc condition consists of two logically different requirements:

$$\max_{1 \le j \le 21} \lambda_j(u) = 4, \qquad \forall Q \in \mathcal{P}\backslash K(u)\ \exists j:\ Q \in D_j,\ \lambda_j(u) = 4.$$

The first condition tests the arc degree, and the second tests inclusion-maximality. Although there are $3^7 = 2187$ pair-compatible point-selection configuration, no assumption of statistical sampling, group orbit reduction, or coding-theoretic structure is required for their examination.

## 5. CONSTRUCTION AND CLASSIFICATION OF COMPLETE (14,4)-ARCS CONTAINING THE FANO SUBPLANE

The cardinality bound is reached exactly when the seven exterior pairs each contribute one point. This special case permits a direct geometric proof of completeness once the global line-occupancy condition is satisfied.

**Proposition 5.1** If $K$ is a fourteen-point degree-four arc containing $F$, then $|K \cap P_i| = 1$ for all $i$. In particular, the pair-compatible fourteen-point candidate family has $2^7 = 128$ members.

*Proof.* By Theorem 4.1 there can be at most one chosen point in each of the seven pairs. Since $|K\backslash F| = 7$, equality must hold for every pair. Each pair

then admits two choices, independently before the non-Fano line restrictions are tested. ▫

**Theorem 5.2** Every fourteen-point pair-compatible set $K \supseteq F$ containing no five-secant is a complete $(14,4)$-arc.

*Proof.* Each of the seven Fano lines contains its three points of $F$ and exactly one selected exterior point. Hence all seven are four-secants, so the set has degree exactly four. Let $Q \in \mathcal{P}\backslash K$. Since all pairs are singly occupied, $Q$ is the unselected member of a unique pair $P_i$. Its Fano line contains $Q$ together with four points of $K$ and is therefore a four-secant through $Q$. All omitted points are blocked; Proposition 2.4 establishes completeness. ▫

**Theorem 5.3** Exactly sixteen labelled complete $(14,4)$-arcs contain $F$. Each has secant distribution

$$(t_0, t_1, t_2, t_3, t_4) = (0,0,7,0,14).$$

*Proof.* Proposition 5.1 parameterizes all candidates by $u \in \{1,2\}^7$. Each Fano line contains four selected points, so five-secants can occur only on non-Fano lines. By Corollary 3.4, an assignment is allowed precisely when, on each of those fourteen lines, at most three of its four exterior points are selected. For each of the $128$ assignments, test the fourteen non-Fano lines in Table 6 and reject any selection containing all four exterior points of one of these lines. Exactly sixteen selections survive; they are listed in full in Table 2. As each has no five-secant, Theorem 5.2 gives completeness. Direct evaluation on the twenty-one lines yields the stated secant vector for each survivor. The secant-counting identities $\sum t_i = 21$, $\sum i t_i = 70$ and $\sum \binom{i}{2} t_i = 91$ are satisfied, consistently with $|K| = 14$. Exhaustiveness follows because every eligible fourteen-point arc appears among the $128$ point-selection configuration and each point-selection configuration is tested exactly once. ▫

Table 2. The sixteen exterior-pair selections yielding complete (14,4)-arcs containing the Fano subplane F.

The seven ordered exterior pairs $P_1, \ldots, P_7$ are defined in Table 1. Each selection specifies exactly one point from each pair: entry 1 denotes its first point, and entry 2 denotes its second point.

| No. | Exterior-pair selection | No. | Exterior-pair selection |
|---|---|---|---|
| 1 | (1,1,1,1,2,1,1) | 9 | (2,1,1,1,2,2,2) |
| 2 | (1,1,1,2,1,2,2) | 10 | (2,1,1,2,1,1,1) |
| 3 | (1,1,2,1,1,2,1) | 11 | (2,1,2,1,1,1,2) |
| 4 | (1,1,2,2,2,1,2) | 12 | (2,1,2,2,2,2,1) |
| 5 | (1,2,1,1,1,1,2) | 13 | (2,2,1,1,1,2,1) |
| 6 | (1,2,1,2,2,2,1) | 14 | (2,2,1,2,2,1,2) |

| | | | |
|---|---|---|---|
| 7 | (1,2,2,1,2,2,2) | 15 | (2,2,2,1,2,1,1) |
| 8 | (1,2,2,2,1,1,1) | 16 | (2,2,2,2,1,2,2) |

Each exterior-pair selection uniquely determines a fourteen-point extension of F. The sixteen selections in Table 2 are exactly those satisfying the degree-four arc condition on all twenty-one ambient lines. By Theorem 5.2, every resulting extension is complete.

For example, the first selection $u = (1,1,1,1,2,1,1)$ determines the seven exterior points

$$S = \{N_4, N_8, N_{16}, N_{11}, N_{19}, N_{20}, N_{13}\}.$$

The corresponding complete fourteen-point arc is $K = F \cup S$.

**Example 5.4** Consider the first exterior-pair selection in Table 2,

$$u = (1,1,1,1,2,1,1).$$

According to the ordering of the pairs in Table 1, this selection chooses the following seven exterior points:

$$S = \{N_4, N_8, N_{16}, N_{11}, N_{19}, N_{20}, N_{13}\}.$$

Adjoining these points to the seven-point Fano subplane gives

$$K_{14} = F \cup S$$
$$= \{N_1, N_2, N_3, N_4, N_6, N_7, N_8, N_{10}, N_{11}, N_{13}, N_{15}, N_{16}, N_{19}, N_{20}\}.$$

A direct evaluation of the twenty-one ambient lines yields the secant distribution

$$(t_0(K_{14}), t_1(K_{14}), t_2(K_{14}), t_3(K_{14}), t_4(K_{14})) = (0,0,7,0,14).$$

Thus no ambient line contains five points of $K_{14}$, while fourteen lines are four-secants. To verify completeness explicitly, the seven points outside $K_{14}$ and one incident four-secant for each are recorded in Table 3.

Table 3. Pointwise four-secant completeness certificates for the (14,4)-arc $K_{14}$ in Example 5.4.

| Omitted point | Four-secant witness | Four selected points on the witness line |
|---|---|---|
| $N_5$ | $D_1$ | $N_1, N_2, N_3, N_4$ |
| $N_9$ | $D_2$ | $N_1, N_6, N_7, N_8$ |
| $N_{12}$ | $D_{15}$ | $N_4, N_7, N_{16}, N_{20}$ |
| $N_{14}$ | $D_{16}$ | $N_4, N_8, N_{11}, N_{15}$ |
| $N_{17}$ | $D_3$ | $N_1, N_{10}, N_{15}, N_{16}$ |
| $N_{18}$ | $D_4$ | $N_1, N_{11}, N_{13}, N_{20}$ |
| $N_{21}$ | $D_8$ | $N_2, N_8, N_{13}, N_{16}$ |

For example, $D_1$ contains the omitted point $N_5$ together with the four selected points $N_1$, $N_2$, $N_3$ and $N_4$. Hence adjoining $N_5$ would create a five-secant. The same argument applies to every row of Table 3. In particular,

$$\forall\, Q \in \mathcal{P} \setminus K_{14}, \qquad \exists\, D_j : Q \in D_j \text{ and } |D_j \cap K_{14}| = 4.$$

Therefore no point of the complement can be adjoined without violating the degree-four condition. By Proposition 2.4, $K_{14}$ is complete. □

## 6. CONSTRUCTION AND CLASSIFICATION OF COMPLETE (13,4)-ARCS CONTAINING THE FANO SUBPLANE

For a thirteen-point extension, one exterior pair remains empty. In contrast to the fourteen-point case, the two members of that pair are not automatically blocked by their own Fano line.

**Proposition 6.1** Every thirteen-point degree-four arc containing $F$ selects exactly one point from six of the seven exterior pairs and none from the remaining pair. Thus there are $7 \cdot 2^6 = 448$ pair-compatible candidates.

*Proof.* Six points are added to $F$, and Theorem 4.1 permits at most one from each exterior pair. Exactly six distinct pairs must therefore be occupied. The empty pair is unique, and each of the other six contributes one of its two elements. ▫

**Lemma 6.2** Let $K$ be a thirteen-point pair-compatible set with no five-secant, and suppose $P_h = \{a, b\}$ is the unique empty exterior pair. Then $K$ is a complete $(13,4)$-arc if and only if both $a$ and $b$ are incident with four-secants of $K$.

*Proof.* The six occupied Fano lines have precisely four points of $K$ each; hence the set has degree four. On each occupied exterior pair, the unselected member lies on the four-secant consisting of its Fano line and selected partner. The only omitted points not already certified by these six lines are $a$ and $b$. Proposition 2.4 therefore reduces completeness to the existence of a four-secant through each of these two points. ▫

**Corollary 6.3** The four-secants required in Lemma 6.2 for $a$ and $b$ must be non-Fano lines.

*Proof.* Proposition 3.2 implies that an exterior point lies on exactly one Fano line, namely the line of its own exterior pair. The line of $P_h$ contains only the three selected points of $F$, because $P_h$ is empty. It therefore is a three-secant and cannot block either $a$ or $b$. All other four-secant witnesses are among the fourteen non-Fano lines. ▫

**Theorem 6.4** Exactly 168 labelled complete $(13,4)$-arcs contain the fixed Fano subplane. For each of the seven choices of empty exterior pair, there are 40 candidates without five-secants, of which 24 are complete. Every complete member has secant vector

$$(t_0, t_1, t_2, t_3, t_4) = (0,2,3,7,9).$$

*Proof.* Fix the empty pair $P_h$. The remaining six pairs give $2^6 = 64$ assignments. By Corollary 3.4, the admissibility test rejects any assignment containing all four exterior points of a non-Fano line; the fourteen non-Fano line tests given by Table 6 leave forty assignments for each empty pair. Lemma 6.2 and Corollary 6.3 reduce the remaining completeness test to the two points of $P_h$, each requiring a non-Fano four-secant. For each surviving selection, check all non-Fano lines incident with each of the two points and retain it only if both points lie on four-secants. Exactly twenty-four selections pass these two tests for each empty pair. Table 4 reports the counts for all seven choices of $h$, obtained from the twenty-one ambient lines of Table 6. Since each thirteen-point configuration has exactly one empty pair, these seven classes are disjoint and the total is $7 \cdot 24 = 168$. Every retained arc has the displayed secant vector on the full ambient incidence table. In particular, the identities $\sum t_i = 21$, $\sum i t_i = 65$ and $\sum \binom{i}{2} t_i = 78$ hold, as required for thirteen selected points. The candidate encoding covers all possible $F$-containing thirteen-point arcs, which proves exhaustiveness. ▫

Table 4. Classification of complete (13,4)-arcs according to the unique unoccupied Fano exterior pair.

| Unoccupied exterior pair | Pair-compatible candidates | Candidates without five-secants | Complete (13,4)-arcs |
|---|---|---|---|
| $P_1$ | 64 | 40 | 24 |
| $P_2$ | 64 | 40 | 24 |
| $P_3$ | 64 | 40 | 24 |
| $P_4$ | 64 | 40 | 24 |
| $P_5$ | 64 | 40 | 24 |
| $P_6$ | 64 | 40 | 24 |
| $P_7$ | 64 | 40 | 24 |
| Total | 448 | 280 | 168 |

For each fixed unoccupied exterior pair, the remaining six pairs admit $2^6 = 64$ possible point selections. Applying the line-intersection restrictions to all twenty-one ambient lines yields exactly 40 candidates without five-secants. Among these, 24 satisfy the completeness criterion, as both points of the unoccupied pair are incident with non-Fano four-secants. Since each thirteen-point extension has a unique unoccupied exterior pair, the seven classes are disjoint. Consequently, the classification yields precisely $7.24 = 168$ distinct labelled complete (13,4)-arcs containing the prescribed Fano subplane.

**Example 6.5** Select a verified assignment for which $P_1 = \{N_4, N_5\}$ is empty. The resulting arc is

$$K_{13} = \{N_1, N_2, N_3, N_6, N_7, N_8, N_{10}, N_{11}, N_{14}, N_{15}, N_{16}, N_{18}, N_{20}\}.$$

Its secant vector is $(0,2,3,7,9)$. The next table gives a labelled four-secant witness for each of its eight omitted points. In particular, the witnesses for $N_4$ and $N_5$ are non-Fano lines, as required by Corollary 6.3.

Table 5. Pointwise four-secant completeness certificates for the complete (13,4)-arc in Example 6.5.

| Omitted point | Four-secant witness | Selected points on the witness line |
|---|---|---|
| $N_4$ | $D_{16}$ | $N_8, N_{11}, N_{14}, N_{15}$ |
| $N_5$ | $D_{18}$ | $N_6, N_{14}, N_{16}, N_{18}$ |
| $N_9$ | $D_2$ | $N_1, N_6, N_7, N_8$ |
| $N_{12}$ | $D_6$ | $N_2, N_6, N_{10}, N_{11}$ |
| $N_{13}$ | $D_4$ | $N_1, N_{11}, N_{18}, N_{20}$ |
| $N_{17}$ | $D_3$ | $N_1, N_{10}, N_{15}, N_{16}$ |
| $N_{19}$ | $D_7$ | $N_2, N_7, N_{15}, N_{18}$ |
| $N_{21}$ | $D_{10}$ | $N_3, N_6, N_{15}, N_{20}$ |

Table 5 provides an explicit completeness certificate for the constructed $(13,4)$-arc. Each of the eight points outside $K_{13}$ lies on an ambient line containing exactly four selected points. In particular, the two points of the unoccupied exterior pair $P_1 = \{N_4, N_5\}$ are blocked by the non-Fano lines $D_{16}$ and $D_{18}$, respectively. The remaining six external points are also incident with four-secants, as shown in the table. Consequently, adjoining any external point produces a five-secant, and $K_{13}$ is complete by Proposition 2.4.

## 7. EXHAUSTIVE CLASSIFICATION AND VERIFICATION OF FANO-CONTAINING COMPLETE (K,4)-ARCS

The geometric restrictions established in the preceding sections reduce the classification problem to $3^7$ possible exterior-pair selections. Each selection is uniquely determined by choosing either no point or exactly one point from each of the seven exterior pairs. The resulting enumeration is exhaustive and reproducible from the twenty-one ambient lines of $PG(2,4)$. The procedure distinguishes the degree-four arc condition from completeness under point addition and requires neither probabilistic sampling nor assumptions concerning projective equivalence.

Let $u = (u_1, \dots, u_7) \in \{0,1,2\}^7$ represent an exterior-pair selection, where $u_i = 0$ indicates that no point is selected from $P_i$, while $u_i = 1$ and $u_i = 2$ indicate selection of its first and second point, respectively. Denote the corresponding point set by $K(u)$. For each projective point $Q$, define the membership indicator

$$x_Q(u) = \begin{cases} 1, & Q \in K(u), \\ 0, & Q \notin K(u). \end{cases}$$

For each ambient line $D_j$, $1 \le j \le 21$, define its intersection number by

$$\lambda_j(u) = \sum_{Q \in D_j} x_Q(u) = |D_j \cap K(u)|.$$

The candidate $K(u)$ is a complete $(k, 4)$-arc if and only if

$$\max_{1 \le j \le 21} \lambda_j(u) = 4$$

and

$$\forall Q \notin K(u), \quad \exists j \in \{1, \ldots, 21\}: \; Q \in D_j \text{ and } \lambda_j(u) = 4.$$

The first condition guarantees that no ambient line contains five selected points and that at least one four-secant exists. The second condition ensures that every external point is blocked by a four-secant, which is equivalent to completeness.

**Algorithm 7.1 (Exhaustive Fano Exterior-Pair Enumeration).**

**Input.** The seven ordered exterior pairs $P_1, \ldots, P_7$ and the twenty-one ambient lines $D_1, \ldots, D_{21}$.

1. For each exterior-pair selection $u \in \{0,1,2\}^7$, construct the uniquely determined point set $K(u)$.
2. Compute $\lambda_j(u)$ for every ambient line $D_j$, $1 \le j \le 21$.
3. Reject $K(u)$ if $\max_j \lambda_j(u) \neq 4$.
4. For each external point $Q \notin K(u)$, determine whether there exists a line $D_j$ containing $Q$ such that $\lambda_j(u) = 4$.
5. Retain $K(u)$ if and only if every external point satisfies the four-secant condition. Record its cardinality, selected point set, secant distribution, and four-secant witnesses.

### *7.1. Complete Incidence Model and Verification of the Ambient Plane*

The exhaustive classification of complete $(k, 4)$ -arcs containing the prescribed Fano subplane requires the complete point–line incidence structure of $PG(2,4)$. Although the seven Fano lines determine the exterior-pair restrictions, the remaining fourteen ambient lines impose additional conditions essential to arc admissibility and completeness. Accordingly, the verification uses all twenty-one lines of the plane.

Let $\mathcal{P} = \{N_1, \ldots, N_{21}\}$ denote the point set, and let $\mathcal{L} = \{D_1, \ldots, D_{21}\}$ be the ambient line set. The incidence structure follows from the homogeneous coordinates over $GF(4)$ introduced in Section 2.1. A projective point $N_i = (x_i, y_i, z_i)$ belongs to the line $\ell = [a, b, c]$ precisely when

$$ax_i + by_i + cz_i = 0.$$

Applying this relation to the normalized projective points and distinct projective lines gives Table 6. The incidence checks are exact: each line contains five points, each point lies on five lines, and any two distinct points lie on exactly one line. The seven Fano lines of Table 1 occur among these twenty-one ambient lines.

For any candidate point set $K \subseteq \mathcal{P}$, the occupancy of line $D_j$ is defined by

$$\lambda_j(K) = |K \cap D_j|, \quad 1 \leq j \leq 21.$$

The candidate is a degree-four arc exactly when its maximum line occupancy is four:

$$max_{1 \leq j \leq 21} \lambda_j(K) = 4.$$

By the external four-secant criterion, such an arc is complete if and only if every omitted point has a witness line:

$$\forall Q \in \mathcal{P} \setminus K, \exists j \in \{1, \ldots, 21\}: Q \in D_j \ and \ \lambda_j(K) = 4.$$

Thus Table 6 supports both the exclusion of five-secants and the independent pointwise certification of completeness. Each entry gives the five points incident with the indicated line.

Table 6. The twenty-one ambient lines of PG(2,4) in the fixed labelled incidence realization.

| Line | Five incident points | Line | Five incident points |
|---|---|---|---|
| $D_1$ | $\{N_1, N_2, N_3, N_4, N_5\}$ | $D_{12}$ | $\{N_3, N_8, N_{12}, N_{17}, N_{18}\}$ |
| $D_2$ | $\{N_1, N_6, N_7, N_8, N_9\}$ | $D_{13}$ | $\{N_3, N_9, N_{11}, N_{16}, N_{19}\}$ |
| $D_3$ | $\{N_1, N_{10}, N_{15}, N_{16}, N_{17}\}$ | $D_{14}$ | $\{N_4, N_6, N_{13}, N_{17}, N_{19}\}$ |
| $D_4$ | $\{N_1, N_{11}, N_{13}, N_{18}, N_{20}\}$ | $D_{15}$ | $\{N_4, N_7, N_{12}, N_{16}, N_{20}\}$ |
| $D_5$ | $\{N_1, N_{12}, N_{14}, N_{19}, N_{21}\}$ | $D_{16}$ | $\{N_4, N_8, N_{11}, N_{14}, N_{15}\}$ |
| $D_6$ | $\{N_2, N_6, N_{10}, N_{11}, N_{12}\}$ | $D_{17}$ | $\{N_4, N_9, N_{10}, N_{18}, N_{21}\}$ |
| $D_7$ | $\{N_2, N_7, N_{15}, N_{18}, N_{19}\}$ | $D_{18}$ | $\{N_5, N_6, N_{14}, N_{16}, N_{18}\}$ |
| $D_8$ | $\{N_2, N_8, N_{13}, N_{16}, N_{21}\}$ | $D_{19}$ | $\{N_5, N_7, N_{11}, N_{17}, N_{21}\}$ |
| $D_9$ | $\{N_2, N_9, N_{14}, N_{17}, N_{20}\}$ | $D_{20}$ | $\{N_5, N_8, N_{10}, N_{19}, N_{20}\}$ |
| $D_{10}$ | $\{N_3, N_6, N_{15}, N_{20}, N_{21}\}$ | $D_{21}$ | $\{N_5, N_9, N_{12}, N_{13}, N_{15}\}$ |
| $D_{11}$ | $\{N_3, N_7, N_{10}, N_{13}, N_{14}\}$ | | |

The incidence data in Table 6 satisfy the defining parameters of $PG(2,4)$: each of the twenty-one lines has five incident points, every point belongs to five

lines, and every pair of distinct points determines one line. These consistency checks validate the fixed labelled model used in the exhaustive enumeration. In particular, the seven Fano lines induce the exterior-pair restrictions, while the other fourteen lines impose the additional constraints. Hence the table supplies the complete input for the admissibility tests, secant distributions, and external four-secant certificates established in the classification.

**Proposition 7.1** Algorithm 7.1 returns every complete $(k, 4)$-arc containing $F$ exactly once and returns no inadmissible or incomplete configuration.

*Proof.* By Corollary 4.2, every pair-compatible extension of $F$ has a unique encoding $u$. By Theorem 4.1, every degree-four extension containing $F$ is pair-compatible, so none lies outside the enumerated domain. Step 3 verifies that the maximum line-intersection number is exactly four, thereby excluding five-secants while ensuring the existence of at least one four-secant; Steps 4–5 are equivalent to completeness by Proposition 2.4. Hence all and only eligible configurations are retained, and the encoding ensures uniqueness. ▫

Table 7. Exhaustive Classification of Exterior-Pair-Compatible Extensions Containing the Fixed Fano Subplane.

| k | Pair-compatible | Five-secant-free candidates | Complete |
|---|---|---|---|
| 7 | 1 | 1 | 0 |
| 8 | 14 | 14 | 0 |
| 9 | 84 | 84 | 0 |
| 10 | 280 | 280 | 0 |
| 11 | 560 | 546 | 0 |
| 12 | 672 | 588 | 0 |
| 13 | 448 | 280 | 168 |
| 14 | 128 | 16 | 16 |
| Total | 2187 | 1809 | 184 |

The number of exterior-pair-compatible extensions of cardinality $k = 7 + r$ is given by $\binom{7}{r}2^r$, since $r$ of the seven exterior pairs must be selected and each selected pair contributes one of its two points. The second numerical column records candidates containing no five-secant, as determined by the complete twenty-one-line incidence test. The final column counts those candidates that additionally possess at least one four-secant and satisfy the external-point completeness criterion. The exhaustive classification yields precisely 168 complete (13,4)-arcs and 16 complete (14,4)-arcs, with no complete extensions of other cardinalities.

Verification protocol. Each of the $3^7 = 2187$ exterior-pair selections is constructed from the seven ordered pairs of Table 1. The line occupancy is obtained by counting selected labels in every row of Table 6. The initial filter

accepts exactly those candidates whose maximum occupancy equals four. A candidate is retained as complete only if every omitted point belongs to at least one line having occupancy four. The full cardinality-by-cardinality results appear in Table 7. For thirteen-point candidates, the six unselected partners of occupied pairs are automatically blocked; only the two points of the empty pair require additional witnesses. These explicit finite tests provide the computational component of the classification and do not assume projective orbit reduction.

**Theorem 7.2** Let $F$ be the prescribed Ceva-induced Fano subplane in $PG(2,4)$. Then every complete $(k,4)$-arc containing $F$ has cardinality $k = 13$ or $k = 14$. More precisely, there exist exactly 168 labelled complete $(13,4)$ -arcs and 16 labelled complete $(14,4)$ -arcs containing $F$. Consequently, the total number of complete degree-four arcs containing the fixed Fano subplane is

$$168 + 16 = 184.$$

No complete $(k,4)$-arc containing $F$ exists for any other cardinality.

**Proof.** By Theorem 4.1, every degree-four arc containing $F$ contains at most one point from each of the seven exterior pairs. Since $F$ consists of seven points, every such arc satisfies

$$7 \leq k \leq 14.$$

Furthermore, Proposition 7.1 establishes that Algorithm 7.1 examines every exterior-pair-compatible extension exactly once and retains precisely those satisfying both the degree-four arc condition and the external four-secant completeness criterion.

The exhaustive classification summarized in Table 7 shows that no complete $(k,4)$ -arc containing $F$ exists for $7 \leq k \leq 12$. For the remaining cardinalities, Theorem 6.4 establishes exactly 168 complete $(13,4)$-arcs, while Theorem 5.3 establishes exactly 16 complete $(14,4)$-arcs.
Since the enumeration is exhaustive and the families have different cardinalities, they are disjoint and account for every complete degree-four arc containing $F$. Therefore, the total number is

$$168 + 16 = 184.$$

This completes the classification. ▫

## 8. DISCUSSION

The classification obtained in this study demonstrates how the incidence structure of an embedded Fano subplane governs the construction and completeness of degree-four arcs in $PG(2,4)$. The seven Fano lines partition the fourteen exterior points into seven disjoint pairs, imposing a natural restriction on admissible extensions and establishing the upper bound $k \leq 14$. However, the exterior-pair restriction alone does not characterize all degree-four arcs containing $F$. The fourteen non-Fano ambient lines introduce additional geometric constraints, as established in Proposition 3.3 and Corollary 3.4. Consequently, the classification is determined by the interaction between the internal incidence structure of the Fano subplane and the remaining lines of the ambient projective plane.

A significant geometric distinction arises between the complete thirteen- and fourteen-point extensions. For an admissible fourteen-point extension, exactly one exterior point is selected from each of the seven pairs. Each omitted point therefore lies on a Fano line containing four selected points, making completeness an immediate consequence of the subplane geometry. In contrast, a thirteen-point extension necessarily leaves one exterior pair unoccupied. Although the six occupied pairs automatically provide four-secant witnesses for their omitted partners, the two points of the unoccupied pair require additional blocking by non-Fano lines. Thus, completeness in the thirteen-point family is governed by a localized incidence condition involving only two external points. This distinction provides a geometric explanation for the different completeness mechanisms of the two families.

The exhaustive classification further reveals that the prescribed Fano subplane is contained in precisely 168 labelled complete $(13,4)$-arcs and 16 labelled complete $(14,4)$ -arcs. These multiplicities arise from the combined application of the exterior-pair restrictions, the line-intersection conditions, and the external four-secant completeness criterion. While the geometric arguments determine the structure of admissible selections and reduce the completeness tests, the exact numerical multiplicities are established by exhaustive verification using all twenty-one ambient lines. The resulting classification therefore combines structural geometric reasoning with a finite and reproducible incidence-based procedure.

The results concern complete degree-four arcs containing the fixed Ceva-induced Fano subplane and do not constitute a classification of all complete $(k,4)$-arcs in $PG(2,4)$. Within this prescribed geometric setting, the analysis

demonstrates how embedded subplane configurations can provide effective structural restrictions and completeness criteria for finite-projective arc classification.

## 9. CONCLUSION

This study establishes a complete classification of degree-four arcs containing the prescribed Ceva-induced Fano subplane in $PG(2,4)$. By examining the incidence relations of the embedded subplane and its exterior points, the investigation develops geometric restrictions that characterize admissible extensions and provide explicit criteria for completeness. The seven exterior point pairs determine the possible selection patterns, while the remaining ambient lines impose the additional conditions required for the construction of complete arcs.

The classification identifies exactly 168 labelled complete $(13,4)$-arcs and 16 labelled complete $(14,4)$ -arcs, with no complete extensions of other cardinalities containing the fixed Fano subplane. The two families exhibit distinct geometric completeness mechanisms: admissible fourteen-point extensions are automatically complete, whereas completeness of thirteen-point extensions depends on the four-secant blocking of the two points in the unique unoccupied exterior pair. These results are established through geometric arguments supported by exhaustive incidence verification over all twenty-one lines of the ambient plane.

The findings demonstrate that the embedded Fano structure provides more than an initial configuration for arc construction; its incidence properties directly determine the geometric restrictions and local completeness conditions governing the resulting families. Thus, the classification contributes a configuration-based characterization of complete degree-four arc extensions in $PG(2,4)$, combining explicit geometric constructions with exact finite verification.

## REFERENCES

Akça, Z., Bayar, A., & Ekmekçi, S. (2020). On the intuitionistic fuzzy projective Menelaus and Ceva's conditions. Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics, 69(1), 891–899. https://doi.org/10.31801/cfsuasmas.567753

Altıntaş Kahriman, E., & Bayar, A. (2025a). Complete (k,3)-arcs containing complete quadrangle vertices in PG(2,4): Structural analysis based on diagonal point inclusion. Eskişehir Technical University Journal of Science and Technology A –

Applied Sciences and Engineering, 26(4), 478–485. https://doi.org/10.18038/estubtda.1754684

Altıntaş Kahriman, E., & Bayar, A. (2025b). Construction and classification of complete (k,3)-arcs from a Ceva 6-figure in PG(2,4). Hagia Sophia Journal of Geometry, 7(2), 46–51. https://izlik.org/JA69PZ64XH

Bayar, A., & Ekmekçi, S. (2014). On the Menelaus and Ceva 6-figures in the fibered projective planes. Abstract and Applied Analysis, 2014, Article 803173. https://doi.org/10.1155/2014/803173

Benítez, J. (2007). A unified proof of Ceva and Menelaus' theorems using projective geometry. Journal for Geometry and Graphics, 11(1), 39–44.

Hirschfeld, J. W. P. (1998). Projective geometries over finite fields (2nd ed.). Oxford University Press.

Hirschfeld, J. W. P., & Pichanick, E. V. D. (2016). Bounds for arcs of arbitrary degree in finite Desarguesian planes. Journal of Combinatorial Designs, 24(4), 184–196. https://doi.org/10.1002/jcd.21426

Hirschfeld, J. W. P., & Thas, J. A. (2016). General Galois geometries. Springer. https://doi.org/10.1007/978-1-4471-6790-7

Kaplan, N., Kimport, S., Lawrence, R., Peilen, L., & Weinreich, M. (2017). Counting arcs in projective planes via Glynn's algorithm. Journal of Geometry, 108, 1013–1029. https://doi.org/10.1007/s00022-017-0391-1

Kaya, R., & Çiftçi, Ş. (1985). On Menelaus and Ceva 6-figures in Moufang projective planes. Geometriae Dedicata, 19(3), 295–296. https://doi.org/10.1007/BF00149369

# Sol-Gel Processing and Electrospinning Technique: A Comparative Review of Advanced Methods in Nanomaterial Production

**Fatih SEVİM[1*]**
**Ömer LAÇİN[2]**
**Fatih DEMİR[3]**
**Semet ÇELİK[4]**

1. *Corresponding Author; Prof. Dr.; Atatürk University, Faculty of Engineering, Department of Chemical Engineering. fsevim@atauni.edu.tr, ORCID: 0000-0002-0805-8499
2. Prof. Dr.; Atatürk University, Faculty of Engineering, Department of Chemical Engineering. olacin@atauni.edu.tr, ORCID: 0000-0002-5276-3056
3., Prof. Dr.; Atatürk University, Faculty of Engineering, Department of Chemical Engineering. fatihdemir@atauni.edu.tr, ORCID: 0000-0002-0264-1074
4. Prof. Dr.; Atatürk University, Faculty of Engineering, Department of Civil Engineering. scelik@atauni.edu.tr, ORCID: 0000-0002-5276-3056

## ABSTRACT

This chapter presents a comparative examination of the principles, advantages, limitations, and application areas of sol-gel processing and electrospinning, two techniques that are widely employed in nanomaterial production. Both methods play an important role in the synthesis of nanostructured materials, yet they offer distinct advantages and constraints. Sol-gel processing is generally preferred for the production of inorganic and composite materials, whereas electrospinning is particularly prominent in biomedical and energy-related applications. Owing to distinctive properties such as high specific surface area, tunable porosity, and functionality, nanomaterials possess broad application potential spanning biomedicine to energy systems. The processing steps, mechanisms, advantages, and limitations of both methods are analyzed in light of the relevant literature, and hybrid approaches such as sol-gel-assisted electrospinning are also discussed.

Nanotechnology has, in recent years, driven revolutionary advances across numerous fields, including materials science, biomedical engineering, and energy systems. At the heart of these developments lies the controlled production of structures at the nanometer scale. In this context, sol-gel processing and electrospinning are among the most widely used and effective techniques for nanomaterial production. This chapter examines in detail the fundamental principles, production mechanisms, advantages, limitations, and application areas of both methods.

The findings indicate that sol-gel and electrospinning techniques possess complementary characteristics, and that hybrid systems combining the two are expected to become increasingly widespread in the future.

*Keywords – Sol-gel processing, electrospinning method, nanotechnology, ceramic nanofiber production methods*

## INTRODUCTION

### *Development and Importance of Nanotechnology*

One of the most significant developments in science and technology since the final quarter of the twentieth century has been the ability to produce materials in a controlled manner at the nanometer scale. A nanometer (nm) corresponds to one-billionth of a meter, a length scale close to atomic and molecular dimensions. Materials produced at this scale exhibit physical, chemical, mechanical, optical, and electrical properties that differ substantially from those of their macro-scale counterparts. As particle size decreases, the surface-to-volume ratio increases, quantum size effects emerge, and surface atoms become dominant — factors that allow nanomaterials to

display markedly higher performance than conventional materials (Guozhong, 2004; Rao et al., 2004).

The concept of nanotechnology was first proposed in 1959 by the Nobel laureate physicist Richard P. Feynman. In his lecture entitled "There's Plenty of Room at the Bottom," Feynman envisioned those new materials could be produced through the individual manipulation of atoms, an idea that laid the foundation of modern nanotechnology (Poole & Owens, 2003). The subsequent development of advanced characterization techniques such as scanning tunnelling microscopy (STM) and atomic force microscopy (AFM) made atomic-scale imaging and manipulation possible, accelerating the growth of nanoscience.

Today, nanotechnology lies at the intersection of numerous disciplines, including materials science, chemistry, physics, biology, electronics, energy engineering, and medicine. Owing to advantages such as high specific surface area, tunable pore structure, superior mechanical strength, enhanced optical properties, and high chemical activity, nanomaterials play a critical role in a wide range of technological applications (Kumar & Kumbhat, 2016).

***Fundamental Properties of Nanomaterials***

Nanomaterials are structures with at least one dimension in the range of 1–100 nm. As particle size decreases, the physicochemical behavior of the material changes substantially. The principal reasons for this can be summarized as follows:

- An increase in the proportion of surface atoms
- An increase in free surface energy
- The emergence of quantum confinement effects
- An increase in diffusion rate
- An increase in chemical reactivity
- Changes in electronic band structure

Owing to these features, nanomaterials offer advantages such as:

- High adsorption capacity
- Enhanced catalytic activity
- High electrical conductivity
- Superior mechanical strength
- Tunable optical properties

Nanomaterials can be classified according to their morphology as shown in Table 1 (Xia et al., 2003; Guozhong, 2004).

Table 1. Classification of nanostructures by dimensionality

| Dimensionality | Nanostructure |
|---|---|
| 0-Dimensional | Nanoparticles, quantum dots |
| 1-Dimensional | Nanowires, nanofibers, nanotubes |
| 2-Dimensional | Graphene, thin films |
| 3-Dimensional | Nanoporous structures, aerogels |

Each morphology offers advantages for different applications. For instance, nanofibers are preferred in filtration and biomedical applications owing to their high surface area, whereas nanoparticles are widely used in catalysis and drug-delivery systems.

### *Approaches to Nanomaterial Production*

Nanomaterial production methods are generally examined under two fundamental approaches.

#### *Top-Down Approach*

In this approach, bulk materials are reduced to the nanoscale through mechanical or physical processes. Principal methods include:

- Ball milling
- Lithography
- Laser ablation
- Mechanical grinding

The advantage of top-down methods is that they can be applied at an industrial scale; however, the formation of crystal defects and limited control over particle size are among their significant disadvantages (Cushing et al., 2004).

#### *Bottom-Up Approach*

In the bottom-up approach, atoms, ions, or molecules are assembled in a controlled manner to form nanostructures. Methods within this group include:

- Sol-Gel
- Electrospinning

- Hydrothermal synthesis
- Solvothermal synthesis
- Chemical vapor deposition (CVD)
- Atomic layer deposition (ALD)

Bottom-up methods allow more homogeneous structures to be obtained and enable precise control of particle size (Xia et al., 2003; Cushing et al., 2004).

***Significance of Sol-Gel and Electrospinning Techniques***

Although numerous methods have been developed for nanomaterial synthesis, sol-gel processing and electrospinning have, in recent years, been among the most extensively investigated techniques (Hench & West, 1990; Xue et al., 2019). This is primarily because both methods are low-cost, applicable from laboratory to industrial scale, and capable of producing nanostructured materials of diverse composition. The sol-gel method is particularly favored for the production of metal oxide ceramics, thin films, aerogels, and porous structures, whereas electrospinning offers unique advantages for the continuous production of nanofibers (Brinker & Scherer, 1990; Ramakrishna et al., 2005). Hybrid processes developed by combining the two methods enable the production of high-performance ceramic nanofibers and functional composite structures (Larsen & Skotak, 2008; Sevim & Sevimli, 2014).

***Current Research Trends***

Recent studies indicate that, in nanomaterial production, sustainability has become as important a criterion as high performance. Approaches such as environmentally friendly solvents, low-temperature synthesis methods, the use of biodegradable polymers, and the reduction of energy consumption have become focal points of research. In addition, the use of artificial intelligence and machine-learning algorithms in process optimization is becoming increasingly widespread. Owing to the large number of experimental parameters involved, data-driven modelling approaches — particularly in sol-gel and electrospinning processes — contribute to determining optimum production conditions in a shorter time.

## OVERVIEW OF NANOMATERIAL PRODUCTION TECHNOLOGIES

The physical and chemical properties of nanomaterials depend not only on their chemical composition but also, to a large extent, on their production method. Two nanomaterials with the same chemical formula may exhibit marked differences in crystal structure, particle size, porosity, surface area, degree of crystallinity, and morphology when produced using different synthesis routes. The synthesis method is therefore one of the most important factors determining the final performance of a nanomaterial. Production techniques available today offer different advantages in terms of cost, production capacity, energy consumption, environmental impact, and product quality. The choice of an appropriate method is directly related to the targeted application: for example, chemical vapor deposition (CVD) is preferred for high-purity thin films, the sol-gel method for porous metal oxides, and electrospinning for continuous nanofiber production.

### *Fundamental Approaches to Nanomaterial Production*

Nanomaterial synthesis is generally divided into two fundamental strategies: Top-Down and Bottom-Up.

#### *Top-Down Approach*

In the top-down approach, a bulk material is reduced to the nanoscale through mechanical, physical, or chemical processes. This approach can be regarded as the nanoscale application of conventional production techniques. Its most important advantage is its applicability at industrial scale, and it is particularly preferred for the large-scale production of metal alloys and ceramics. However, this approach also has several disadvantages:

- Formation of crystal defects
- High energy consumption
- Wide particle-size distribution
- Irregular morphology
- Surface damage

For these reasons, top-down methods may be limited in applications requiring high crystallinity and homogeneous morphology.

***Bottom-Up Approach***

The bottom-up approach relies on the controlled assembly of atoms, ions, or molecules to form nanostructures. Proceeding in a manner similar to natural crystallization processes, this approach is today the most widely used synthesis strategy. The principal bottom-up methods are:

- Sol-Gel
- Hydrothermal synthesis
- Solvothermal synthesis
- Chemical vapor deposition (CVD)
- Atomic layer deposition (ALD)
- Electrodeposition
- Microemulsion method
- Electrospinning

The main advantages of the bottom-up approach are:

- Atomic-level control
- Low defect density
- Homogeneous structure
- Narrow particle-size distribution
- High specific surface area
- Controlled pore structure

For this reason, bottom-up methods are predominantly preferred for the production of advanced functional nanomaterials (Xia et al., 2003).

***Chemical Synthesis Techniques***

Chemical methods constitute the synthesis group with the widest range of applications in the controlled production of nanomaterials.

***Sol-Gel Method***

The sol-gel process is based on the formation of a three-dimensional network structure through the hydrolysis and condensation reactions of metal alkoxides or metal salts (Brinker & Scherer, 1990; Hench & West, 1990). The most important advantages of this method are:

- Low processing temperature

- Chemical homogeneity
- High purity
- Compositional control
- Thin-film production
- Ability to form porous structures

The sol-gel method is particularly used for the production of the following materials: $SiO_2$, $TiO_2$, ZnO, $Al_2O_3$, $ZrO_2$, $Fe_2O_3$.

### *Multicomponent ceramics*

#### *Hydrothermal Synthesis*

Hydrothermal synthesis is carried out in sealed autoclave systems under high temperature and high pressure. This method is widely used for the production of nanocrystals, nanowires, nanorods, and layered structures.

Advantages: high crystallinity, low defect density, morphology control, single-crystal production.

Disadvantages: long reaction time, the need for pressurized systems, difficulty of scale-up.

Solvothermal Method

Solvothermal synthesis is a variant of the hydrothermal method carried out using organic solvents. The use of organic solvents enables smaller particle sizes, better morphology control, and the synthesis of different crystal phases. This method is particularly preferred for the synthesis of metal–organic framework (MOF) structures, quantum dots, and various metal oxide nanoparticles.

### *Physical Production Techniques*

#### *Chemical Vapor Deposition (CVD)*

CVD relies on the decomposition of gas-phase precursor compounds at high temperatures to form a thin film on a solid surface. Its principal application areas include graphene, carbon nanotubes, diamond thin films, semiconductors, and protective coatings.

Advantages: very high purity, dense coating, homogeneous film, industrial applicability.

Disadvantages: high temperature requirement, complex equipment, high investment cost.

***Atomic Layer Deposition (ALD)***

ALD is one of the most precise thin-film production techniques, providing atomic-level thickness control; only a single atomic layer forms in each cycle. This enables ultrathin films, high coating homogeneity, and the coating of complex surfaces. ALD is particularly widely used in microelectronics and energy-storage systems.

***The Place of Electrospinning in Nanomaterial Production***

Unlike other synthesis methods, electrospinning enables the direct, continuous production of nanofibers (Greiner & Wendorff, 2007; Xue et al., 2019). The advantages offered by nanofibers include very high surface area, low density, controlled porosity, flexible structure, and ease of surface functionalization. Owing to these properties, electrospinning is widely used in tissue engineering, drug-delivery systems, filtration, sensors, battery separators, supercapacitor electrodes, and catalysis.

## *Comparison of Production Methods*

Table 2. Comparison of major nanomaterial production methods

| Property | Sol-Gel | Electrospinning | Hydrothermal | CVD | ALD |
|---|---|---|---|---|---|
| Processing temperature | Low | Room temperature | Medium–High | High | Medium |
| Product type | Powder, film, gel | Nanofiber | Crystal | Thin film | Thin film |
| Size control | Very good | Very good | Good | Very good | Excellent |
| Porosity | Very high | Very high | Medium | Low | Low |
| Scalability | High | Medium–High | Medium | High | Medium |
| Cost | Low | Low–Medium | Medium | High | Very high |

In this section, the fundamental synthesis approaches used in nanomaterial production have been systematically examined. Although top-down methods provide high production capacity, they present certain limitations in terms of size control and crystal defects. Bottom-up approaches, in contrast, are particularly prominent in the production of advanced

functional nanomaterials owing to their atomic-level structural control and high degree of homogeneity.

Among the chemical synthesis methods, the sol-gel process enables the production of high-purity ceramic and oxide materials at low temperature, whereas electrospinning stands out for its success in continuous nanofiber production. Thin-film techniques such as CVD and ALD offer high quality and precision but are generally reserved for more specialized applications owing to their investment costs and processing temperatures.

This comparison sets the stage for the detailed discussion, in the following sections, of why sol-gel processing and electrospinning are among the most important nanomaterial production methods in contemporary nanotechnology.

***Sol-Gel Processing Method***

The sol-gel method is one of the most widely used chemical production techniques for the synthesis of nanostructured ceramics, metal oxides, thin films, coatings, composites, and porous materials (Brinker & Scherer, 1990). Owing to its ability to form molecularly homogeneous mixtures, to yield high-purity products at low processing temperatures, and to allow precise control over different chemical compositions, this method occupies an important place in materials science and nanotechnology (Hench & West, 1990; Pierre, 1998). It is particularly widely used for the production of metal-oxide-based nanomaterials such as silica ($SiO_2$), titania ($TiO_2$), zirconia ($ZrO_2$), alumina ($Al_2O_3$), zinc oxide (ZnO), and iron oxide ($Fe_2O_3$).

The sol-gel process is based on the formation of a three-dimensional network structure through the hydrolysis and condensation reactions of metal alkoxides or metal salts in the liquid phase (Livage et al., 1988). During this process, the initially fluid sol gradually increases in viscosity, forming a continuous network structure and giving rise to a gel. Drying the gel and subjecting it to appropriate heat treatment yields high-purity nanostructured ceramic or metal oxide materials.

One of the most important advantages of the sol-gel method is that the chemical composition can be controlled at the molecular level. Whereas conventional solid-state synthesis methods require prolonged milling and sintering at high temperatures to obtain a homogeneous mixture, in the sol-gel

method the precursor compounds are mixed at the atomic level, allowing homogeneous products to be produced at lower temperatures. This reduces energy consumption while also enabling the synthesis of finer-grained materials with higher surface area.

Today, the sol-gel method has an extremely broad range of applications, including photocatalysis, environmental technologies, energy-storage systems, biomaterials, sensors, optical coatings, and adsorption. It continues to attract considerable research interest owing to the advantages it offers in the production of porous structures and nanomaterials with high surface area.

### *Historical Development of the Sol-Gel Method*

The foundations of the sol-gel method were laid in the mid-nineteenth century through studies on the chemical preparation of silica-based glasses. The earliest systematic studies began with research on the polymerization behavior of silicic acid; however, the true development of the method became possible only in the second half of the twentieth century, with advances in metal alkoxide chemistry (Brinker & Scherer, 1990).

From the 1980s onward, growing demand for high-purity thin films in the electronics, optics, and ceramics industries accelerated the development of sol-gel technology (Klein, 1988). In subsequent years, with the rise of nanotechnology, the method came to be widely used not only for thin-film production but also for the synthesis of nanoparticles, nanocomposites, aerogels, and nanofibers.

More recently, the sol-gel process has been combined with methods such as electrospinning, hydrothermal synthesis, 3D-printing technologies, and additive manufacturing, enabling the development of hybrid production techniques (Larsen & Skotak, 2008; Sanchez et al., 2005). These hybrid approaches enhance both morphological control and functional properties.

### *Basic Principle of the Sol-Gel Process*

The sol-gel method is a multi-step synthesis process based on solution chemistry. In general, the process consists of the following steps.

1. Preparation of the precursor solution

Metal alkoxides or metal salts are dissolved in suitable solvents, most commonly low-molecular-weight alcohols such as ethanol, methanol, or isopropanol. Homogeneity of the solution is established at this stage.

2. Hydrolysis reaction

Metal alkoxide molecules react with water to form hydroxyl groups. This reaction is the first step in the formation of the network structure.

$$M(OR)_n + H_2O \rightarrow HO–M(OR)_{n-1} + R–OH \qquad (1)$$

3. Condensation reaction

The hydroxylated species react with one another to form metal–oxygen–metal (M–O–M) bonds, and the three-dimensional inorganic network begins to develop.

$$M–OH + RO–M \rightarrow M–O–M + ROH \qquad (2)$$

4. Gel formation

Once the network structure reaches a critical density, the system loses its fluidity and a gel forms.

5. Aging

Condensation reactions continue within the gel. The pore structure is strengthened, and mechanical strength increases.

6. Drying

The solvent present in the pores is removed. Drying conditions directly affect the pore structure of the final product.

7. Heat treatment (sintering/calcination)

Organic residues are removed, and the crystal structure develops further. This stage also enhances the mechanical strength of the material.

This process can be summarized by the flow diagram shown in Figure 1.

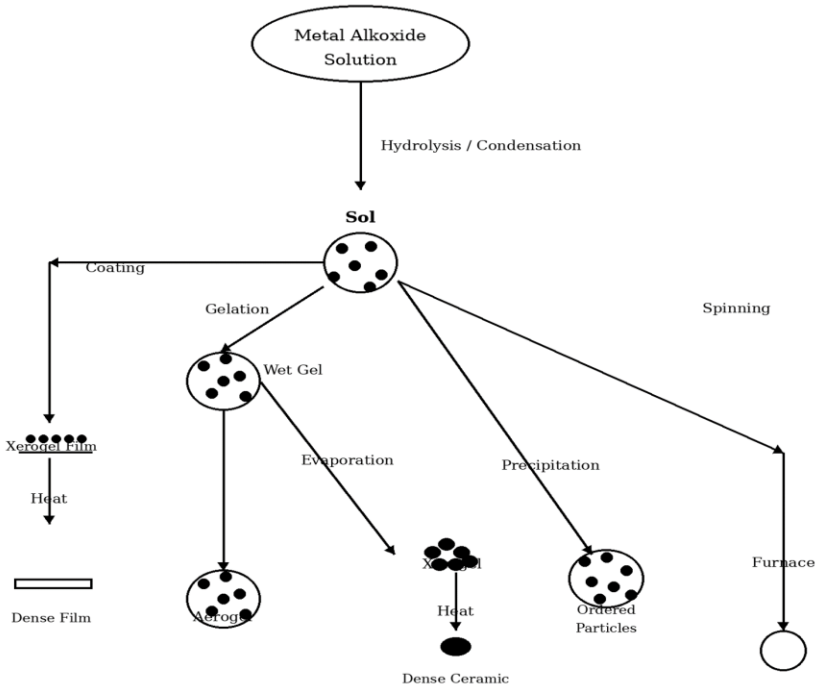



Figure 1. Principal processing steps of the sol-gel process.

***The Concept and Structure of the Sol***

A "sol" is defined as a system in which colloidal-sized solid particles are stably dispersed within a liquid phase. The size of particles within a sol typically ranges from 1 to 100 nm. Because these particles undergo continuous Brownian motion, they display a certain degree of stability against sedimentation.

The principal factors determining the stability of sol systems are:

- Particle size
- Zeta potential
- pH
- Ionic strength
- Solvent properties
- Temperature
- Electrolyte concentration

As long as the electrostatic repulsive forces at the particle surface remain sufficient, the system remains stable. However, as pH changes, ionic

strength increases, or condensation reactions proceed, the particles begin to approach one another and form a three-dimensional network structure. This stage represents the critical point at which the sol system transitions to gelation (Brinker & Scherer, 1990).

### *Precursors Used in the Sol-Gel Process*

The success of sol-gel synthesis depends directly on the chemical properties of the selected precursor compounds. Precursors are the fundamental components that participate in hydrolysis and condensation reactions to form the final metal oxide network (Livage et al., 1988).

The most commonly used precursors are metal alkoxides, with the general formula $M(OR)_x$, where M represents the metal atom, OR the alkoxy group, and x the number of ligands, which depends on the valence electron count of the metal. Common examples are listed in Table 3.

Table 3. Common metal alkoxide precursors and their resulting oxides

| Precursor | Final product |
|---|---|
| TEOS (tetraethyl orthosilicate) | $SiO_2$ |
| TMOS (tetramethyl orthosilicate) | $SiO_2$ |
| Titanium isopropoxide | $TiO_2$ |
| Aluminium sec-butoxide | $Al_2O_3$ |
| Zirconium propoxide | $ZrO_2$ |
| Zinc acetate | ZnO |
| Iron(III) nitrate | $Fe_2O_3$ |

The most important advantages of metal alkoxides are their high purity and controllable hydrolysis behavior. However, the extreme moisture sensitivity of some alkoxides, together with their high cost, has led to the use of metal salts as alternative precursor compounds.

### *Hydrolysis Reactions*

The first chemical step underlying the sol-gel method is the hydrolysis reaction. At this stage, metal alkoxides (M–OR) react with water molecules, replacing the alkoxy groups (–OR) with hydroxyl groups (–OH). The resulting

metal hydroxide species subsequently participate in condensation reactions to form the three-dimensional metal–oxygen network.

$M–OR + H_2O \rightarrow M–OH + R–OH$ (3)

where M is the metal ion (Si, Ti, Zr, Al, etc.), OR is the alkoxy group, and ROH is the alcohol formed as a reaction product.

When hydrolysis is complete, all alkoxy groups are replaced by hydroxyl groups:

$M(OR)_n + nH_2O \rightarrow M(OH)_n + nROH$ (4)

The rate of the hydrolysis reaction is strongly influenced by the amount of water, the type of solvent, temperature, pH, the presence of a catalyst, and the chemical structure of the metal alkoxide used.

***Acid-Catalyzed Hydrolysis***

Under acidic conditions, hydrolysis begins with the protonation of the alkoxy group by hydrogen ($H^+$) ions. The protonated alkoxy group is more readily displaced, facilitating the attack of a water molecule on the metal center. This mechanism is characterized by the following features:

- The hydrolysis rate is high
- The condensation reaction is relatively slow
- More linear and homogeneous polymer chains form
- Gel structures with smaller pore diameters are obtained
- The mechanism is preferred for thin-film production

Structures formed under acid catalysis are generally dense and of low porosity. For this reason, they are widely used in the production of optical coatings, thin films, and transparent ceramics (Hench & West, 1990).

***Base-Catalyzed Hydrolysis***

Under basic conditions, hydroxide ions ($OH^-$) accelerate the reaction. In this case, hydrolysis proceeds at a moderate rate while condensation occurs very rapidly, producing densely branched polymer networks with large pores and higher overall porosity. Base-catalyzed systems are particularly suitable for producing structures used as adsorbents, catalyst supports, membranes, and aerogels.

Table 4. Comparison of acid- and base-catalyzed sol-gel systems

| Property | Acid catalysis | Base catalysis |
|---|---|---|
| Hydrolysis rate | Very high | Moderate |
| Condensation | Slower | Very fast |
| Structure | Linear | Branched |
| Pore size | Small | Large |
| Density | High | Lower |
| Applications | Thin films, optical materials | Adsorbents, catalysts, membranes |

## Condensation Reactions

As a result of hydrolysis, the metal hydroxide species react with one another to form metal–oxygen–metal (M–O–M) bonds. These reactions are termed condensation and can occur through two fundamental mechanisms.

***Water Elimination***

Water is released through the combination of two hydroxyl groups:

$M–OH + HO–M \rightarrow M–O–M + H_2O$

This reaction is one of the most important mechanisms in the development of the ceramic network structure.

***Alcohol Elimination***

An alcohol is formed through the reaction of a hydroxyl group with an alkoxy group:

$M–OH + M–OR \rightarrow M–O–M + ROH$

This mechanism is particularly common in systems in which hydrolysis has not proceeded to completion.

### *Formation of the Three-Dimensional Network Structure*

As condensation reactions progress, the resulting metal–oxygen bonds combine with one another to form a continuous three-dimensional network. This structure imparts high mechanical strength, high thermal stability, and high porosity. Initially, only clusters a few nanometers in size form; over time, these clusters link together to produce a continuous gel network spanning the entire solution (Brinker & Scherer, 1990).

***Gelation***

Gelation is the critical stage at which the sol system loses its fluid character and begins to behave as an elastic solid. At this point, the viscosity of the solution rises sharply, the polymer network structure spans the entire volume, and the particles form continuous connections. Gel time is one of the most important parameters of the process; it is affected by temperature, pH, the water-to-alkoxide ratio, solvent type, catalyst amount, and the coordination structure of the metal ion. Controlled gelation is critical for achieving homogeneous pore distribution and high mechanical strength.

***Aging***

Chemical reactions do not cease entirely once the gel has formed; condensation reactions within the network structure continue for a certain period, a process termed aging. During aging, the chains thicken, new M–O–M bonds form, mechanical strength increases, shrinkage occurs, and pore size is rearranged. Aging time can range from a few hours to several weeks. Prolonged aging generally results in a more durable gel, a more ordered pore structure, and higher crystallinity, whereas excessive aging can lead to pore closure and a reduction in surface area.

## *Factors Affecting Sol-Gel Chemistry*

***Water-to-Alkoxide Ratio (R Value)***

This is one of the most important parameters. A low R value results in slow hydrolysis, smaller particles, and a more controlled structure, whereas a high R value leads to rapid hydrolysis, large clusters, and premature gelation.

***pH***

Under acidic conditions, linear chain formation and small pores are favoured; under basic conditions, branched network structures and high porosity result.

***Temperature***

An increase in temperature accelerates hydrolysis, speeds up condensation, and shortens gel time. Excessive temperature, however, may lead to irregular pore formation.

***Solvent Type***

The polarity, evaporation rate, and viscosity of the solvent significantly affect the pore structure of the final product. Ethanol, methanol, and isopropanol are the most commonly used solvents.

In this section, the fundamental chemical reactions of the sol-gel method have been examined in detail. The hydrolysis stage generates the active species required for condensation by hydroxylating the metal alkoxides, while condensation reactions lead to the formation of metal–oxygen–metal bonds and the development of the three-dimensional network structure. Gelation and aging processes, in turn, play a decisive role in the mechanical strength, porosity, and microstructure of the resulting network.

Through careful control of process parameters, pore size, surface area, particle morphology, and crystal structure can be optimized for the intended application. For this reason, the sol-gel method is regarded as one of the most flexible and widely used chemical production techniques for the synthesis of high-performance nanomaterials (Pierre, 1998).

## ELECTROSPINNING TECHNIQUE

Electrospinning is an electrohydrodynamic method that enables the production of continuous nano/microfibers from polymer solutions or melts under a high electric field (Greiner & Wendorff, 2007; Li & Xia, 2004). Fiber diameter typically ranges from 50 nm to 5 μm, and fibers well below a few hundred nanometers can be obtained under appropriate process conditions. Owing to advantages such as high specific surface area, tunable porosity, low density, and ease of surface functionalization, the method is widely used in filtration, biomedical engineering, sensors, and energy-storage systems (Xue et al., 2019; Bhardwaj & Kundu, 2010).

### *Basic Operating Principle*

A standard electrospinning system consists of a high-voltage power supply, a syringe pump, a metal needle (spinneret), and a collector. The polymer solution forms a droplet at the needle tip. Under the effect of the electric field, the charge density at the surface of the droplet increases, and once the surface tension falls below a critical threshold, a Taylor cone forms. The charged jet ejected from the apex of the cone elongates as it travels toward the collector, undergoes instability motions, and solidifies into a fiber as the solvent evaporates (Li & Xia, 2004).

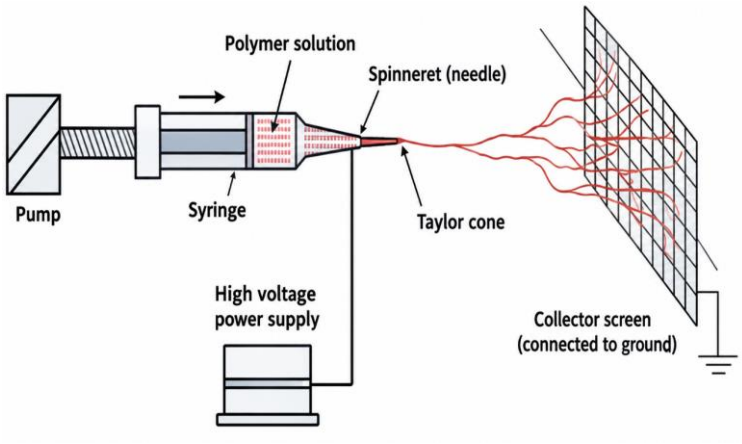



Figure 2. Schematic representation of the electrospinning system.

An electrospinning apparatus essentially consists of three main components:

- High-voltage power supply
- Feeding unit (syringe, metal needle, etc.)
- Collector (conductive plate, rotating drum, etc.)

***Jet Formation and Instability Mechanisms***

The jet initially travels along a straight path; subsequently, electrostatic repulsion gives rise to whipping instability. This motion causes substantial elongation of the jet, thereby reducing fiber diameter. Excessively rapid evaporation of the solvent can lead to porous fibers, whereas excessively slow evaporation can cause the fibers to adhere to one another (Li & Xia, 2004; Bhardwaj & Kundu, 2010).

***Solution Parameters***

Table 5. Effect of solution parameters on fiber formation

| Parameter | General effect |
|---|---|
| Polymer concentration | Low → bead formation; high → thicker fibers |
| Viscosity | Insufficient → discontinuous jet; adequate → continuous fiber |

| | |
|---|---|
| Conductivity | Increase → greater jet elongation, finer fiber |
| Surface tension | High → jet formation is hindered |
| Solvent volatility | Determines solidification rate and pore structure |

***Process and Environmental Parameters***

Applied voltage: generally, 10–30 kV; too low a voltage prevents jet formation, while too high a voltage increases instability.

Flow rate: a high flow rate can lead to thick fibers and bead formation.

Needle-to-collector distance: must allow sufficient time for solvent evaporation (typically 10–20 cm).

Temperature and humidity: humidity particularly affects surface porosity in hydrophilic systems, while temperature lowers viscosity and accelerates evaporation. (Bhardwaj & Kundu, 2010)

## Advanced Electrospinning Approaches

Table 6. Advanced electrospinning configurations and their principal advantages (Huang et al., 2003; Xue et al., 2019)

| **Method** | **Main advantage** |
|---|---|
| Coaxial | Core–shell fibers, controlled drug release |
| Emulsion | Multi-stage fiber structures, biomolecule encapsulation |
| Needleless | High production rate, industrial scalability |
| Melt electrospinning | Solvent-free production, environmental advantage |
| Near-field | Precise fiber placement and patterning |

***Electrospinning Combined with Sol-Gel Chemistry***

A common approach in ceramic nanofiber production is to blend a sol containing metal alkoxides with a polymer carrier, and electrospin the resulting mixture. The composite fibers obtained are subsequently freed of the polymer through drying and calcination, leaving behind networks of ceramic

nanofibers such as $TiO_2$, ZnO, $SiO_2$, or $ZrO_2$. This hybrid method combines the compositional control offered by sol-gel chemistry with the fiber-morphology advantages of electrospinning (Larsen & Skotak, 2008; Sevim & Sevimli, 2014).

***Applications***

Biomedical: tissue scaffolds, wound dressings, controlled drug release (Kim et al., 2005).

Filtration: high-efficiency membranes for PM2.5 and microorganism retention.

Energy: battery separators, electrodes, supercapacitors, and fuel-cell components (Yanilmaz et al., 2016; Aravindan et al., 2013).

Sensors and catalysis: gas sensors, photocatalytic coatings, and adsorbents.

***Advantages and Limitations***

Advantages:

- Simple and relatively low-cost equipment
- Very high surface area and porosity
- Broad range of usable materials (polymers, composites, ceramic precursors)
- Precise control of fiber diameter and morphology

Limitations:

- Low production rate of conventional single-needle systems
- Solvent safety and vapor emissions
- Sensitivity to environmental conditions (particularly humidity)
- Difficulty in maintaining long-term process stability

Electrospinning is one of the most flexible and effective methods for continuous nanofiber production. Fiber morphology can be precisely tuned through the combined control of solution properties, applied voltage, flow rate, needle-to-collector distance, and environmental conditions. Advanced approaches such as coaxial, emulsion, and needleless electrospinning enhance the functionality and scalability of the method. In particular, electrospinning combined with sol-gel chemistry offers a powerful hybrid strategy for producing ceramic nanofibers with high surface area, with considerable

potential in energy, environmental, and biomedical applications (Larsen et al., 2024; Larsen & Gonzalez, 2026).

## COMPARISON OF SOL-GEL AND ELECTROSPINNING TECHNIQUES, HYBRID APPROACHES, AND CURRENT APPLICATIONS

Among the methods used in nanomaterial synthesis, sol-gel processing and electrospinning are two important technologies with a wide range of applications, spanning from laboratory scale to industrial production. Although both methods enable the controlled production of nanostructured materials, they differ significantly in terms of production mechanism, resulting product morphology, process parameters, and application areas.

The sol-gel method relies on chemical reactions to form molecularly homogeneous metal oxide networks, whereas electrospinning is an electrohydrodynamic process that enables the continuous production of nanofibers from polymer solutions under a high electric field. For this reason, the two methods are generally regarded not as alternatives to one another but as complementary technologies in many applications.

### *Comparison of Sol-Gel and Electrospinning Techniques*

#### *Production Mechanism*

The sol-gel method is a chemical synthesis process that begins at the molecular level through hydrolysis and condensation reactions and results in the formation of a three-dimensional inorganic network structure. The final product typically takes the form of nanoparticles, thin films, coatings, or porous ceramics. Electrospinning, in contrast, is based on the transformation of a polymer solution into fine jets by electrostatic forces generated under high voltage, with continuous fibers obtained as the solvent evaporates. The method therefore directly generates fiber morphology.

#### *Product Morphology*

Products obtained via the sol-gel method:

- Nanoparticles
- Thin films
- Aerogels
- Xerogels

- Coatings
- Ceramic blocks

Products obtained via electrospinning:

- Nanofibers
- Microfibers
- Core–shell fibers
- Porous fibers
- Multilayer fiber networks

For this reason, electrospinning offers a clear advantage in applications requiring fibrous structures.

***Process Conditions***

The sol-gel process is generally initiated at room temperature, although calcination is subsequently applied to promote crystallization. Electrospinning, in most cases, can be carried out at room temperature and does not require additional heat treatment during fiber formation; however, in the production of ceramic nanofibers, subsequent calcination is applied to remove the polymer carrier.

***Production Capacity***

The sol-gel method is well suited to working with large-volume solutions and is therefore advantageous for the production of nanoparticles in large quantities. In conventional single-needle electrospinning systems, production rates are low; however, multi-needle and needleless electrospinning systems have substantially increased industrial-scale production capacity (Xue et al., 2019).

Table 7. Comparison of sol-gel and electrospinning techniques

| Property | Sol-Gel | Electrospinning |
|---|---|---|
| **Production principle** | Chemical reaction | Electrohydrodynamic |
| **Product type** | Nanoparticle, film, gel | Continuous nanofiber |
| **Size control** | Very good | Very good |
| **Porosity** | High | Very high |
| **Fiber production** | No | Yes |
| **Thin-film production** | Very suitable | Limited |
| **Industrial scale** | High | Medium–High |
| **Cost** | Low | Low–Medium |
| **Processing temperature** | Low | Room temperature |

### *Hybrid Sol-Gel/Electrospinning Approach*

One of the most important synthesis strategies developed in recent years is the combination of sol-gel and electrospinning techniques within a single production process. In this approach, sol solutions prepared from metal alkoxides are mixed with suitable polymers and electrospun. The resulting composite fibers are converted, through controlled drying and calcination, into fully inorganic ceramic nanofibers (Larsen & Skotak, 2008; Yanilmaz et al., 2016). Through this method, it is possible to obtain, within the same structure, high purity, controlled fiber diameter, high porosity, large surface area, and superior mechanical and thermal stability.

### *Current Applications*

#### *Water Treatment and Adsorption*

Nanomaterials based on $TiO_2$, $Fe_3O_4$, $SiO_2$, and $Al_2O_3$ synthesized via the sol-gel method are used as effective adsorbents for the removal of heavy metal ions, organic pollutants, and dyes. Nanofiber membranes produced via electrospinning, owing to their high porosity, offer high permeability and low-pressure drop-in applications such as microfiltration, ultrafiltration, and oil/water separation. Nanofibers functionalized via sol-gel chemistry further enhance adsorption capacity, providing a significant advantage in hybrid treatment systems (Larsen et al., 2024).

#### *Photocatalysis*

Metal oxides such as $TiO_2$, ZnO, and $WO_3$ are widely used for the degradation of organic pollutants under solar or UV light. The sol-gel method enables these oxides to be produced with high purity and controlled crystal phases, while electrospinning organizes photocatalysts on high-surface-area fiber structures, thereby increasing the amount of active surface available (Larsen et al., 2024).

#### *Energy Storage*

In lithium-ion batteries, sodium-ion batteries, and supercapacitors, the surface area and ion-transport characteristics of the electrode material are of great importance. Nanocrystalline metal oxides synthesized via the sol-gel method provide high electrochemical activity, while electrospun nanofiber electrodes shorten electron-transport pathways, thereby improving charge/discharge performance (Aravindan et al., 2013; Yanilmaz et al., 2016; Larsen & Gonzalez, 2026).

***Sensor Technologies***

In metal-oxide-based gas sensors, the surface area of the nanostructure directly affects sensitivity. Thin films produced via the sol-gel method and electrospun nanofiber networks exhibit high sensitivity in the detection of $NO_2$, $NH_3$, $H_2$, and volatile organic compounds (Wang, 2004).

***Biomedical Applications***

Electrospun nanofibers, owing to their ability to mimic the natural extracellular matrix (ECM), are widely used in tissue engineering. Bioceramics synthesized via the sol-gel method, such as hydroxyapatite and bioactive glass, play an important role in bone regeneration and implant coatings. Hybrid systems can simultaneously enhance biocompatibility and mechanical strength (Kim et al., 2005).

### *Future Perspectives*

The following trends are expected to become increasingly prominent in the future of nanomaterial production technologies:

Process optimization through artificial intelligence and machine learning

Environmentally friendly solvents and sustainable production, in line with green-chemistry principles

Multifunctional nanocomposites

Hybrid production methods integrated with 3D printing

Smart, stimuli-responsive nanofiber systems

Continuous, automated industrial production lines

These developments are expected both to accelerate the commercialization of laboratory research and to contribute to the development of next-generation products in the energy, health, and environmental technology sectors.

## CONCLUSION

In this book chapter, sol-gel processing and electrospinning techniques have been comparatively examined from the standpoint of nanomaterial production. The sol-gel method stands out for its advantages in molecular-level compositional control, low-temperature synthesis, and the production of

high-purity metal oxides, while electrospinning offers outstanding performance particularly in biomedical, filtration, and energy applications owing to its continuous nanofiber production, high specific surface area, and tunable porosity.

The hybrid use of the two methods brings together the strengths of both technologies, enabling the development of high-performance ceramic nanofibers and functional composite materials. Today, the use of such hybrid structures is increasingly observed in fields such as photocatalysis, water treatment, sensor technologies, energy storage, and tissue engineering.

In the future, together with sustainable production approaches, data-driven process optimization, and artificial-intelligence-assisted design methods, sol-gel and electrospinning techniques are expected to be used even more widely at the industrial scale. These developments will expand the application areas of nanotechnology across different disciplines and make significant contributions to the development of high-value-added, next-generation materials.

The rapid advances taking place in nanotechnology have made the controlled production of high-performance, functional nanomaterials one of the central focuses of scientific research. In this context, sol-gel processing and electrospinning are among the most important nanomaterial production methods, owing to their synthetic flexibility, product quality, and broad range of applications. Each method offers distinct advantages, and the choice of production technique depends on the chemical composition, morphology, and intended application of the target material.

The sol-gel method offers significant advantages owing to its ability to form molecularly homogeneous mixtures, to enable the synthesis of high-purity metal oxide and ceramic materials at low temperature, to allow precise control of chemical composition, and to succeed in the production of thin films, coatings, nanoparticles, and porous structures. Through careful control of hydrolysis and condensation reactions, properties such as particle size, crystal structure, porosity, and surface area can be optimized for the intended application. However, shrinkage and crack formation during drying, together with the high cost of certain metal alkoxides, remain among the principal limitations of the method.

The electrospinning technique stands out as one of the most effective methods for continuous nanofiber production. Fiber networks with high specific surface area, tunable pore structure, and applicability to a wide range of polymer, ceramic, or composite systems provide significant advantages in biomedical engineering, filtration, sensor technologies, energy-storage systems, and environmental applications. The ability to precisely tailor fiber morphology through control of solution properties, electric field strength, flow rate, and environmental conditions is one of the method's principal strengths. Nevertheless, the limited production rate of conventional single-needle systems and the sensitivity of the process to environmental conditions remain key disadvantages to be considered in industrial applications.

The comparative assessment presented here indicates that these two methods should be regarded not as alternatives but as complementary technologies. The sol-gel method excels in compositional control and the production of high-purity metal oxides, while electrospinning offers a unique advantage in the formation of three-dimensional fiber networks. Hybrid sol-gel/electrospinning approaches developed in recent years combine the strengths of both methods, enabling the successful production of ceramic nanofibers with high surface area, functional composites, and multipurpose nanostructures (Sevim & Sevimli, 2014). These hybrid systems have produced notable results in advanced applications such as photocatalysis, heavy-metal adsorption, gas sensing, lithium-ion batteries, supercapacitors, and tissue engineering.

Current research indicates that sustainability and digitalization are gaining increasing importance in nanomaterial production processes. The use of environmentally friendly solvents, the reduction of energy consumption, the development of bio-based polymers, and the minimization of waste generation constitute the core components of green production approaches. In addition, modelling methods based on artificial intelligence and machine learning offer significant opportunities for optimizing process parameters, accelerating experimental studies, and reducing production costs.

In conclusion, sol-gel processing and electrospinning are among the most powerful production platforms in contemporary nanotechnology. The use of these methods, either individually or in hybrid combination, makes an important contribution to the development of high-performance, application-

specific nanomaterials. With the future development of sustainable production technologies, smart manufacturing systems, data-driven process optimization, and multifunctional nanocomposites, these two methods are expected to find even broader application in both scientific research and industrial practice. Accordingly, an increase in interdisciplinary research aimed at further developing sol-gel and electrospinning techniques is expected to make important contributions to innovative solutions across many fields, particularly energy, environment, health, and advanced materials technologies.

## REFERENCES


Aravindan, V., Gnanaraj, J., Lee, Y.-S., & Madhavi, S. (2013). Li-ion conducting ceramic materials for electrochemical energy storage. Journal of Applied Electrochemistry, 43, 795–805.

Bhardwaj, N., & Kundu, S. C. (2010). Electrospinning: A fascinating fiber fabrication technique. Biotechnology Advances, 28(3), 325–347.

Brinker, C. J., & Scherer, G. W. (1990). Sol-gel science: The physics and chemistry of sol-gel processing. Academic Press.

Cao, G., & Wang, Y. (2011). Nanostructures and nanomaterials (2nd ed.). World Scientific.

Cushing, B. L., Kolesnichenko, V. L., & O'Connor, C. J. (2004). Recent advances in the liquid-phase syntheses of inorganic nanoparticles. Chemical Reviews, 104(9), 3893–3946.

Greiner, A., & Wendorff, J. H. (2007). Electrospinning: A fascinating method for the preparation of ultrathin fibers. Angewandte Chemie International Edition, 46, 5670–5703.

Guozhong, C. (2004). Nanostructures and nanomaterials: Synthesis, properties and applications. Imperial College Press.

Hench, L. L., & West, J. K. (1990). The sol-gel process. Chemical Reviews, 90(1), 33–72. https://doi.org/10.1021/cr00099a003

Huang, Z.-M., Zhang, Y.-Z., Kotaki, M., & Ramakrishna, S. (2003). A review on polymer nanofibers by electrospinning and their applications. Composites Science and Technology, 63, 2223–2253.

Kim, H. W., Song, J. H., & Kim, H. E. (2005). Nanofiber generation of gelatin–hydroxyapatite biomimetics. Journal of Biomedical Materials Research Part B: Applied Biomaterials.

Klein, L. C. (Ed.). (1988). Sol-gel technology for thin films, fibers, preforms, electronics and specialty shapes. Noyes Publications.

Kumar, A., & Kumbhat, S. (2016). Essentials in nanoscience and nanotechnology. Wiley.

Larsen, G., & Gonzalez, D. (2026). Twenty years of making inorganic ultrafine fibers via electrospinning and sol–gel techniques II: An update on battery and fuel cell applications. Chemical Engineering Communications. https://doi.org/10.1080/00986445.2025.2611386

Larsen, G., & Skotak, M. (2008). Co-solvent mediated fiber diameter and fiber morphology control in electrospinning of sol-gel formulations. Journal of Non-Crystalline Solids, 354, 5547–5554. https://doi.org/10.1016/j.jnoncrysol.2008.08.002

Larsen, G., Gonzalez, D., Noriega, S., & Ragusa, J. (2024). Twenty years of making inorganic ultrafine fibers via electrospinning and sol–gel techniques I: An update on catalysis, photocatalysis, and adsorption R&D opportunities. Chemical Engineering Communications. https://doi.org/10.1080/00986445.2024.2347915

Li, D., & Xia, Y. (2004). Electrospinning of nanofibers: Reinventing the wheel? Advanced Materials, 16(14), 1151–1170.

Livage, J., Henry, M., & Sanchez, C. (1988). Sol-gel chemistry of transition metal oxides. Progress in Solid State Chemistry, 18(4), 259–341. https://doi.org/10.1016/0079-6786(88)90005-2

Lu, X., Wang, C., & Wei, Y. (2009). One-dimensional composite nanomaterials: Synthesis by electrospinning and their applications. Small, 5(21), 2349–2370. https://doi.org/10.1002/smll.200900445

Pierre, A. C. (1998). Introduction to sol-gel processing. Springer.

Poole, C. P., & Owens, F. J. (2003). Introduction to nanotechnology. Wiley.

Ramakrishna, S., Fujihara, K., Teo, W. E., Lim, T. C., & Ma, Z. (2005). An introduction to electrospinning and nanofibers. World Scientific.

Rao, C. N. R., Müller, A., & Cheetham, A. K. (Eds.). (2004). The chemistry of nanomaterials. Wiley-VCH.

Sanchez, C., Belleville, P., Popall, M., & Nicole, L. (2011). Applications of advanced hybrid materials. Chemical Society Reviews, 40(2), 696–753.

Sanchez, C., Julián, B., Belleville, P., & Popall, M. (2005). Applications of hybrid organic–inorganic nanocomposites. Journal of Materials Chemistry, 15, 3559–3592. https://doi.org/10.1039/B509097K

Sevim, F., & Sevimli, E. (2014). Structure investigations of $NiO/Al_2O3/B_2O_3/SiO_2$ composite based on PVA using sol-gel processing and the electrospinning technique. Asian Journal of Chemistry, 26(8).

Wang, Z. L. (2004). Zinc oxide nanostructures: Growth, properties and applications. Journal of Physics: Condensed Matter, 16(25), R829.

Xia, Y., Yang, P., Sun, Y., et al. (2003). One-dimensional nanostructures: Synthesis, characterization, and applications. Advanced Materials.

Xia, Y., Yin, Y., Lu, Y., & McLellan, J. (2003). Template-assisted self-assembly of nanostructured materials. Advanced Functional Materials.

Xue, J., Wu, T., Dai, Y., & Xia, Y. (2019). Electrospinning and electrospun nanofibers: Methods, materials, and applications. Chemical Reviews, 119(8), 5298–5415. https://doi.org/10.1021/acs.chemrev.8b00593

Yanilmaz, M., Lu, Y., Zhu, J., & Zhang, X. (2016). Silica/polyacrylonitrile hybrid nanofiber membrane separators via sol-gel and electrospinning techniques for lithium-ion batteries. Journal of Power Sources, 313, 205–212. https://doi.org/10.1016/j.jpowsour.2016.02.089

# Experimental Methodology in Magnetic Resonance

**Mehmet Zafer KÖYLÜ**

Prof. Dr.; Dicle Üniversitesi Fen Fakültesi Fizik Bölümü zkoylu@dicle.edu.tr ORCİD No: 0000-0002-3540-8204

## 1. Experimental Methodology in Magnetic Resonance

The experimental process in magnetic resonance centers on detecting induced transitions between energy levels established within a magnetic system exposed to a highly intense external magnetic field. These transitions are driven secondarily by a weak radiofrequency (RF) or microwave (MD) field. Consequently, the experimental setup for both Nuclear Magnetic Resonance (NMR) and Electron Spin Resonance (EPR) relies on three primary components.

- **Magnetic Field Source:** A system engineered to generate the high-intensity magnetic field applied to the sample. Water-cooled electromagnets with iron cores are conventionally utilized for this purpose. More recently, superconducting electromagnets have also been adopted, despite their requirement for a continuous supply of liquid helium.
- **RF or MD Source:** This component delivers the requisite energy to stimulate transitions between the discrete energy levels. In NMR applications, an oscillator operating in the radiofrequency spectrum is employed to establish a relatively weak field inside a coil wrapped around the spin system. For EPR, a klystron is used to generate an electromagnetic field that acts upon the spin system positioned inside a resonant cavity.
- **Detection System:** A specialized configuration designed to sense the energy absorbed by the magnetic system, typically utilizing crystal detectors for monitoring.

A magnetic resonance spectrometer comprising these three fundamental blocks can be custom-built to align with specific research objectives, technical standards, or budgetary constraints. Nevertheless, the underlying physical principles remain identical across all designs.

### 1.1 NMR Spektrometresi

A basic schematic of an NMR spectrometer is presented in Figure 1. The spin system is placed inside a coil characterized by a self-inductance coefficient $L_0$, and this entire assembly is positioned between the poles of an electromagnet. An electrical current from the RF source is driven through the coil. A resistor, $R_0$, is integrated to transform the RF source into a constant-current supply.

Upon achieving the resonance condition, the energy absorbed by the spin system induces a drop in both the quality factor (Q-factor) and the wave impedance of the parallel $L_0C_0$ circuit. This reduction manifests as a corresponding decrease in the output voltage monitored by an RF voltmeter connected across the circuit. The resulting voltage drop, recorded as an amplitude variation on the voltmeter, traces out the definitive NMR spectrum.

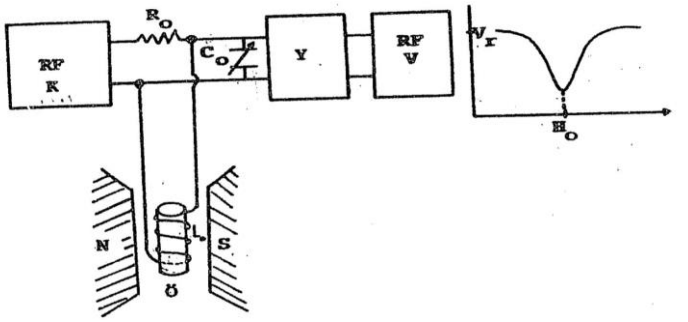



Figure 1. A basic NMR spectrometer. RFK: RF Source, Y: Amplifier, RFV: RF Voltmeter, Ö: Sample Tube.

In both EPR and NMR spectroscopy, the primary objective is to visually monitor the energy absorbed by the spin system on an oscilloscope or to map it onto paper using a chart recorder. To accomplish this, both methodologies standardly employ an alternating current magnetic field paired with the Phase-Sensitive Detector (EDA) technique.

## 2. Equation of Motion in a Non-Isolated Spin System

Under the influence of a constant external field $H_0$ applied along the Z' axis, the magnetic moments within a spin system align such that the net magnetization vector possesses a component exclusively along the Z' direction, denoted here as $M_Z$ (Figure 2a). Consequently, no net magnetization component exists within the transverse (X'-Y') plane.

Let us assume an external field $H_1$ is transiently applied along the X' axis. This external perturbation alters the orientation of the magnetic moments, causing the net magnetization vector to project a transient $M_{Y'}$ component along the Y' direction (Figure 2b).

Next, we terminate the $H_1$ field and evaluate the subsequent relaxation dynamics of these magnetic moments.

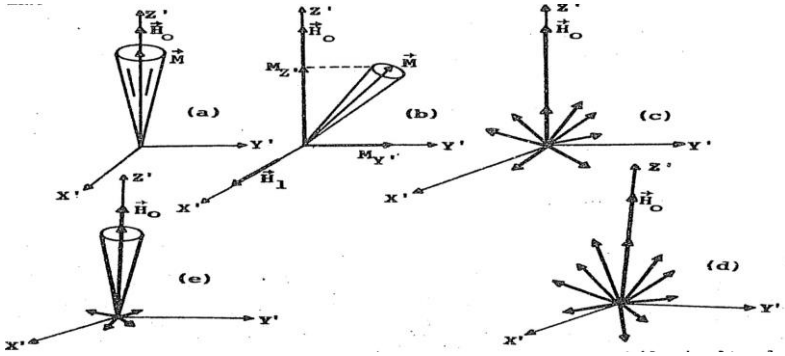



Figure 2. Behavior of magnetic moments under the influence of an RF field and relaxation processes. a) In the absence of the $H_1$ field, b) Following the application of the $H_1$ field, c-e) After the termination of the $H_1$ field.

Because the magnetic dipoles are non-isolated, they actively exchange energy among themselves, leading to a dephasing distribution within the (X'-Y') plane. Furthermore, if the applied external magnetic field exhibits spatial inhomogeneities, each individual magnetic moment experiences a slightly modified local field strength. As a direct consequence, they undergo Larmor precession at varying frequencies.

Due to the combination of these two phenomena, the transverse magnetization component $M_{Y'}$ begins to decay immediately after the $H_1$ field is turned off. Driven by this mechanism, known as the spin-spin interaction, $M_{Y'}$ asymptotically approaches zero over time (Figures 2c and 2d).

Concurrently, because the magnetic dipoles are coupled to their surroundings, they interact directly with the lattice. They dissipate a portion of their absorbed energy into the lattice to re-establish thermal equilibrium. This process, designated as the spin-lattice interaction, forces the longitudinal component of the magnetization vector along the Z' axis to recover toward its thermal equilibrium value, $M_0$. After a characteristic time interval, $M_Z$ fully recovers to $M_0$ while the transverse component $M_{Y'}$ simultaneously decays to zero (Figure 2e).

Therefore, any comprehensive analysis tracking the motion of the net magnetization vector must explicitly account for these relaxation processes. The seminal theoretical formulation of this behavior was introduced by Felix Bloch. In 1946, he became the first scientist to mathematically model these relaxation mechanisms by introducing the fundamental differential equations that now bear his name: the Bloch Equations.

### 2.1 Relaxation Times $T_1$ and $T_2$

The relaxation constants $T_1$ and $T_2$ can be explicitly defined by utilizing the Bloch equations expressed in the rotating frame of reference

$$dM_{X'} / dt = \Delta\omega\, M_{Y'} - M_{X'} / T_2 \quad (1)$$

$$dM_{Y'} / dt = -\Delta\omega\, M_{X'} + \omega_1 M_{Z'} - M_{Y'} / T_2 \quad (2)$$

$$dM_{Z'} / dt = -\omega_1 M_{Y'} - (M_{Z'} - M_0) / T_1 \quad (3)$$

If a 90° RF pulse is delivered to the system and subsequently terminated exactly at the resonance condition, equations (1), (2), and (3) simplify directly to:

$$dM_{X'} / dt = -M_{X'} / T_2 \quad (4)$$

$$dM_{Y'} / dt = -M_{Y'} / T_2 \quad (5)$$

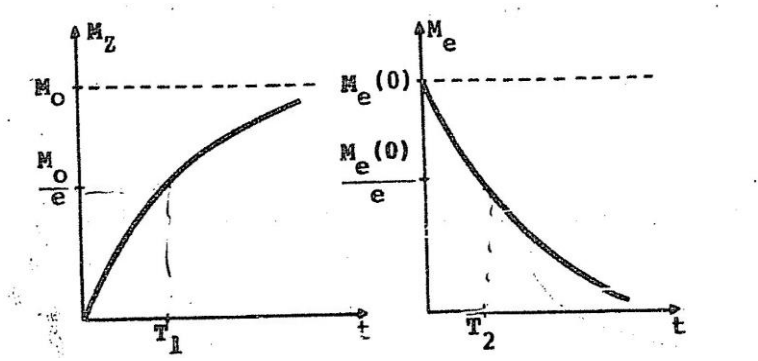
$$dM_{Z'} / dt = -(M_{Z'} - M_0) / T_1 \quad (6)$$

Taking the exact instance when the 90° pulse is cut off as t=0, and assuming that $M_Z = 0$ at this boundary condition (t=0), the analytical solution to equation (6) yields:

$$M_Z(t) = M_0\,(1 - e^{-t/T1}) \quad (7)$$

This exact relationship serves to define the longitudinal relaxation time, $T_1$, which is physically identical to the spin-lattice relaxation time. Specifically, since the application of a 90° pulse rotates the net magnetization entirely into the transverse (X'-Y') plane, $M_Z$ becomes zero. Following the termination of this pulse at t=0, the characteristic time required for $M_Z$ to recover back to a value of $M_0/e$ is defined as the longitudinal relaxation time $T_1$.

Figure 3a illustrates this recovery curve, showing how $M_Z$-which drops to zero due to the 90° pulse-restores itself back toward its equilibrium state $M_0$ once the excitation ceases. From a physical perspective, this curve dictates the rate at which energy absorbed from the RF or MD source is transferred directly to the surrounding lattice.




**(a)** **(b)**

Figure 3. Following the termination of the 90° pulse: a) The recovery of $M_Z$ back to its equilibrium state, b) The decay of $M_e$ back toward its baseline value.

Furthermore, letting $M_{X'}(0)$ and $M_{Y'}(0)$ represent the initial transverse components of the magnetization vector along the X' and Y' axes at the immediate end of the pulse (t=0), the mathematical solutions to equations (4) and (5) are given by:

$$M_{X'}(t) = M_{X'}(0) \cdot e^{-t/T2} \quad (8)$$

$$M_{Y'}(t) = M_{Y'}(0) \cdot e^{-t/T2} \quad (9)$$

Consequently, the total net transverse magnetization at any given time t within the rotating coordinate frame can be expressed as:

$$M_e = M_e(0)e^{-t/T_2} \quad (10)$$

Where the initial magnitude is defined by:

$$Me(0) = \{ (M^2_{X'})(0) + (M^2_{Y'})(0) \}^{1/2} \quad (11)$$

The transverse relaxation time $T_2$, which is physically equivalent to the spin-spin relaxation time, is defined as the time interval required for Me to decay to Me(0)/e after the 90° pulse is cut off. This exponential decay of Me toward zero is depicted in Figure 3b.

Thus far, we have evaluated the relaxation response of the net magnetization vector in the rotating frame after the excitation field is removed under resonance conditions. The corresponding dynamics can also be transformed and examined within the stationary laboratory frame of reference. For instance, if Me is oriented along the laboratory X axis at t=0, the time-dependent evolution of the transverse components can be mapped via equations (8) and (9) as follows:

$$M_X(t) = M_e(0) \cdot e^{-t/T_2} \cdot \cos(\omega_0 t) \quad (12)$$

$$M_Y(t) = M_e(0) \cdot e^{-t/T_2} \cdot \sin(\omega_0 t) \quad (13)$$

Meanwhile, the longitudinal component retains the form given in equation (7):

$$M_Z(t) = M_0 \cdot (1 - e^{-t/T_1}) \quad (14)$$

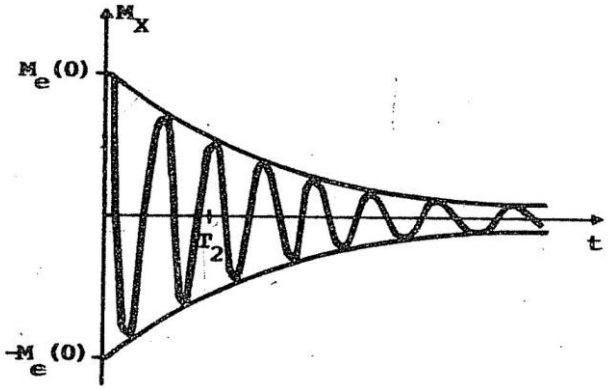



Figure 4. Time-dependent variation of the $M_X$ component following the termination of the 90° pulse.

These equations demonstrate that when subjected to relaxation mechanisms, the path traced by the magnetic system in laboratory coordinates is a damped harmonic motion rather than a simple periodic rotation. As modeled in Figure 4, the time-dependent oscillation of the MX component is bounded by an envelope curve tracking an amplitude variation equal to $Me(0)e^{-t/T2}$.

## 2.2 Free Induction Decay (FID)

In conventional continuous-wave (CW) spectrometer systems, the structural setup and components are designed to apply the RF or MD field as an uninterrupted, ongoing wave. In other words, CW techniques track the behavior of the magnetization vector by keeping the spin system under continuous electromagnetic exposure.

Alternatively, one can probe the system by applying a transient, timed electromagnetic pulse—such as a 90° or 180° pulse—and monitoring the response after it is cut off. For example, if a 90° pulse is briefly applied and then terminated, the $M_X$ and $M_Y$ components will decay in the absence

of the driving field, while the $M_Z$ component recovers back toward its baseline thermal equilibrium state.

Instruments engineered to apply these short interval 90° or 180° pulses to evaluate the post-pulse magnetization dynamics are classified as Pulsed Spectrometers. The primary hardware blocks of a pulsed spectrometer are outlined in Figure 5.

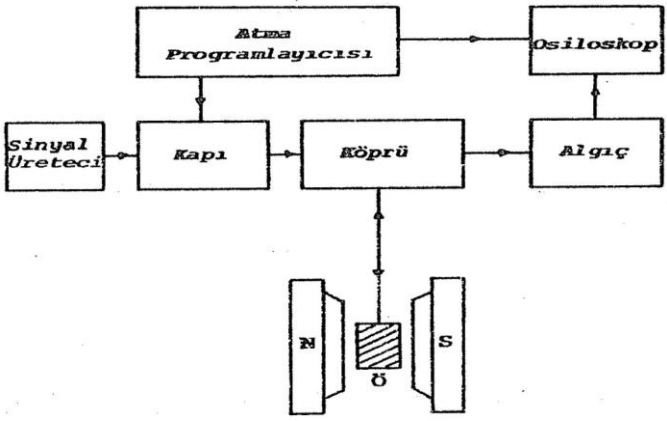



Figure 5. Main functional blocks of a Pulsed Spectrometer.

The signal generator produces continuous-wave RF or MD energy. A gating circuit situated at its output works in tandem with a pulse programmer to shape pulses of adjustable widths and specific intervals (Figure 6). In general terms, a pulse is parameterized by its duration or width, defined by:

$$\tau_a = 2\pi/\gamma H_1 \qquad (15)$$

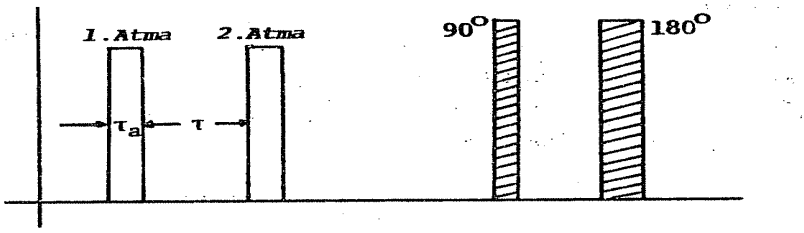



Figure 6. Defining characteristics of an electromagnetic pulse.

As illustrated in Figure 6, the pulses are separated by an adjustable delay time τ. The physical difference between a 90° pulse and a 180° pulse is that the pulse width of the 180° configuration is exactly twice the duration of

the 90° configuration.

In pulsed spectroscopy, a critical criterion for the applied 90° or 180° pulses is that:

$$\tau_a << T_1, T_2 \qquad (16)$$

This condition ensures that the magnetic system behaves as a free system during excitation; meaning relaxation effects are negligible while the pulse is active. Once the 90° pulse is cut off, the magnetization residing in the (X-Y) plane undergoes an exponential decay governed by $T_2$. In the rotating frame, this is quantified by the decay of Me via equation (10), whereas in the laboratory frame, it is traced by the decay of the $M_X$ or $M_Y$ components via equations (12) and (13) (Figure 3).

When a receiver coil is oriented within the (X-Y) plane, this precessing transverse magnetization induces an electrical signal oscillating at the Larmor frequency $\omega_0$. This configuration enables direct tracking of the $M_X$ decay, from which the $T_2$ relaxation time can be experimentally extracted (Figure 3b).

This diagnostic approach closely resembles the dual-coil induction method introduced by Bloch, which is widely known as the nuclear induction method for tracking magnetization. Because the detection window opens only after the excitation pulse has completely ceased, this specific technique is called Free Induction Decay (FID).

If the sample under analysis consists of a single, homogenous spin species and the pulse is delivered exactly at the resonance frequency, the generated FID signal behaves as a mathematically pure, exponentially decaying function (Figure 7). At t=0, the 90° pulse is applied, and it is cut off the instant the magnetization vector aligns fully along the Y' axis (Figure 7a). The subsequent free decay of $M_{Y'}$ is captured as an FID signal using a receiver coil aligned with the Y' axis (Figure 7b).

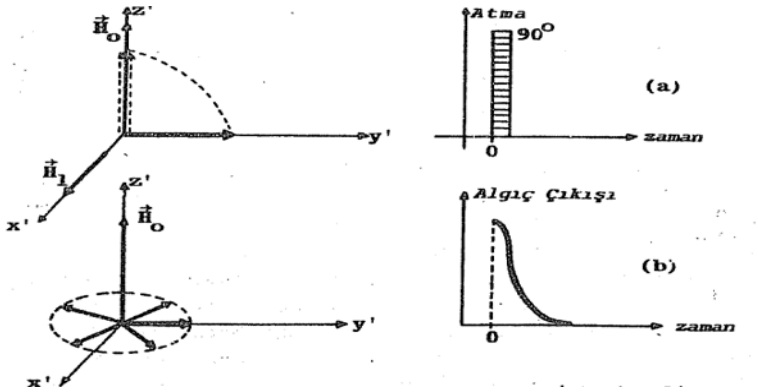



Figure 7. Schematic representation of a Free Induction Decay (FID) signal.

Since the exponential decay envelope of the FID signal is intrinsically linked to the transverse relaxation, it can be directly analyzed to determine $T_2$. However, if the external magnetic field acting on the spin system contains spatial inhomogeneities, the effective decay time constant behaves according to the following modified relationship:

$$1/T^*_2 = 1/T_2 + ½\gamma\Delta H \qquad (17)$$

Here, $\Delta H$ serves as a quantitative measure of the spatial inhomogeneity present within the external magnetic field.

### 2.3 The Inversion Recovery Method

To measure relaxation without field-homogeneity artifacts, an inversion recovery sequence consisting of a 180° - τ - 90° pulse train can be implemented to reconstruct the true relaxation profile. A schematic diagram of this sequence is provided in Figure 8.

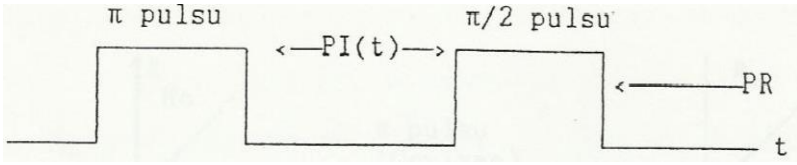



Figure 8. Pulse sequence diagram for the Inversion Recovery method.

The time-dependent trajectories of the magnetization vectors mapping out this Inversion Recovery sequence are detailed in Figure 9. The sequential steps of this process can be broken down as follows: First, a π (180°) pulse acts upon the initial thermal equilibrium magnetization $M_0$, inverting it by 180° to point along the -Z axis, yielding $-M_0$ (Figure 9b). Following this inversion, the system naturally seeks to recover toward its original equilibrium state (Figure 9a).

After a variable delay time $t_1$ (the pulse interval), the recovering magnetization vector reaches a value of $M(t_1)$ (Figure $9c_1$). At this exact moment, a π/2 (90°) readout pulse is applied to rotate this longitudinal magnetization into the X-Y plane, where it is monitored as signal $d_1$. By repeating this entire sequence for a range of distinct delay times satisfying $t_1 < t_2 < t_3 < 5t_1$, one can map out successive magnetization vectors ($c_1$, $c_2$, $c_3$) and record their corresponding signal intensities ($d_1$, $d_2$, $d_3$). Plotting these signal responses as a function of the delay time t yields the curve displayed in Figure 10.

Mathematically, the magnetization vector tracks the following exponential equation:

$$M(t)=M_0\{1-2e^{-t/T_1}\} \quad (18)$$

This governing equation models the longitudinal recovery of, $M_0$ back toward equilibrium after being inverted along the, Z axis by the 180° pulse, which is derived directly by solving the differential equation:

$$dM_Z/dt=-(M_Z-M_0)/T_1 \quad (19)$$

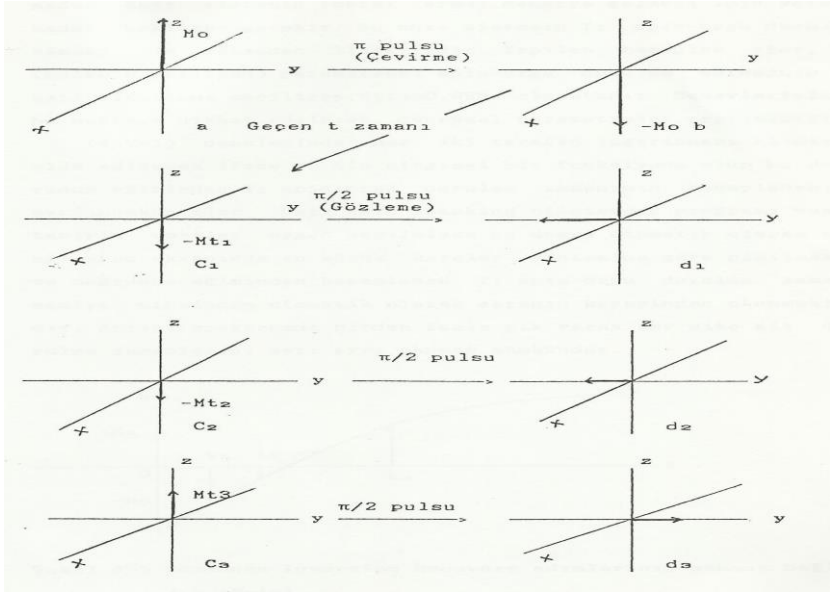



Figure 9. Vector trajectories of the net magnetization vector M during an Inversion Recovery sequence.

When these recovering magnetization vectors are tipped into the X-Y plane by the π/2 monitoring pulse, the initial response observed is the FID signal. As shown in Figure 11, this transient FID signal decays freely over time, capturing the relaxation dynamics.

Critically, after executing one pulse sequence, a sufficient delay period must be observed before initiating the next sequence to allow the spin system to return fully to thermal equilibrium. This necessary wait time is deeply dependent on the spin-lattice relaxation time ($T_1$) of the sample. Theoretical calculations demonstrate that if the Pulse Repetition (PR) delay parameter is set to approximately 5 times the spin-lattice relaxation time (5 × T1), the magnetization successfully restores to $M(t) = 0.99\ M_0$.

By taking the natural logarithm of both sides of Equation (18), the resulting expression becomes a linear function of the delay time t. The slope of this straight line can then be used to calculate the spin-lattice relaxation time $T_1$. Utilizing specialized Auto Stacking software routines, if the input data points are appropriately sampled, this linear fit is automatically performed on the oscilloscope screen via the method of least squares. The calculated $T_1$ value is then directly displayed in seconds in the corner of the screen. Furthermore, if the spectrum contains multiple distinct peaks, the software can resolve and display the individual relaxation times for each peak independently.

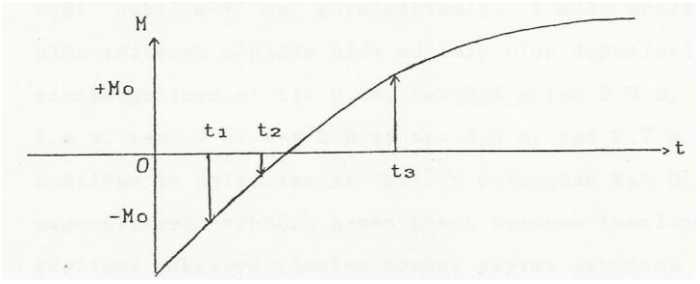



Figure 10. Time-dependence curve reconstructed from observed Inversion Recovery steps.

Figure 11a plots the transient variation of the induced magnetization amplitude over time after the driving RF pulse has ceased. It is important to emphasize that the raw time-domain decay shown in Figure 11a does not constitute a conventional NMR spectrum. To obtain a standard spectrum, a Fourier transform (FT) algorithm must be applied to convert this time-dependent data into a frequency-dependent profile, enabling the final signal to be visualized on a frequency scale (Figure 11b).

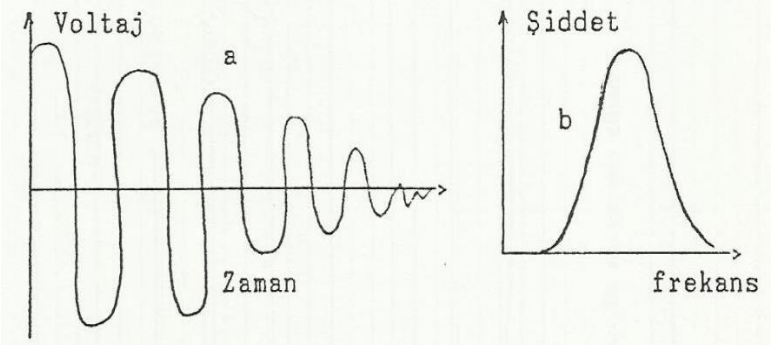



Figure 11. a) The free induction decay of the NMR signal, b) The processed NMR spectrum obtained via Fourier transformation.

## REFERENCES


Abragam, A. 1961. The Principles of Nuclear Magnetism, Oxford, The Clarendon Press. London.

Andrew, E.R. 1958. Nuclear Magnetic Resonance. Cambridge University Press,

Apaydın, F., 1996. Magnetik Rezonans, Hacettepe Üniversitesi, 9-15, 93-106, Ankara.

Assenheim, H.M. 1968. Introduction to Electron Spin Resonance. Adam Hilger, London.

Bahçeli, S. 1996. Çekirdek Manyetik Rezonansın Temelleri, Ankara. Balcı, M. 2004. Nükleer Manyetik Rezonans, ODTÜ Yayıncılık, Ankara.

Bloembergen, N., Purcell, E.M. and Paund, R.V. 1948. Phys. Rev. 73.679. Bloch, F. 1946. Phys.Rev., 70.460.

Bloch, F., Hansen, W.W. and Packard, M. 1946. Phys.Rev. 70.485.

Bloom, S. 1957. Effects of Radiation Damping on Spin Dynamics, J. Appl. Phys. 28, 800.

Bluyssen, H., Dymanus, A., Reuss, J. and Verhoeven, J. 1967. Spin Rotation Constants in $H_2O$, HDO and $D_2O$, Physics Letters. 25, 584-585.

Eisenstadt, M. 1985. NMR Relaxation of Protein and Water Protons in Diamagnetic Hemoglobin Solutions, Biochemistry 21, 434-41.

Gültekin, D.H., Gore, J.C. 2004. Temperature Dependence of Nuclear Magnetization and Relaxation.

Hornak, J.P., 1997-1999. The Basic of NMR, Rochester Istitute of Technology, New York, USA.

Mao, X.A., Wu, D.H. and Ye, C.H., 1994. Radiation Damping Effects on Spin Lattice Relaxation

Mao, X.A., Wu, D.H. and Ye, C.H., 1994. Nuclear Magnetic Resonance Line Shape Theory the Presence of Radiation Damping, Phys. Rev. 49, 22.

Mims, W.B. 1972. Electron Spin Echoes, in Electron Paramagnetic Resonance, S. Geschwind, Ed. Plenum, Bölüm 4, New York.

Poole Jr., C.P. 1967. Electron Spin Resonance. Interscience Pub. New York. Poole Jr., C.P. and

Farach, H.A. 1971. Relaxation in Magnetic Resonance. Academic Inc. New York.

Portis, A.M. 1953. Phys. Rev., 91.1071. Powles, J.G., Smith, D.W.G. 1964. Protein Spin Relaxation in Water. Phys. Letter, 9, 239.

Purcell, E.M., Torrey, H.C. and Paund, R.V. 1946. Phys Rev.

Sherry, A.D., Birnbaum, E.R., Darnall, D.W., 2004. The Binding of Manganese (II) to Bovine Serum Albumin. Department of Chemistry, New Mexico State University, Las Cruces, New Mexico 88003, USA.

Slichter, C.P. 1936. Principles of Magnetic Resonance. Harper and Row, New York.

Slichter, C.P. 1963. Principles of Magnetic Resonance, Harper and Row , Bölüm 2, New York.

Sykes, B.D., Hull, W.E. and Snyder G.H. 1978. Experimental Evidence for the Role of Cross Relaxation Times in Proton Nuclear Magnetic Resonance Spin Lattice Relaxation Time Measurements in Proteins, Biophysical Journal, 21, 137.

Walker, S. and Straw, 1961. Spectroscopy. Vol. I and II. Chapman and Hall Ltd. Weger, M. 1960.

# Conducting Polymer Effects in Supercapacitors

**Murat ATES[1,2]**

1- Prof. Dr.; Tekirdağ Namık Kemal Üniversitesi Fen-Edebiyat Fakültesi Kimya Bölümü. mates@nku.edu.tr ORCID No: 0000.-0002-1806-0330.
2- Prof. Dr.; Nanochem Polymer Energy Müh. Dan. Tic. LTD. Company Silahtaraga Mh., University 1st street, Number: 13/1 Z242/1, 59860, Tekirdag, Turkiye.

## ABSTRACT

Supercapacitors or ultracapacitors can be used for the future supercapacitor devices due to their long-life cycle, easy synthesis, non-toxicity, and economic (Maslana et al., 2020:1349, Ghiyasiyan-Arani and Salavati-Niasari, 2019:23057, Zhang et al, 2020:3599).

Conducting polymers (CPs) are one of the classes of used electrode components in supercapacitor applications. In this review article, a few conducting polymers, such as poly(3,4-ethylenedioxythiophene), polypyrrole, polycarbazole, polyaniline and polythiophene are presented in their various composite materials for supercapacitor electrode materials.

CPs and metal oxides supply higher specific capacitance than carbonaceous materials because of Faradaic reactions (pseudocapacitance) existing in the bulk of material. However, they have limited power density. So, in many studies, both PCs and EDLC mechanisms have been used together for electrode preparation.



*Keywords – Supercapacitor, conducting polymer, pseudocapacitance, polypyrrole, PEDOT.*


## INTRODUCTION

Supercapacitors have more attention because of high power density, fast charge–discharge capability, eco-friendly nature (Peng et al., 2026:287). The hybrid electrode materials have been used to obtain a high electrochemical performance because of their both EDLC and PC mechanisms (Uppugalla et al., 2014:242, Male et al., 2015:231). Supercapacitors have advantages at high specific capacitance, high power density, long-term stability, and fast charge-discharge capability (Viswanathan et al., 2017:483).

Supercapacitors show two main mechanisms, EDLC and PCs processes. The electrical and surface chemical properties of electrode materials affect the electrochemical performance of the supercapacitors (Im et al., 2017:160).

Since electrode materials are the determining factor for supercapacitor performance, conductive polymers such as PANI, and PPy have recently garnered significant interest due to their exceptional attributes, including low-molecular weight, nano-scale, after doped process as a good electrical conductivity, and high thermal and long-term stability. They can be used as electrode components for supercapacitor applications (Ding et al., 2026:971).

### *Poly(3,4-ethylenedioxthiophene) (PEDOT)*

PEDOT is used together with $TiO_2$ nanoparticles and rGO by emulsion polymerization technique in different weight ratios of $TiO_2$ and PEDOT (Gulercan et al., 2019:867). In addition, PEDOT:PSS used together with $MnO_2$ and rGO as a composite form prepared by a two-step process. The electrochemical performance of supercapacitor device shows a specific capacitance of 169.1 F/g at 0.2 A/g. Moreover, it has a cycle stability with capacitance retention of 66.2% at 2 A/g after 2000 cycles (Yan et al., 2014:53). EDOT was electropolymerized on Ni foam (NF) electrode using pulse-modulated mode. It has a potential range 0.0-1.4 V and given a specific capacitance of 39.6 F/g at 0.5 A/g. It has a capacitance retention of 92.5% at 4 A/g for 20.000 cycles (Huang et al., 2027:118323).

### *Polypyrrole (PPy)*

PPy can be preferrable due to its good electrical conductivity, high electrical and optical properties, etc. (Ates, 2011:1).

Perovskite transition metal oxides, such as $LaMnO_3$@CC-PPy materials were synthesized by electrodeposition method in two-step (Sun et al., 2022:601). PPy was added to enhance the electrical conductivity of pure $LaMnO_3$ material. The supercapacitor presented a good specific capacitance of 862 $F.g^{-1}$ at 1 $A.g^{-1}$ with capacitance retention rate of 75% at 10 $A.g^{-1}$. Moreover, it shows good cycle stability of the electrodes after 3.000 cycles at 10 $A.g^{-1}$ with a retention rate of 66%.

Polypyrrole is doped with various electrolytes for supercapacitor applications due to its higher electrical conductivity, structural flexibility, long-term stability and high specific capacitance. However, its usage in supercapacitors is low due to its worse cycling performance (Alias et al., 2026:122462). The chemical synthesis mechanism was presented in **Figure 1**.

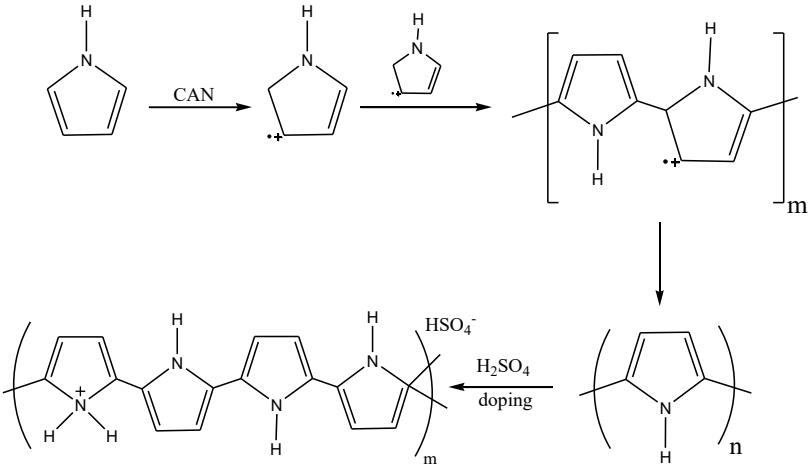



Figure 1. PPy synthesis mechanism.

Although polypyrrole exhibits exceptional electrochemical performance, it has negative properties such as long-term stability and easy degradation during more repeated charge-discharge cycles (Thali et al., 2025:e202502298). Kalavani et al. reported that an ACP@PANI-co-PPy copolymer at $[ANI]_o/[Py]_o$=1:3 electrode showed a specific capacitance of 65 F/g at 50 mV/s, an energy density of 6.7 Wh/kg, and a power density of 312.2 W/kg. Furthermore, it exhibited 99.5% stability after 2000 cycles (Kalavani et al., 2026:19288). Manikant et al. developed poly(aniline-co-pyrrole/$SnO_2$ (PAPSn)) composites with three different $SnO_2$ ratios using an in-situ polymerization technique. In the symmetric electrode system, the highest specific capacitance was calculated as 670 F/g and power density of 696 W/kg, and an energy density of 35 Wh/kg. In addition, it shows a capacitance retention of 85% after 3000 charge-discharge processes (Manikant et al., 2025: 106573).

Rahman et al. synthesized copolymers using various conductive polymers. The highest specific capacitance was obtained as $C_{sp}$= 420 F/g for poly(ANI-co-Py). Compared to other copolymers, the PANI-PPy copolymer exhibited 97.8% capacitive retention after 100 charge/discharge cycles (Rahman et al., 2025:253). Dong et al. synthesized a Ni-$MnO_2$ / PANI-co-PPy electrode material, with a polymer to monomer ratio of 1:5.The highest specific capacitance was measured as 445.49 F/g at 2 mV/s and maintained 61.65% stability after 5000 cycles (Dong at al., 2020: 2051007).

### *Polycarbazole (PCz)*

Polycarbazole (PCz) and its derivatives are synthesized chemically or electrochemically in literature. The electrochemical polymerization supplies a controlled and reprodicible coating on the electrode (Ates et al. 2008:247). PCz was synthesized in two different oxidizing agents (APS and

$KMnO_4$) by controlled interfacial polymerization. It is important to control rate of the electrolyte diffusion. PCz-APS shows better electrolyte diffusion than PCz-$KMnO_4$ at both glassy carbon and platinum electrodes (Kumar et al., 2015:2001).

Table 1: The electrochemical performance reported for some PCz-based electrodes.

| Composite | Enhancement strategy | $C_{sp}$ / F/g | Reference |
|---|---|---|---|
| Hyper-crosslinked microporous PCz | Chemical activation using $K_2CO_3$ | 215 at 0.5 A/g | (Raju et al., 2025:7273) |
| PIn-Cz-NiO | İn-situ method | 155.19 at 5mV/s | (Raghavendra et al., 2025:4300) |
| PCz/CdS/α-$Fe_2O_3$ | using in-situ chemical reactions | 634.14 at 3 mV/s | (Gunasekaran et al., 2025:86) |
| rGO/Zn/PCz | in-situ chemical polym. | 33.88 at 10 mV/s | (Ates et al., 2018:3261) |

***Polyaniline (PANI)***

PANI can be used as electrode component for supercapacitors due to its high electrochemical activity, long-term stability, low-price, easy synthesis, and biocompatibility (Singu et al., 2014;1995; Shayeh et al., 2015;20446). PANI has charges from polaron and bipolaron obtained from synthesis of polymer. So, it mostly depends on the electrolyte ions that called as dopant (Jundale et al., 2013:3526). In literature, there are many studies including PANI as a nanocomposite, components of rGO, PANI, and PVA nanofibers with different initial monomer concentrations of rGO (Chen et al., 2020;5958). It has a specific capacitance of 174 $F \times g^{-1}$. However, the specific capacitance of PANI/PVA electrode was measured as 105 $F \times g^{-1}$. PANI has been used in many energy storage systems because of its wide range of conductivity from insulating to metallic regime, low-cost, easy synthesis, and good environmental stability (Khan et al., 2012).

PANI is generally synthesized by both chemical and electrochemical oxidation polymerization of aniline monomer. In literature, PANI and its additives enable composite material formation, which seems to be a good way to obtain higher electrochemical performance of supercapacitor (Gholivand et al., 2015). There are many characterization techniques

including spin coating, chemical or electropolymerization, etc. In literature, rGO was modified with PANI as a composite form using in-situ chemical oxidative polymerization with surfactant of *N*-propane sulfonic acid (AnS) at 0 ºC. The mass ratio of $w_{rGO}/w_{ANI}$ was taken as 1:1, 1:2, and 1:5. The highest specific capacitance was calculated as 1019 F/g at 1 mV/s by CV method and a capacitive retention of 80% after 100 cycles (Grigoras ve Solonaru, 2017).

PANI is used as an electrode component of supercapacitor due to its high specific pseudocapacitance, good redox efficiency, superior electrical conductivity, and the economic accessibility of aniline monomers (Das et al., 2026:113268).

### *Polythiophene (PTh)*

The specific capacitance of polythiophene (PTh)/carbon nitride (GCN)/vanadium pentaoxide ($V_2O_5$) was calculated as 720 F/g in 3 M KOH at 5 A/g. The average specific energy and specific power were calculated as 32 Wh/kg and 3291 W/kg, respectively. The capacitance retention was obtained as only 96% of the initial capacitance after 5.000 cycles (Ganeson and Xavier, 2024:117101).

## RESULTS AND DISCUSSION

Conducting polymers such as PEDOT, PPy, PCz, PANI and PTh were presented on the usage of supercapacitor applications. They have pseudocapacitor mechanisms on supercapacitor device evaluations. They enhance the electrical conductivity of electrode materials. Conducting polymers have a covalent bond with carbaneous materials such as graphene, active carbon, carbon nanotubes, etc. and metal oxides (NiO, $MnO_2$, $TiO_2$, etc.). As a result, they reduce the prices of electrode materials and enhance the electrical conductivity of the materials.

## REFERENCES

Alias, S., Balakrishnan, R., Lazovic, S., Vinodan, K., George, S.C., Kalarikkal, N.(2026). Advanced in polypyrrole hybrid electrodes for supercapacitors: From molecular design to flexible energy storage, J. Energy Storage, 168, 122462.

Ates, M. (2011). Review study of electrochemical impedance spectroscopy and equivalent electrical circuits of conducting polymers on carbon surfaces, Prog. Org. Coat., 71, 1-10.

Ates, M., Caliskan, S., Özten, E. (2018). Supercapacitor study of reduced graphene oxide / Zn nanoparticle / polycarbazole electrode active materials and equivalent circuit models, J. Solid State Electrochem., 22(10), 3261-3271.

Ates, M., Castillo, J., Sarac, A.S., Schuhmann, W. (2008). Carbon fiber microelectrodes electrocoated with polycarbazole and poly(carbazole-co-p-tolylsulfonyl pyrrole) films for the detection of dopamine in presence of ascorbic acid, Microchim. Acta, 160, 247-251.

Chen, Z., Jiang, Y., Xin, B., Jiang, S., Liu, Y., Lin, L. (2020). Electrochemical analysis of conducting reduced graphene oxide / polyaniline / polyvinyl alcohol nanofibers as supercapacitor electrodes, J. Mater. Sci.: Mater. Electronics, 31, 5958-5965.

Das, S., Banerjee, A., Pal, P., Rudra, S., Nandi, U., & Ghosh, A. (2024). Hydrothermally synthesized gadolinium doped molybdenum disulfide for electrochemical supercapacitor applications. Journal of Energy Storage, 99, Part:A, 113268.

Ding, Z., You, M., Xin, B., & Newton, A. (2026). Architecting triple synergy: MXene-enhanced short carbon nanofibers interlaced with polyaniline for supercapacitors, Ionics, 32(1), 971–986.

Dong, J., Yang, Q., Zhao, Q., Hou, Z., Zhou, Y., & Zhang, R. (2020). Poly(aniline-co-pyrrole)-coated Ni-doped manganese dioxide as electrode materials for supercapacitors. Functional Materials Letters, 13(02), 2051007.

Ganeson, R., Xavier, J.R. (2024). Fabrication of polythiophene / graphitic carbon nitride/$V_2O_5$ nanocomposite for high-performance supercapacitor electrode, Mater. Sci. @ Eng. B, 300:Article number: 11701.

Ghitasiyan-Arani, M., Salavati-Niasari, M. (2019). New nanocomposites based on Li-Fe-Mn double spinel and carbon self-doped graphitic carbon nitrides with synergistic effect for electrochemical hydrogen storage application, Ind. Eng. Chem. Res., 58(51), 23057-23067.

Gholivand, MB., Heydari, H., Abdolmaleki, A., Hosseini, H. (2015). Nanostructured CuO/PANI composite as supercapacitor electrode material, Materials Science in Semiconductor Processing, 30, 157-161.

Grigoras, M., Solonaru, A.M. (2017). Water soluble polyaniline/graphene composites as materials for energy storage applications, Express Polym. Lett. 11(2), 127-139.

Gulercan, D., Commandeur, D., Chen, Q., Sarac, A.S. (2019). A ternary PEDOT-$TiO_2$-reduced graphene oxide nanocomposite for supercapacitor applications, Macromol. Res., 27(9), 867-875.

Gunasekaran, R., Charles, J., Gopal, S.K. (2025). Investigations on electrochemical activity of poly(carbazole)/Cadmium sulfide / Hematite Iron (III) oxide (PCz/CdS/α-$Fe_2O_3$) nanocomposite electrode for supercapacitors, J. Polym. Sci., 63, 86-108.

Huang, Z.T., Chen, Y.C., Xiao, Y.M., Lin, J.Y. (2027). Pulse modulated electropolymerization of poly(3,4-ethylenedioxythiophene) thin films for symmetric supercapacitors, Synth. Met., 322, 118323.

Im, S., Park, Y.R., Park, S., Kim, H.J., Doh, J.H., Kwon, K., Hong, W.G., Kim, B., Yang, W.S., Kim, T.Y. (2017). Nanoparticle intercalation-induced interlayer-gap opened graphene-polyaniline nanocomposite for enhanced supercapacitive performances, Appl. Surf. Sci., 412, 160-169.

Jundale, D.M., Navale, S.T., Khuspe, G.D., Dalavi, D.S., Patil, P.S., Patil, V.B. (2013). Polyaniline-CuO hybrid nanocomposites: Synthesis, structural, morphological, optical and electrical transport studies, J. Mater. Sci. Mater. Electron. 24, 3526-2535.

Kalavani, S., Kiani, A., Alizadeh, A., & Ahmadi, E. (2026). Nature-inspired composite consisting of activated carbon derived from date palm kernel and poly(aniline-co-pyrrole) copolymer as electrode for energy storage devices. Scientific Reports, 16, 19288.

Khan, MMR., Wee, YK., Mahmood, WAK. (2012). Effect of CuO on the morphology and conductivity properties of PANI nanofibers, Synth. Met., 162, 1065-1072.

Kumar, A., Tiwari, M., Prakash, R. (2015). Electrochemical study of interfacially synthesized polycarbazole with different oxidants, Chem ElectroChem, 2, 2001-2010.

Male, U., Srivasan, P., Singu, B.S. (2015). Incorporation of polyaniline nanofibers on graphene oxide by interfacial polymerization pathway for supercapacitor, Int. Nano Lett., 5, 231-240.

Maslana, K., Kalericzuk, R.J., Zieliriska, R., Mijowska, E. (2020). Synthesis and characterization of nitrogen-doped carbon nanotubes derived from g-$C_3N_4$, Materials, 13(6), 1349.

Manikant, Moharana, S., Sutar, A.K., & Maharana, T. (2025). Poly(aniline-co-pyrrole)/$SnO_2$ nanocomposites: Innovative electrode materials for enhanced capacitance for symmetric supercapacitor applications. Reactive and Functional Polymers, 219, 106573.

Peng, M., Cui, Z., Peng, W., Liu, J. (2026). Advances in fluorine-free MXene for electrochemical energy storage, J. Mater. Sci. & Technol., 271, 287-306.

Rahman, M.M., Alam, I., Hossen, M.R., Azim, F., Anjum, N., Faruk, M.O., Rahman, M.M., & Okoli, O. I. (2025). Facile synthesis of conductive copolymers and its supercapacitor application. Journal of Composites Science, 9(5), 253.

Raghavendra, B., Sankarappa, T., Malge, A., Heerasingh, M., Jamadar, P. (2025). Structural, optical, electrical and electrochemical studies of transitional metal oxides doped PIn/PCz copolymer composites, J. Inorganic and Organometallic Polymers and Materials, 35(6), 4300-4314.

Raju, C., Kalla, R.M.N., Rao, S.S., Doraswamy, S., Alshehri, S.A., Lee, J.W. (2025). Carbazole-based porous polymers for utilization in supercapacitors and adsorption of bromothymol blue dye, Ionics, 31(7), 7273-7282.

Shayeh, JS, Norouzi, P., Ganjali, M.R. (2015). Studying the supercapacitive behavior of a polyaniline / nano-structural manganese dioxide composite using first Fourier transform continues cycle voltammetry, RSC Adv., 5(26), 20446-20452.

Singu, B.S., Male, U., Srinivasan, P., Pabba, S. (2014). Use of surfactant in aniline polymerization with $TiO_2$ to PANI:$TiO_2$ for supercapacitor performance, J Solid State Electrochem. 18, 1995-2003.

Sun, X., Hao, Z., Zeng, F., Xu, J., Nan, H., Meng, Z., Yang, J., Shi, W., Zeng, Y., Hu, X., Tian, H. (2022). Coaxial cable-like dual conductive channel strategy in polypyrrole coated perovskite lanthanum manganite for high-performance asymmetric supercapacitors, J. Colloid Interf. Sci., 610, 601-609.

Thali, B.G., Agrahari, D.S., & Kamble, R.M. (2025). Current advancement in electrochemical potential of polypyrrole-based composite materials for supercapacitor application: A Mini Review. ChemistrySelect. 10(18), e202502298.

Uppugalla, S., Male, U., Srinivasan, P. (2014). Design and synthesis of heteroatom doped carbon/polyaniline hybrid material for high performance electrode in supercapacitor application, Electrochim. Acta, 146, 242-248.

Viswanathan, A., Shetty, A.N. (2017). Facile in-situ single step chemical synthesis of reduced graphene oxide polyaniline nanocomposite and its electrochemical performance for supercapacitor application, Electrochim. Acta, 257, 483-493.

Yan, D., Liu, Y., Li, Y., Zhuo, R., Wu, Z., Ren, P., Li, S., Wang, J., Yan, P., Geng, Z. (2014). Synthesis and electrochemical properties of $MnO_2$/rGO/PEDOT:PSS ternary composite electrode material for supercapacitors, Mater Letters, 127, 53-55.

Zhang, X., Liao, H., Liu, X., Shang, R., Zhou, Y., Zhou, Y. (2020). Graphitic carbon nitride nanosheets made by different methods as electrode material for supercapacitors, Ionics, 26, 3599-3607.

# Maximum Likelihood Estimation of Stress–Strength Reliability for Series and Parallel Systems under Lomax and Power Function Distributions

**Sevcan DEMİR ATALAY**[1]
**Ruslan SHUKUR**[2]

1- Prof. Dr. : Ege Üniversitesi Fen Fakültesi İstatistik Bölümü. sevcan.demir@ege.edu.tr ORCID No: 0000-0002-3117-0825
2- Doktora Öğrencisi: Ege Üniversitesi Fen Bilimleri Enstitüsü İstatistikAnabilimdalı. ruslansukur@gmail.com ORCID No: 0000-0002-3020-2020

## ABSTRACT

Stress–strength reliability (SSR) is a fundamental concept in reliability theory for systems exposed to random stresses. In this study, explicit expressions for the SSR of series and parallel systems are derived under special cases of the Lomax and Power Function distributions. Maximum likelihood estimators of the system reliability are obtained, and they are investigated through Monte Carlo simulation using mean square error as a performance criterion. It is shown that, under parameter equality conditions, the system reliability becomes independent of the distributional parameters and depends solely on the number of components.

*Keywords –Lomax Distribution, Power Function Distribution, Stress – Strength Reliability, Series and Parallel System, Monte Carlo Simulation*

## INTRODUCTION

Stress-strength reliability (SSR) is a widely used concept in reliability theory for quantifying the probability that a system's strength exceeds the applied stress. It has been extensively studied under various distributional assumptions and system configurations due to its importance in engineering and stochastic system analysis. In recent years, increasing attention has been given to the estimation of SSR for complex systems, including series and parallel structures, using likelihood-based methods.

The main contribution of this paper is the derivation of explicit SSR expressions for series and parallel systems under special cases of the Lomax and Power Function distributions, together with the development of maximum likelihood estimators and an investigation of their finite-sample performance through simulation.

Many researchers have conducted studies on system reliability and developed statistical and probabilistic models. Stress-strength system reliability, which plays an important role in reliability analysis, is one of them. Let $Y$ be the stress variable of a component that is subjected to, and $X$ be the strength variable, which is the ability of this component to overcome stress. The probability of the strength exceeding the stress is defined as the reliability of the component. This idea was introduced by Birnbaum (1956) and developed by Birnbaum and McCarty (1958). The latter paper is the first to include $P(Y < X)$ in its title. The formal term "stress - strength" appears in the title of Church and Harris (1970). The mathematical expression,

theory, applications and developments of the topic are extensively studied in "The Stress-Strength Model and its Generalizations" by Kotz et al. (2003).

Let the independent variables $X$ and $Y$ be the strength and stress variables, respectively, with cumulative distribution functions $F_X(x)$ and $G_Y(y)$. SSR is defined as

$$R = P(X > Y) = \iint_{y<x} d\,F_X(x) dG_Y(y) \qquad (1)$$

This expression was also defined by Johnson (1988) as follows:

$$R = P(X > Y) = \int_0^\infty P(X > Y|Y = y) g(y) dy = \int_0^\infty \bar{F}(y) dG(y) \qquad (2)$$

where $> 0$ , $y > 0$ and $\bar{F}(x) = 1 - F(x)$.

Many attempts have been made to compute and estimate the probability $R = P(X > Y)$ in the literature. This probability has been used frequently in the calculation of reliability for various system structures, and in point and interval estimations for reliability using parametric or non-parametric techniques. Although studies on $R$ are mainly theoretical, there are also real-life applications in various fields such as physics, engineering, genetics, quality control, psychology, economics, medicine, pedagogy, the pharmaceutical industry, agriculture and biology. Place et al. (1999) used the stress–strength model for the reliability prediction of helicopter transmission systems. Adimari and Chiogna (2006) applied the stress-strength model in the medicine by using a data set on the detection of carriers of Duchenne Muscular Dystrophy. Eryılmaz (2008a) mentioned that stress–strength concept can also be effectively used in various applications, including medicine and psychology although it is frequently considered in the reliability context. An example of a two-treatment comparison was also included in that study. In the study by Eryılmaz (2008b) it was mentioned that a bridge consisting of $n$ cables survives if at least $k$ consecutive cables survive a common random stress available in an environment. Domma and Giordano (2012) defined households whose annual consumption exceeds their annual income as financially fragile and used stress-strength models to measure the financial fragility of households. Hejazi (2017) developed and used stress-strength models for the optimization of $(s, S)$ inventory systems.

SSR has been considered in the literature for various system structures such as series, parallel, $k - out - of - n$ and consecutive $k - out - of - n$ systems, series-parallel and parallel-series systems. Since the stress $Y$ and strength $X$ are random variables, reliability can only be calculated if the

probability distributions of $X$ and $Y$ are known. Therefore, SSR has been studied in the literature for some specific distributions. In these studies, the cases where $X$ and $Y$ follow the same type of distribution were examined and generally focused on calculating and estimating stress – strength reliability for exponential, normal, gamma, Poisson, Weibull, log normal and pareto distributions. Bhattacharyya and Johnson (1974) formulated a stress-strength model for $s$ of $k$ system consisting of identical components. Chung (1982) presented reliability expressions for five stress-strength models of parallel systems with non-identical units. Necsulescu & Krausz (1986) evaluated the SSR of a parallel system. Kunchur & Munoli (1993) considered estimation of SSR of a multicomponent parallel system based on the exponential distribution. Hanagal (1996) obtained the estimation of system reliability in the series system under the assumption of multivariate Pareto distribution. Paul and Uddin (1997) estimated SSR under the assumption of distributions of strength variables are different. Hanagal (1998) evaluated the reliability of a two-component parallel system by considering gamma, Weibull and Pareto distributions. In another study, Hanagal (2003) considered the estimation of SSR of a series system for Gamma, Weibull and Pareto distributions. Eryılmaz (2008a) evaluated the reliability of series, parallel and $n-from-k$ systems using bivariate normal distribution, bivariate Pareto distribution and bivariate FGM distributions. Kundu & Gupta (2006) and Kundu & Raqab (2009) used Weibull distribution, Saracoglu and Kaya (2007) used Gompertz distribution, Eryılmaz (2010) used exponential distribution and Rezai et al. (2014) used Pareto distribution in their studies. Rao and Kantam (2010) employed maximum likelihood estimation (MLE) to estimate the reliability of a multicomponent parallel and series system within a stress-strength model assuming log-logistic distributions. A similar study was done by Rao (2012) for generalized exponential distribution. Hejazi (2017) presented a comparative table of some studies on the SSR. In this table, the studies are classified according to the distribution of the stress and strength variables and whether the variables are one, two or multi-component. Ashok et al. (2019) evaluated the reliability of multi-component series and parallel systems where the stress and strength components follow the Weibull distribution and they obtained a general reliability statement for multi-

component systems. Hamad and Salman (2021), on the other hand, studied reliability estimation and obtained a SSR model for one-component and two-components series systems using the Power Lomax distribution.

In parallel to these studies, one of the aims of this study is to contribute to the theory field by obtaining SSR models of series and parallel systems consisting of $n$-component when both the stress and strength random variables follow the special case of Lomax distribution and Power Function distribution. Both of these distributions have recently been studied in the literature.

The Lomax distribution was first proposed by Lomax in 1954. This distribution is also called the Pareto II distribution. It was originally introduced for modeling business failure data and moreover it has been applied in various fields such as reliability modeling and life testing, engineering, actuarial and biological sciences, medicine, modelling the size of cities and income & wealth data. The distribution has been studied by many researchers in the literature. Harris (1968) used the Lomax distribution for income and wealth data, Atkinson and Harrison (1978) used it for modeling business failure data while Corbelini et al. (2007) used it for queuing problems and to model firm size. It has also found application in modeling the distribution of the sizes of computer files on servers (Holland et al., 2006).

The probability density (pdf) and cumulative distribution functions (cdf) of a random variable $X$ with the Lomax distribution are given below.

$$f(x) = \frac{\alpha}{\lambda}(1+\frac{x}{\lambda})^{-(\alpha+1)} \quad , x > 0 \, , \alpha, \lambda > 0 \tag{3}$$

$$F(x) = 1 - (1+\frac{x}{\lambda})^{-\alpha} \tag{4}$$

Here, $\alpha\ (> 0)$ and $\lambda\ (> 0)$ are the scale and shape parameters, respectively. In this paper, we will study the special case of this distribution where $\lambda = 1$. The probability density function for this case is given by

$$f(x,\alpha) = \begin{cases} \alpha(1+x)^{-(\alpha+1)} & , \ 0 < x < \infty \\ 0 & , \quad o.w \end{cases} \tag{5}$$

and the cumulative distribution function is given by

$$F(x,\alpha) = 1 - (1+x)^{-\alpha} \ , \quad 0 < x < \infty \tag{6}$$

The graph of $f(x,\alpha)$ is given in Figure 1.

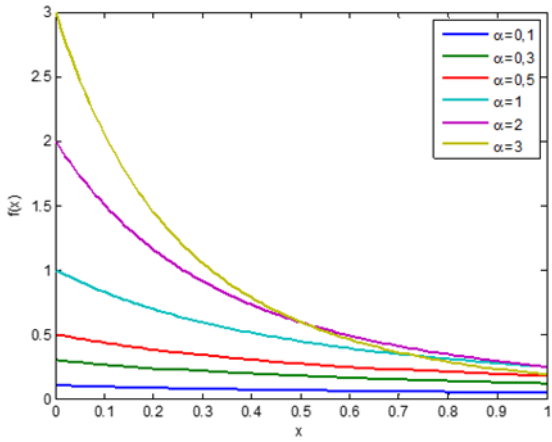



Figure 1. Graph of the pdf of the Lomax distribution when $\lambda = 1$

The Power Function (PF) distribution is a special case of the uniform, beta, and Pearson type I distributions. This distribution is generally applied in modeling of failure data, income data, and lifetime data. Meniconi and Barry (1996) proposed the two-parameter PF distribution as a simple alternative to the exponential distribution when it comes to modeling failure data related to electrical components.

The pdf and cdf of two-parameter PF distribution with scale parameter $\beta$ and shape parameter $\theta$ are, respectively, given by

$$f(x,\beta,\theta) = \begin{cases} \beta\theta^{\beta}x^{\beta-1}, & 0 < x < \theta^{-1} \\ 0, & 0.w \end{cases}, \quad \beta,\theta > 0 \qquad (7)$$

and

$$F(x,\beta,\theta) = \theta^{\beta}x^{\beta},\ 0 < x < \theta^{-1}. \qquad (8)$$

In this paper, we will study the special case of PF distribution where $\theta = 1$. The pdf and cdf of this distribution is given in equation (9) and (10), respectively.

$$f(x,\beta) = \begin{cases} \beta x^{\beta-1}, & 0 < x < 1 \\ 0, & 0.w \end{cases} \qquad (9)$$

$$F(x,\beta) = x^{\beta}, \quad 0 < x < 1 \qquad (10)$$

The graph of $f(x,\beta)$ is given in Figure 2.

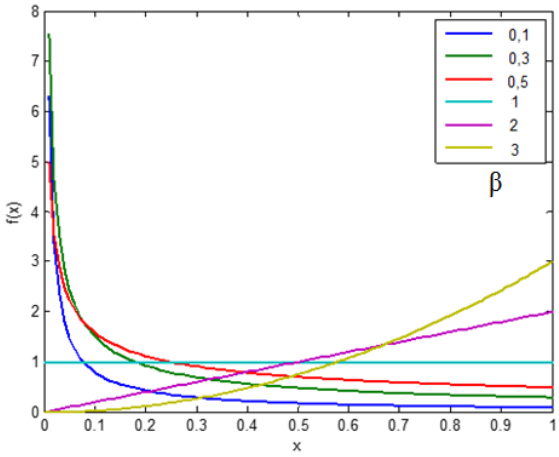



Figure 2. Graph of the pdf of the Power Function distribution when $\theta = 1$

In Section 2 of this study, the SSR of series and parallel systems when both stress and strength random variables follow the special cases of Lomax and PF distributions are obtained. The relationship between the reliability of parallel and series systems for the two distributions is shown. In Section 3, ML estimates of the series and parallel system reliability under the special cases of the Lomax and PF distributions considered in the study are obtained. Monte Carlo simulation is performed to see the performance of these estimates. Mean square error (MSE) values are calculated as a performance measure. Section 4 is the conclusion.

## SSR OF SERIES AND PARALLEL SYSTEMS FOR THE SPECIAL CASE OF LOMAX AND PF DISTRIBUTION

The SSR of the series system is denoted as $R_s = P(X_{(1)} > Y)$ and SSR of the parallel system is denoted as $R_p = P(X_{(n)} > Y)$. Here, $X_{(1)}$ and $X_{(n)}$ denote the smallest order statistic and the largest order statistic, respectively (Bhattacharya and Roychowdhury, 2013).

Let's consider a series system consisting of $n$ components whose random strengths are $X_1, X_2, \ldots, X_n$, respectively. Assume that all components are subjected to the same stress variable $Y$. Let the cumulative distribution function of the stress and strength variables be $G(y)$ and $F(x)$, respectively. Using equation (2), SSR of series system is defined as

$$R_s = P\left(X_{(1)} > Y\right) = \int_0^\infty P(X_{(1)} > Y \mid Y = y) dG(y). \quad (11)$$

Using $P\left(X_{(1)} > y\right) = P(\min(X_1, X_2, \dots, X_n) > y) = P(X_1 > y, X_2 > y, \dots, X_n > y)$ equations, SSR of series system with $n$ components can be calculated with equation (12)

$$R_s = P\left(X_{(1)} > Y\right) = \int_0^\infty [1 - F_X(y)]^n dG(y) \quad (12)$$

Let $X_1, X_2, \dots, X_n$ independent and identically distributed random variables, i.e. $X_i \sim Lomax(\alpha_1)$. Also, $Y \sim Lomax(\alpha_2)$ and the cdf of strength and stress variables are $F(x, \alpha_1) = 1 - (1+x)^{-\alpha_1}$ and $G(y, \alpha_2) = 1 - (1+y)^{-\alpha_2}$, respectively. Using these distribution functions in equation (12), $R_{s_n}^{Lomax}$ is written as

$$R_{s_n}^{Lomax} = P\left(X_{(1)} > Y\right) = \int_0^\infty (1 - 1 + (1+y)^{-\alpha_1})^n \, \alpha_2 (1+y)^{-(\alpha_2+1)} dy$$
$$= \int_0^\infty \alpha_2 (1+y)^{-(n\alpha_1+\alpha_2+1)} dy \quad .$$

By using $u = (1+y)$, $du = dy$ transform, when the distributions of stress and strength variables follow the special case of Lomax distribution, the SSR of series system subjected to a common stress ($R_{s_n}^{Lomax}$) is calculated with equation (13)

$$R_{s_n}^{Lomax} = \frac{\alpha_2}{n\alpha_1 + \alpha_2} \quad (13)$$

Let $X_1, X_2, \dots, X_n$ independent and identically distributed random variables, i.e. $X_i \sim PF(\beta_1)$. Also, $Y \sim PF(\beta_2)$ and the cdf of strength and stress variables are $F(x, \beta_1) = x^{\beta_1}$ and $G(y, \beta_2) = y^{\beta_2}$, respectively. Using equations (9) and (10) in equation (12), the reliability of a series system is written as

$$R_{s_n}^{PF} = P\left(X_{(1)} > Y\right) = \int_0^1 (1 - y^{\beta_1})^n \, \beta_2 y^{\beta_2 - 1} dy.$$

Using Binomial expansion, we can write

$$(1 + (-y^{\beta_1}))^n = \sum_{k=0}^n \binom{n}{k} 1^{n-k} (-y^{\beta_1})^k.$$

Then,

$$R_{s_n}^{PF} = \int_0^1 \left( \sum_{k=0}^n \binom{n}{k} 1^{n-k} (-1)^k (y^{\beta_1})^k \right) \beta_2 y^{\beta_2 - 1} dy$$
$$= \sum_{k=0}^n \binom{n}{k} 1^{n-k} (-1)^k \, \beta_2 \int_0^1 y^{k\beta_1 + \beta_2 - 1} \, dy.$$

After some calculations, the SSR of series systems subjected to a common stress when both stress and strength follow the PF distribution ($R_{s_n}^{PF}$) is calculated with equation (14).

$$R_{s_n}^{PF} = \sum_{k=0}^{n}(-1)^k \frac{\binom{n}{k}\beta_2}{k\beta_1+\beta_2} \tag{14}$$

And now, let's consider a parallel system consisting of $n$ components whose random strengths are $X_1, X_2, \dots, X_n$, respectively, and assume that all components are subjected to a common stress $Y$. Let the cumulative distribution function of the stress and strength variables be $G(y)$ and $F(x)$, respectively. Using equation (2), the SSR of parallel system can be defined as

$$R_p = P\big(X_{(n)} > Y\big) = \int_0^{\infty} P(X_{(n)} > Y \mid Y = y)dG(y).$$

Using

$$P\big(X_{(n)} > y\big) = 1 - P\big(X_{(n)} \le y\big) = 1 - P(\max(X_1, X_2, \dots, X_n) \le y)$$
$$= 1 - P(X_1 \le y, X_2 \le y, \dots, X_n \le y)$$

equations, SSR of parallel systems with $n$ components is calculated with

$$R_p = P\big(X_{(n)} > Y\big) = \int_0^{\infty}(1 - [F_X(y)]^n)dG(y). \tag{15}$$

$X_1, X_2, \dots, X_n$ independent and identically distributed random variables, i.e. $X_i \sim Lomax(\alpha_1)$, $Y \sim Lomax(\alpha_2)$ and cumulative distribution functions of strength and stress variables are $F(x_i, \alpha_1) = 1 - (1+x)^{-\alpha_1}$ and $G(y, \alpha_2) = 1 - (1+y)^{-\alpha_2}$, respectively. Substituting these distribution functions in (15), $R_{p_n}^{Lomax}$ is written as

$$R_{p_n}^{Lomax} = \int_0^{\infty}(1 - [1 - (1+y)^{-\alpha_1}]^n)\,\alpha_2(1+y)^{-(\alpha_2+1)}dy. \tag{16}$$

Using binomial expansion

$$(1 + (-(1+y)^{-\alpha_1}))^n = \sum_{k=0}^{n}\binom{n}{k}1^{n-k}(-(1+y)^{-\alpha_1})^k$$

and substituting the last expression in (16),

$$R_{p_n}^{Lomax} = \int_0^{\infty}\left(1 - [\sum_{k=0}^{n}\binom{n}{k}1^{n-k}(-1)^k((1+y)^{-\alpha_1})^k]\right)\alpha_2(1+y)^{-(\alpha_2+1)}dy$$

is obtained. Using the transformation $u = (1+y),\ du = dy$ and after some calculations, the SSR of the parallel systems subjected to a common stress when both stress and strength follow the Lomax distribution ($R_{p_n}^{Lomax}$) is calculated with

$$R_{p_n}^{Lomax} = 1 - \sum_{k=0}^{n}(-1)^k \frac{\binom{n}{k}\alpha_2}{k\alpha_1+\alpha_2} \quad . \qquad (17)$$

And finally, $X_1, X_2, \dots, X_n$ independent and identically distributed random variables, i.e. $X_i \sim PF(\beta_1)$ and $Y \sim PF(\beta_2)$. The cdf of strength and stress variables are $F(x,\beta_1) = x^{\beta_1}$ and $G(y,\beta_2) = y^{\beta_2}$, respectively. Writing these cdfs in equation (15), $R_{p_n}^{PF}$ is calculated with

$$R_{p_n}^{PF} = P\left(X_{(n)} > Y\right) = \int_0^1 (1-(y^{\beta_1})^n)\,\beta_2 y^{\beta_2-1} dy.$$

After some calculations, the SSR of parallel systems subjected to a common stress when both stress and strength follow the PF distribution ($R_{p_n}^{PF}$) is calculated using (18)

$$R_{p_n}^{PF} = 1 - \frac{\beta_2}{n\beta_1+\beta_2}\,. \qquad (18)$$

According to these results, for $\alpha_1 = \beta_1$ and $\alpha_2 = \beta_2$ and, we can write

$$R_{S_n}^{PF} = 1 - R_{P_n}^{Lomax}$$

and

$$R_{P_n}^{PF} = 1 - R_{S_n}^{Lomax}$$

## ESTIMATION OF SSR OF SERIES AND PARALLEL SYSTEMS FOR THE SPECIAL CASE OF LOMAX AND PF DISTRIBUTIONS

In this section, we consider the estimation of $R_{S_n}^{Lomax}$, $R_{p_n}^{Lomax}$, $R_{S_n}^{PF}$, and $R_{P_n}^{PF}$ using the ML estimation method. Let $X_1, X_2, \dots, X_n$ be a random sample of size $n$ from $Lomax(\alpha_1)$ with the pdf $f(x,\alpha_1) = \alpha_1(1+x)^{-(\alpha_1+1)}$ and $Y_1, Y_2, \dots, Y_m$ is a random sample of size m from $Lomax(\alpha_2)$ with the pdf $f(y,\alpha_2) = \alpha_2(1+y)^{-(\alpha_2+1)}$. Then the log – likelihood function

$$L(x,y,\alpha_1,\alpha_2) = {\alpha_1}^n \prod_{i=1}^{n}(1+x_i)^{-(\alpha_1+1)}\, {\alpha_2}^m \prod_{j=1}^{m}(1+y_j)^{-(\alpha_2+1)} \qquad (19)$$

By taking the derivatives of the likelihood function $L(x,y,\alpha_1,\alpha_2)$ with respect to the parameters $\alpha_1$ and $\alpha_2$ parameters and setting them equal to zero, the ML estimators of these parameters are obtained as follows:

$$\hat{\alpha}_1 = \frac{n}{\sum_{i=1}^{n}\ln(1+x_i)} \quad , \quad \hat{\alpha}_2 = \frac{m}{\sum_{j=1}^{m}\ln(1+y_j)}\,. \qquad (20)$$

Using the invariance property of the ML method, the maximum likelihood estimators of $\hat{R}_{s_n}^{Lomax}$ and $\hat{R}_{p_n}^{Lomax}$ are obtained as follows:

$$\hat{R}_{s_n}^{Lomax} = \frac{\hat{\alpha}_2}{n\hat{\alpha}_1+\hat{\alpha}_2} \text{ and } \hat{R}_{p_n}^{Lomax} = 1 - \sum_{k=0}^{n}(-1)^k \frac{\binom{n}{k}\hat{\alpha}_2}{k\hat{\alpha}_1+\hat{\alpha}_2}. \quad (21)$$

Similarly, let $X_1, X_2, \dots, X_n$ be a random sample of size $n$ from PF($\beta_1$) with the pdf $f(x,\beta_1) = \beta_1$ and $Y_1, Y_2, \dots, Y_m$ be a random sample of size m from $PF(\beta_2)$ with the pdf $f(y,\beta_2) = \beta_2 y^{\beta_2-1}$. Then the log – likelihood function

$$L(x,y,\beta_1,\beta_2) = {\beta_1}^n \prod_{i=1}^{n} x_i^{\beta_1-1} {\beta_2}^m \prod_{j=1}^{m} y_j^{\beta_2-1}. \quad (22)$$

By taking the derivatives of the likelihood function $L(x,y,\beta_1,\beta_2)$ with respect to the parameters $\beta_1$ and $\beta_2$ parameters and setting them equal to zero, the ML estimators of these parameters are obtained as follows:

$$\hat{\beta}_1 = -\frac{n}{\sum_{i=1}^{n}\ln x_i} \quad , \quad \hat{\beta}_2 = -\frac{m}{\sum_{j=1}^{m}\ln y_j}. \quad (23)$$

Using the invariance property of ML method, the ML estimator of $\hat{R}_{s_n}^{PF}$ and $\hat{R}_{p_n}^{PF}$ are obtained as follows:

$$\hat{R}_{s_n}^{PF} = \sum_{k=0}^{n}(-1)^k \frac{\binom{n}{k}\hat{\beta}_2}{k\hat{\beta}_1+\hat{\beta}_2} \text{ and} \hat{R}_{p_n}^{PF} = 1 - \frac{\hat{\beta}_2}{n\hat{\beta}_1+\hat{\beta}_2}. \quad (24)$$

***Simulation Study***

A Monte Carlo simulation was performed for ML estimates of the stress - strength reliability of the series and parallel systems under the assumption that the stress and strength random variables conform to the special case of Lomax and PF distributions. A code was written in the Matlab program for the simulation study.

First, random numbers of $n$ and $m$ sizes were derived from the $U\sim[0,1]$ distribution. These two sets of random numbers were transformed into Lomax distributions with the desired parameters. Thus, sample values for the stress and strength variables were obtained. These sample values were used in the equations given in (20) to obtain ML estimates ($\hat{\alpha}_1$, $\hat{\alpha}_2$) of the parameters of the stress and strength variables conforming to the Lomax distribution. These estimated parameter values were used in equation (21) to obtain $\hat{R}_{s_n}^{Lomax}$ and $\hat{R}_{p_n}^{Lomax}$. All of these steps were also performed to obtain $\hat{R}_{S_n}^{PF}$, and $\hat{R}_{P_n}^{PF}$. The simulation study was performed with 10,000 repetitions.

The MSE value was calculated with the formula (25) using the estimate of the system reliability value obtained in each repetition.

$$MSE(\hat{T}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{T}_i - T)^2 \qquad (25)$$

Here, $T$ is the actual value, and $\hat{T}_i$ is the estimated value obtained at the $ith$ repetition for $i = 1,2,\dots,N$. Considering different parameters $(\alpha_1, \alpha_2, \beta_1, \beta_2)$, different sample sizes $(n, m)$, different number of components $(k)$, MSE values of $\hat{R}_{s_n}^{Lomax}$ and $\hat{R}_{p_n}^{Lomax}$ were obtained and given in Table 1 and ML estimated values of $\hat{R}_{S_n}^{PF}$, and $\hat{R}_{P_n}^{PF}$ were obtained and given in Table 2.

Table1. MSE of $\hat{R}_{s_n}^{Lomax}$ and $\hat{R}_{p_n}^{Lomax}$ for different parameters

| | | MSE for $\hat{R}_{s_n}^{Lomax}$ | | | | | MSE for $\hat{R}_{p_n}^{Lomax}$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $k$ | $R_{s_k}^{Lomax}$ | $(n,m)$ | | | | $R_{p_k}^{Lomax}$ | $(n,m)$ | | | |
| $\alpha_1 = 0{,}2$ $\alpha_2 = 0{,}5$ | | (10, 10) | (20, 20) | (50, 50) | (100, 100) | | (10, 10) | (20, 20) | (50, 50) | (100, 100) |
| 1 | 0,7143 | 0,00855 | 0,00421 | 0,00168 | 0,00085 | 0,7143 | 0,00841 | 0,00433 | 0,00164 | 0,00083 |
| 5 | 0,3333 | 0,00985 | 0,00486 | 0,00201 | 0,00095 | 0,9716 | 0,00212 | 0,00071 | 0,00023 | 0,00011 |
| 10 | 0,2000 | 0,00563 | 0,00269 | 0,00105 | 0,00051 | 0,9929 | 0,00071 | 0,00020 | 0,00004 | 0,00002 |
| 15 | 0,1429 | 0,00358 | 0,00161 | 0,00063 | 0,00030 | 0,9971 | 0,00038 | 0,00008 | 0,00001 | 0,00001 |
| $\alpha_1 = 0{,}7$ $\alpha_2 = 0{,}2$ | | | | | | | | | | |
| 1 | 0,2222 | 0,00639 | 0,00313 | 0,00124 | 0,00060 | 0,2222 | 0,00633 | 0,00317 | 0,00119 | 0,00060 |
| 5 | 0,0541 | 0,00073 | 0,00030 | 0,00011 | 0,00005 | 0,4514 | 0,01797 | 0,00914 | 0,00365 | 0,00186 |
| 10 | 0,0278 | 0,00021 | 0,00009 | 0,00003 | 0,00002 | 0,5422 | 0,02056 | 0,01065 | 0,00440 | 0,00222 |
| 15 | 0,0187 | 0,00010 | 0,00004 | 0,00001 | 0,00001 | 0,5899 | 0,02104 | 0,01137 | 0,00464 | 0,00235 |
| $\alpha_1 = 0{,}5$ $\alpha_2 = 0{,}5$ | | | | | | | | | | |
| 1 | 0,5000 | 0,01196 | 0,00612 | 0,00246 | 0,00126 | 0,5000 | 0,01193 | 0,00627 | 0,00252 | 0,00123 |
| 5 | 0,1667 | 0,00433 | 0,00208 | 0,00079 | 0,00038 | 0,8333 | 0,01192 | 0,00595 | 0,00228 | 0,00115 |
| 10 | 0,0909 | 0,00180 | 0,00078 | 0,00029 | 0,00014 | 0,9091 | 0,00858 | 0,00402 | 0,00140 | 0,00068 |
| 15 | 0,0625 | 0,00090 | 0,00040 | 0,00014 | 0,00007 | 0,9375 | 0,00681 | 0,00282 | 0,00102 | 0,00045 |
| $\alpha_1 = 0{,}2$ $\alpha_2 = 0{,}2$ | | | | | | | | | | |
| 1 | 0,5000 | 0,01181 | 0,00611 | 0,00246 | 0,00123 | 0,5000 | 0,01184 | 0,00604 | 0,00252 | 0,00122 |
| 5 | 0,1667 | 0,00431 | 0,00212 | 0,00079 | 0,00039 | 0,8333 | 0,01192 | 0,00599 | 0,00236 | 0,00116 |
| 10 | 0,0909 | 0,00179 | 0,00076 | 0,00029 | 0,00014 | 0,9091 | 0,00871 | 0,00390 | 0,00144 | 0,00069 |
| 15 | 0,0625 | 0,00095 | 0,00041 | 0,00014 | 0,00007 | 0,9375 | 0,00666 | 0,00282 | 0,00099 | 0,00046 |

| $\alpha_1 = 0{,}8$ $\alpha_2 = 0{,}8$ | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,5000 | 0,01179 | 0,00616 | 0,00249 | 0,00125 | 0,5000 | 0,01182 | 0,00619 | 0,00252 | 0,00128 |
| 5 | 0,1667 | 0,00440 | 0,00210 | 0,00079 | 0,00040 | 0,8333 | 0,01193 | 0,00586 | 0,00229 | 0,00117 |
| 10 | 0,0909 | 0,00178 | 0,00078 | 0,00028 | 0,00014 | 0,9091 | 0,00847 | 0,00384 | 0,00143 | 0,00072 |
| 15 | 0,0625 | 0,00091 | 0,00041 | 0,00015 | 0,00007 | 0,9375 | 0,00671 | 0,00284 | 0,00098 | 0,00048 |
| $\alpha_1 = 0{,}5$ $\alpha_2 = 0{,}8$ | | | | | | | | | | |
| 1 | 0,6154 | 0,01079 | 0,00548 | 0,00224 | 0,00113 | 0,6154 | 0,01099 | 0,00558 | 0,00223 | 0,00115 |
| 5 | 0,2424 | 0,00706 | 0,00346 | 0,00137 | 0,00068 | 0,9246 | 0,00565 | 0,00258 | 0,00092 | 0,00045 |
| 10 | 0,1379 | 0,00342 | 0,00157 | 0,00059 | 0,00029 | 0,9704 | 0,00307 | 0,00112 | 0,00034 | 0,00015 |
| 15 | 0,0964 | 0,00196 | 0,00085 | 0,00032 | 0,00015 | 0,9836 | 0,00181 | 0,00065 | 0,00016 | 0,00007 |

Table 2. MSE of $\hat{R}_{S_n}^{PF}$, and $\hat{R}_{P_n}^{PF}$ for different parameters

| | | MSE for $\hat{R}_{s_n}^{PF}$ | | | | | MSE for $\hat{R}_{p_n}^{PF}$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\beta_1 = 0{,}2$ $\beta_2 = 0{,}5$ | $R_{S_k}^{PF}$ | (n,m) | | | | $R_{p_k}^{PF}$ | (n,m) | | | |
| $k$ | | (10, 10) | (20, 20) | (50, 50) | (100, 100) | | (10, 10) | (20, 20) | (50, 50) | (100, 100) |
| 1 | 0,2857 | 0,00849 | 0,00433 | 0,00167 | 0,00084 | 0,2857 | 0,00853 | 0,00406 | 0,00163 | 0,00086 |
| 5 | 0,0284 | 0,00213 | 0,00080 | 0,00023 | 0,00010 | 0,6667 | 0,00973 | 0,00496 | 0,00195 | 0,00102 |
| 10 | 0,0071 | 0,00071 | 0,00020 | 0,00005 | 0,00002 | 0,8000 | 0,00565 | 0,00277 | 0,00105 | 0,00053 |
| 15 | 0,0029 | 0,00038 | 0,00008 | 0,00001 | 0,00001 | 0,8571 | 0,00351 | 0,00160 | 0,00063 | 0,00030 |
| $\beta_1 = 0{,}7$ $\beta_2 = 0{,}2$ | | | | | | | | | | |
| 1 | 0,7778 | 0,00635 | 0,00317 | 0,00120 | 0,00059 | 0,7778 | 0,00656 | 0,00322 | 0,00120 | 0,00061 |
| 5 | 0,5486 | 0,01762 | 0,00906 | 0,00366 | 0,00187 | 0,9459 | 0,00071 | 0,00031 | 0,00011 | 0,00005 |
| 10 | 0,4578 | 0,02059 | 0,01082 | 0,00442 | 0,00217 | 0,9722 | 0,00022 | 0,00008 | 0,00003 | 0,00002 |
| 15 | 0,4101 | 0,02118 | 0,01096 | 0,00474 | 0,00236 | 0,9813 | 0,00009 | 0,00004 | 0,00001 | 0,00001 |
| $\beta_1 = 0{,}5$ $\beta_2 = 0{,}5$ | | | | | | | | | | |
| 1 | 0,5000 | 0,01194 | 0,00603 | 0,00250 | 0,00127 | 0,5000 | 0,01179 | 0,00621 | 0,00247 | 0,00124 |
| 5 | 0,1667 | 0,01207 | 0,00591 | 0,00235 | 0,00117 | 0,8333 | 0,00451 | 0,00210 | 0,00081 | 0,00040 |
| 10 | 0,0909 | 0,00821 | 0,00384 | 0,00147 | 0,00069 | 0,9091 | 0,00166 | 0,00076 | 0,00029 | 0,00014 |
| 15 | 0,0625 | 0,00641 | 0,00267 | 0,00100 | 0,00047 | 0,9375 | 0,00093 | 0,00040 | 0,00015 | 0,00007 |
| $\beta_1 = 0{,}2$ $\beta_2 = 0{,}2$ | | | | | | | | | | |
| 1 | 0,5000 | 0,01179 | 0,00621 | 0,00245 | 0,00124 | 0,5000 | 0,01168 | 0,00614 | 0,00248 | 0,00124 |
| 5 | 0,1667 | 0,01176 | 0,00594 | 0,00237 | 0,00119 | 0,8333 | 0,00434 | 0,00208 | 0,00082 | 0,00039 |
| 10 | 0,0909 | 0,00843 | 0,00383 | 0,00146 | 0,00069 | 0,9091 | 0,00179 | 0,00079 | 0,00027 | 0,00014 |
| 15 | 0,0625 | 0,00637 | 0,00278 | 0,00099 | 0,00046 | 0,9375 | 0,00091 | 0,00039 | 0,00015 | 0,00007 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\beta_1 = 0{,}8$ $\beta_2 = 0{,}8$ | | | | | | | | | | |
| 1 | 0,5000 | 0,01196 | 0,00609 | 0,00246 | 0,00124 | 0,5000 | 0,01206 | 0,00610 | 0,00248 | 0,00124 |
| 5 | 0,1667 | 0,01225 | 0,00590 | 0,00237 | 0,00115 | 0,8333 | 0,00444 | 0,00203 | 0,00078 | 0,00039 |
| 10 | 0,0909 | 0,00827 | 0,00399 | 0,00138 | 0,00071 | 0,9091 | 0,00174 | 0,00077 | 0,00028 | 0,00014 |
| 15 | 0,0625 | 0,00655 | 0,00275 | 0,00099 | 0,00048 | 0,9375 | 0,00091 | 0,00039 | 0,00015 | 0,00007 |
| $\beta_1 = 0{,}5$ $\beta_2 = 0{,}8$ | | | | | | | | | | |
| 1 | 0,3846 | 0,01114 | 0,00566 | 0,00216 | 0,00112 | 0,3846 | 0,01089 | 0,00547 | 0,00222 | 0,00109 |
| 5 | 0,0754 | 0,00577 | 0,00252 | 0,00094 | 0,00043 | 0,7576 | 0,00710 | 0,00344 | 0,00139 | 0,00069 |
| 10 | 0,0296 | 0,00291 | 0,00110 | 0,00035 | 0,00016 | 0,8621 | 0,00333 | 0,00155 | 0,00059 | 0,00029 |
| 15 | 0,0164 | 0,00192 | 0,00061 | 0,00017 | 0,00007 | 0,9036 | 0,00197 | 0,00086 | 0,00031 | 0,00015 |

From Table 1 and Table 2, it is seen that as the number of components increases, the reliability of the parallel system increases while the reliability of the series system decreases. For the same $\alpha_2$, $R_{s_n}^{Lomax}$ and $R_{p_n}^{Lomax}$ decrease when $\alpha_1$ increases. On the contrary, for the same $\beta_2$, $R_{s_n}^{PF}$ and $R_{p_n}^{PF}$ increase when $\beta_1$ increases. Whereas, for the same $\alpha_2$, $R_{s_n}^{Lomax}$ and $R_{p_n}^{Lomax}$ increase when $\alpha_1$ decreases. On the contrary, for the same $\beta_2$, $R_{s_n}^{PF}$ and $R_{p_n}^{PF}$ decrease when $\beta_1$ decreases. When the parameters are equal ($\alpha_1 = \alpha_2, \beta_1 = \beta_2$), the reliability of the system becomes independent of the parameters, i.e. when the parameters are equal, the reliability value depends on the number of components.

## CONCLUSION

In this study, the SSR expressions for series and parallel systems are derived under special cases of the Lomax and PF distributions. Maximum likelihood estimators for the proposed reliability measures are obtained, and their finite-sample performance is examined through Monte Carlo simulation. The results demonstrate that the estimators exhibit stable behavior and that, under parameter equality conditions, the resulting reliability becomes independent of the distributional parameters and depends solely on the system size. Furthermore, as can be seen from Table 1 and Table 2, MSE values decrease as the sample size increases.

The proposed framework may be extended to other system configurations and distributional assumptions, which could be considered in future studies.

## REFERENCE


Adimari, G., and Chiogna, M. (2006). Partially parametric interval estimation of Pr{Y >X}. *Computational Statistics & Data Analysis* 51, 1875 – 1891.

Ashok, P.Swathi, N., Devi, M.T., and Maheswari, T.S.U. (2019). Reliability for a Multicomponent System Using Mixture of Two Weibull Distributions, *International Journal of Recent Technology and Engineering*, Vol.7, 5S2, 556-559.

Atkinson, A., and Harrison, A. (1978), Distribution of personal wealth in Britain. *Cambridge University Press*, Cambridge

Bhattacharyya, G.K, and Johnson, R. A. (1974), Estimation of reliability in a multicomponent stress-strength model. *Journal of the American Statistical Association*, 69(348), 966–970.

Bhattacharya, D. and Roychowdhury, S. (2013), Reliability of a Coherent System in a Multicomponent Stress-Strength Model, *American Journal of Mathematical and Management Sciences*, 32(1), 40-52.

Birnbaum, Z.W. (1956), On a use of Mann-Whitney statistics. *Proc. Third Berkeley Symp. in Math. Statist. Probab.,* Vol. 1, 13-17, University of California Press, Berkeley, CA.

Birnbaum, Z., McCarty, R. et al. (1958), A distribution-free upper confidence bound for $[Y < X]$, based on independent samples of $X$ and $Y$. *Annals of Mathematical Statistics*, 29(2), 558–562.

Chung, W.K. (1982), Some stress-strength reliability models, *Microelectronics Reliability*, 22(2), 277-280.

Church, J. D., and Harris, B. (1970), The estimation of reliability from stress-strength relationships. *Technometrics,* 12(1), 49–54.

Domma, F., and Giordano, S. (2012), A stress-strength model with dependent variables to measure household financial fragility, *Stat. Methods Appl.,* vol. 21, 375-389.

Eryilmaz, S. (2008a), Multivariate stress–strength reliability model and its evaluation for coherent structures, *Journal of Multivariate Analysis,* 99, 1878–1887.

Eryılmaz, S. (2008b), Consecutive k-out-of-n: G system in stress-strength setup, *Communications in Statistics-Simulation and Computation*, 37(3), 579-589.

Eryılmaz, S. (2010), On system reliability in stress-strength setup, *Statistics and Probability Letters*, vol. 80, 834-839.

Hamad, A.M. and Salman, B. (2021), On estimation of the stress-strength reliability on POLO distribution function, *Ain Shams Engineering Journal*, 4037-4044.

Hanagal, D. D. (1996), Estimation of system reliability under bivariate Pareto distribution. *Parishankyan Samikkha*, 3,13-18.

Hanagal, D. D. (1998), Estimation of system reliability in stress-strength models for distributions useful in life testing, *IAPQR Transactions*, 23(1), 61-65.

Hanagal, D. (2003), Estimation of system reliability in multicomponent series stress–strength models. *Jounal of Indian Statistical Association* 41, 1–7.

Harris, C.M. (1968), The Pareto Distribution as a Queue Service Discipline. *Operations Research*, 16(2), 307-313.

Hejazi, T. H. (2017), A multiresponse model for reliability-based simulation optimization in systems subjected to random external stresses*, Qual Reliab. Engng Int*., vol.33, 1225–1233.

Holland, O., Golaup, A. and Aghvami, A. H. (2006), Traffic characteristics of aggregated module downloads for mobile terminal reconfiguration, *IEE Proc. Commun*. 135, 683-690.

Johnson, R.A. (1988), 3 Stress-strength models for reliability, *Handbook of Statistics*, Vol.7, 27-54.

Kotz, S., and Pensky, M. (2003), The stress-strength model and its generalizations: theory and applications. Singapore: World Scientific.

Kunchur, S.H., and Munoli S.B. (1993), Estimation of reliability for a multicomponent survival stress-strength model based on exponential distributions, *Communications in Statistics - Theory and Methods*, 22(3), 769-779.

Kundu, D., and Gupta, R. D. (2006), Estimation of [$Y < X$] for weibull distributions. *IEEE Transactions on Reliability*, 55(2), 270–280.

Kundu, D., and Raqab, M.Z. (2009), Estimation of R = P {Y < X} for three-parameter Weibull distribution. *Statistics & Probability Letters* 79, 1839–1846.

Meniconi, M. and Barry, DM. (1996), The power function distribution: a useful and simple distribution to assess electrical component reliability. *Microelectron Reliab.*, 36(9), 1207-1212

Necsulescu, D.S., and Krausz A.S. (1986), A Multi-Step Stress-Strength Model of a Parallel System, *IEEE Transactions On Reliability*, R-35(1), 119-123.

Paul, R. K. and Uddin, B., 1997, Estimation of reliability of stress-strength model with non-identical component strengths, *Microelectronics and Reliability*, 37(6), 923-927.

Place, C.S., Strutt, J.E., Allsopp, K., Irving, P.E. and Trille C. (1999), Reliability Prediction Of Helicopter Transmission Systems Using Stress–Strength Interference With Underlying Damage Accumulation, *Quality And Reliability Engineering International*, 15, 69–78.

Rao, G. S. (2012), Estimation of reliability in multicomponent stress-strength model based on generalized exponential distribution. Colomb. J. Stat. 35(1), 67–76.

Rao, G. S., Kantam, R. R.L. (2010), Estimation of reliability in multicomponent stress-strength model: Log-logistic distribution. Electron. J. Appl. Stat. Anal. 3(2), 75–84.

Saracoglu, B., and Kaya, M.F. (2007), Maximum likelihood estimation and confidence intervals of system reliability for Gomperts distribution in stress–strength models. *Selcuk Journal of Applied Mathematics* 8, 25–36.

# Comparative Histomorphology and Ultrastructural Organization of the Digestive System in Selected Vector Mosquito Larvae

**Sıla DAŞ[1]**
**Yücel BAŞIMOĞLU KOCA[2]**

1-Master's Degree Student; Aydın Adnan Menderes University, Graduate School of Natural and Applied Sciences, Department of Biology. 2510300111@stu.adu.edu.tr ORCID NO: 0009-0004-1950-4345
2- Professor Dr.; Aydın Adnan Menderes University, Faculty of Science, Department of Biology. ykoca@adu.edu.tr ORCID NO: 0000-0002-1031-923X.

## ABSTRACT

Mosquitoes are major vectors of numerous arboviruses and parasitic pathogens and pose a significant threat to global public health. Consequently, the aquatic larval stage represents a key target for vector control strategies. This chapter presents a semi-systematic review aimed at comparing the normal histological and ultrastructural organization of the digestive system in fourth-instar larvae of four medically important mosquito species: *Aedes aegypti*, *Aedes albopictus*, *Culex pipiens*, and *Culex quinquefasciatus*. The literature search was conducted using the ScienceDirect, PubMed, and Google Scholar databases, focusing on studies that investigated the digestive system of healthy larvae using light microscopy and transmission electron microscopy.

The reviewed literature indicates that the general anatomical organization of the larval digestive tract is largely conserved across the species examined, although individual regions exhibit structural specializations associated with their physiological functions. The foregut is characterized by a chitinous intima that facilitates food transport and provides mechanical protection. The midgut, which is the principal site of digestion and nutrient absorption, is distinguished by a single-layered cuboidal or columnar epithelium, prominent microvilli, and metabolically active epithelial cells. The gastric caeca increase the absorptive surface area, whereas the peritrophic matrix forms a semipermeable barrier that helps separate food contents from the midgut epithelium and contributes to epithelial protection. Structural specializations of the hindgut are associated with ion transport and osmoregulation. The comparative assessment highlights substantial similarities in midgut morphology between *Ae. aegypti* and *Ae. albopictus*. Studies of *C. pipiens* and *C. quinquefasciatus* provide complementary insights into the structural organization of different regions of the larval digestive system.

In conclusion, this review provides a comparative framework for understanding the normal histological and ultrastructural organization of the digestive system in vector mosquito larvae. It offers a foundation for future research in vector histopathology and the development of targeted bioinsecticides.



*Keywords: Mosquito larvae; digestive system; histomorphology; ultrastructure; midgut; peritrophic matrix; vector control.*

## 1. INTRODUCTION

Mosquitoes are among the most important vector organisms involved in the transmission of numerous infectious diseases affecting both human and animal health through the pathogens they harbor. Owing to their major epidemiological significance and central role in vector control programs worldwide, these mosquitoes have been the focus of extensive biological and ecological research. Taxonomically, *Aedes aegypti*, *Aedes albopictus*, *Culex pipiens*, and *Culex quinquefasciatus* belong to the family Culicidae within the order Diptera, class Insecta, and phylum Arthropoda (Demirsoy, 1997; Klowden, 2013). All four species are classified within the subfamily *Culicinae*, one of the most medically important mosquito lineages. Within this subfamily, *Aedes aegypti* and *Aedes albopictus* are members of the genus *Aedes*, whereas *Culex pipiens* and *Culex quinquefasciatus* belong to the genus *Culex* (Herms, 1950; Christophers, 1960; Harbach, 2011:1). Among these species, *Aedes aegypti* and *Aedes albopictus* are recognized as highly efficient vectors of arboviral diseases, including dengue, Zika, chikungunya, and yellow fever, whereas *Culex pipiens* and *Culex quinquefasciatus* serve as principal vectors of West Nile virus and the lymphatic filariasis parasite *Wuchereria bancrofti* (Herms, 1950; Busvine, 1980; Demirsoy, 1997).

The biology of mosquito larvae holds critical importance in entomological and vector control research, particularly due to their aquatic developmental life cycle. The larval digestive system exhibits a linear organization extending from the mouth to the pharynx, esophagus, foregut, midgut, ileum, colon, and rectum (Wigglesworth, 1942:56; Christophers, 1960). Although species within the same genus exhibit high morphological and ecological similarity, at the intergeneric level, significant microscopic variations have been reported in digestive system structures such as the foregut, midgut, and hindgut of larvae (Snodgrass, 1959:1; Soffar, 2012:145; Cabeça et al., 2025:85). These taxonomic and anatomical differences are of critical importance in terms of shaping the natural microbiota of larvae and determining the limits of their cellular responses to targeted botanical or biological insecticides (Silva et al., 2008:663; Yu et al., 2015:1006).

Within this linear system, the midgut is not only a structure responsible for digestion and absorption processes but also constitutes the primary biological barrier against environmental factors. Gastric caeca, located in the anterior region of the midgut contribute to nutrient processing by increasing the digestive surface area. The peritrophic matrix (PM) present in this region

acts as a semi-permeable structure surrounding the food bolus and serves as an important defense mechanism protecting the epithelial tissue against mechanical damage and pathogen invasion (Richards & Richards, 1971:2253; Billingsley & Rudin, 1992:87; Eisemann & Binnington, 1994:15). The PM is composed of chitin microfibrils, proteins, and glycoproteins, forming a complex structure that regulates the spatial and functional organization of digestion (Stam et al., 1978:175; Lehane, 1997:525; Tellam et al., 1999:87).

Modern histological and ultrastructural studies have demonstrated that the midgut epithelium of mosquito larvae exhibits a highly specialized and metabolically active structure. Epithelial cells contain dense microvilli, an advanced endoplasmic reticulum, and secretory granules, playing an active role in digestion and substance transport (Silva et al., 2008:663; Lemos et al., 2018:1). Furthermore, this tissue represents a highly sensitive target for environmental toxins and biological agents.

In recent years, histopathological and ultrastructural studies have revealed that plant extracts and biological agents induce significant morphological alterations in the midgut tissues of mosquito larvae. These effects include epithelial cell degeneration, loss of microvilli, and disruption of peritrophic matrix integrity (Al-Mehmadi & Al-Khalaf, 2010:77; Al-Doaiss et al., 2021:274; Arafa & Salama, 2021:105; Yu et al., 2015:1006).

In this context, the present study aims to comparatively demonstrate the normal histological and morphological characteristics of the foregut, midgut, and hindgut structures in the larvae of four important vector species (*Aedes aegypti*, *Aedes albopictus*, *Culex pipiens*, and *Culex quinquefasciatus*). The findings are expected to provide a fundamental reference for future histopathological investigations and the development of biological control strategies.

## 2. MATERIALS AND METHODS

### 2.1. Literature Selection, Data Synthesis And Analysis Criteria

This study was conducted within the framework of a semi-systematic review approach. The data used were obtained through an extensive literature survey performed in the ScienceDirect, PubMed, and Google Scholar databases. In the literature screening process, histological and ultrastructural studies addressing the larval digestive system and anal gill structures of *Aedes aegypti*, *Aedes albopictus*, *Culex pipiens*, and *Culex quinquefasciatus* were evaluated.

The inclusion criterion for studies was defined as the presence of clearly identifiable histological and ultrastructural data derived from control groups of fourth-instar (final-stage) larvae of the relevant species. Toxicity-focused studies were included only when morphological data from healthy control groups were explicitly and clearly reported. This approach aimed to enable comparative analysis of normal tissue organization independent of pathological effects.

The literature data obtained from selected studies were systematically classified within a species-based comparative analysis framework. During the analysis process, intestinal structures (epithelial morphology, microvilli architecture, cellular organization, and the peritrophic matrix (PM)), as well as anal gill structures, were considered the primary evaluation parameters of morphological and histological organization.

**2.2 Investigated Species and Biological Stages**

For comparative analysis, fourth-instar (final-stage) larvae, in which morphological differentiation is most pronounced, were considered. This developmental stage was selected due to its suitability for the most accurate evaluation of epithelial organization and ultrastructural features. The species included in the analysis belong to the highly vector-competent genera *Aedes* and *Culex*, namely *Aedes aegypti*, *Aedes albopictus*, *Culex pipiens*, and *Culex quinquefasciatus* (Silva et al., 2008:663; Al-Doaiss et al., 2021:274).

**2.3. Comparative Parameters**

To enable a standardized evaluation of interspecific morphological variation, epithelial cell organization, microvilli density and integrity, nuclear localization, cytoplasmic granule distribution, peritrophic matrix (PM), and anal gill structure were defined as the primary analytical parameters.

**2.4. Microscopic Evaluation Standards**

In this review, both classical morphological descriptions reported in the literature and recent histological and ultrastructural findings from experimental studies were evaluated using a holistic approach. Only studies containing clearly defined, methodologically reliable morphological data derived from control groups and enabling interspecific comparison were included in the analysis. In this context, structural information provided in

classical sources was integrated with control-group data from experimental studies to establish a reliable reference framework for the normal organization of larval tissues, independent of pathological effects.

In light microscopic histological evaluations reported in the literature, Hematoxylin and Eosin (H&E) and Periodic Acid–Schiff (PAS) staining techniques are most commonly used for the assessment of tissue integrity and cellular morphology. While H&E staining allows the examination of general tissue organization and nuclear morphology, the PAS reaction is particularly important for the identification of carbohydrate-rich structures and the peritrophic matrix (PM).

For ultrastructural analyses, Transmission Electron Microscopy (TEM) data served as the primary reference, allowing for a detailed evaluation of epithelial cell organization, microvilli integrity, mitochondrial density, and rough endoplasmic reticulum (rER) structure (Silva et al., 2008:663; Cabeça et al., 2025:85). This methodological framework allows interspecific comparisons to be assessed using standardized and comparable morphological criteria.

## 3. GENERAL ANATOMY AND ORGANIZATION OF THE LARVAL DIGESTIVE SYSTEM

The general body plan of mosquito larvae consists of three main regions: head, thorax, and abdomen. The abdomen is typically composed of eight distinct segments and terminates in the posterior ninth abdominal segment (terminal anal segment). Located in this terminal region, osmoregulatory structures (anal papillae) play a vital role in maintaining ionic and water balance between the larva and its external environment (Snodgrass, 1959:1; Soffar, 2012:145).

In larvae, the digestive system exhibits a continuous tubular structure extending from the mouth opening to the anus. Anatomically and embryologically, this system is divided into three main regions: foregut, midgut, and hindgut (Jiraungkoorskul & Jiraungkoorskul, 2015:15; Al-Doaiss et al., 2021:274; Soffar, 2012:145) **(Figure 1).**

In vector mosquito larvae, the digestive tract is not an anatomically isolated structure; rather, it is closely organized and coordinated with accessory tissues such as the fat body, muscle tissue, and salivary glands located beneath the exoskeleton. This tissue integration is considered one of

the key morphological indicators of preserved functional capacity in healthy control larvae (Lemos et al., 2018:1) **(Figure 2).**

In fourth-instar larvae of *Culex pipiens*, the anatomical relationship between different body systems and overall tissue integrity constitutes a fundamental basis for morphological characterization of the species. As observed in histological sections, the transition between the foregut and midgut reflects the continuity of tissue architecture in *Culex pipiens*. The fat body occupying large areas of the body cavity plays a central role in energy storage and metabolic regulation. In addition, the exoskeletal cuticle, which forms the outermost layer, and the muscular layer, which supports the internal organs, ensure the structural stability of the body plan. The comparison of this complex tissue organization with other species provides critical data for both taxonomic differentiation and the analysis of cellular responses to environmental factors (Al-Doaiss et al., 2021:274) **(Figure 3).**

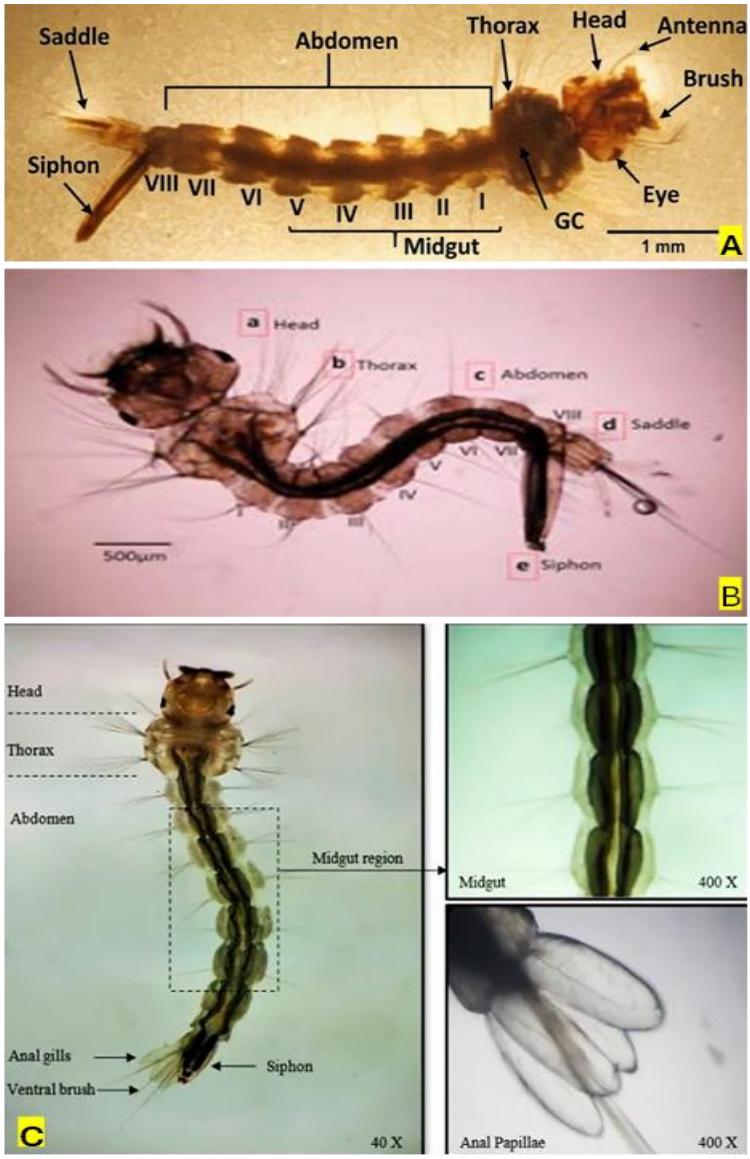



Figure 1. General morphological and histological characteristics of mosquito larvae. **(A)** General appearance of a *Culex pipiens* larva observed under a stereomicroscope (×10). The larval body is composed of three main regions: the head, thorax, and abdomen. The abdomen consists of eight distinct segments. The midgut (mesenteron), which represents the principal component of the digestive system, originates from the gastric caeca (GC) located in the first abdominal segment. The midgut extends through the second and third abdominal segments, while the posterior midgut occupies the fourth and fifth segments. The sixth to eighth abdominal segments constitute the hindgut (proctodeum).

**(B)** Light microscopic image of a *Culex quinquefasciatus* larva (×40). The head, thorax, and abdominal regions are clearly distinguishable, together with the respiratory siphon and the protective saddle structure located in the posterior region.
**(C)** Microscopic images of a fourth-instar (*IV instar*) *Aedes aegypti* larva. The general morphology (×40) illustrates the head, thorax, abdominal segments, and midgut structure. The posterior region contains the anal gills, ventral brush, and respiratory siphon. Higher magnification (×400) reveals detailed features of the midgut epithelium as well as the anal papillae.
**(D)** Microscopic appearance of a fourth-instar (*IV instar*) *Aedes albopictus* larva. The head, thorax, and abdominal regions are readily observed, together with the respiratory siphon and anal papillae located at the posterior end. These morphological and histological observations were interpreted in accordance with previous reports (Jiraungkoorskul & Jiraungkoorskul, 2015:15; Al-Doaiss et al., 2021:274; Sharma et al., 2022; Soviana et al., 2023).




Figure 2. Superficial transverse section through the thoracic region of an *Aedes aegypti* larva showing the organization of the gastric caeca and adjacent tissues (H&E, ×200). The section reveals the spatial arrangement of the exoskeleton (EXO), fat body (FB), gastric caeca (GC), muscle tissue (MU), and salivary gland (SG) (Lemos et al., 2018:1).

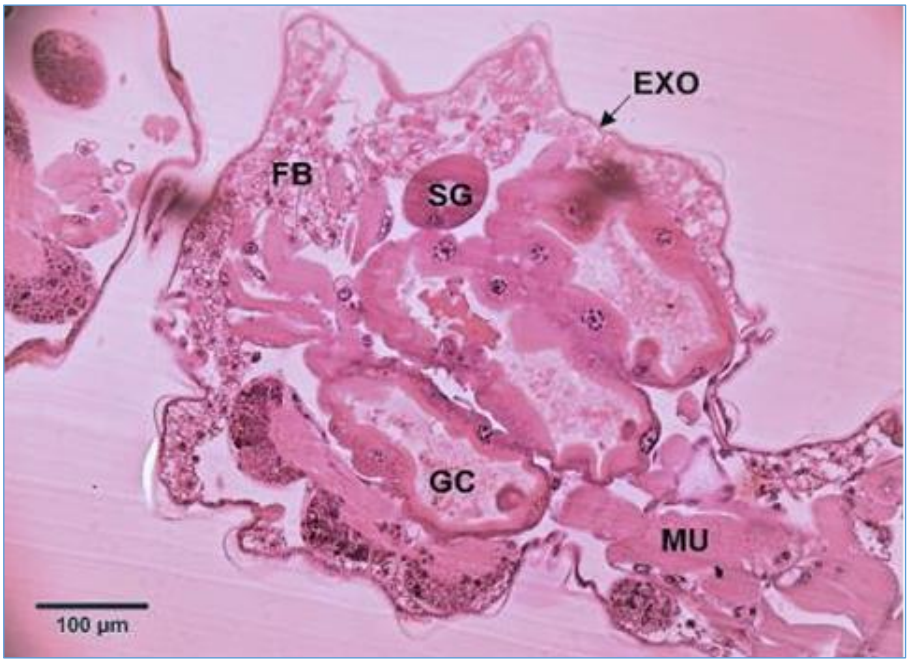



Figure 3. Light microscopic appearance of the foregut and midgut regions in a longitudinal section of a fourth-instar (*IV instar*) *Culex pipiens* larva. The epithelial layer, fat body (FB), exoskeleton (EXO), and muscle tissue (Ms) are observed with well-preserved histological organization, indicative of normal tissue architecture (H&E, ×40) (Al-Doaiss et al., 2021:274).

In adult *Culex pipiens*, body organization exhibits advanced differentiation associated with terrestrial adaptation and reproduction while maintaining the fundamental larval tissue plan. In longitudinal sections of female individuals, flight muscles, ovaries, and all segments of the digestive tract are clearly observable, while abdominal ganglia and the heart reflect intersystemic integration. In male individuals, testicular structures and genital ducts of the reproductive system are histologically distinguishable. These morphological transformations from larval to adult stages represent structural

adaptations maintained throughout the biological life cycle of the species (Gregor et al., 2022:836) **(Figure 4).**

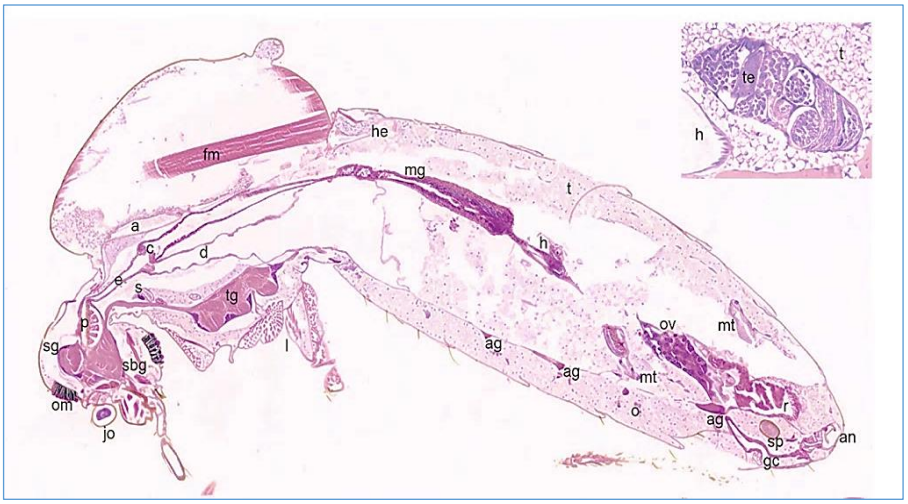



Figure 4. Normal anatomical structure of adult *Culex pipiens* specimens stained with hematoxylin and eosin (H&E). Main image: longitudinal section of an adult female; inset: genital duct system of an adult male. Anatomical structures identified include the aorta (a), abdominal ganglia (ag), anus (an), cardia (c), diverticulum (d), esophagus (e), flight muscles (fm), gonoduct (gc), hindgut (h), heart (he), Johnston's organ (jo), legs (l), midgut (mg), Malpighian tubules (mt), oenocytes (o), ommatidia (om), ovary (ov), pharynx (p), rectum (r), salivary gland (s), subesophageal ganglion (sbg), supraesophageal ganglion (sg), spermatheca (sp), trophocytes (t), testes (te), and thoracic ganglion (tg) (Gregor et al., 2022:836).

### 3.1. Foregut

The foregut constitutes the first portion of the alimentary canal, extending from the oral cavity through the pharynx and esophagus to the proventriculus (cardia) (Journal of Vector Borne Diseases, 2018). Embryologically of ectodermal origin, this structure functions as a protective barrier against the mechanical effects of food particles due to the chitinous intima layer lining its internal surface, thereby safeguarding the underlying epithelial tissue.

In larvae, food intake is achieved via a highly developed pharyngeal pump located in the head region, where rhythmic contractions of the pharyngeal muscles generate negative pressure, enabling the uptake of water and organic particles into the alimentary canal (Lemos et al., 2018:1). The

esophagus functions as a narrow tubular transition region that transports ingested material toward the midgut, while the surrounding circular and longitudinal muscle fibers facilitate the conduction of food along the canal (Silva et al., 2008:663; Lemos et al., 2018:1).

The proventriculus (cardia), which forms the boundary between the esophagus and the midgut, regulates food passage through a valve-like structure and clearly demarcates the morphological separation between these two regions (Journal of Vector Borne Diseases, 2018). Histological sections obtained from fourth-instar larvae clearly demonstrate the spatial proximity of the esophagus, proventriculus, and gastric caeca located in the anterior midgut (Lemos et al., 2018:1) **(Figure 5).**

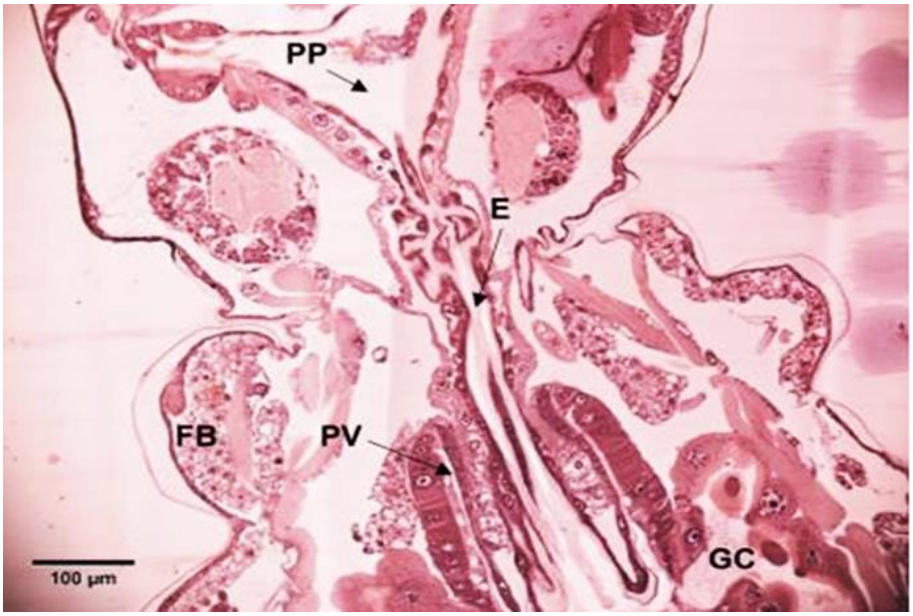



Figure 5. General appearance of a longitudinal section through the head and thoracic regions of a fourth-instar (*IV instar*) *Aedes aegypti* larva (H&E, ×200). The section shows the esophagus (E), fat body (FB), gastric caeca (GC), pharyngeal pump (PP), and proventriculus (PV), illustrating the structural organization of the anterior digestive tract and associated tissues (Lemos et al., 2018:1).

At the genus-level anatomical organization, examinations conducted by Al-Mehmadi (2010:125) on *Culex quinquefasciatus* larvae have shown that, in healthy individuals, epithelial cells of the foregut exhibit a regular organizational pattern in both anterior and posterior regions, with preserved

tissue integrity. This organized structure indicates mechanical stability in food intake and transport processes **(Figure 6).**

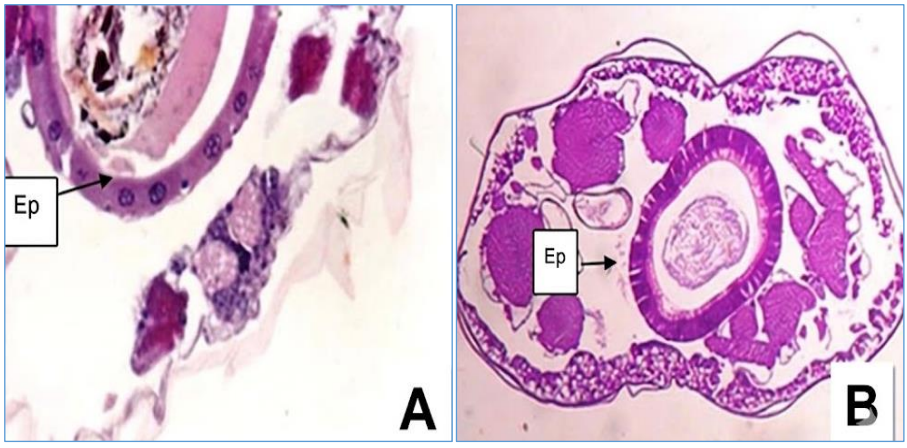



Figure 6. Histological appearance of foregut epithelial cells (Ep.) in healthy (control group) third-instar *Culex quinquefasciatus* larvae. Transverse sections obtained from the anterior (A) and posterior (B) regions demonstrate the normal morphology and organization of the foregut epithelium (H&E, ×200) (Al-Mehmadi, 2010:125).

### 3.1.1. Ultrastructural Features of the Foregut

Transmission electron microscopy (TEM) observations indicate that foregut epithelial cells exhibit a relatively simple internal organization compared with those of the midgut. The most prominent characteristic of these cells is the presence of a thick, electron-dense chitinous cuticular layer (intima) at the apical surface, rather than microvilli (Soffar, 2012:145). Regarding intracellular organelle distribution, a limited number of mitochondria and rough endoplasmic reticulum (rER) profiles have been observed, consistent with a comparatively low metabolic activity (Soffar, 2012:145; Silva et al., 2008:663).

### 3.2. Midgut (Mesenteron)

The midgut (mesenteron) represents the most physiologically active region of the larval digestive tract. In this region, digestion is mediated through the synthesis and secretion of digestive enzymes, as well as the absorption of nutrients. Embryologically of endodermal origin, the midgut differs from the foregut and hindgut in that it lacks a chitinous intima layer. Its luminal surface is lined by a single-layered epithelium composed of

prismatic or cuboidal cells, including differentiated cell types responsible for digestion, secretion, and absorption.

Although the general organization is conserved in *Culex* species, the extension of the midgut along abdominal segments and its lumen volume have been reported to be greater than those observed in *Aedes* species. These structural differences provide an important morphological basis for evaluating cellular differentiation and regional functional specialization within the midgut epithelium, the most metabolically active region of the digestive system (Al-Doaiss et al., 2021:274) **(Figure 7).**

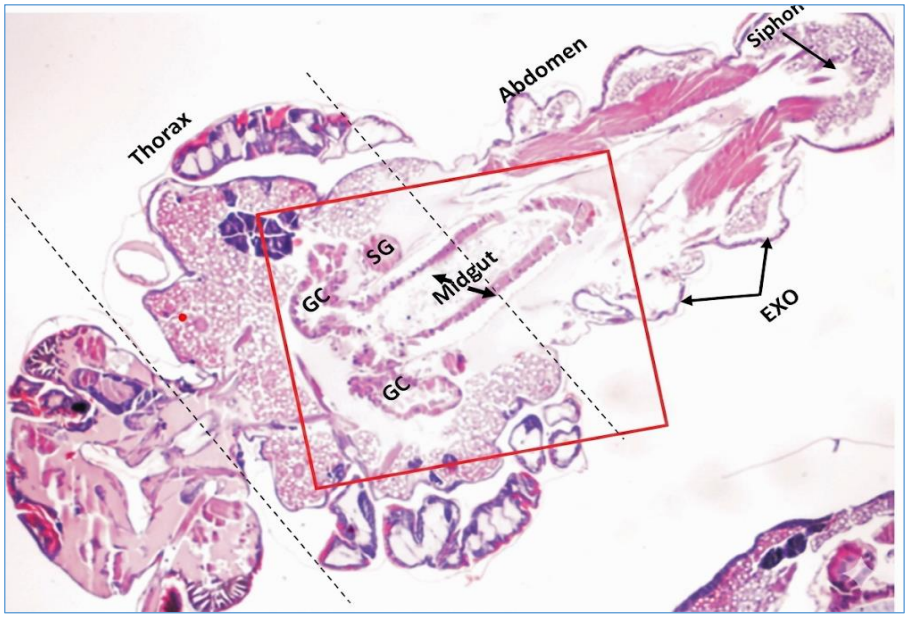



Figure 7. Whole-body longitudinal section of a fourth-instar (*IV instar*) *Culex pipiens* larva (H&E, ×40). The image reveals the extensive midgut with a wide lumen extending throughout the abdominal segments, together with the pharyngeal pump (PP), esophagus (O), fat body (FB), gastric caeca (GC), and salivary glands (SG), demonstrating the overall organization of the larval digestive system and associated tissues (Al-Doaiss et al., 2021:274).

In the anterior region of the digestive system, the gastric caeca (GC), as extensions of the mesenteron, exhibit a characteristic organization. Histological sections reveal that these structures possess a lobular morphology that increases the digestive surface area. The epithelial layer consists of a single layer of prismatic cells with prominent and typically centrally located

nuclei (N). The apical surface of these cells is characterized by a dense microvillar layer (MICRO), also referred to as the brush border, which maximizes the contact surface with luminal contents and enhances absorption efficiency. This structural organization represents a fundamental functional unit for nutrient uptake and enzymatic activity during the final larval instar, which corresponds to the most active feeding and growth phase (Al-Mehmadi & Al-Khalaf, 2010:77) **(Figure 8).** In fourth-instar larvae of *Aedes aegypti*, the midgut epithelium, particularly in the gastric caeca region, is characterized by a well-developed brush border structure that enhances nutrient absorption.

In this region, epithelial cells predominantly exhibit a columnar morphology (Lemos et al., 2018:1) **(Figure 9).**




Figure 8. Transverse section of the gastric caecum (GC) of a *Culex quinquefasciatus* larva. In this control-group specimen, the microvilli (MICRO) lining the luminal surface of the epithelial cells and the prominent nuclei (N) are clearly visible, reflecting normal histological architecture and cellular integrity (H&E, ×400) (Al-Mehmadi & Al-Khalaf, 2010:77)




Figure 9. Histological organization of the brush border (BB) and epithelial cells (EC) within the gastric caecal region of *Aedes aegypti* (H&E, ×1000). The image highlights the distinct brush-border architecture, epithelial cell arrangement, nuclei (N), and surrounding muscle tissue (MU), illustrating the normal microanatomical structure of the gastric caeca (Lemos et al., 2018:1).

The epithelial cells of the main midgut region exhibit either cuboidal or columnar morphology. Within this region, small regenerative cells (RC), which play a role in maintaining epithelial integrity and supporting cellular renewal, are distributed among the digestive cells. Two distinct luminal compartments are recognized between the epithelial layer and the peritrophic matrix (PM): the ectoperitrophic space (ECS) and the endoperitrophic space (ENS) (Lemos et al., 2018:1) **(Figure 10).**

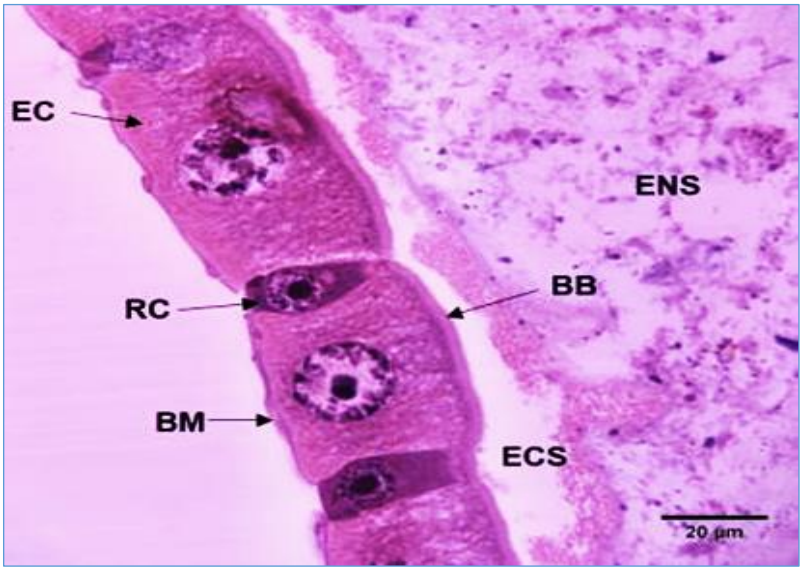



**Figure 10.** Cuboidal epithelial cells (EC) and interspersed regenerative cells (RC) in the mesenteron region of *Aedes aegypti*. BM: basal membrane; ECS/ENS: peritrophic matrix regions (H&E, ×1000) (Lemos et al., 2018:1).

In the anterior region of the digestive system, the gastric caeca are located within the first abdominal segment, whereas the midgut (mesenteron) extends through the second and third abdominal segments in the central region and continues into the fourth and fifth abdominal segments in the posterior region (Lemos, 2018) **(Figure 11).**

The general body organization of *Culex quinquefasciatus* larvae is characterized by a histologically integrated and structurally intact arrangement of the head, thoracic, and abdominal regions. In healthy control specimens, the midgut epithelium is characteristically composed of a single layer of digestive cells. These cells possess a highly developed brush border or microvillar layer on their luminal surface, while their cytoplasm contains acidophilic regions indicative of active metabolic processes (Jiraungkoorskul & Jiraungkoorskul, 2015:15) **(Figure 12).** The cytoplasm of the epithelial cells generally exhibits a heterogeneous, granular organization, and the nucleus is typically located in the central region of the cell (Lemos et al., 2018:1) **(Figure 13).**

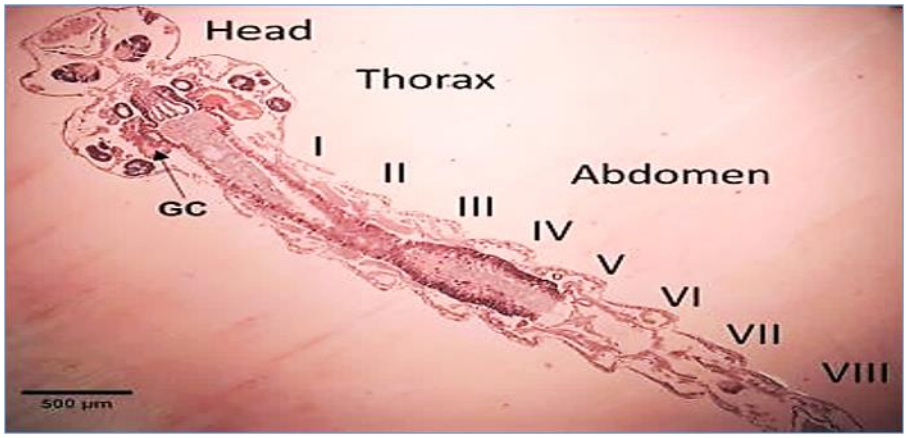



Figure 11. Histological appearance of a whole-body longitudinal section of *Aedes aegypti* larva (H&E, ×40). The head, thorax, and abdominal segments are clearly distinguished, with abdominal segments I–VIII observed in sequence. Histologically, the first segment contains the gastric caeca (GC), the second and third segments correspond to the mid-region of the mesenteron, the fourth and fifth segments represent the posterior midgut, and the sixth to eighth segments constitute the hindgut region (Lemos et al., 2018:1).




Figure 12. Histological overview of a whole-body longitudinal section of *Culex quinquefasciatus* larva. The head, thorax, and abdomen are clearly observed, and the overall anatomical organization appears intact and consistent with normal morphology in the control group. (H&E, ×10). (Jiraungkoorskul & Jiraungkoorskul, 2015:15).

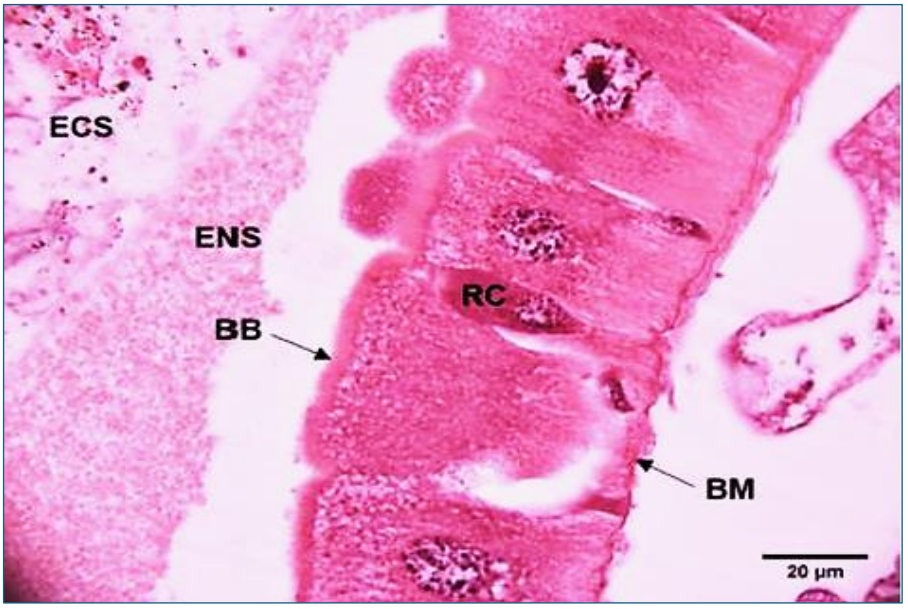



Figure 13. Heterogeneous cytoplasmic organization and centrally located nuclei in cuboidal and columnar epithelial cells of *Aedes aegypti*. The apical brush border (BB), basal membrane (BM), interspersed regenerative cells (RC), and the ectoperitrophic (ECS) and endoperitrophic (ENS) spaces are clearly distinguished within the epithelial architecture. (H&E, ×1000). (Lemos et al., 2018:1).

### 3.2.1. Ultrastructural Features of the Midgut

Transmission electron microscopy (TEM) analyses of midgut epithelial cells reveal a highly complex and distinctly polarized intracellular organization. At the apical surface of the cell, a dense microvillar layer is observed, particularly prominent in cells exhibiting a characteristic "mace-shaped" morphology, which are primarily associated with absorptive function. This structure is interpreted as an adaptive specialization for increasing the absorptive surface area (Silva et al., 2008:663) **(Figures 14, 15).**

Within the microvilli, actin filament bundles provide structural stability and maintain the integrity of the apical membrane domain. In the basal region of the cells, prominent plasma membrane invaginations and densely packed mitochondrial clusters are present, supporting ion transport and active transport processes (Silva et al., 2008:663) **(Figure 16).**

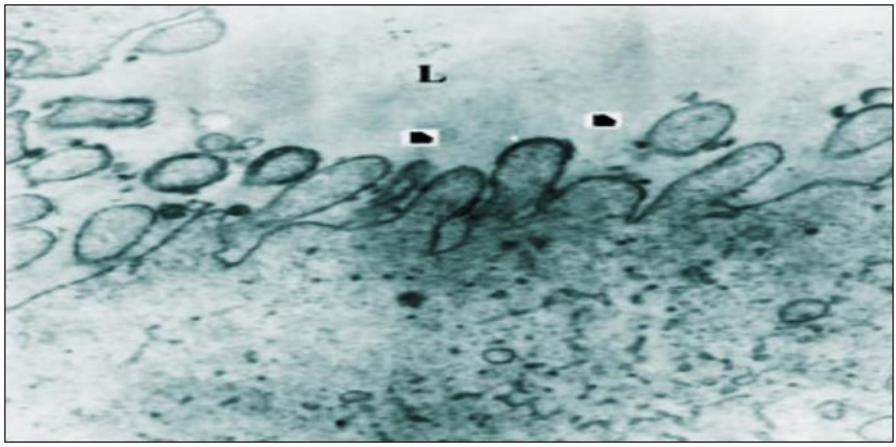



Figure 14. Electron micrograph of the apical region of club-shaped cells and the microvillus (MV) boundary in the midgut of *Aedes albopictus* (TEM, ×20,000). L: lümen. (Silva et al., 2008:663).

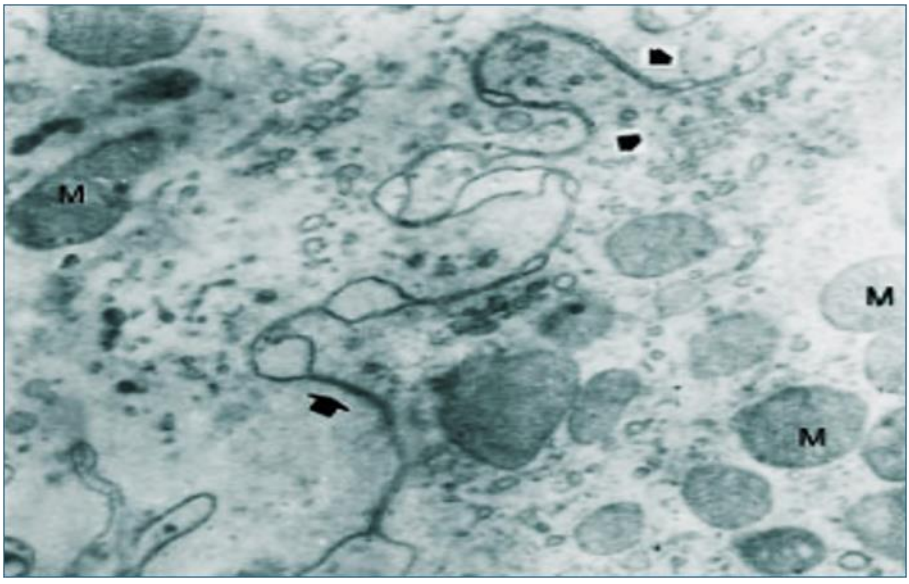



Figure 15. Club-shaped epithelial cell morphology with expanded apical regions facing the lumen in *Aedes albopictus* larvae. The section shows regional differentiation along the midgut lumen (L), including normal epithelial organization in the anterior region (a), characteristic cell groups exhibiting apically enlarged, club-shaped morphology toward the posterior region (b), and transitional epithelial cells between these regions (c). (H&E, ×40). (Silva et al., 2008:663).

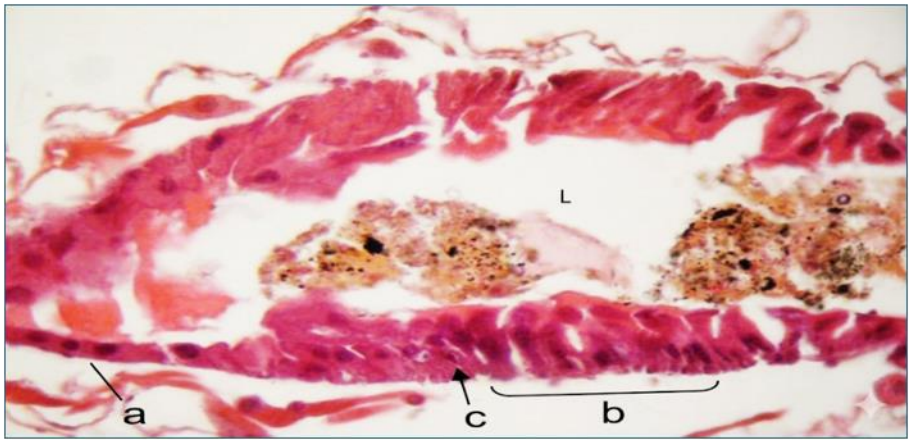



Figure 16. Electron micrograph of the basal surface and cytoplasmic organization of epithelial cells in *Aedes albopictus*. The image shows numerous mitochondria (M) with distinct cristae, along with junctional sites of adjacent lateral plasma membranes and membrane infoldings (arrowheads), indicating normal cellular organization. (TEM, ×30,000). (Silva et al., 2008:663).

In the cytoplasm, rough endoplasmic reticulum (rER), responsible for protein and enzyme synthesis, is arranged in parallel lamellae extending from the perinuclear region to the mid-basal cytoplasm. It becomes particularly concentrated in the apical region, where it plays an active role in the production of digestive enzymes. The Golgi apparatus is closely associated with the rough endoplasmic reticulum; its structure consists of flattened, parallel-stacked cisternae that facilitate vesicle formation (Yıkılmaz & Deveci, 2004:9) **(Figure 17).**

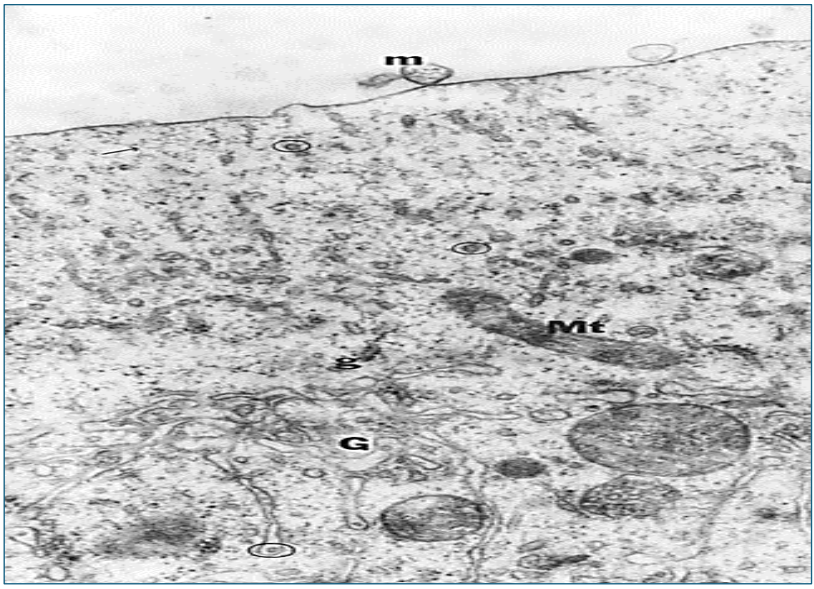



Figure 17. Ultrastructural features of midgut cells in *Culex pipiens*. The apical microvilli (m), Golgi complex (G), rough endoplasmic reticulum (g), and mitochondria with distinct cristae (Mt) are observed. Additionally, cytoskeletal filaments in the apical cytoplasm (arrows), together with small vesicles and vesicular structures clustered around the Golgi apparatus (circles), are clearly visible. (TEM, ×42,000). (Yıkılmaz & Deveci, 2004:9).

Lysosomes and endosomal structures are present in the cytoplasm as small, electron-dense vesicles and are involved in both the degradation of digestive products and the regulation of intracellular waste. Septate junctions between adjacent cells play a critical role in maintaining tissue integrity (Al-Doaiss et al., 2021:274) **(Figures 18, 19).**

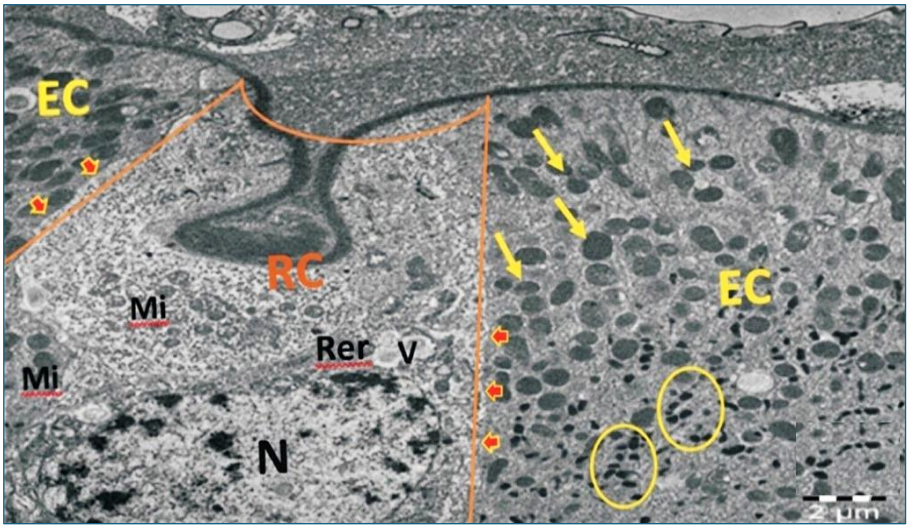



Figure 18. Ultrastructural organization and regenerative capacity of the midgut epithelium in *Culex pipiens* larvae. The image shows the organization of main epithelial cells (EC), which maintain tissue integrity, and a basally located regenerative cell (RC) responsible for epithelial renewal. In control specimens, the regenerative cell exhibits a centrally located oval nucleus (N), mitochondria (Mi), rough endoplasmic reticulum (Rer), and vesicular structures (V). Abundant mitochondrial clusters (yellow arrows), lysosome-like dense granules (yellow circles), and intercellular junctional complexes (red arrowheads) in epithelial cells confirm the structural stability of the tissue. (TEM, ×1,500). (Al-Doaiss et al., 2021:274).

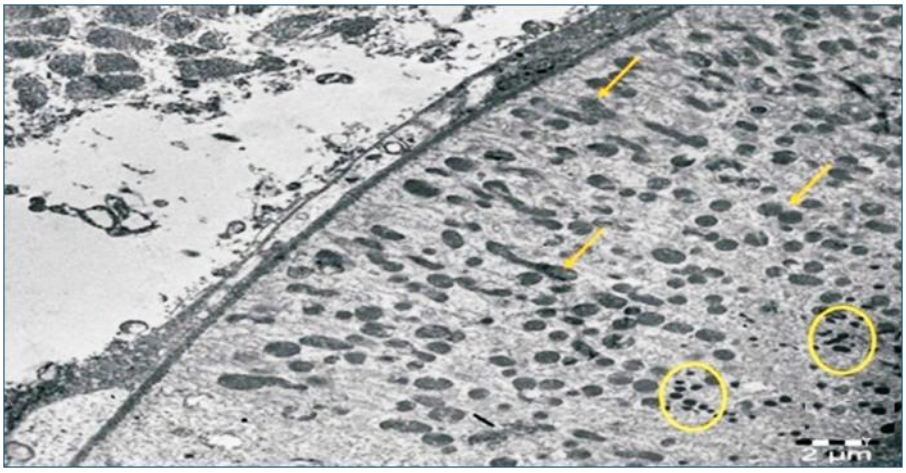


Figure 19. Transmission electron micrograph of a longitudinal section of the midgut of a fourth-instar (*IV instar*) *Culex pipiens* larva. The basal region of the simple cuboidal/columnar epithelial cells, enriched with mitochondrial clusters (yellow arrows) and lysosome-like dense granules (yellow circles), is observed together with the basal lamina surrounding the organ, indicating normal ultrastructural organization. (TEM, ×1,500). (Al-Doaiss et al., 2021:274).

### 3.2.2. Peritrophic Matrix and Luminal Organization

The structural integrity and continuity of the peritrophic matrix (PM) are critical for the protection of the midgut epithelium and the regulation of digestive processes. Under light microscopy, the PM appears as a homogeneous and continuous layer, whereas transmission electron microscopy (TEM) reveals it as a semipermeable network composed of fine fibrils. The ultrastructural organization of the PM functions as an important barrier regulating the distribution of digestive enzymes and food particles within the intestinal lumen. Rather than being a passive barrier, the PM is a dynamic structure whose thickness varies depending on feeding status and environmental conditions (Soffar, 2012:145; Lemos et al., 2018:1).

The midgut lumen contains a heterogeneous mixture of organic particles, microorganisms, and detritus, reflecting the filter-feeding behavior of larvae (Al-Doaiss et al., 2021:274) (Figure 20). The transition zone between the foregut and midgut represents a critical anatomical boundary where ingested material first comes into contact with the PM and where the digestive process is initiated. In terms of digestive system organization, *Aedes aegypti* and *Aedes albopictus* exhibit a high degree of similarity. In both species, a single-layered prismatic epithelium, dense microvilli, and a well-defined basal lamina are characteristic features. For this reason, *Aedes aegypti* is widely

used as a model organism in midgut histological studies and is accepted as a reference for *Aedes albopictus* (Al-Mehmadi, 2010:125; Soffar, 2012:145; Lemos et al., 2018:1; Rudin & Hecker, 1979:497; Clements, 1992; Billingsley & Lehane, 1996:3; Klowden, 2013).

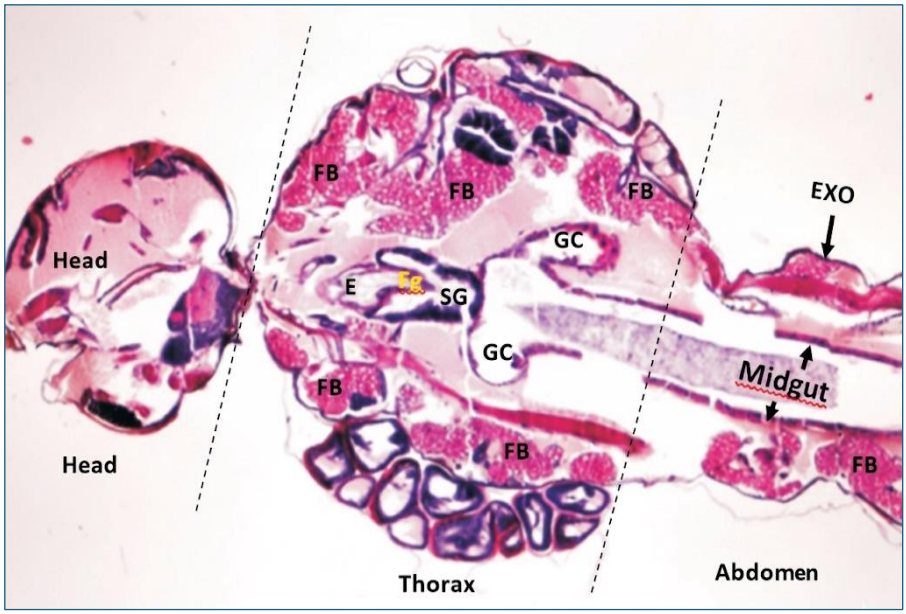



Figure 20. Whole-body longitudinal section of a fourth-instar (*IV instar*) *Culex pipiens* larva (H&E, ×40). The general anatomical organization of the head (Head), thorax (Thorax), and abdomen (Abdomen), demarcated by dashed lines, is shown together with the junction between the foregut (Fg) and midgut (Midgut). Histologically, the esophagus (E), gastric caeca (GC), salivary glands (SG), fat body occupying the internal spaces (FB), and the external exoskeleton (EXO) are observed in a well-preserved and organized structural arrangement (Al-Doaiss et al., 2021:274).

### 3.3. Hindgut

The hindgut begins at the point where the Malpighian tubules open into the alimentary canal and extends to the anus, forming a tubular structure (Jiraungkoorskul & Jiraungkoorskul, 2015:15). Responsible for the excretion of digestive waste and the osmoregulatory processes of the larva, this region consists of large cuboidal epithelial cells of ectodermal origin (Al-Doaiss et al., 2021:274). The luminal surface is lined by a thin, permeable chitinous cuticle layer specialized to allow water and ion exchange (Rudin & Hecker,

1979:497; Soffar, 2012:145). Strong circular muscle layers extending along the hindgut wall function as sphincter-like structures that enable the controlled elimination of waste material (defecation) (Jiraungkoorskul & Jiraungkoorskul, 2015:15).

### 3.3.1. Ultrastructural Features of the Hindgut

Particularly in the rectal region, the hindgut epithelium exhibits a complex ultrastructural architecture specialized for water and ion reabsorption. TEM sections reveal that the basal and lateral plasma membranes form deep invaginations to maximize surface area. Within these membrane infoldings, numerous large mitochondria clusters are present, providing the energy required for active transport processes (Rudin & Hecker, 1979:497).

## 4. CONCLUSION

This review synthesizes the digestive system organization of major vector mosquito larvae, including *Aedes aegypti*, *Aedes albopictus*, *Culex pipiens*, and *Culex quinquefasciatus*, with particular emphasis on histological and ultrastructural characteristics during the fourth larval instar. Although the general body plan is highly conserved among these species, accumulated evidence demonstrates that each region of the digestive tract exhibits a remarkable degree of structural and functional specialization closely associated with feeding physiology, osmoregulatory processes, and vectorial competence.

The foregut displays a conserved anatomical organization across all examined species. The chitinous intima provides mechanical protection, whereas the pharynx–esophagus complex facilitates the efficient transport of ingested food particles. At the ultrastructural level, foregut epithelial cells exhibit lower metabolic activity and a less complex organelle organization compared with those of the midgut, supporting the view that the primary role of this region is the conduction rather than the absorption of nutrients.

The midgut represents the most highly specialized and metabolically active compartment of the larval digestive system. Across all species, common ultrastructural features include a single-layered cuboidal to columnar epithelium, a dense microvillar border, abundant mitochondria, and a well-

developed rough endoplasmic reticulum network. Gastric caeca markedly increase the absorptive surface area, thereby enhancing digestive efficiency. In particular, the peritrophic matrix (PM) emerges as a dynamic and multifunctional structure involved in food compartmentalization, epithelial protection, and the regulation of interactions with microbial communities and environmental factors. Collectively, these characteristics establish the midgut as the principal physiological and biochemical interface between the larva and its external environment.

Comparative analyses reveal notable interspecific differences, particularly in *Culex pipiens* and *Culex quinquefasciatus*, where the midgut lumen is generally broader and epithelial organization extends more extensively throughout the abdominal segments. In contrast, *Aedes aegypti* and *Aedes albopictus* exhibit a high degree of structural similarity, further supporting the use of *A. aegypti* as a representative model organism for investigations of mosquito midgut histology. The hindgut is primarily specialized for osmoregulation and ion reabsorption. Prominent basal membrane infoldings and mitochondria-rich epithelial cells reflect the substantial energetic demands associated with active transport processes, highlighting the critical role of this region in maintaining ionic and osmotic homeostasis.

In conclusion, the digestive system of mosquito larvae should be regarded not merely as an organ system involved in nutrition and metabolism, but also as a multifunctional biological interface mediating interactions with environmental toxins, xenobiotics, and pathogens. The histological and ultrastructural framework presented in this review provides a robust foundation for future studies in vector histopathology, bioinsecticide development, and integrated vector management strategies. In particular, the midgut epithelium and peritrophic matrix emerge as key targets for interventions aimed at disrupting larval development and reducing vectorial capacity.

## REFERENCES

Al-Doaiss, A. A., Al-Mekhlafi, F. A., Abutaha, N. M., Al-Keridis, L. A., Shati, A. A., Al-Kahtani, M. A., & Alfaifi, M. Y. (2021). Morphological, histological and ultrastructural characterisation of *Culex pipiens* (Diptera: Culicidae) larval midgut. *African Entomology*, *29*(1), 274–288. https://doi.org/10.4001/003.029.0274

Al-Mehmadi, R. M. (2010). Histological studies on the foregut of *Culex quinquefasciatus* (Say) larvae. *Journal of the Egyptian Society of Parasitology*, *40*(1), 125–134.

Al-Mehmadi, R. M., & Al-Khalaf, A. A. (2010). Larvicidal and histological effects of *Melia azedarach* extract on *Culex quinquefasciatus* Say larvae (Diptera: Culicidae). *Journal of King Saud University - Science*, *22*(2), 77–85. https://doi.org/10.1016/j.jksus.2010.02.004

Arafa, A. A., & Salama, S. M. (2021). Histopathological and ultrastructural changes in the midgut of *Aedes aegypti* larvae treated with botanical extracts. *Egyptian Academic Journal of Biological Sciences. A, Entomology*, *14*(1), 105–118.

Billingsley, P. F., & Lehane, M. J. (1996). Structure and ultrastructure of the insect midgut. In M. J. Lehane & P. F. Billingsley (Eds.), *Biology of the insect midgut* (pp. 3–30). Springer. https://doi.org/10.1007/978-94-009-1519-0_1

Billingsley, P. F., & Rudin, W. (1992). The role of the mosquito peritrophic membrane in bloodmeal digestion and infectivity. *Review of Medical and Veterinary Entomology*, *80*, 87–109.

Busvine, J. R. (1980). *Insects and hygiene: The biology and control of insect pests of medical and domestic importance*. Chapman & Hall.

Cabeça, T. K., Valotto, C. F., & Azeredo-Oliveira, M. T. V. (2025). Comparative midgut morphology of *Aedes aegypti* and *Aedes albopictus* fourth-instar larvae. *Brazilian Journal of Biology*, *85*, e290849. https://doi.org/10.1590/1519-6984.290849

Christophers, S. R. (1960). *Aedes aegypti (L.) the yellow fever mosquito: Its life history, bionomics and structure*. Cambridge University Press.

Clements, A. N. (1992). *The biology of mosquitoes: Volume 1: Development, nutrition and reproduction*. Chapman & Hall.

Demirsoy, A. (1997). *Entomoloji: Böceklerin genel yapısı ve biyolojisi*. Meteksan Yayınları.

Eisemann, C. H., & Binnington, K. C. (1994). The peritrophic membrane: Its formation, structure, chemical composition and function in relation to interactions between pathogens and hosts. *International Journal for Parasitology*, *24*(1), 15–26. https://doi.org/10.1016/0020-7519(94)90055-8

Gregor, K. M., Becker, S. C., Hellhammer, F., Schön, K., Baumgärtner, W., & Puff, C. (2022). Histochemical staining techniques in *Culex pipiens* and *Drosophila melanogaster* (Diptera) with a comparison to mammals. *Veterinary Pathology*, *59*(5), 836–849. https://doi.org/10.1177/03009858221088786

Harbach, R. E. (2011). Classification within the cosmopolitan genus *Culex* (Diptera: Culicidae): The foundation for molecular systematics and phylogenetic

research. *Acta Tropica*, *120*(1-2), 1–14. https://doi.org/10.1016/j.actatropica.2011.06.005

Herms, W. B. (1950). *Medical entomology: With special reference to the health and well-being of man and animals* (4. baskı). Macmillan Company.

Jiraungkoorskul, K., & Jiraungkoorskul, W. (2015). Larvicidal and histopathological effects of *Cassia siamea* leaf extract against *Culex quinquefasciatus*. *Tropical Life Sciences Research*, *26*(2), 15–25. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4729405/

Klowden, M. J. (2013). *Physiological systems in insects* (3. baskı). Academic Press.

Lehane, M. J. (1997). Peritrophic matrix structure and function. *Annual Review of Entomology*, *42*(1), 525–550. https://doi.org/10.1146/annurev.ento.42.1.525

Lemos, A. B., Adam, F. C., Moura, K. R. S., Moraes, L. B., & Silva, O. S. (2018). Histological and histochemical characterization of the midgut of healthy *Aedes aegypti* larvae. *Annual Research & Review in Biology*, *22*(1), 1–15. Makale ARRB.37443. https://doi.org/10.9734/ARRB/2018/37443

Richards, A. G., & Richards, P. A. (1971). Origin and composition of the peritrophic membrane of the mosquito, *Aedes aegypti*. *Journal of Insect Physiology*, *17*(11), 2253–2275. https://doi.org/10.1016/0022-1910(71)90183-1

Rudin, W., & Hecker, H. (1979). Functional morphology of the midgut of *Aedes aegypti* L. (Insecta, Diptera) during blood digestion. *Cell and Tissue Research*, *200*(3), 497–503. https://doi.org/10.1007/BF00236412

Silva, V. C., Pinheiro, N. L., Scherer, P. O., Falcão, S. S., Ribeiro, V. R., Mendes, R. M. M., Chaves, R., Cardozo-de-Almeida, M., & Santos-Mallet, J. R. (2008). Histology and ultrastructure of *Aedes albopictus* larval midgut infected with *Bacillus thuringiensis* var. *israelensis*. *Microscopy Research and Technique*, *71*(9), 663–668. https://doi.org/10.1002/jemt.20605

Snodgrass, R. E. (1959). *The anatomical life of the mosquito*. Smithsonian Miscellaneous Collections, *139*(8), 1–87.

Soffar, S. (2012). Histological study on the larvae of *Culex pipiens* (Linnaeus, 1758) (Diptera: Culicidae). *Parasitologists United Journal*, *5*(2), 145–159.

Stam, P. A., Landolt, P. J., & Germida, J. J. (1978). The peritrophic membrane of the *Culex pipiens* larva: Formation and structure. *Journal of Invertebrate Pathology*, *32*(2), 175–184. https://doi.org/10.1016/0022-2011(78)90196-1

Tellam, R. L., Wijffels, G., & Willadsen, P. (1999). Peritrophic matrix proteins. *Insect Biochemistry and Molecular Biology*, *29*(2), 87–101. https://doi.org/10.1016/S0965-1748(98)00123-4

Wigglesworth, V. B. (1942). The storage of protein, fat, glycogen and uric acid in the fat body and other tissues of mosquito larvae. *Journal of Experimental Biology*, *19*(1), 56–77. https://doi.org/10.1242/jeb.19.1.56

Yıkılmaz, M. S., & Deveci, Ö. (2004). Structure of midgut and peritrophic matrix in the last instar larvae of *Culex pipiens* (Diptera: Culicidae). *Gazi Üniversitesi Fen Bilimleri Dergisi*, *17*(2), 9–19.

Yu, K. X., Wong, C. L., Ahmad, R., & Jantan, I. (2015). Larvicidal activity, inhibition effect on development, histopathological alteration and morphological aberration induced by seaweed extracts in *Aedes aegypti* (Diptera: Culicidae). *Asian Pacific Journal of Tropical Medicine*, *8*(12), 1006–1012. https://doi.org/10.1016/j.apjtm.2015.11.011