PREFACE TO THE EDITION

It is with great pleasure that we present the forthcoming issue of the **International Journal of Pure Science Research Studies (IJPSRS)**, bringing together a diverse collection of scholarly contributions that reflect the depth, breadth, and evolving character of contemporary research in the pure sciences. The articles featured in this issue demonstrate how fundamental scientific theories, mathematical frameworks, and biological mechanisms continue to provide essential foundations for understanding increasingly complex natural and computational phenomena.

The issue opens with a review of Compressed Sensing, examining the mathematical principles underlying sparse signal recovery, the conditions required for successful reconstruction, and the algorithms that make such recovery computationally feasible. The discussion of stability under noisy measurements, together with applications in magnetic resonance imaging and radio astronomy, illustrates the productive relationship between theoretical mathematics and practical scientific problems.

The biological sciences are represented by a detailed examination of genetic regulation of photoperiod sensitivity in flowering plants. Focusing particularly on Arabidopsis thaliana, the article explores the interaction between light perception, circadian rhythms, and molecular regulators such as CONSTANS and FLOWERING LOCUS T. The study also highlights the broader significance of understanding flowering mechanisms for crop adaptation in the context of changing environmental and climatic conditions.

The issue further turns to fundamental questions in modern physics through a review of neutrino oscillations. By examining flavour mixing, mass-squared differences, mixing angles, and the continuing search for charge-parity violation and the absolute neutrino mass scale, the article provides an overview of both the established foundations and unresolved questions in neutrino physics. The discussion underscores how experimental observations continue to challenge and refine our understanding of particle physics.

A significant mathematical dimension of this issue is represented by the article on persistent homology and topological data analysis. By connecting mathematical topology with computational data analysis, the contribution demonstrates how topological features can be extracted across different scales and applied in areas ranging from neuroscience and materials science to genomics and machine learning. Its attention to persistence diagrams, stability, and emerging computational approaches reflects the growing importance of mathematical structures in contemporary data-driven research.

The issue also features a comprehensive review of optimal transport, tracing its development from the classical formulation of Monge and the mathematical contributions of Kantorovich to modern computational approaches. The discussion of Wasserstein distances, entropic regularisation, and numerical methods demonstrates the continuing evolution of this mathematical framework and its expanding applications in image retrieval, generative modelling, domain adaptation, and biological data analysis.

Completing this issue is a review of mycorrhizal symbioses, exploring the cellular mechanisms, ecological functions, and responses of plant–fungus associations to global environmental change. The article connects molecular interactions and nutrient exchange with wider questions of soil carbon, nitrogen cycling, plant productivity, and ecosystem responses to warming, nitrogen deposition, and land-use change. It highlights the importance of microscopic biological partnerships in shaping larger ecological processes.

Taken together, the contributions in this issue illustrate the remarkable diversity of contemporary pure science while revealing important connections across disciplines. From sparse mathematical representations and topological structures to quantum phenomena, genetic regulation, mathematical optimisation, and ecological symbiosis, these studies demonstrate how fundamental inquiry continues to expand the boundaries of scientific understanding.

We hope that this issue of IJPSRS will provide researchers, academicians, students, and scientific professionals with valuable perspectives on emerging developments in the pure sciences and encourage further investigation, interdisciplinary dialogue, and scholarly collaboration.

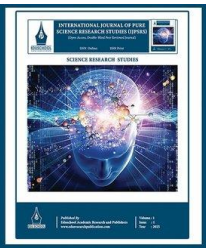
We sincerely thank all the authors, reviewers, and editorial contributors whose efforts have made this issue possible. We look forward to continuing our commitment to disseminating rigorous and meaningful research through the International Journal of Pure Science Research Studies.

Dr. Sandhya E

Chief editor

CONTENTS

INTERNATIONAL JOURNAL OF PURE SCIENCE RESEARCH STUDIES (IJPSRS)
(Open Access, Double-Blind Peer Reviewed Journal)
ISSN Online:



Compressed Sensing: Sparse Recovery Theory, Algorithms and Practical Applications

Sandhya E

Associate Professor, Department of Basic Sciences & Humanities, Jyothi Engineering College, Thrissur, India

Article Information

Received: 23rd April 2026
Received in revised form: 25th May 2026
Accepted: 27th June 2026
Available online: 14th August 2026

Volume: 2
Issue: 3
DOI: https://doi.org/10.5281/zenodo.21915660

Abstract

Classical sampling theory fixes the rate at which a signal must be measured by its bandwidth alone. Compressed sensing replaces that criterion with one based on structure: a signal that is sparse in some known basis can be reconstructed exactly from a number of linear measurements proportional to its sparsity and only logarithmic in its ambient dimension. This paper reviews the conditions under which such recovery is possible, the algorithms that achieve it, and the settings in which the theory has been applied. Uniqueness requires that the measurement matrix act almost isometrically on sparse vectors, a property that random matrices possess with high probability but that is difficult to certify for any given matrix. Under that property the combinatorial search for the sparsest solution can be replaced by minimisation of the sum of absolute values, a convex program, and the reconstruction is stable when measurements are noisy and the signal is only approximately sparse. Two numerical experiments carried out for this review illustrate the sharp boundary between success and failure as sparsity and the number of measurements are varied. Applications in magnetic resonance imaging and radio astronomy are examined together with the limits of the approach.

Keywords: Compressed Sensing, Sparse Recovery, Restricted Isometry, Basis Pursuit, Orthogonal Matching pursuit.

.

1. INTRODUCTION

The sampling theorem associated with Nyquist and Shannon states that a signal with no frequency content above a given limit is determined by samples taken at twice that limit. The result is exact and, read as a requirement rather than a sufficient condition, misleading. It assumes nothing about the signal except its bandwidth, whereas signals of practical interest are far from arbitrary: photographs compress well in a wavelet basis, and audio compresses well in a local cosine basis, precisely because most of their transform coefficients are negligible.

Conventional practice acquires everything and then discards most of it during compression. Compressed sensing asks whether the discarded measurements need to have been taken at all. Two papers published in 2006 showed that they do not.[1,2] If a signal has at most s non-zero coefficients in a known basis, then it can be recovered exactly from a number of suitably chosen linear measurements proportional to s times the logarithm of the ratio of dimension to sparsity, which is far fewer than the dimension when the signal is genuinely sparse.

The result rests on two ingredients that are worth separating. The first is a geometric condition on the measurement operator ensuring that distinct sparse signals produce distinguishable measurements. The second is an algorithmic fact: the natural but intractable problem of finding the sparsest consistent solution can, under that condition, be replaced by a convex program that is solvable in polynomial time. The sections that follow treat each in turn, then examine algorithms, numerical behaviour and applications.

2. CONDITIONS FOR RECOVERY

Write the measurement process as a matrix with m rows and n columns acting on an unknown vector, with m much smaller than n. The system is underdetermined and has infinitely many solutions, so recovery is possible only if the solution set contains exactly one sufficiently sparse vector.

The elementary condition is stated in terms of the spark of the matrix, the smallest number of linearly dependent columns. Any vector with fewer than half that many non-zero entries is the unique sparsest solution. Computing the spark requires examining every subset of columns, so this criterion is of theoretical value only. A weaker but computable substitute is the mutual coherence, the largest inner product between distinct normalised columns; guarantees expressed through coherence are easy to check and pessimistic, since coherence cannot fall below a bound determined by the matrix dimensions.

The condition that produced the strongest results is the restricted isometry property introduced by Candès and Tao.[3] A matrix satisfies it at order s with constant delta if it changes the length of every s-sparse vector by a factor lying between one minus delta and one plus delta. In other words, the matrix acts almost like an isometry on every s-dimensional coordinate subspace at once. When the constant at order 2s is small enough, distinct s-sparse vectors cannot produce the same measurements, and the convex relaxation described below recovers the correct one.

Verifying the property for a specific matrix is computationally hard, which is the awkward point in the theory. What can be shown is that certain random constructions satisfy it with high probability. Matrices with independent Gaussian or subgaussian entries do so once the number of rows exceeds a constant multiple of s times the logarithm of the ratio of n to s , and matrices formed from randomly selected rows of a discrete Fourier transform do so with an additional logarithmic factor .[4,5] The Fourier case matters in practice, because many measurement systems can select frequencies but cannot apply an arbitrary dense matrix to a physical signal.

3. CONVEX RELAXATION AND ITS GUARANTEES

Seeking the vector with fewest non-zero entries subject to the measurement constraints is a combinatorial problem. Replacing the count of non-zero entries by the sum of absolute values yields a convex program, known as basis pursuit, that can be recast as a linear program .[6] The substitution is not a heuristic in this setting: under the restricted isometry condition the two problems have the same solution.

Geometry explains why the sum of absolute values promotes sparsity. Its unit ball is a cross-polytope whose extreme points lie on the coordinate axes, so expanding this ball until it first meets the affine solution set typically produces contact at a vertex, which is a sparse vector. The Euclidean ball has no such structure and yields a dense solution.

The guarantees extend beyond exactly sparse signals and noiseless measurements, which is essential for applications. If measurements are corrupted by noise of bounded energy and the signal is compressible rather than sparse, the solution of the corresponding constrained problem lies within a distance of the true signal controlled by the sum of the noise level and the error of the best s-term approximation .[7] The bound degrades gracefully, so nothing catastrophic happens when the assumptions hold only approximately, which is the usual situation. Comprehensive treatments of these results are available in book form .[8]

4. ALGORITHMS

Three families of solvers are in common use, and they trade accuracy against cost differently. Table 1 summarises them.

Convex programming approaches solve the relaxed problem directly. Interior point methods are accurate and scale poorly; first-order methods based on iterative shrinkage are the practical choice for large problems, and the accelerated variant converges at a rate proportional to the inverse square of the iteration count .[9]

Greedy methods build the support one element at a time. Orthogonal matching pursuit selects at each step the column most correlated with the current residual, projects the measurements onto the span of the selected columns and repeats .[10] It is simple and fast, and its guarantees are weaker than those of convex relaxation, though for many random ensembles the practical performance is comparable. Compressive sampling matching pursuit adds and prunes several elements per iteration and recovers guarantees of the same order as basis pursuit .[11]

Thresholding methods iterate a gradient step followed by retention of the largest entries. Iterative hard thresholding is the simplest of these and converges under a restricted isometry condition .[12] Approximate message passing adds a correction term derived from a statistical physics analysis, which removes the correlation between iterates and yields performance that matches the theoretical phase transition of convex relaxation at a fraction of the cost.[13]

Fig. 1 reports a numerical experiment carried out for this review, in which random Gaussian matrices were used to measure random sparse vectors of dimension 200 and orthogonal matching pursuit was applied to recover them, with twenty independent trials at each combination of sparsity and number of measurements. The transition from reliable to unreliable recovery is sharp and follows a boundary that is close to linear in the two

parameters, matching the phase transition behaviour predicted for this class of problem.[14] Fig. 2 shows a single instance from the recoverable region: eighty measurements of a 256-dimensional vector with ten non-zero entries, recovered to a relative error at the level of machine precision.

The experiment also illustrates the practical reading of such diagrams. At a sparsity of ten, recovery was reliable once about sixty measurements were available, which is roughly six times the sparsity and less than a third of the ambient dimension. At a sparsity of thirty, reliability required close to one hundred and fifty measurements, so the ratio of measurements to non-zero entries falls as sparsity increases, in line with the logarithmic term in the theory. A design should sit well above the transition rather than on it, since the region where the success probability is intermediate is also the region where reconstruction is most sensitive to noise.

Fig 1: Empirical probability of exact recovery by orthogonal matching pursuit as a function of sparsity and number of measurements, computed for this review with twenty trials per parameter combination.

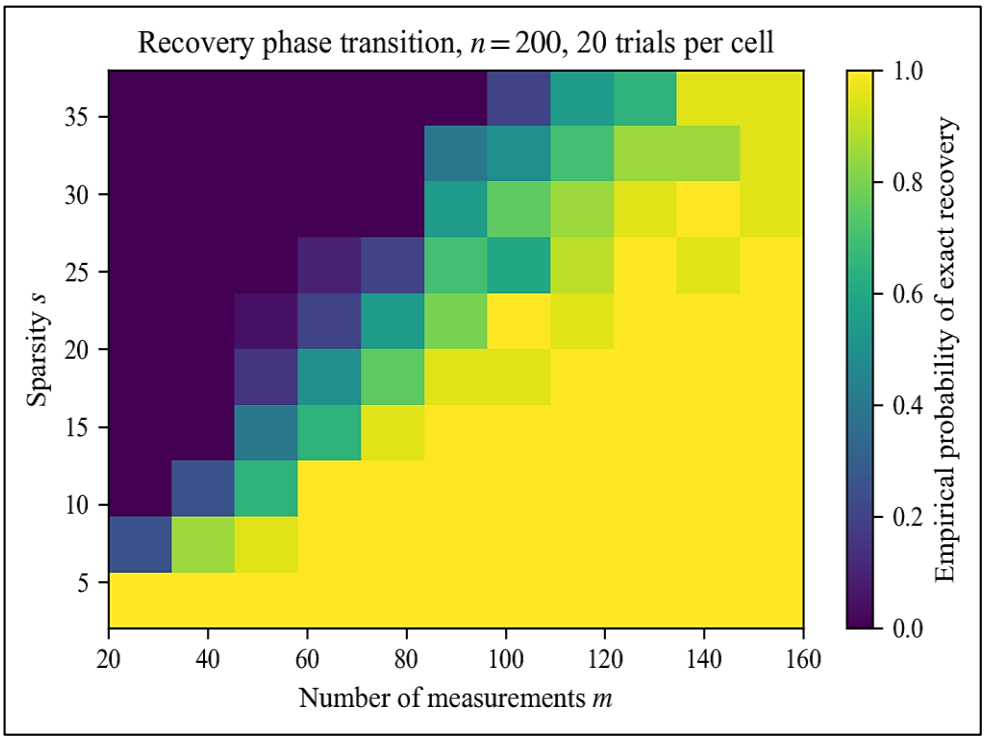
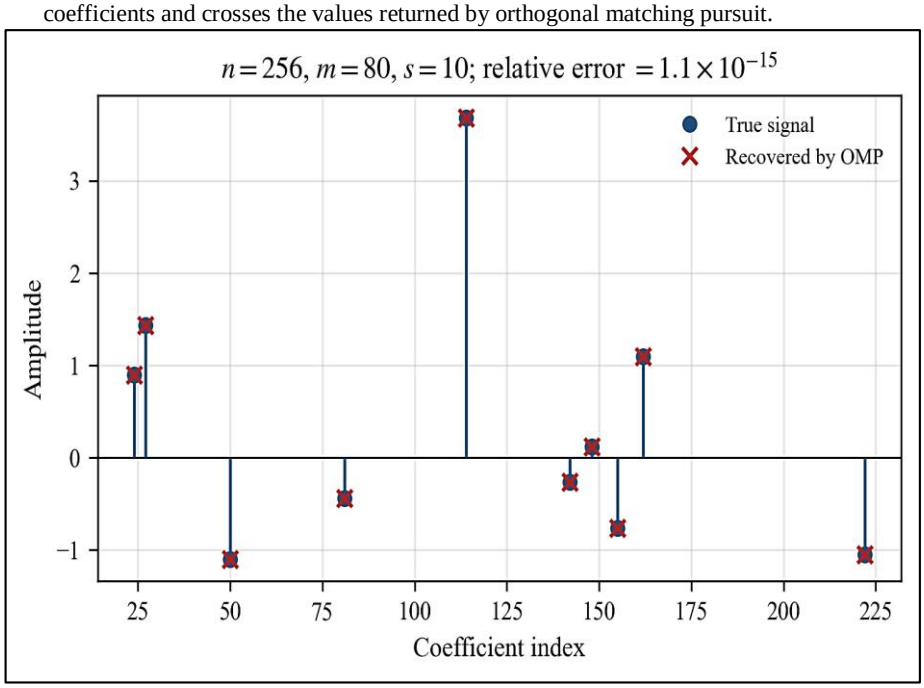


Table 1. Algorithms for sparse recovery

Algorithm	Type	Cost per iteration	Guarantee
Basis pursuit	Convex program	Depends on solver	Strongest, under restricted isometry [1,3]
Iterative shrinkage	First-order convex	One matrix product	Converges to the convex optimum [9]
Orthogonal matching pursuit	Greedy	One product and a least-squares fit	Weaker, competitive in practice [10]
Compressive sampling matching pursuit	Greedy	Several matrix products	Comparable order to basis pursuit [11]
Iterative hard thresholding	Thresholding	One matrix product	Requires small isometry constant [12]
Approximate message passing	Thresholding with correction	One matrix product	Attains the predicted transition [13]

Fig 2: A single recovery instance from the successful region of Fig. 1. Circles mark the true coefficients and crosses the values returned by orthogonal matching pursuit.



5. APPLICATIONS

Magnetic resonance imaging is the application where the idea has changed clinical practice. The scanner samples the Fourier transform of the image along trajectories in frequency space, and acquisition time is proportional to the number of samples taken. Medical images are compressible in wavelet bases, and randomised undersampling of frequency space produces artefacts that are incoherent and therefore removable by sparse reconstruction .[15] Paediatric imaging benefited first, since shorter scans reduce the need for sedation, and reconstruction of this type is now standard on commercial scanners.[16]

Optical imaging provided an early demonstration of a different kind. A single-pixel camera modulates the incoming image with a sequence of random patterns using a micromirror array and records only the total intensity for each pattern; the image is then reconstructed from far fewer measurements than pixels [17]. The device is of limited practical value at visible wavelengths, where detector arrays are cheap, but the principle matters at wavelengths where arrays are expensive.

Radio interferometry poses the same mathematical problem. An array samples the Fourier transform of the sky brightness at positions determined by the baselines between antennas, leaving large gaps. Reconstruction under sparsity assumptions is one of the methods used in the imaging of the black hole shadow reported by the Event Horizon Telescope, where several independent reconstruction pipelines were compared precisely because the inverse problem is underdetermined.[18]

The same circle of ideas extends to low-rank recovery. Replacing the count of non-zero entries by the rank of a matrix and the sum of absolute values by the sum of singular values gives conditions under which a matrix is recoverable from a small random sample of its entries, a result that underlies collaborative filtering and many problems in system identification.[19]

6. STRUCTURED SPARSITY AND THE DESIGN OF MEASUREMENTS

Plain sparsity ignores information that is usually available. The large wavelet coefficients of a natural image are not scattered at random; they cluster along edges and form connected trees across scales. Model-based compressive sensing restricts the search to signals whose support obeys such a structure, replacing the projection onto the set of sparse vectors by projection onto the structured model. The number of measurements needed then depends on the number of admissible supports rather than on all subsets of a given size, which removes the logarithmic factor and reduces the requirement in practice.[20] Similar arguments apply to block sparsity, where non-zero entries occur in groups, and to signals that are sparse in a redundant dictionary rather than an orthonormal basis.

The design of the measurement operator is equally consequential and is often outside the experimenter's control. Physical systems typically permit only structured randomness: partial Fourier sampling in magnetic resonance and radio imaging, random convolution in radar, and subsampled Hadamard patterns in optics. These ensembles satisfy the isometry condition with worse constants than fully random matrices, and the sampling

pattern matters. Uniform random selection of Fourier frequencies performs poorly on images, whose energy is concentrated at low frequencies; variable density schemes that sample the centre of frequency space densely and the periphery sparsely perform much better, and this behaviour has been given a theoretical basis in terms of local coherence.[21] Practical sampling patterns in medical imaging follow this principle rather than uniform randomness.

A related point concerns the sparsifying transform. Guarantees are stated for a basis in which the signal is sparse, and the choice of that basis is a modelling decision that determines how many measurements are needed. Total variation, wavelets and learned dictionaries all serve, and the reconstruction quality obtained in an application usually depends more on this choice than on which of the algorithms in Table 1 is used to solve the resulting problem.

7. LIMITS AND COMMON MISUNDERSTANDINGS

Three points are frequently overstated. First, the theory assumes that measurements can be designed, and in many settings they cannot. If the physics dictates the measurement operator, and if that operator is coherent with the sparsity basis, the guarantees do not apply, however sparse the signal may be.

Second, sparsity is a modelling assumption rather than an observed fact. Natural signals are compressible, meaning their sorted coefficients decay, rather than exactly sparse, and the recovery error then contains a term proportional to the tail that no algorithm can remove. Reported reconstructions at very high undersampling factors often rely on additional structure, such as smoothness across coil channels or across time, and attributing the result to sparsity alone misstates what is doing the work.

Third, noise sets a floor. The stable recovery bounds are proportional to the noise level with a constant that grows as the isometry constant approaches its critical value, so a system operating near the recovery boundary amplifies noise substantially even though it does not fail outright.[7] Practical designs stay well inside the boundary shown in Fig. 1.

Reconstruction methods trained on data have displaced sparse optimisation in several imaging applications, since a learned prior represents the statistics of a class of images better than any fixed basis. What such methods retain from this theory is the sampling strategy: incoherent measurement remains the reason the reconstruction problem is solvable at all.

8. CONCLUSION

Compressed sensing established that the number of measurements needed to determine a signal depends on its information content rather than on its bandwidth. The mathematical statement is precise: under a restricted isometry condition satisfied by suitable random operators, minimising the sum of absolute values recovers any sufficiently sparse vector exactly, and recovers compressible vectors with an error controlled by the noise and by the approximation tail. The algorithmic landscape is mature, with convex, greedy and thresholding methods offering different balances of guarantee and cost, and the empirical phase transition is sharp enough to serve as a design rule. The practical influence has been greatest where measurement is expensive, in medical imaging and radio astronomy. The main open problems are the difficulty of certifying the isometry property for structured operators and the integration of learned priors with the guarantees that make the classical theory trustworthy.

REFERENCES

1. Candès EJ, Romberg J, Tao T. 2006. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans Inf Theory. 52(2):489–509.
2. Donoho DL. 2006. Compressed sensing. IEEE Trans Inf Theory. 52(4):1289–306.
3. Candès EJ, Tao T. 2005. Decoding by linear programming. IEEE Trans Inf Theory. 51(12):4203–15.
4. Baraniuk R, Davenport M, DeVore R, Wakin M. 2008. A simple proof of the restricted isometry property for random matrices. Constr Approx. 28(3):253–63.
5. Rudelson M, Vershynin R. 2008. On sparse reconstruction from Fourier and Gaussian measurements. Commun Pure Appl Math. 61(8):1025–45.
6. Chen SS, Donoho DL, Saunders MA. 1998. Atomic decomposition by basis pursuit. SIAM J Sci Comput. 20(1):33–61.
7. Candès EJ, Romberg JK, Tao T. 2006. Stable signal recovery from incomplete and inaccurate measurements. Commun Pure Appl Math. 59(8):1207–23.
8. Foucart S, Rauhut H. 2013. A mathematical introduction to compressive sensing. New York (NY): Birkhäuser.
9. Beck A, Teboulle M. 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J Imaging Sci. 2(1):183–202.
10. Tropp JA, Gilbert AC. 2007. Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans Inf Theory. 53(12):4655–66.
11. Needell D, Tropp JA. 2009. CoSaMP: iterative signal recovery from incomplete and inaccurate samples. Appl Comput Harmon Anal. 26(3):301–21.
12. Blumensath T, Davies ME. 2009. Iterative hard thresholding for compressed sensing. Appl Comput Harmon Anal. 27(3):265–74.
13. Donoho DL, Maleki A, Montanari A. 2009. Message-passing algorithms for compressed sensing. Proc Natl Acad Sci USA. 106(45):18914–9.

14. Donoho DL, Tanner J. 2010. Precise undersampling theorems. Proc IEEE. 98(6):913–24.
15. Lustig M, Donoho D, Pauly JM. 2007. Sparse MRI: the application of compressed sensing for rapid MR imaging. Magn Reson Med. 58(6):1182–95.
16. Vasanawala SS, Alley MT, Hargreaves BA, Barth RA, Pauly JM, Lustig M. 2010. Improved pediatric MR imaging with compressed sensing. Radiology. 256(2):607–16.
17. Duarte MF, Davenport MA, Takhar D, et al. 2008. Single-pixel imaging via compressive sampling. IEEE Signal Process Mag. 25(2):83–91.
18. Event Horizon Telescope Collaboration. 2019. First M87 Event Horizon Telescope results. IV. Imaging the central supermassive black hole. Astrophys J Lett. 875(1).
19. Candès EJ, Recht B. 2009. Exact matrix completion via convex optimization. Found Comput Math. 9(6):717–72.
20. Baraniuk RG, Cevher V, Duarte MF, Hegde C. 2010. Model-based compressive sensing. IEEE Trans Inf Theory. 56(4):1982–2001.
21. Krahmer F, Ward R. 2014. Stable and robust sampling strategies for compressive imaging. IEEE Trans Image Process. 23(2):612–22.

INTERNATIONAL JOURNAL OF PURE SCIENCE RESEARCH STUDIES (IJPSRS)
(Open Access, Double-Blind Peer Reviewed Journal)
ISSN Online:

Genetic Regulation of Photoperiod Sensitivity in Flowering Plants

Kesavan K

Associate Professor and Head, Department of Aquaculture, MES Asmabi College, Vemballur, Kerala, India

Article Information

Received: 28th April 2026
Received in revised form: 30th May 2026
Accepted: 3rd July 2026
Available online: 14th August 2026

Volume: 2
Issue: 3
DOI: https://doi.org/10.5281/zenodo.21915909

Abstract

Photoperiodism represents a critical adaptive mechanism enabling plants to synchronize flowering with optimal environmental conditions. This review examines the molecular and genetic foundations underlying photoperiod sensitivity in flowering plants, with particular emphasis on the model organism Arabidopsis thaliana. The photoperiodic flowering pathway integrates light signals through photoreceptors with endogenous circadian rhythms to regulate the expression of key floral integrators. Central to this process is the CONSTANS (CO) protein, which acts as a critical molecular switch whose stability and activity are regulated by both light quality and photoperiod length. Under long-day conditions, CO accumulates during the light period and activates transcription of FLOWERING LOCUS T (FT), encoding a mobile florigen signal that promotes flowering. This pathway involves complex regulatory networks including GIGANTEA (GI), FLAVIN-BINDING KELCH REPEAT F-BOX 1 (FKF1), and various E3 ubiquitin ligases that modulate CO protein stability. Understanding these genetic mechanisms holds significant implications for crop improvement, particularly in adapting agricultural species to changing climate conditions and expanding their geographical range. This review synthesizes current knowledge of photoperiod sensing mechanisms and highlights emerging areas for future research in plant developmental biology.

Keywords: Photoperiodism, Photoperiod sensitivity, Arabidopsis thaliana, Circadian rhythm, Photoreceptors, Florigen.

1. INTRODUCTION

The transition from vegetative to reproductive development represents one of the most critical decision points in the plant life cycle. For many flowering plant species, this developmental switch is tightly regulated by environmental cues, particularly day length or photoperiod. Photoperiodism, first described by Garner and Allard in 1920, enables plants to measure seasonal changes through detection of day length variations, thereby synchronizing flowering with favorable conditions for pollination and seed development.[1,2] This adaptive mechanism has profound implications for plant fitness, agricultural productivity, and species distribution across latitudinal gradients.

Plants can be classified into distinct photoperiodic response categories based on their flowering requirements. Long-day plants (LDPs) flower when day length exceeds a critical threshold, typically during late spring and summer in temperate regions. Conversely, short-day plants (SDPs) initiate flowering when nights exceed a critical duration, generally in late summer or autumn. Day-neutral plants exhibit flowering that is independent of photoperiod. This classification system, while useful, oversimplifies the complex quantitative relationships between photoperiod and flowering time observed in nature.[3] The molecular dissection of

photoperiodic pathways has revealed that the same core genetic components can be configured differently across species to generate diverse photoperiodic responses.

Research using the model organism Arabidopsis thaliana, a facultative long-day plant, has provided fundamental insights into the genetic architecture of photoperiodic flowering. Over the past three decades, forward and reverse genetic approaches have identified numerous genes essential for photoperiod perception and response.[4] These discoveries have revealed that photoperiodic flowering involves the integration of at least four major pathways: the photoperiod pathway, the vernalization pathway, the autonomous pathway, and the gibberellin pathway. While these pathways can function independently, they converge on a small number of floral integrator genes that ultimately determine flowering time.[5]

This review focuses specifically on the genetic regulation of photoperiod sensitivity, examining the molecular mechanisms by which plants measure day length and translate this information into appropriate developmental responses. We discuss the critical roles of photoreceptors, circadian clock components, and the photoperiod-specific regulatory network centered on CONSTANS and FLOWERING LOCUS T. Furthermore, we explore how natural variation in photoperiod pathway genes contributes to local adaptation and consider the applied implications of this knowledge for crop improvement and climate change adaptation.

2. MOLECULAR COMPONENTS OF THE PHOTOPERIOD PATHWAY

2.1. Photoreceptors and Light Quality Sensing

The perception of light signals initiating photoperiodic responses depends on multiple photoreceptor families, each sensitive to different wavelengths of light. Phytochromes (PHY) primarily absorb red and far-red light, existing in two photointerconvertible forms: the red light-absorbing Pr form and the far-red light-absorbing Pfr form.[6] In Arabidopsis, five phytochrome genes (PHYA-PHYE) encode photoreceptors with overlapping but distinct functions. PHYB plays a particularly important role in photoperiodic flowering, promoting flowering under long days through stabilization of CONSTANS protein in the light.[7]

Cryptochromes (CRY), which absorb blue and UV-A light, also contribute significantly to photoperiodic flowering regulation. CRY2 promotes flowering under long days by enhancing CO protein stability and transcriptional activity.[8,7]

The flavin-binding ZEITLUPE (ZTL) family proteins, including ZTL itself, FLAVIN-BINDING KELCH REPEAT F-BOX 1 (FKF1), and LOV KELCH PROTEIN 2 (LKP2), represent another class of blue light receptors crucial for circadian clock function and photoperiodic flowering.[9] FKF1 specifically interacts with GIGANTEA in a light-dependent manner to degrade CYCLING DOF FACTOR proteins, thereby releasing CO transcription from repression.[10]

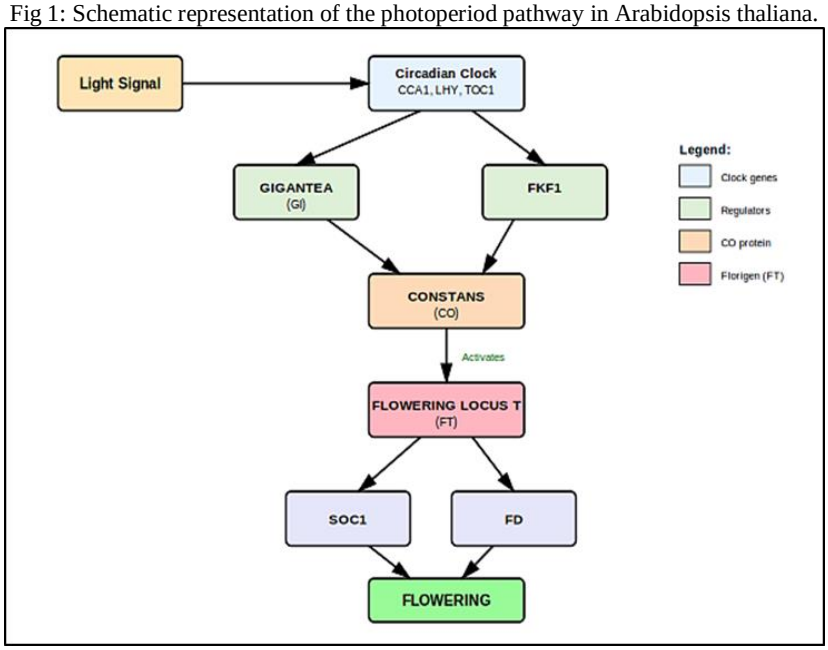
2.2. The Circadian Clock Connection

A functional circadian clock is essential for photoperiodic time measurement in plants. The clock generates endogenous approximately 24-hour rhythms that are entrained by environmental light-dark cycles. In Arabidopsis, the core oscillator comprises multiple interlocking transcriptional-translational feedback loops.

The morning-expressed MYB transcription factors CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) and LATE ELONGATED HYPOCOTYL (LHY) repress evening genes, while the evening-expressed TIMING OF CAB EXPRESSION 1 (TOC1) and related PSEUDO-RESPONSE REGULATOR (PRR) proteins repress morning genes.[11] This reciprocal regulation generates sustained oscillations in gene expression throughout the day-night cycle.

The circadian clock controls photoperiodic flowering by regulating the rhythmic expression of key photoperiod pathway genes, most notably CONSTANS. CO mRNA accumulates predominantly in the evening due to direct transcriptional activation by the clock component GIGANTEA and repression by CDF proteins, which are themselves clock-regulated.[12] Under long-day conditions, this timing ensures that CO expression peaks when light is still present, allowing CO protein to accumulate and activate FT transcription. Under short days, peak CO expression occurs in darkness, when CO protein is rapidly degraded, preventing FT activation.[13] This mechanism, termed 'external coincidence,' represents the molecular basis of photoperiod discrimination in Arabidopsis.

Fig 1: Schematic representation of the photoperiod pathway in Arabidopsis thaliana.

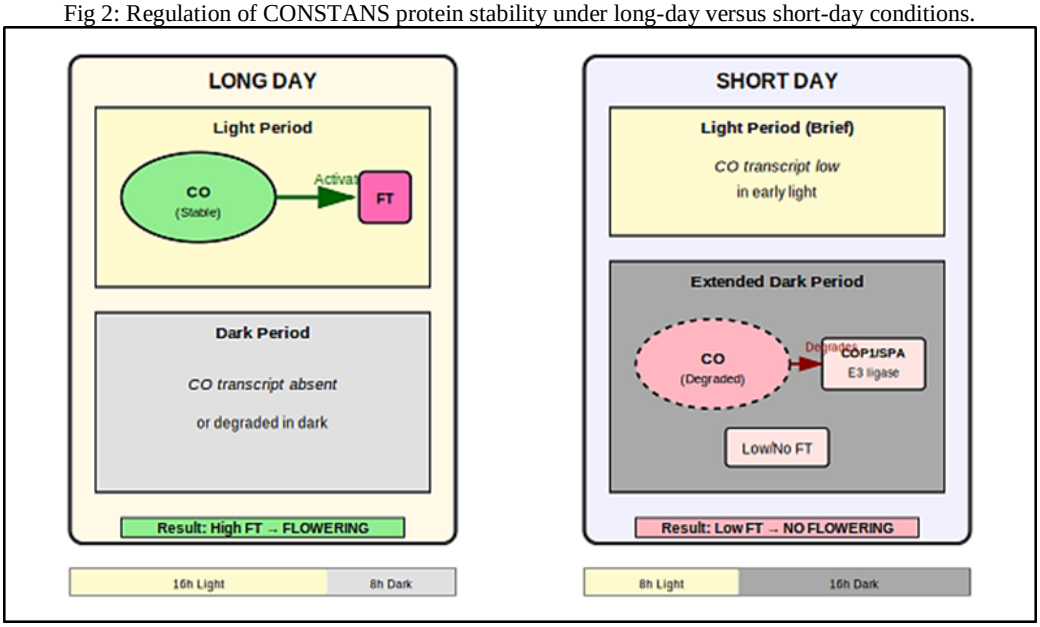


Light signals are perceived by photoreceptors and integrated with the circadian clock. The clock regulates expression of GIGANTEA (GI) and FKF1, which control CONSTANS (CO) accumulation. CO activates FLOWERING LOCUS T (FT), which then promotes flowering through downstream transcription factors SOC1 and FD. Adapted from Song et al..[3]

2.3. CONSTANS: The Central Photoperiod Switch

The CONSTANS gene encodes a zinc-finger transcription factor that serves as the primary output of the photoperiod measurement system. CO protein contains two conserved domains: a B-box zinc finger domain at the N-terminus involved in protein-protein interactions, and a CCT (CO, CO-like, TOC1) domain at the C-terminus essential for nuclear localization and transcriptional activation.[14, 15] The CO protein acts as a transcriptional activator of FLOWERING LOCUS T, directly binding to conserved regulatory elements in the FT promoter.[16] CO protein stability and activity are exquisitely regulated by light through multiple mechanisms. Under darkness, CO protein is targeted for degradation by the COP1/SPA E3 ubiquitin ligase complex, which mediates CO ubiquitination and subsequent proteasomal degradation.[17,18] In the light, phytochromes and cryptochromes antagonize COP1 activity, stabilizing CO protein. Additionally, the PHYTOCHROME-DEPENDENT LATE-FLOWERING (PHL) protein promotes CO degradation specifically in the dark through an alternative degradation pathway.[19] This complex regulation ensures that CO protein accumulates only when its expression coincides with light, which occurs specifically under long-day photoperiods.

Fig 2: Regulation of CONSTANS protein stability under long-day versus short-day conditions.



Under long days (left panel), CO expression peaks during the light period, allowing CO protein stabilization by photoreceptors and subsequent FT activation. Under short days (right panel), CO expression occurs predominantly in darkness when CO protein is degraded by COP1/SPA E3 ligase complexes, preventing FT induction. This differential CO protein accumulation underlies photoperiodic discrimination. Based on Valverde et al. [7] and Liu et al..[17]

2.4. FLOWERING LOCUS T and the Florigen Concept

FLOWERING LOCUS T represents the long-sought floral stimulus or 'florigen' whose existence was predicted by classical grafting experiments. FT encodes a small globular protein belonging to the phosphatidylethanolamine-binding protein (PEBP) family.[20, 21] FT is expressed predominantly in leaf vascular tissue under inductive long-day conditions, and the FT protein moves through the phloem to the shoot apical meristem where it initiates flowering.[22, 23] This mobile signal function explains the classical observation that exposure of even a single leaf to inductive photoperiods can trigger flowering of the entire plant.

At the shoot apex, FT protein interacts with the bZIP transcription factor FD to form a transcriptional activation complex.[24,25] This FT-FD complex directly activates floral meristem identity genes including SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1) and APETALA1 (AP1), which encode MADS-box transcription factors that specify floral fate and initiate flower development.[26] The discovery of FT as the mobile flowering signal represented a major breakthrough in plant developmental biology and confirmed the existence of a systemic floral induction mechanism operating across plant organs.

2.5. Genetic Variation and Natural Diversity in Photoperiod Sensitivity

Natural populations of plant species exhibit substantial variation in photoperiod sensitivity, reflecting local adaptation to diverse environments across latitudinal and altitudinal gradients. In Arabidopsis, naturally occurring accessions display a broad range of flowering time responses to photoperiod, from highly photoperiod-sensitive ecotypes that flower late under short days to day-neutral accessions that flower early regardless of photoperiod.[27] Quantitative trait locus (QTL) mapping and genome-wide association studies have identified major effect loci underlying this natural variation, with polymorphisms in photoperiod pathway genes frequently contributing to flowering time diversity.[5]

The FRIGIDA (FRI) and FLOWERING LOCUS C (FLC) genes represent major determinants of natural flowering time variation, though their primary function relates to vernalization rather than photoperiod. However, the photoperiod pathway gene CONSTANS also exhibits functional polymorphism in natural populations. In some rapid-cycling accessions, CO alleles carry mutations that enhance FT activation, contributing to early flowering.[28] Additionally, variation in circadian clock genes affects photoperiod sensitivity by altering the phase of CO expression relative to the light-dark cycle.[29]

In crop species, photoperiod sensitivity has been a major target of both conscious and unconscious selection during domestication and modern breeding. The expansion of crop cultivation into higher latitudes required modification of photoperiod responses to ensure flowering under longer summer days. In rice (Oryza sativa), a short-day plant, mutations in the rice ortholog of CONSTANS-like genes such as Hd1 (Heading date 1) reduce photoperiod sensitivity, allowing cultivation at northern latitudes where traditional varieties would remain vegetative.[30] Similarly, in wheat and barley, allelic variation at the PHOTOPERIOD 1 (PPD1) locus, encoding a PRR gene, determines photoperiod sensitivity and has been crucial for spring growth habit and adaptation to diverse environments.[31]

3. AGRICULTURAL APPLICATIONS AND CROP IMPROVEMENT

Understanding the genetic basis of photoperiod sensitivity offers numerous opportunities for crop improvement and agricultural innovation. Manipulation of flowering time through modification of photoperiod pathway genes enables breeders to develop varieties optimized for specific geographic regions and growing seasons. In soybean (Glycine max), extensive work has identified numerous E genes (E1-E10) that regulate photoperiod sensitivity and maturity, allowing development of cultivars adapted to latitudes ranging from the tropics to northern temperate zones.[32,33] Precise tuning of photoperiod response through allelic selection at these loci enables optimization of yield potential and avoidance of frost damage.

Climate change poses significant challenges to agriculture through altered temperature regimes and seasonal patterns. Modification of photoperiod sensitivity can help crops maintain productivity under changing environmental conditions. Research has demonstrated that reducing photoperiod sensitivity through mutations in CO-like genes or FT regulators can enable crops to flower within a broader window of environmental conditions, providing a buffer against climate variability.[34] However, such modifications must be carefully balanced against potential penalties to yield and biomass accumulation that may result from premature flowering.

Emerging genomic technologies including gene editing via CRISPR-Cas systems offer unprecedented precision in modifying photoperiod pathway genes. Recent studies have successfully used targeted mutagenesis of PRR genes in rice and other crops to fine-tune flowering time while minimizing unintended pleiotropic

effects.[35] Such approaches enable creation of novel allelic series that can be deployed in breeding programs to develop varieties with optimal flowering behavior for specific production environments. Furthermore, synthetic biology approaches may eventually allow construction of entirely novel photoperiod sensing circuits tailored to specific agricultural applications.

4. CONCLUSION

The genetic regulation of photoperiod sensitivity in flowering plants represents one of the best-understood developmental timing mechanisms in biology. Decades of research, primarily using Arabidopsis as a model system, have revealed a sophisticated molecular network that integrates environmental light signals with endogenous circadian rhythms to control flowering time. The identification of key components including photoreceptors, clock genes, CONSTANS, and FLOWERING LOCUS T, along with detailed understanding of their regulatory interactions, provides a comprehensive framework for understanding photoperiodic time measurement.

This mechanistic understanding has translated directly into agricultural applications. The discovery that natural variation in photoperiod pathway genes contributes to local adaptation has informed breeding strategies for optimizing flowering time in diverse crop species. Moreover, the conservation of core photoperiod signaling components across flowering plant species suggests that insights gained from model organisms can be effectively applied to crops. As climate change alters growing conditions worldwide, the ability to manipulate photoperiod sensitivity through both conventional breeding and biotechnological approaches will become increasingly valuable for maintaining agricultural productivity.

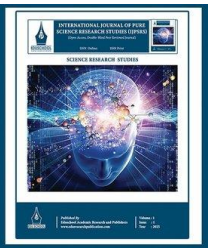
Future research directions include deeper investigation of species-specific modifications to the core photoperiod pathway that generate diverse flowering responses across the plant kingdom. While the basic architecture of the photoperiod pathway appears conserved, variation in gene copy number, expression patterns, and protein-protein interactions creates a rich substrate for evolutionary diversification. Understanding how these variations arise and how they can be harnessed for crop improvement represents an exciting frontier in plant biology. Additionally, integration of photoperiod signaling with other environmental response pathways, particularly those involving temperature and water availability, remains incompletely understood and deserves further investigation to develop crops resilient to multiple environmental stresses.

REFERENCES

1. Garner WW, Allard HA. 1920. Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants. *J Agric Res.* 18:553–606.
2. Thomas B, Vince-Prue D. 1997. *Photoperiodism in plants.* 2nd ed. San Diego (CA): Academic Press.
3. Song YH, Shim JS, Kinmonth-Schultz HA, Imaizumi T. 2015. Photoperiodic flowering: time measurement mechanisms in leaves. *Annu Rev Plant Biol.* 66:441–464. doi:10.1146/annurev-arplant-043014-115555.
4. Andrés F, Coupland G. 2012. The genetic basis of flowering responses to seasonal cues. *Nat Rev Genet.* 13(9):627–639. doi:10.1038/nrg3291.
5. Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F. 2010. Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genet.* 6(5):e1000940. doi:10.1371/journal.pgen.1000940.
6. Rockwell NC, Su YS, Lagarias JC. 2006. Phytochrome structure and signaling mechanisms. *Annu Rev Plant Biol.* 57:837–858. doi:10.1146/annurev.arplant.56.032604.144208.
7. Valverde F, Mouradov A, Soppe W, Ravenscroft D, Samach A, Coupland G. 2004. Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. *Science.* 303(5660):1003–1006. doi:10.1126/science.1091761.
8. Guo H, Yang H, Mockler TC, Lin C. 1998. Regulation of flowering time by *Arabidopsis* photoreceptors. *Science.* 279(5355):1360–1363. doi:10.1126/science.279.5355.1360.
9. Somers DE, Schultz TF, Milnamow M, Kay SA. 2000. ZEITLUPE encodes a novel clock-associated PAS protein from *Arabidopsis*. *Cell.* 101(3):319–329. doi:10.1016/S0092-8674(00)80841-7.
10. Imaizumi T, Schultz TF, Harmon FG, Ho LA, Kay SA. 2005. FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in *Arabidopsis*. *Science.* 309(5732):293–297. doi:10.1126/science.1110586.
11. Greenham K, McClung CR. 2015. Integrating circadian dynamics with physiological processes in plants. *Nat Rev Genet.* 16(10):598–610. doi:10.1038/nrg3976.
12. Sawa M, Nusinow DA, Kay SA, Imaizumi T. 2007. FKF1 and GIGANTEA complex formation is required for day-length measurement in *Arabidopsis*. *Science.* 318(5848):261–265. doi:10.1126/science.1146994.
13. Imaizumi T, Kay SA. 2006. Photoperiodic control of flowering: not only by coincidence. *Trends Plant Sci.* 11(11):550–558. doi:10.1016/j.tplants.2006.09.004.
14. Putterill J, Robson F, Lee K, Simon R, Coupland G. 1995. The CONSTANS gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell.* 80(6):847–857. doi:10.1016/0092-8674(95)90288-0.
15. Robson F, Costa MM, Hepworth SR, Vizir I, Piñeiro M, Reeves PH, Putterill J, Coupland G. 2001. Functional importance of conserved domains in the flowering-time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic plants. *Plant J.* 28(6):619–631. doi:10.1046/j.1365-313x.2001.01163.x.

16. Tiwari SB, Shen Y, Chang HC, Hou Y, Harris A, Ma SF, McPartland M, Hymus GJ, Adam L, Marion C, Belachew A, Repetti PP, Reuber TL, Ratcliffe OJ. 2010. The flowering time regulator CONSTANS is recruited to the FLOWERING LOCUS T promoter via a unique cis-element. *New Phytol.* 187(1):57–66. doi:10.1111/j.1469-8137.2010.03251.x.
17. Liu LJ, Zhang YC, Li QH, Sang Y, Mao J, Lian HL, Wang L, Yang HQ. 2008. COP1-mediated ubiquitination of CONSTANS is implicated in cryptochrome regulation of flowering in *Arabidopsis*. *Plant Cell.* 20(2):292–306. doi:10.1105/tpc.107.057281.
18. Jang S, Marchal V, Panigrahi KC, Wenkel S, Soppe W, Deng XW, Valverde F, Coupland G. 2008. *Arabidopsis* COP1 shapes the temporal pattern of CO accumulation conferring a photoperiodic flowering response. *EMBO J.* 27(8):1277–1288. doi:10.1038/emboj.2008.68.
19. Endo M, Tanigawa Y, Murakami T, Araki T, Nagatani A. 2013. PHYTOCHROME-DEPENDENT LATE-FLOWERING accelerates flowering through physical interactions with phytochrome B and CONSTANS. *Proc Natl Acad Sci USA.* 110(44):18017–18022. doi:10.1073/pnas.1310631110.
20. Kardailsky I, Shukla VK, Ahn JH, Dagenais N, Christensen SK, Nguyen JT, Chory J, Harrison MJ, Weigel D. 1999. Activation tagging of the floral inducer FT. *Science.* 286(5446):1962–1965. doi:10.1126/science.286.5446.1962.
21. Kobayashi Y, Kaya H, Goto K, Iwabuchi M, Araki T. 1999. A pair of related genes with antagonistic roles in mediating flowering signals. *Science.* 286(5446):1960–1962. doi:10.1126/science.286.5446.1960.
22. Corbesier L, Vincent C, Jang S, Fornara F, Fan Q, Searle I, Giakountis A, Farrona S, Gissot L, Turnbull C, Coupland G. 2007. FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. *Science.* 316(5827):1030–1033. doi:10.1126/science.1141752.
23. Jaeger KE, Wigge PA. 2007. FT protein acts as a long-range signal in *Arabidopsis*. *Curr Biol.* 17(12):1050–1054. doi:10.1016/j.cub.2007.05.008.
24. Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, Ikeda Y, Ichinoki H, Notaguchi M, Goto K, Araki T. 2005. FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science.* 309(5737):1052–1056. doi:10.1126/science.1115983.
25. Wigge PA, Kim MC, Jaeger KE, Busch W, Schmid M, Lohmann JU, Weigel D. 2005. Integration of spatial and temporal information during floral induction in *Arabidopsis*. *Science.* 309(5737):1056–1059. doi:10.1126/science.1114358.
26. Lee J, Lee I. 2010. Regulation and function of SOC1, a flowering pathway integrator. *J Exp Bot.* 61(9):2247–2254. doi:10.1093/jxb/erq098.
27. Koornneef M, Alonso-Blanco C, Peeters AJ, Soppe W. 2004. Genetic control of flowering time in *Arabidopsis*. *Annu Rev Plant Biol.* 55:141–172. doi:10.1146/annurev.arplant.55.031903.141644.
28. Hagenblad J, Nordborg M. 2002. Sequence variation and haplotype structure surrounding the flowering time locus FRI in *Arabidopsis thaliana*. *Genetics.* 161(1):289–298. doi:10.1093/genetics/161.1.289.
29. Mendez-Vigo B, de Andrés MT, Ramiro M, Martínez-Zapater JM, Alonso-Blanco C. 2010. Temporal analysis of natural variation for the rate of leaf production and its relationship with flowering initiation in *Arabidopsis thaliana*. *J Exp Bot.* 61(6):1611–1623. doi:10.1093/jxb/erq032.
30. Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, Sasaki T. 2000. Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene CONSTANS. *Plant Cell.* 12(12):2473–2484. doi:10.1105/tpc.12.12.2473.
31. Beales J, Turner A, Griffiths S, Snape JW, Laurie DA. 2007. A pseudo-response regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat (*Triticum aestivum* L.). *Theor Appl Genet.* 115(5):721–733. doi:10.1007/s00122-007-0603-4.
32. Watanabe S, Xia Z, Hideshima R, Tsubokura Y, Sato S, Yamanaka N, Takahashi R, Anai T, Tabata S, Kitamura K, Harada K. 2012. A map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering. *Genetics.* 188(2):395–407.
33. Lu S, Zhao X, Hu Y, Liu S, Nan H, Li X, Fang C, Cao D, Shi X, Kong L, Su T, Zhang F, Li S, Wang Z, Yuan X, Cober ER, Weller JL, Liu B, Hou X, Kong F. 2020. Natural variation at the soybean J locus improves adaptation to the tropics and enhances yield. *Nat Genet.* 52(8):773–779.
34. Jung C, Müller AE. 2009. Flowering time control and applications in plant breeding. *Trends Plant Sci.* 14(10):563–573. doi:10.1016/j.tplants.2009.07.005.
35. Endo M, Shimizu H, Nohales MA, Araki T, Kay SA. 2016. Tissue-specific clocks in *Arabidopsis* show asymmetric coupling. *Nature.* 515(7527):419–422. doi:10.1038/nature13919.
36. Bouché F, Lobet G, Tocquin P, Périlleux C. 2016. FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*. *Nucleic Acids Res.* 44(D1):D1167–D1171. doi:10.1093/nar/gkv1054.

INTERNATIONAL JOURNAL OF PURE SCIENCE RESEARCH STUDIES (IJPSRS)
(Open Access, Double-Blind Peer Reviewed Journal)
ISSN Online:



Neutrino Oscillations: Evidence, Parameters and Open Questions in Neutrino Physics

Vimala George

Associate Professor, Deptartment of Physics, At. Xavier's College for Women (Autonomous), Aluva, India

Article Information

Received: 1st May 2026
Received in revised form: 3rd June 2026
Accepted: 6th July 2026
Available online: 14th August 2026

Volume: 2
Issue: 3
DOI: https://doi.org/10.5281/zenodo.21916073

Abstract

Neutrinos change flavour as they propagate, and this behaviour requires them to carry non-zero mass, a property the original Standard Model did not provide. This review sets out the quantum-mechanical description of flavour mixing, summarises the observations that established it, and assesses what remains unresolved. Atmospheric measurements at Super-Kamiokande, the neutral-current measurement of the solar flux at the Sudbury Neutrino Observatory, and the reactor antineutrino deficit recorded by KamLAND together fixed the two dominant mass-squared splittings and their mixing angles. Short-baseline reactor experiments later measured the third mixing angle with percent-level precision, which opened the search for leptonic charge-parity violation now pursued by long-baseline accelerator beams. Global analyses constrain five of the six oscillation parameters to a few percent, while the sign of the larger splitting, the octant of the atmospheric angle, and the value of the charge-parity phase remain uncertain. Oscillation data fix only mass differences, so the absolute scale is bounded instead by tritium endpoint spectroscopy and by cosmological structure. Whether the neutrino is its own antiparticle is being tested through searches for neutrinoless double beta decay. Experiments now under construction should settle the mass ordering and provide a first measurement of the phase.

Keywords: Neutrino Oscillation, PMNS Matrix, Mass Ordering, Leptonic CP Violation, Neutrino Mass.

.

1. INTRODUCTION

The neutrino was introduced as a bookkeeping device to rescue energy conservation in beta decay and remained, for four decades, a particle assumed to be massless. That assumption was written into the Standard Model of particle physics, which contains no right-handed neutrino field and therefore no renormalisable mass term for neutral leptons. The first sign of trouble came from the Homestake experiment, which counted roughly a third of the electron neutrinos that solar models predicted should arrive from the core of the Sun.[1] For thirty years the deficit was attributed by many to shortcomings in astrophysical modelling rather than to particle properties.

A second anomaly appeared in underground detectors built to look for proton decay. Cosmic-ray interactions in the upper atmosphere produce muon and electron neutrinos in a ratio close to two to one, a prediction that follows from the pion and muon decay chain and is insensitive to the details of the primary cosmic-ray flux. Detectors consistently recorded fewer muon neutrinos than expected, and the deficit grew with the distance travelled through the Earth. In 1998 the Super-Kamiokande collaboration reported a zenith-angle dependence of that deficit which no detector systematic could reproduce and which matched the pattern expected from flavour oscillation .[2]

Confirmation from the solar sector followed. The Sudbury Neutrino Observatory measured both the electron-neutrino flux and the flux of all three active flavours from boron-8 decay in the Sun, and found that the total agreed with the solar model while the electron component did not.[3] The neutrinos had not disappeared; they

had changed flavour. The reactor experiment KamLAND then observed the same parameters at work in a laboratory beam of electron antineutrinos, recording an energy-dependent deficit over baselines of about 180 km .[4] The three results together closed the argument.

This paper reviews the formalism that describes flavour change, the measurements that determine its parameters, and the questions that remain. The emphasis is on what the data now establish, what they constrain only weakly, and which of the experiments under construction are expected to resolve each remaining ambiguity.

2. THEORETICAL DESCRIPTION OF FLAVOUR CHANGE

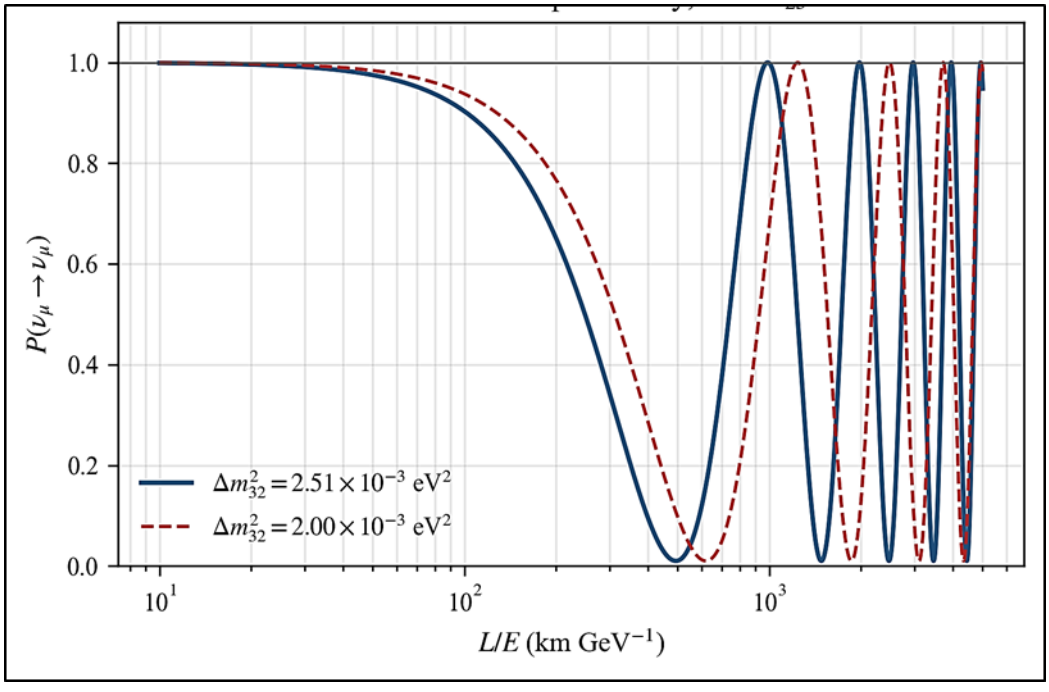
Flavour oscillation follows from a single assumption: the neutrino states produced in weak interactions are not states of definite mass. The idea was proposed by Pontecorvo in the context of neutrino and antineutrino transitions and developed for flavour mixing by Maki, Nakagawa and Sakata.[5,6] A neutrino created together with a charged lepton of flavour a is a superposition of mass states, related by a unitary matrix now named after those authors and Pontecorvo.

Because the mass states carry slightly different phases as they propagate, the superposition that arrives at a detector is no longer the state that was produced. For two flavours the survival probability takes a compact form in which the oscillation phase depends on the mass-squared difference, the distance travelled and the neutrino energy through the ratio of baseline to energy. Written in practical units, the phase equals 1.267 times the mass-squared difference in square electronvolts multiplied by that ratio in kilometres per gigaelectronvolt. Fig. 1 shows the resulting survival probability for muon neutrinos at two values of the atmospheric splitting; the position of the first minimum is what a long-baseline experiment measures most precisely, and a ten percent change in the splitting moves it visibly.

The three-flavour description requires three mixing angles, two independent mass-squared differences and one phase that can generate charge-parity violation in the lepton sector. If neutrinos are their own antiparticles, two further phases enter, but these do not affect oscillation probabilities. The two splittings differ by a factor of roughly thirty, which conveniently separates the experimental programme: the smaller splitting governs solar and long-baseline reactor measurements, and the larger one governs atmospheric and accelerator measurements.

Propagation through matter modifies the picture. Electron neutrinos undergo coherent forward scattering on electrons through the charged current, which the other flavours do not, and the resulting potential shifts the effective mixing angle. Wolfenstein derived the effect, and Mikheyev and Smirnov showed that a neutrino crossing a region of falling density can convert almost completely if it passes through a resonance. [7,8]The solar neutrino deficit is the clearest example: high-energy boron-8 neutrinos leave the Sun predominantly in the second mass state, so their survival probability is set by the mixing angle rather than by an oscillation phase. The same effect makes the sign of the larger splitting accessible to experiments with long baselines through the Earth.

Fig 1: Muon-neutrino survival probability as a function of baseline divided by energy, computed for two values of the atmospheric mass-squared splitting. Long-baseline experiments tune their beam energy and baseline to sit near the first minimum.



3. THE EXPERIMENTAL RECORD

Atmospheric neutrinos provided the first compelling evidence. Super-Kamiokande, a 50 kiloton water Cherenkov detector, exploited the fact that neutrinos arriving from above travel some tens of kilometres while those arriving from below travel up to the diameter of the Earth. The upward-going muon-neutrino rate was suppressed by about half, the downward rate was not, and the electron-neutrino rate matched expectation in both directions .[2] The pattern identified the dominant channel as transitions from muon to tau flavour.

The solar programme reached its conclusion through a measurement that was insensitive to flavour. Neutral-current interactions on deuterium occur with equal strength for all three active flavours, so the Sudbury Neutrino Observatory could compare the total active flux with the electron-flavour flux measured through charged-current interactions in the same detector. The total agreed with the standard solar model while the electron component accounted for roughly a third of it .[3] KamLAND then detected reactor antineutrinos from the Japanese power network and observed a spectral distortion consistent with the same parameters, which demonstrated that the phenomenon is a property of neutrinos and not of the Sun .[4]

The third mixing angle was the last to be measured. Its size sets the amplitude of electron-neutrino appearance in an accelerator beam and therefore controls access to the charge-parity phase. Daya Bay used eight functionally identical detectors at several distances from six reactor cores, which cancelled the flux uncertainty that had limited earlier attempts, and reported a non-zero value with high significance in 2012 .[9] RENO published a consistent result within weeks.[10] The angle turned out to be large enough that appearance experiments became practical, which reshaped the field.

Long-baseline accelerator experiments now dominate the search for the remaining unknowns. T2K directs a beam from the Japan Proton Accelerator Research Complex to Super-Kamiokande, 295 km away, and compares neutrino with antineutrino running; its data disfavour a substantial region of the phase near the charge-parity conserving values.[11] NOvA observes a beam from Fermilab over a baseline of 810 km and provides complementary sensitivity to the mass ordering through the matter effect.[12] The two experiments prefer somewhat different regions of parameter space, and reconciling them is an active question.

4. PRESENT PARAMETER DETERMINATIONS

Global analyses combine solar, atmospheric, reactor and accelerator data with a common treatment of systematic uncertainties. Table 1 lists representative values from one such analysis for the normal mass ordering.[13] Five quantities are known to a few percent. The solar angle and the smaller splitting are fixed mainly by solar data combined with KamLAND; the third angle is fixed almost entirely by short-baseline reactor experiments; and the atmospheric splitting is constrained by both accelerator and atmospheric samples.

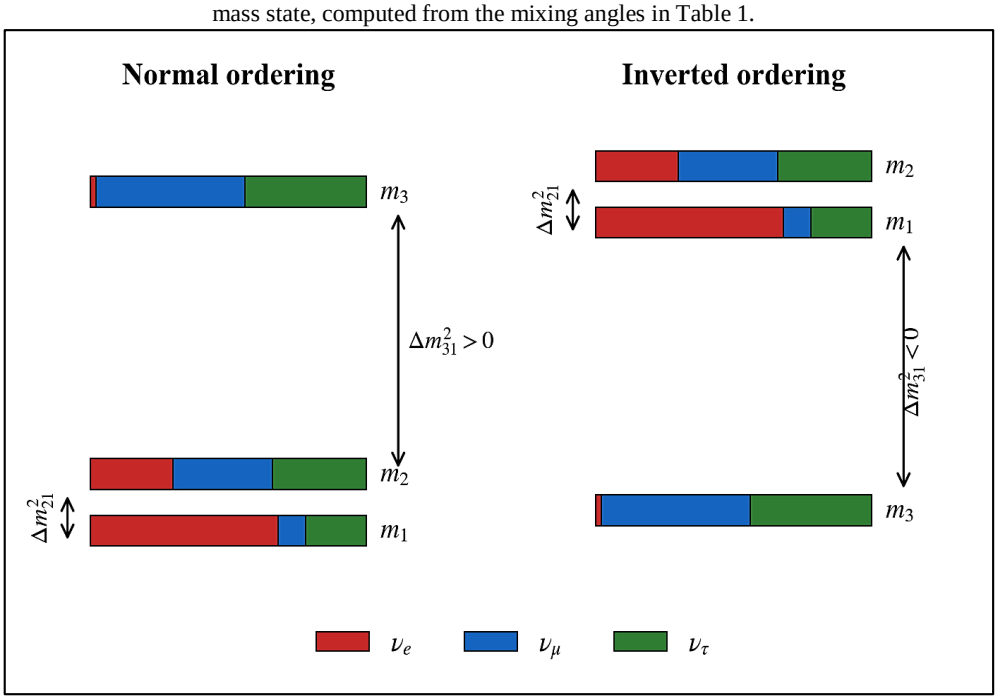
Two features of the table deserve comment. First, the atmospheric angle is consistent with maximal mixing, and whether it lies above or below 45 degrees, the so-called octant question, is not resolved. Second, the phase that governs charge-parity violation is quoted with an uncertainty of many tens of degrees, and its allowed region depends noticeably on which data sets are included. Values differ modestly between the published global fits, and quoted intervals should be read as analysis-dependent rather than as settled numbers.

Table 1. Representative global-fit values of the three-flavour oscillation parameters for the normal mass ordering Values follow the global analysis of reference.[13]

Parameter	Best fit	Three-sigma range	Principal constraint
$\sin^2\theta_{12}$	0.303	0.270 to 0.341	Solar data with KamLAND
$\sin^2\theta_{23}$	0.572	0.406 to 0.620	Atmospheric and accelerator
$\sin^2\theta_{13}$	0.02203	0.02029 to 0.02391	Short-baseline reactors
Δm^2_{21} (10^{-5} eV^2)	7.41	6.82 to 8.03	KamLAND with solar data
Δm^2_{31} (10^{-3} eV^2)	+2.511	+2.428 to +2.597	Accelerator and atmospheric
δ (degrees)	197	108 to 404	Accelerator appearance channels

The sign of the larger splitting remains open. Two arrangements are possible, illustrated in Fig. 2. In the normal ordering the state with the smallest electron content is the heaviest; in the inverted ordering it is the lightest. The distinction matters well beyond oscillation physics, because the inverted arrangement implies a minimum effective mass for neutrinoless double beta decay that current detectors are close to reaching, whereas the normal ordering allows that observable to be arbitrarily small.

Fig 2: The two possible mass orderings. Bar shading shows the approximate flavour content of each mass state, computed from the mixing angles in Table 1.



5. QUESTIONS THAT OSCILLATION DATA CANNOT ANSWER

Oscillation probabilities depend on differences of squared masses, so they say nothing about the absolute scale. Three independent approaches bound it. Kinematic measurements examine the endpoint of the tritium beta spectrum, where a non-zero mass removes the last few electronvolts of phase space. The KATRIN spectrometer has pushed this method below the electronvolt level and reported an upper limit of 0.8 eV at 90 percent confidence, with later analyses of larger data sets tightening the bound further .[14] The method is model independent, which is its main strength.

Cosmology provides a much tighter but model-dependent constraint. Massive neutrinos stream freely in the early universe and suppress the growth of structure on small scales, so the sum of the masses can be inferred from the cosmic microwave background combined with measurements of galaxy clustering. Analyses of Planck data with baryon acoustic oscillation measurements give an upper bound of about 0.12 eV on the sum.[15] The number assumes the standard cosmological model, and it would relax if that model were extended.

Whether the neutrino is a Dirac particle, distinct from its antiparticle, or a Majorana particle that is its own antiparticle, is the deepest of the open questions. A Majorana mass term violates lepton number and would allow a nucleus to undergo double beta decay without emitting neutrinos. No such decay has been seen. GERDA reported no candidate events in germanium-76 with essentially zero background , and KamLAND-Zen has set the strongest limits using xenon-136 dissolved in liquid scintillator, reaching into the region of effective masses predicted for the inverted ordering .[16,17] Interpretation of these limits in terms of mass depends on nuclear matrix elements, which differ between calculations by roughly a factor of three.

A separate question concerns whether additional neutrino states exist that do not couple to the weak interaction. The LSND experiment reported an excess of electron antineutrinos consistent with a splitting near one square electronvolt , and MiniBooNE later reported an excess of low-energy events in a similar region . [18,19] MicroBooNE, using liquid-argon imaging that separates electrons from photons event by event, found no corresponding excess of electron-neutrino interactions.[20] The tension is not fully resolved, and a global sterile-neutrino interpretation now sits uncomfortably with reactor and atmospheric disappearance data.

6. EXPERIMENTS NOW UNDER CONSTRUCTION

Three facilities are expected to resolve most of what remains. The Deep Underground Neutrino Experiment will send a wide-band beam from Fermilab to liquid-argon detectors 1300 km away in South Dakota. The baseline is long enough for the matter effect to separate the two orderings decisively, and the wide spectrum allows the oscillation pattern to be measured at several energies at once, which breaks degeneracies between the phase and the ordering .[21] Hyper-Kamiokande, a detector with roughly eight times the fiducial volume of its predecessor, will use the existing Japanese beam over a shorter baseline where matter effects are small, so its measurement of the phase is comparatively clean.[22]

JUNO takes an entirely different route. A 20 kiloton liquid scintillator detector placed 53 km from two reactor complexes will resolve the fine interference between the two splittings in the antineutrino energy spectrum. That interference pattern depends on the ordering without any reliance on matter effects or on the value of the phase, which makes the measurement complementary to the accelerator programme.[23] The experiment will also improve the precision on the solar parameters by roughly an order of magnitude.

For the Majorana question, the next generation of double beta decay searches aims at tonne-scale isotope masses with backgrounds low enough to cover the whole inverted-ordering region. If the ordering turns out to be normal and the lightest mass is small, a null result would remain compatible with Majorana neutrinos, so a definitive answer may require both a positive signal and independent information on the mass scale.

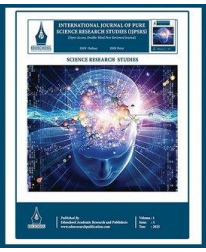
7. CONCLUSION

Flavour oscillation is now measured rather than inferred. The mixing angles and the two mass-squared splittings are known with a precision that would have seemed implausible when the solar deficit was first reported, and the framework that describes them requires only the assumption that flavour states mix. What remains unknown is structurally important: the ordering of the masses, the size of charge-parity violation among leptons, the absolute mass scale, and the Dirac or Majorana character of the particle. Each of these bears on why neutrino masses are so much smaller than those of the charged fermions, and the last two bear on whether lepton number is conserved at all. The experiments now being built address them directly, and the next decade should convert several of these open questions into measurements.

REFERENCES

1. Davis R Jr, Harmer DS, Hoffman KC. 1968. Search for neutrinos from the Sun. Phys Rev Lett. 20(21):1205-1209.
2. Fukuda Y, Hayakawa T, Ichihara E, et al. 1998. Evidence for oscillation of atmospheric neutrinos. Phys Rev Lett. 81(8):1562-1567.
3. Ahmad QR, Allen RC, Andersen TC, et al. 2002. Direct evidence for neutrino flavor transformation from neutral-current interactions in the Sudbury Neutrino Observatory. Phys Rev Lett. 89(1):011301.
4. Eguchi K, Enomoto S, Furuno K, et al. 2003. First results from KamLAND: evidence for reactor antineutrino disappearance. Phys Rev Lett. 90(2):021802.
5. Pontecorvo B. 1968. Neutrino experiments and the problem of conservation of leptonic charge. Sov Phys JETP. 26:984-988.
6. Maki Z, Nakagawa M, Sakata S. 1962. Remarks on the unified model of elementary particles. Prog Theor Phys. 28(5):870-880.
7. Wolfenstein L. 1978. Neutrino oscillations in matter. Phys Rev D. 17(9):2369-2374.
8. Mikheyev SP, Smirnov AY. 1985. Resonance amplification of oscillations in matter and spectroscopy of solar neutrinos. Sov J Nucl Phys. 42(6):913-917.
9. An FP, Bai JZ, Balantekin AB, et al. 2012. Observation of electron-antineutrino disappearance at Daya Bay. Phys Rev Lett. 108(17):171803.
10. Ahn JK, Chebotaryov S, Choi JH, et al. 2012. Observation of reactor electron antineutrinos disappearance in the RENO experiment. Phys Rev Lett. 108(19):191802.
11. Abe K, Akutsu R, Ali A, et al. 2020. Constraint on the matter-antimatter symmetry-violating phase in neutrino oscillations. Nature. 580(7803):339-344.
12. Acero MA, Adamson P, Aliaga L, et al. 2019. First measurement of neutrino oscillation parameters using neutrinos and antineutrinos by NOvA. Phys Rev Lett. 123(15):151803.
13. Esteban I, González-García MC, Maltoni M, Schwetz T, Zhou A. 2020. The fate of hints: updated global analysis of three-flavor neutrino oscillations. J High Energy Phys. 2020(9):178.
14. Aker M, Beglarian A, Behrens J, et al. 2022. Direct neutrino-mass measurement with sub-electronvolt sensitivity. Nat Phys. 18(2):160-166.
15. Aghanim N, Akrami Y, Ashdown M, et al. 2020. Planck 2018 results. VI. Cosmological parameters. Astron Astrophys. 641.
16. Agostini M, Araujo GR, Bakalyarov AM, et al. 2020. Final results of GERDA on the search for neutrinoless double-β decay. Phys Rev Lett. 125(25):252502.
17. Abe S, Asami S, Eizuka M, et al. 2023. Search for the Majorana nature of neutrinos in the inverted mass ordering region with KamLAND-Zen. Phys Rev Lett. 130(5):051801.
18. Aguilar A, Auerbach LB, Burman RL, et al. 2001. Evidence for neutrino oscillations from the observation of antineutrino-electron appearance in an antineutrino-muon beam. Phys Rev D. 64(11):112007.
19. Aguilar-Arevalo AA, Brown BC, Bugel L, et al. 2021. Updated MiniBooNE neutrino oscillation results with increased data and new background studies. Phys Rev D. 103(5):052002.
20. Abratenko P, Alrashed M, An R, et al. 2022. Search for an excess of electron neutrino interactions in MicroBooNE using multiple final-state topologies. Phys Rev Lett. 128(24):241801.
21. Abi B, Acciarri R, Acero MA, et al. 2020. Deep Underground Neutrino Experiment (DUNE), far detector technical design report, volume II: DUNE physics. J Instrum. 15(8).
22. Abe K, Abe Ke, Aihara H, et al. 2018. Hyper-Kamiokande design report. arXiv:1805.04163 [physics.ins-det].
23. An F, An G, An Q, et al. 2016. Neutrino physics with JUNO. J Phys G Nucl Part Phys. 43(3):030401.

INTERNATIONAL JOURNAL OF PURE SCIENCE RESEARCH STUDIES (IJPSRS)
(Open Access, Double-Blind Peer Reviewed Journal)
ISSN Online:



Persistent Homology in Topological Data Analysis: Theory and Applications

E Julie

Associate Professor, Department Of Zoology, Providence Women's College (Autonomous) Calicut, Kerala, India

Article Information

Received: 6th February 2026
Received in revised form: 9th March 2026
Accepted: 10th April 2026
Available online: 14th May 2026

Volume: 2
Issue: 2
DOI: https://doi.org/10.5281/zenodo.21916424

Abstract

Persistent homology is a computational methodology for extracting topological features of data across a range of scales. Originating from the work of Edelsbrunner, Letscher and Zomorodian in 2002, it has developed into a mature branch of topological data analysis with applications in neuroscience, materials science, genomics, and machine learning. This paper reviews the mathematical foundations of persistent homology, describes its computational realisation via simplicial filtrations and persistence diagrams, and surveys representative applications. Stability theorems and the bottleneck distance provide a principled metric on persistence diagrams, enabling statistical inference and integration with machine-learning pipelines. Current research directions include multi-parameter persistence, differentiable persistence modules, and scalable algorithms for large data.

Keywords: Persistent Homology, Topological Data Analysis, Simplicial Complex, Persistence Diagram; Filtration, Bottleneck Distance.

.

1. INTRODUCTION

Topological data analysis (TDA) is an emerging methodology at the intersection of algebraic topology, statistics, and computer science. Its central objective is to recover the shape of data connected components, loops, voids, and higher-dimensional analogues from finite point-cloud samples. Classical Betti numbers identify these features for a fixed topological space, but real datasets are subject to noise, sampling artefacts, and ambiguity in the choice of scale. A single fixed scale cannot distinguish meaningful topology (signal) from small-scale fluctuations (noise), and no single scale is generally 'correct' for a given dataset.

Persistent homology resolves this ambiguity by tracking topological features across a filtration of spaces indexed by a scale parameter, distinguishing long-lived features (signal) from short-lived features (noise).[1,4] The output is a compact, stable summary a persistence diagram or barcode that encodes the birth and death of topological features across the scale range. This multi-scale representation has proved to be a powerful feature extractor for complex data and has been successfully applied in domains including neuroscience, materials science, genomics, image analysis, and machine learning.

Over two decades TDA has developed a coherent theoretical framework interleaving distances, stability theorems, and decomposition results for persistence modules and practical software tools that have driven applications in diverse scientific domains.[2,5] The field has matured into a recognised branch of applied mathematics with dedicated conferences (SoCG's Applied Topology track, ATMCS), journals (Journal of Applied and Computational Topology, Foundations of Computational Mathematics), and open-source software infrastructure (GUDHI, Ripser, giotto-tda, scikit-tda). The present paper reviews the mathematical foundations,

surveys computational techniques, and illustrates applications across scientific domains. Section 2 treats simplicial complexes and filtrations; Section 3 presents persistence modules and diagrams; Section 4 addresses stability; Section 5 covers computational aspects; Section 6 surveys applications; and Section 7 discusses current research directions.

2. SIMPLICIAL COMPLEXES AND FILTRATIONS

A simplicial complex K on a vertex set V is a collection of simplices (a k-simplex being an ordered (k+1)-tuple of vertices) closed under the face relation: if $\sigma \in K$ and τ is a face of σ, then $\tau \in K$. The topological realisation |K| is obtained by gluing geometric simplices along their common faces; its singular (or simplicial) homology groups H_k (K; F) with coefficients in a field F encode the k-dimensional 'holes' of the complex, with dimensions:

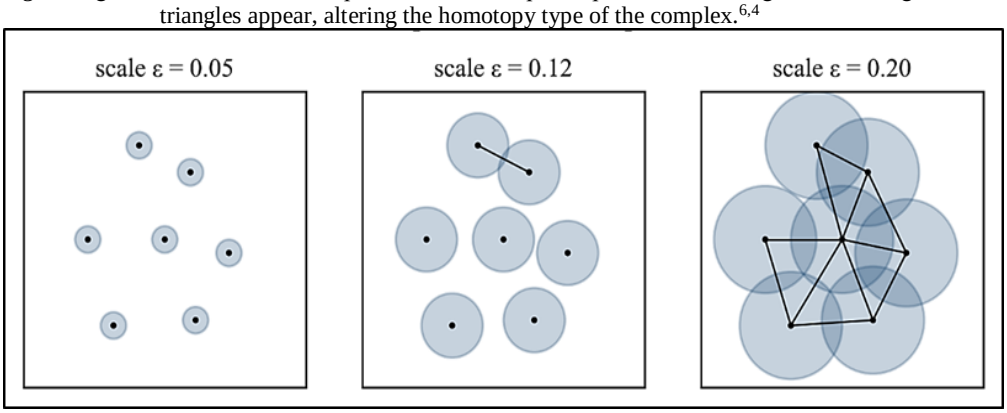
$$\beta_k = \dim H_k(K; F) \tag{1}$$

the Betti numbers. β_0 counts connected components, β_1 counts independent loops not bounding a surface, β_2 counts two-dimensional voids, and so on.

Geometric constructions from point clouds $\{x_1, \ldots, x_n\}$ in a metric space include the Čech complex Č(ε), consisting of simplices $\{x_{i_0}, \ldots x_{i_k}\}$ whose closed balls of radius ε have nonempty intersection; the Vietoris–Rips complex VR(ε), consisting of simplices whose pairwise distances are at most 2ε; the Delaunay complex, the dual of the Voronoi diagram; and the alpha complex, restricting the Čech filtration to Delaunay simplices. The Nerve Theorem links the Čech complex to the homotopy type of the union of balls: if the balls and their pairwise intersections are contractible, the Čech complex is homotopy-equivalent to their union.[6] In practice, the Vietoris–Rips complex is preferred for computational efficiency despite being geometrically larger than the Čech complex, because its construction requires only pairwise distances rather than higher-order intersections. For point clouds in high-dimensional metric spaces, sparse approximations (Sparse Rips, alpha sparse) reduce complex size while preserving persistence up to bounded perturbations.

A filtration is a nested sequence $K_0 \subseteq K_1 \subseteq \ldots \subseteq K_n$ of simplicial complexes, typically indexed by an increasing scale ε (Fig. 1). As ε grows, new topological features (connected components, loops, voids) are born and others die as components merge or loops are filled. More generally, a filtration may be indexed by any totally ordered set, and different physical or data-derived quantities sublevel sets of functions, time in dynamical systems, or biological scales produce filtrations beyond the distance-based construction. Sublevel-set filtrations of real-valued functions on a fixed domain connect persistent homology to Morse theory; the persistence pairing of critical points generalises the Morse–Smale complex and underpins applications in scalar-field visualisation.

Fig 1: Stages of a Čech/Vietoris–Rips filtration over a planar point cloud. As ε grows, new edges and triangles appear, altering the homotopy type of the complex.[6,4]



3. PERSISTENCE MODULES AND DIAGRAMS

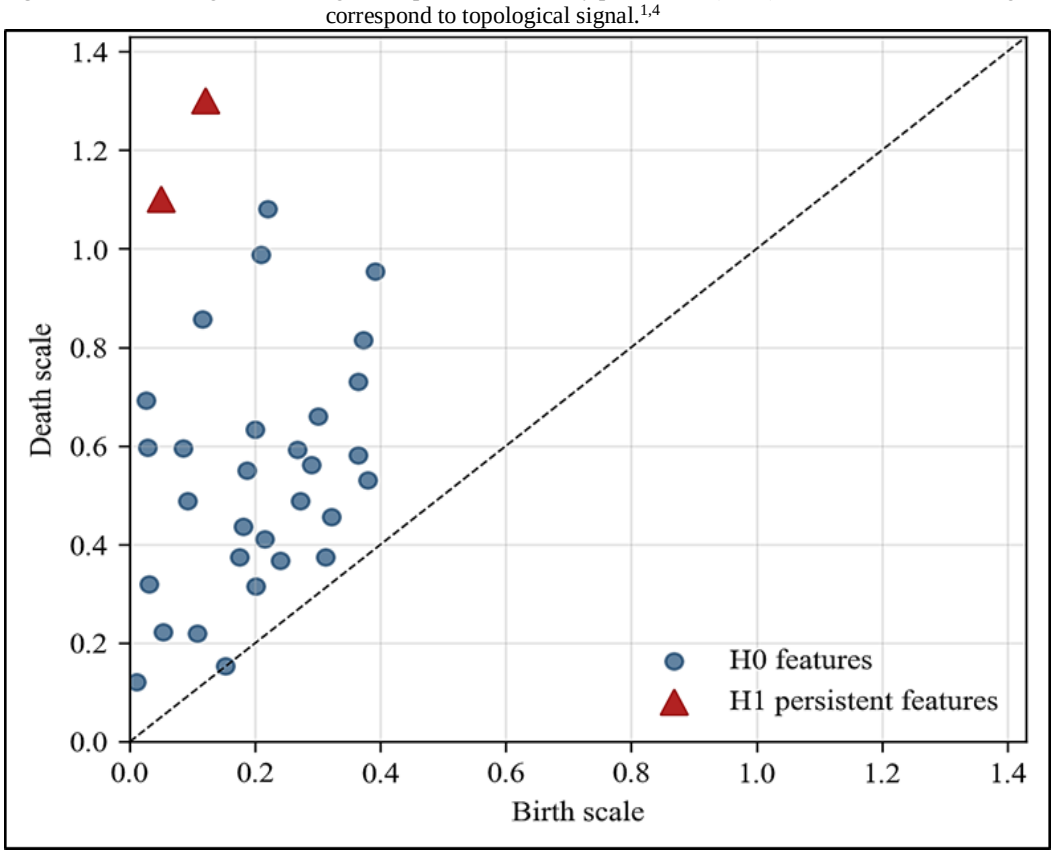
Applying simplicial homology H_k with coefficients in a field F to each stage of a filtration yields a persistence module: a family of vector spaces $\{H_k(K_i)\}$ indexed by i, together with induced linear maps $H_k(K_i) \to H_k(K_j)$ for $i \leq j$, arising from the inclusion $K_i \subset K_j$. Viewed categorically, a persistence module is a functor from the poset $(\mathbb{N}, \leq)$ (or more generally $(\mathbb{R}, \leq)$) to the category of vector spaces over F. This functorial perspective connects persistent homology to deep structural results in representation theory and category theory.

A foundational decomposition theorem, established by Crawley-Boevey (2015) for persistence modules over the reals and earlier by others for finite-type cases, shows that finite-type persistence modules decompose uniquely (up to isomorphism) into interval modules. An interval module I[b,d) is the persistence module with F at each index in [b,d) and zero elsewhere, with identity maps on the non-zero portion. This decomposition gives a complete invariant in the form of a persistence diagram a multiset of points $(b, d) \in \mathbb{R}^2$ with $b < d$, representing

features born at scale b and dying at scale d.[1,4] The persistence barcode is the equivalent horizontal-bar representation, with each bar [b,d) corresponding to one feature.

The persistence diagram admits several equivalent but practically distinct representations. Beyond the diagram and barcode, the persistence landscape is a sequence of real-valued functions $\lambda_k: \mathbb{R} \to \mathbb{R}$ derived from stacking the interval functions in order of height, producing a stable vectorised representation amenable to L^2 statistics.[3] Persistence images further convert the diagram into a fixed-resolution two-dimensional density, enabling direct use as features in convolutional networks.[10] Fig. 2 shows a representative persistence diagram with long-persistence H_1 features indicating loops.

Fig 2: Persistence diagram: the diagonal separates features by persistence (d − b). Points far from the diagonal correspond to topological signal.[1,4]



4. STABILITY AND DISTANCES

The bottleneck distance $d_B(D, D')$ between persistence diagrams D and D' is the infimum over all bijections $\gamma : D \cup \Delta \to D' \cup \Delta$ (where Δ is the diagonal, representing features of zero persistence) of the supremum over points:

$$P(\|p-\gamma(p)\|_\infty \tag{2}$$

This metric provides the natural setting for persistence-module stability. The foundational Stability Theorem of Cohen-Steiner, Edelsbrunner and Harer asserts that if two filtrations are generated by functions f, g on a common domain, then :

$$d_B(D(f), D(g)) \leq \|f - g\|_\infty \tag{3}$$

small perturbations of input data produce only small perturbations of the persistence diagram.[2]This stability theorem underpins the statistical viability of persistent homology on noisy data.

Wasserstein distances provide L^p analogues more suited to statistical analysis: $d_{w_p}(D, D^1)$ is the infimum over bijections:

$$\gamma = \left(\sum_p \| p - \gamma(p) \| \infty^p\right)^{\frac{1}{p}} \tag{4}$$

Wasserstein distances are absolutely continuous with respect to noise, enabling central-limit-theorem-like statistical analysis. The interleaving distance on persistence modules provides a more general categorical framework in which stability becomes a statement about functors, extending to multi-parameter persistence and

zigzag persistence. The algebraic-stability theorem of Chazal and colleagues sharpens the Cohen-Steiner bound and establishes that interleaving distance equals bottleneck distance for one-parameter persistence modules.

5. COMPUTATIONAL ASPECTS

Computing persistence diagrams reduces to a boundary-matrix reduction. For a filtration with m total simplices, one constructs a boundary matrix D whose columns encode boundary homomorphisms; applying column operations modulo 2 (for Z/2 coefficients) to reduce the matrix to a canonical form reveals persistence pairs through the positions of 'pivots'. The naive reduction algorithm is $O(m^3)$ worst-case but typically $O(m^{2.37})$ or better in practice.[7]

Substantial algorithmic improvements have reduced runtimes by orders of magnitude. The 'apparent pairs' optimisation identifies easy persistence pairs during construction, removing them from the reduction. Matrix-compression techniques use spectral sequences to batch-process independent blocks; the clearing optimisation reuses earlier reduction results for subsequent dimensions; discrete Morse theory pre-processing can dramatically shrink the effective complex size by collapsing acyclic portions. Chunk-based algorithms parallelise reduction across filtration segments. Open-source libraries GUDHI (comprehensive TDA toolkit), Ripser (specialised for Vietoris–Rips), Dionysus (general persistence), PHAT (high-performance persistent-homology algorithm toolbox), and Eirene (Julia-based) now support computation on complexes with millions of simplices on commodity hardware.

For very large point clouds or streaming data, approximation techniques become essential. Sparse Rips filtrations achieve controlled approximation with fewer simplices by restricting to edges between spatially nearby points, preserving the persistence diagram to within bounded error. Lazy witness complexes use a small set of landmark points and approximate persistence via witnessing relationships with the full point cloud. Distance-to-measure (DTM) filtrations replace the Euclidean distance with a robust distance estimated from the empirical measure, substantially improving robustness to outliers.[8] Recent advances in vineyards (time-varying persistence) and zigzag persistence (non-monotonic filtrations) extend the framework to dynamic and bidirectional settings.

6. APPLICATIONS

TDA applications span the sciences. In neuroscience, persistent homology characterises functional-brain-network topology derived from resting-state fMRI, diffusion MRI, and MEG/EEG recordings. Persistent Betti numbers computed from thresholded correlation matrices have been used to identify signatures of Alzheimer's disease, schizophrenia, epilepsy, and autism spectrum disorder. Topological summary features derived from brain connectomes achieve classification accuracies competitive with or superior to conventional graph-theoretic measures such as small-worldness and modularity.[9] The cavities and voids of functional connectivity networks, captured by H_1 and H_2 persistence, encode aspects of brain organisation inaccessible to pairwise connectivity measures.

In materials science, TDA analyses porous-material structure, nanoparticle aggregation, dislocation dynamics, glass transition, and amorphous solid structure. The free-volume distribution characterised by persistent homology correlates with ion-transport properties in battery electrolytes; ring-size distributions of amorphous carbon derive quantitatively from H_1 barcodes; and Kondo-lattice phase transitions exhibit persistent-homology signatures in simulated electron densities. In structural biology, persistent homology describes protein-structure cavities, ligand-binding pockets, and macromolecular assembly.

In genomics, TDA has been used to detect structural variation and haplotype diversity through topological analysis of distance matrices of single-nucleotide polymorphisms. The algebraic-topological nature of recombination producing non-tree-like evolutionary structure makes persistent homology a natural tool for detecting horizontal gene transfer, recombination events, and reticulate evolution. Tumour evolutionary trajectories have similarly been analysed with TDA.

In image analysis, cubical persistent homology on greyscale images provides a topological complement to conventional filtering, detecting filamentary structure (astronomical cosmic web), peaks and basins (geographical terrain), and characteristic holes in digital images. TDA has also been applied to financial time series, where persistent Betti numbers computed from sliding-window embeddings have been used to detect market-regime transitions and financial crises. In machine learning, persistence images and persistence landscapes provide vectorised representations compatible with kernel methods and neural networks.[10] Topological regularisation has been used to enforce connectivity in segmentation outputs and to control the topological genus of generative-model outputs. Table 1 summarises representative applications.

Table 1. Representative application domains for persistent homology. [9,10]

Domain	Data object	Topological feature of interest	Illustrative outcome
Neuroscience	fMRI connectivity graphs	Cycles, cavities	Alzheimer's-stage stratification
Materials science	Atomic coordinates	Voids in amorphous solids	Free-volume characterization
Genomics	SNP/haplotype graphs	Non-tree recombination loops	Detection of recombination events
Image analysis	Cubical complexes on greyscale	Peaks and basins	Astronomical filament detection
Machine learning	Data-cloud features	Global shape descriptors	Topological regularisation of neural nets

7. CURRENT DIRECTIONS

Several active research directions extend persistent homology beyond the classical one-parameter framework. Multi-parameter persistence considers filtrations indexed by two or more parameters for example, scale together with density, or time together with noise level. Unlike one-parameter persistence, multi-parameter persistence modules do not admit a complete discrete invariant analogous to the persistence diagram; however, substantial progress has been made on approximate invariants, rank invariants, and fibered-barcode descriptions. The theory connects to commutative algebra (minimal free resolutions of bigraded modules) and is actively being developed into computational tools.

Differentiable persistent homology integrates persistence computation with gradient-based machine-learning pipelines.[11] Defining a meaningful differential of the persistence diagram with respect to input data allows persistence-based loss functions to be optimised by stochastic gradient descent. This approach has been applied to regularisation of neural networks (topological autoencoders, topology-preserving dimensionality reduction), topological priors for generative models, and in physics-informed learning where topological invariants must be preserved during training. Persistence-based loss functions have been used, for example, to enforce specified Betti numbers in generated meshes or to regularise segmentation networks to produce connected regions.

Zigzag persistence generalises to filtrations with maps in both directions useful for dynamic data where features may appear and disappear in non-monotonic fashion. Applications include analysis of dynamic networks, time-varying sensor data, and evolving biological systems. Finally, persistent-homology-based loss functions have been integrated into generative models (topological generative adversarial networks), physics-informed learning, and protein-structure prediction. Persistence landscapes, persistence images, and persistent Betti numbers provide vectorised representations that embed persistence diagrams into standard machine-learning pipelines with kernel methods, random forests, and deep networks.

8. CONCLUSION

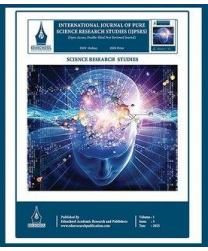
Persistent homology has transitioned from a niche mathematical methodology to a mainstream data-analysis tool within a generation. Its mathematical coherence decomposition theorems, stability results, algebraic foundations combined with scalable computation and an active open-source ecosystem has established persistent homology as a standard component of the modern data-analysis toolkit. Application successes in neuroscience, materials science, genomics, and machine learning have validated the approach empirically, while the theoretical framework continues to deepen through categorical, statistical, and differentiable extensions. The field is now poised for broader integration with deep learning, high-dimensional statistics, and scientific computing, and for continued expansion into application domains where shape and connectivity information is inadequately captured by point-wise or pair-wise descriptors alone.

REFERENCES

1. Edelsbrunner H, Letscher D, Zomorodian A. Topological persistence and simplification. *Discrete & Computational Geometry*. 2002;28(4):511–533.
2. Cohen-Steiner D, Edelsbrunner H, Harer J. Stability of persistence diagrams. *Discrete & Computational Geometry*. 2007;37(1):103–120.
3. Bubenik P. Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research*. 2015;16(3):77–102.
4. Zomorodian A, Carlsson G. Computing persistent homology. *Discrete & Computational Geometry*. 2005;33(2):249–274.
5. Carlsson G. Topology and data. *Bulletin of the American Mathematical Society*. 2009;46(2):255–308.
6. Hatcher A. *Algebraic topology*. Cambridge (UK): Cambridge University Press; 2002.
7. Bauer U. Ripser: efficient computation of Vietoris–Rips persistence barcodes. *Journal of Applied and Computational Topology*. 2021;5(3):391–423.

8. Chazal F, de Silva V, Oudot S. Persistence stability for geometric complexes. *Geometriae Dedicata*. 2014;173(1):193–214.
9. Sizemore AE, Giusti C, Kahn A, Vettel JM, Betzel RF, Bassett DS. Cliques and cavities in the human connectome. *Journal of Computational Neuroscience*. 2018;44(1):115–145.
10. Adams H, Emerson T, Kirby M, et al. Persistence images: a stable vector representation of persistent homology. *Journal of Machine Learning Research*. 2017;18(8):1–35.
11. Carrière M, Chazal F, Ike Y, Lacombe T, Royer M, Umeda Y. PersLay: a neural network layer for persistence diagrams. In: Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS); 2020. p. 2786–2796.

INTERNATIONAL JOURNAL OF PURE SCIENCE RESEARCH STUDIES (IJPSRS)
(Open Access, Double-Blind Peer Reviewed Journal)
ISSN Online:



Optimal Transport: Theory, Numerical Methods and Applications in Applied Mathematics

Assanu Augustine

Assistant Professor, Department of Mathematics, Marian College Kuttikkanam (Autonomous), Kottayam, India

Article Information

Received: 4th May 2026
Received in revised form: 6th June 2026
Accepted: 8th July 2026
Available online: 14th August 2026

Volume: 2
Issue: 3
DOI: https://doi.org/10.5281/zenodo.21916551

Abstract

Optimal transport asks for the cheapest way of rearranging one distribution of mass into another, and the answer supplies a metric on probability measures that respects the geometry of the underlying space. This review traces the subject from the formulation posed by Monge in 1781, through the linear relaxation introduced by Kantorovich, to the characterisation of optimal maps as gradients of convex functions. The resulting Wasserstein distances behave differently from divergences based on pointwise comparison of densities: they remain finite and informative for measures with disjoint support, and the shortest paths they induce displace mass rather than fade one profile into another. The dynamic formulation connects these paths to a fluid mechanical problem and identifies several diffusion equations as gradient flows of familiar energies. Numerical work has changed the field. Entropic regularisation reduces the linear program to matrix scaling iterations that are simple, parallel and differentiable, at the cost of blurring the transport plan by an amount controlled by the regularisation parameter. Semi-discrete and sliced methods offer alternatives when accuracy matters more than speed. Applications now include image retrieval, generative modelling, domain adaptation and the reconstruction of developmental trajectories from single-cell measurements.

Keywords: Optimal Transport, Wasserstein Distance, Kantorovich Duality, Sinkhorn Algorithm, Gradient Flow.

1. INTRODUCTION

The problem has a concrete origin. Monge asked how to move a pile of soil into an embankment of prescribed shape while minimising the total work, where the work of moving a grain equals its mass times the distance travelled .[1] Posed as a search over maps that push one measure onto another, the question resisted analysis for a century and a half, largely because such a map need not exist. If the source is a single atom and the target is spread over two locations, no map can split it.

Kantorovich removed the obstacle by replacing maps with couplings, that is, joint measures whose marginals are the prescribed source and target .[2] Mass may then be split, the feasible set is convex and compact in a suitable topology, and the problem becomes an infinite-dimensional linear program with a well-behaved dual. The relaxation is not merely a technical convenience: it produced the duality theory on which most later analysis rests, and it was developed alongside the work on resource allocation for which its author later shared a Nobel prize in economics.

Interest broadened considerably once it became clear that the minimal cost defines a metric on probability measures with useful geometric properties, and that this metric linearises certain nonlinear partial differential equations. Two comprehensive treatments set out the resulting theory.[3,4] This paper reviews the formulations, the

structure of optimal solutions, the algorithms that made large problems tractable, and the applications that followed.

2. FORMULATIONS AND THE STRUCTURE OF SOLUTIONS

Let two probability measures be given on a metric space together with a cost function specifying the price of moving a unit of mass from one point to another. The Monge problem seeks a measurable map, pushing the first measure onto the second, that minimises the integrated cost. The Kantorovich problem seeks instead a coupling of the two measures minimising the integral of the cost against that coupling. Every map induces a coupling, so the second problem has a value no larger than the first, and under mild conditions the two values coincide.

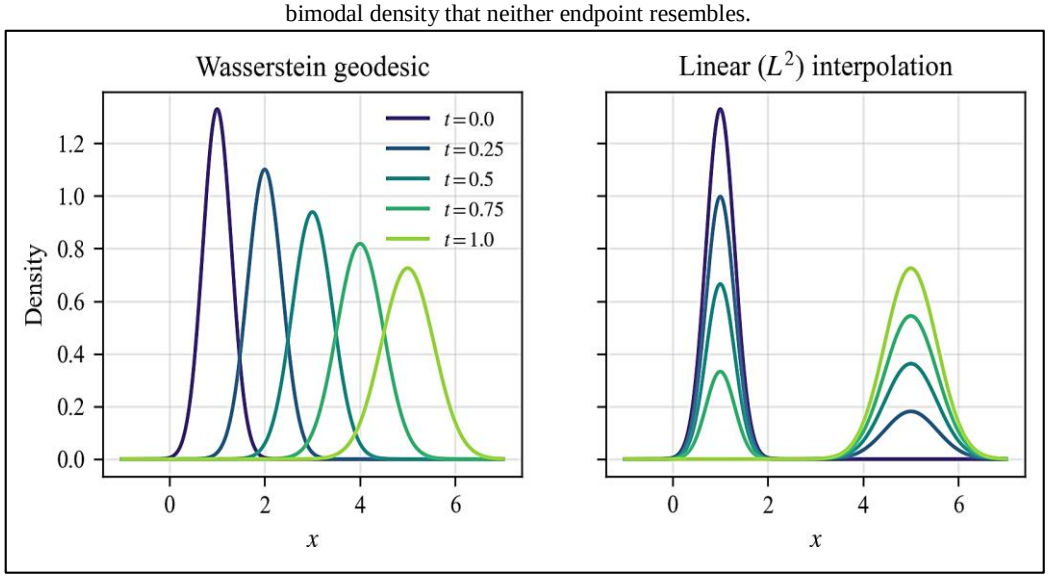
Duality is what makes the relaxed problem workable. The dual seeks a pair of potentials whose sum never exceeds the cost and which maximise the difference of their integrals against the two measures. At the optimum the constraint is active exactly on the support of the optimal coupling, a complementary slackness condition with a clear interpretation as a price system: the potentials are the prices at which mass would be bought at the source and sold at the target.

For quadratic cost on Euclidean space the structure of solutions is completely understood. Brenier proved that if the source measure is absolutely continuous, then the optimal coupling is induced by a map, that map is the gradient of a convex function, and it is unique almost everywhere .[5] The result identifies optimal transport with the polar factorisation of vector fields and connects it to the Monge-Ampère equation, since the Jacobian equation satisfied by the gradient of the potential is of that type.

The minimal cost raised to a suitable power defines the Wasserstein distances. Their most useful feature is that they compare measures through the geometry of the space rather than pointwise. Two narrow bumps separated by a small distance are close in this metric and far apart in any divergence based on overlap of densities. The associated shortest paths reflect the same principle: McCann showed that interpolating between two measures along a Wasserstein geodesic displaces mass continuously.[6] Fig. 1 contrasts this displacement interpolation with ordinary linear interpolation of densities, which creates a spurious second mode instead of moving the first.

One case is worth stating explicitly because it underlies several algorithms. On the real line with a convex cost the optimal map is the monotone rearrangement, obtained by composing the inverse distribution function of the target with the distribution function of the source. The transport cost then reduces to an integral of the distance between quantile functions, so the problem is solved by sorting rather than by optimisation. The observation explains why one-dimensional projections are useful in practice, and it supplies the closed form used in Fig. 1, where the geodesic between two Gaussian measures interpolates the mean and the standard deviation linearly.

Fig 1: Interpolation between two Gaussian densities along a Wasserstein geodesic (left) and by linear combination (right). The geodesic translates and rescales the profile, while the linear path passes through a bimodal density that neither endpoint resembles.



3. THE DYNAMIC FORMULATION AND GRADIENT FLOWS

Benamou and Brenier recast the quadratic problem as a control problem in which a density evolves under a velocity field subject to the continuity equation, and the cost becomes the integrated kinetic energy .[7] The reformulation converts a static assignment into a fluid mechanical problem, which is convex in the pair consisting

of density and momentum and can therefore be attacked by standard splitting methods. It also makes the geodesic structure explicit and provides the natural setting for a Riemannian interpretation of the space of measures.

The consequence with the widest reach concerns evolution equations. Jordan, Kinderlehrer and Otto showed that the Fokker-Planck equation is the gradient flow of the free energy with respect to the quadratic Wasserstein metric, in the precise sense that an implicit scheme alternating between minimising energy and penalising transport distance converges to its solution .[8] The same construction applies to the porous medium equation and to a range of aggregation and diffusion models. The general theory of such flows in metric spaces was developed subsequently, and it now provides both existence proofs and structure-preserving numerical schemes for equations that are difficult to treat by other means.[4]

4. COMPUTATION

Discretising both measures on n points turns the Kantorovich problem into a linear assignment problem. Network simplex and auction algorithms solve it exactly, with worst-case cost that grows roughly as the cube of n, which becomes impractical well before the sizes encountered in imaging or machine learning.

Entropic regularisation changed this. Adding a multiple of the entropy of the coupling makes the objective strictly convex, and the optimality conditions then state that the solution is a diagonal rescaling of a fixed matrix obtained by exponentiating the negative cost. Alternating the two scalings is the Sinkhorn iteration, which involves only matrix-vector products and therefore runs efficiently on parallel hardware.[9] The regularised objective is also differentiable with respect to its inputs, which is why the method was adopted quickly in learning applications.[10] Convergence to a fixed accuracy is achieved in a number of operations that is near linear in the number of entries of the cost matrix.[11]

The price is blurring. Fig. 2 shows plans computed by Sinkhorn iteration for the same pair of one-dimensional densities at three values of the regularisation parameter. At the smallest value the plan concentrates near the graph of the monotone map that solves the exact problem; as the parameter grows the plan spreads and eventually approaches the product of the marginals. Small values recover accuracy but require more iterations and invite numerical underflow, which is handled in practice by working with logarithms.

Other approaches avoid regularisation altogether. Semi-discrete methods exploit the fact that when the target is a finite set of atoms the dual problem is a smooth concave maximisation over their weights, whose gradient is computed from a power diagram; the resulting schemes give exact transport at a cost dominated by geometric computation.[12] Sliced methods project both measures onto random lines, where the one-dimensional problem is solved by sorting, and average the results.[13] Convolutional methods replace the cost matrix by heat kernel applications on a grid or mesh and never form it explicitly.[14] Table 1 compares these options.

Fig 2: Entropic transport plans between two mixtures of Gaussians, computed by Sinkhorn iteration at three values of the regularisation parameter. Support concentrates on the monotone map as the parameter decreases.

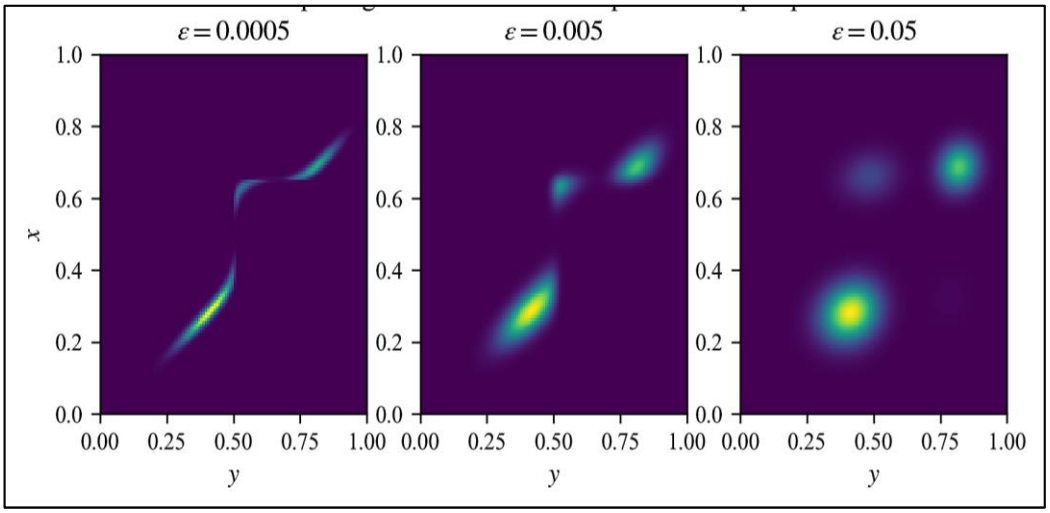


Table 1. Numerical approaches to discrete optimal transport Costs are the scalings usually quoted for each method rather than worst-case bounds.

Method	Typical cost for n points	Output	Suited to
Network simplex	Order n^3 log n	Exact plan	Small problems, exact solutions
Sinkhorn iteration	Order n^2 per iteration	Blurred plan	Large dense problems, learning [9,11]
Semi-discrete	Order n log n per iteration in low dimension	Exact map	Continuous source, discrete target [12]
Sliced projection	Order Ln log n for L direction	Distance only	Fast comparison of point clouds [13]
Convolutional	Linear in grid size	Blurred plan	Measures on grids and meshes [14]

5. STATISTICAL BEHAVIOUR IN HIGH DIMENSION

A practical question is how well the transport cost between two measures can be estimated from samples. The answer is discouraging in high dimension: the expected error of the empirical Wasserstein distance decreases as the sample size raised to the power minus one over the dimension, so the sample size required for a fixed accuracy grows exponentially.[15] This rate reflects the difficulty of covering a high-dimensional space rather than any weakness of the estimator, and it improves when the measures concentrate near a low-dimensional set.

Entropic regularisation improves the statistical picture as well as the computational one. The regularised divergence converges at a rate governed by the regularisation parameter rather than by the dimension, which is one reason regularised quantities are preferred in learning applications even when exact computation would be affordable.[16] The bias introduced by regularisation is partly removed by the debiased form, which subtracts the self-transport terms.

6. APPLICATIONS

Image retrieval provided an early application. Rubner and colleagues compared colour and texture histograms using what they called the earth mover's distance, which is the transport cost with a ground metric on feature space, and reported retrieval performance better than bin-by-bin comparison.[17] The advantage comes from the same property that distinguishes the metric theoretically: neighbouring bins are treated as similar.

In machine learning the metric was adopted for generative modelling. Training a generator by minimising a Wasserstein distance to the data distribution, with the dual potential represented by a constrained network, avoids the vanishing gradients that arise when the model and the data have disjoint support, a situation ordinary divergences cannot distinguish from complete mismatch.[18] Transport also underpins domain adaptation, where a classifier trained on one dataset is applied to another after aligning the two distributions.[19]

A striking recent use is in developmental biology. Single-cell sequencing destroys the cell it measures, so a time course yields unmatched samples from a sequence of distributions rather than trajectories. Treating consecutive time points as source and target of a transport problem recovers probable ancestor and descendant relationships, and applying this to reprogramming experiments identified the developmental routes taken by cells that succeed and by those that do not.[20] The inference relies on the assumption that cells move a small distance in expression space between sampling times, which is testable but not automatic. Barycentres in Wasserstein space provide the corresponding notion of averaging several distributions, and the applied literature collected in reference gives further worked examples.[21,22]

7. EXTENSIONS BEYOND THE BALANCED PROBLEM

Two restrictions in the classical setting limit its use, and both have been relaxed.

The first is that the two measures must have equal total mass. Many applications violate this: cells divide and die between sampling times, image features appear and disappear, and measurement noise creates spurious mass. Unbalanced formulations replace the hard marginal constraints by penalties, usually a divergence between the marginals of the coupling and the prescribed measures, so that mass may be created or destroyed at a price. The scaling algorithms used for the balanced entropic problem extend to this setting with little modification, which is one reason the formulation was adopted quickly .[23] The penalty parameter controls the trade-off between moving mass and discarding it, and results depend on it in ways that are not always reported.

The second restriction is that both measures must live in the same space, since otherwise the cost of moving a point to another is undefined. Comparing a shape in three dimensions with a graph, or two point clouds recorded in incomparable coordinates, requires a different construction. The Gromov-Wasserstein distance compares the internal distance structures instead of the points themselves, minimising over couplings the

discrepancy between pairwise distances in the two spaces.[24] The resulting problem is quadratic and non-convex, so algorithms find local minima, but the formulation gives a principled way to match objects that no rigid alignment relates. It is now used for matching single-cell datasets from different measurement technologies and for correspondence problems in geometry processing.

Both extensions inherit the computational advantages of entropic regularisation and both inherit its bias. Reported distances are therefore comparable only when the regularisation parameter, the penalty weights and the stopping rule are stated, and the literature is not consistent on this point.

8. CONCLUSION

Optimal transport has moved from a specialised topic in analysis to a working tool in computation and statistics within about two decades. The theory is settled for the quadratic cost on Euclidean space, where existence, uniqueness and the structure of the optimal map are known, and the metric it defines has a geometric interpretation that other comparisons between distributions lack. Entropic regularisation removed the computational barrier at the price of a controllable bias, and the alternatives available when that bias is unacceptable are now reasonably mature. The main open difficulty is statistical: estimating transport costs from samples remains expensive in high dimension, and the low-dimensional structure that makes applications work is usually assumed rather than verified. Methods that exploit such structure explicitly are the natural direction for further work.

REFERENCES

1. Monge G. Mémoire sur la théorie des déblais et des remblais. In: Histoire de l'Académie Royale des Sciences de Paris. Paris: Imprimerie Royale; 1781. p. 666–704.
2. Kantorovich LV. On the translocation of masses. Dokl Akad Nauk SSSR. 1942;37(7-8):199–201.
3. Villani C. Optimal transport: old and new. Berlin: Springer; 2009.
4. Ambrosio L, Gigli N, Savaré G. Gradient flows in metric spaces and in the space of probability measures. 2nd ed. Basel: Birkhäuser; 2008.
5. Brenier Y. Polar factorization and monotone rearrangement of vector-valued functions. Commun Pure Appl Math. 1991;44(4):375–417.
6. McCann RJ. A convexity principle for interacting gases. Adv Math. 1997;128(1):153–79.
7. Benamou JD, Brenier Y. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. Numer Math. 2000;84(3):375–93.
8. Jordan R, Kinderlehrer D, Otto F. The variational formulation of the Fokker-Planck equation. SIAM J Math Anal. 1998;29(1):1–17.
9. Cuturi M. Sinkhorn distances: lightspeed computation of optimal transport. Adv Neural Inf Process Syst. 2013;26:2292–300.
10. Peyré G, Cuturi M. Computational optimal transport. Found Trends Mach Learn. 2019;11(5-6):355–607.
11. Altschuler J, Weed J, Rigollet P. Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. Adv Neural Inf Process Syst. 2017;30:1964–74.
12. Mérigot Q. A multiscale approach to optimal transport. Comput Graph Forum. 2011;30(5):1583–92.
13. Bonneel N, Rabin J, Peyré G, Pfister H. Sliced and Radon Wasserstein barycenters of measures. J Math Imaging Vis. 2015;51(1):22–45.
14. Solomon J, de Goes F, Peyré G, et al. Convolutional Wasserstein distances: efficient optimal transportation on geometric domains. ACM Trans Graph. 2015;34(4):66.
15. Weed J, Bach F. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. Bernoulli. 2019;25(4A):2620–48.
16. Genevay A, Chizat L, Bach F, Cuturi M, Peyré G. Sample complexity of Sinkhorn divergences. Proc Mach Learn Res. 2019;89:1574–83.
17. Rubner Y, Tomasi C, Guibas LJ. The earth mover's distance as a metric for image retrieval. Int J Comput Vis. 2000;40(2):99–121.
18. Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. Proc Mach Learn Res. 2017;70:214–23.
19. Courty N, Flamary R, Tuia D, Rakotomamonjy A. Optimal transport for domain adaptation. IEEE Trans Pattern Anal Mach Intell. 2017;39(9):1853–65.
20. Schiebinger G, Shu J, Tabaka M, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. Cell. 2019;176(4):928–43.
21. Agueh M, Carlier G. Barycenters in the Wasserstein space. SIAM J Math Anal. 2011;43(2):904–24.
22. Santambrogio F. Optimal transport for applied mathematicians. Basel: Birkhäuser; 2015.
23. Chizat L, Peyré G, Schmitzer B, Vialard FX. Scaling algorithms for unbalanced optimal transport problems. Math Comput. 2018;87(314):2563–609.
24. Mémoli F. Gromov-Wasserstein distances and the metric approach to object matching. Found Comput Math. 2011;11(4):417–87.

INTERNATIONAL JOURNAL OF PURE SCIENCE RESEARCH STUDIES (IJPSRS)
(Open Access, Double-Blind Peer Reviewed Journal)
ISSN Online:

Mycorrhizal Symbioses: Mechanisms, Ecological Function and Responses to Global Change

Savitha Rabeque C

Assistant Professor, Department of Botany, Providence Women's College Calicut, Kerala, India

Article Information

Received: 8th May 2026
Received in revised form: 10th June 2026
Accepted: 13th July 2026
Available online: 14th August 2026

Volume: 2
Issue: 3
DOI: https://doi.org/10.5281/zenodo.21916724

Abstract

Most land plants acquire mineral nutrients through fungal partners rather than through roots alone. This review examines the cellular basis of that partnership, the differences between the principal mycorrhizal types, and the consequences for plant communities and soil processes. Fossil evidence places arbuscular associations in the earliest land floras, and roughly seventy percent of vascular plant species retain them today. Recognition begins with a reciprocal exchange of diffusible signals, proceeds through a conserved symbiosis signalling pathway shared with nitrogen-fixing associations, and ends in a highly branched fungal structure enclosed by a plant-derived membrane across which phosphate moves inward and carbon moves outward. Work with isotope tracers has shown that the fungus receives both sugars and fatty acids, which it cannot make for itself, and that plants preferentially reward the more productive fungal partners. At the ecosystem scale the type of association predicts how nitrogen is cycled, how much carbon accumulates in soil, and how strongly plant growth responds to rising carbon dioxide. Responses to warming, nitrogen deposition and land conversion differ between types, so shifts in the composition of these associations are likely to alter nutrient cycling in ways that current models capture only in outline.

Keywords: Arbuscular Mycorrhizam Ectomycorrhiza, Symbiosis Signalling, Phosphate Uptake, Soil Carbon.

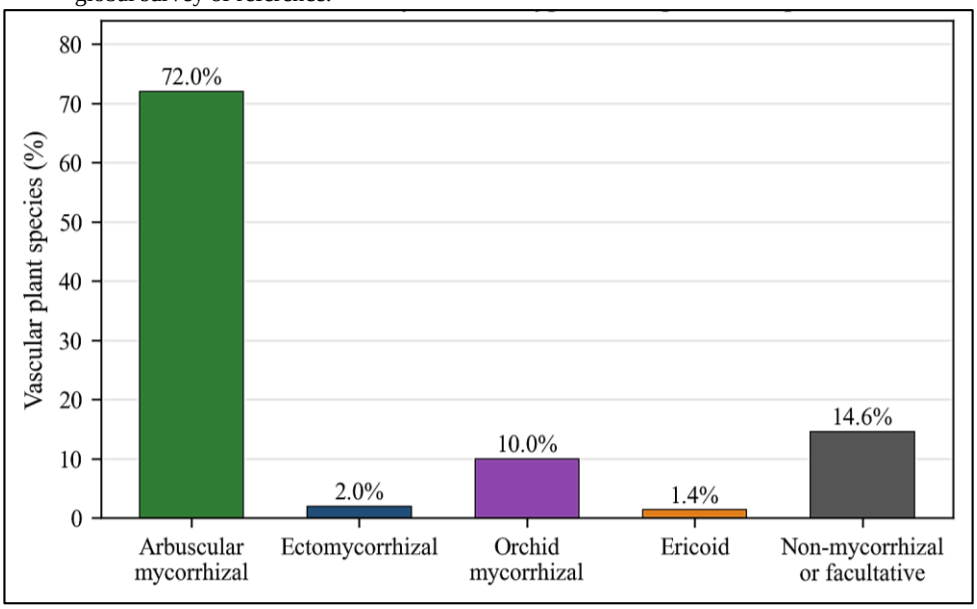
1. INTRODUCTION

The root systems of most terrestrial plants are not, functionally speaking, the organ that absorbs mineral nutrients. That role belongs largely to fungal hyphae that colonise the root and extend into soil pores far too narrow for a root hair to enter. The association is ancient. Fungal structures indistinguishable from modern arbuscules occur in the Rhynie chert, dated to roughly 400 million years, in plants that had no true roots at all .[1] This has led to the view that the symbiosis was a precondition for the colonisation of land rather than a later refinement of it .[2]

The association remains close to universal. Surveys of the vascular flora indicate that around seventy percent of species form arbuscular mycorrhizas, roughly two percent form ectomycorrhizas, and smaller fractions form the specialised orchid and ericoid associations, with the remainder either non-mycorrhizal or inconsistently colonised .[3] Figure 1 shows this distribution. The numerical dominance of the arbuscular type conceals the disproportionate importance of ectomycorrhizal plants, which include most trees of boreal and temperate forests and therefore govern nutrient cycling over large areas of the land surface.

This paper reviews the mechanisms by which these associations operate and the ecological consequences that follow from them. The first sections deal with the diversity of association types and the molecular events that establish them. Later sections consider what the association does at the scale of communities and ecosystems, how it responds to environmental change, and where attempts to use it in agriculture have succeeded or failed.

Fig 1: Approximate share of vascular plant species forming each mycorrhizal type, based on the global survey of reference. [3]



2. THE PRINCIPAL TYPES OF ASSOCIATION

Four associations are usually distinguished, and they differ in fungal partner, in the anatomy of the interface and in the nutrients they deliver. Table 1 summarises the comparison. The arbuscular symbiosis involves fungi of the subphylum Glomeromycotina, which are obligate biotrophs unable to complete their life cycle without a host. Hyphae penetrate the cortical cell wall but not the plasma membrane, and branch into the tree-like arbuscule that gives the type its name. The ectomycorrhizal association evolved repeatedly in the Basidiomycota and Ascomycota, and its hyphae remain between cells, forming a sheath around the root tip and a network in the apoplast .[4]

The functional distinction that matters most is the ability to attack organic material. Arbuscular fungi possess a limited repertoire of degradative enzymes and take up mineral nutrients, principally phosphate, from the soil solution. Many ectomycorrhizal and ericoid fungi retain enzymes inherited from saprotrophic ancestors and can mobilise nitrogen and phosphorus bound in organic matter .[5] This difference underlies much of the ecological contrast discussed later in this paper, since it determines whether the symbiosis competes with free-living decomposers or merely harvests what they release.

Identifying the fungal partners has depended almost entirely on molecular methods, because many taxa sporulate rarely and their mycelium carries few distinguishing characters. Sequencing of ribosomal markers from roots and soil has shown that the number of described arbuscular fungal species is small, in the low hundreds, against the tens of thousands of plant species they serve, so specificity is generally low and one fungal individual may colonise several plant species at once .[3] Ectomycorrhizal fungi are far more numerous and considerably more specific, with lineages restricted to particular host families .[4] The contrast matters for the ecology discussed later, since an arbuscular network can link unrelated plants growing together while an ectomycorrhizal network usually connects members of a single host group.

Table 1. Comparison of the principal mycorrhizal types Compiled from references.[3,4,5]

Type	Fungal lineage	Typical hosts	Interface	Main benefit to plant
Arbuscular	Glomeromycotina	Most herbs, grasses, many tropical trees	Intracellular arbuscular	Phosphate, zinc, water
Ectomycorrhizal	Basidiomycota, Ascomycota	Pinaceae, Fagaceae, Dipterocarpaceae	Sheath and intercellular network	Organic and mineral nitrogen
Ericoid	Ascomycota (Helotiales)	Ericaceae of acid heath and peat soils	Coils in epidermal cells	Nitrogen from recalcitrant organic matter
Orchid	Basidiomycota (Rhizoctonia group)	All Orchidaceae during germination	Intracellular pelotons	Carbon and nutrients to the seedling

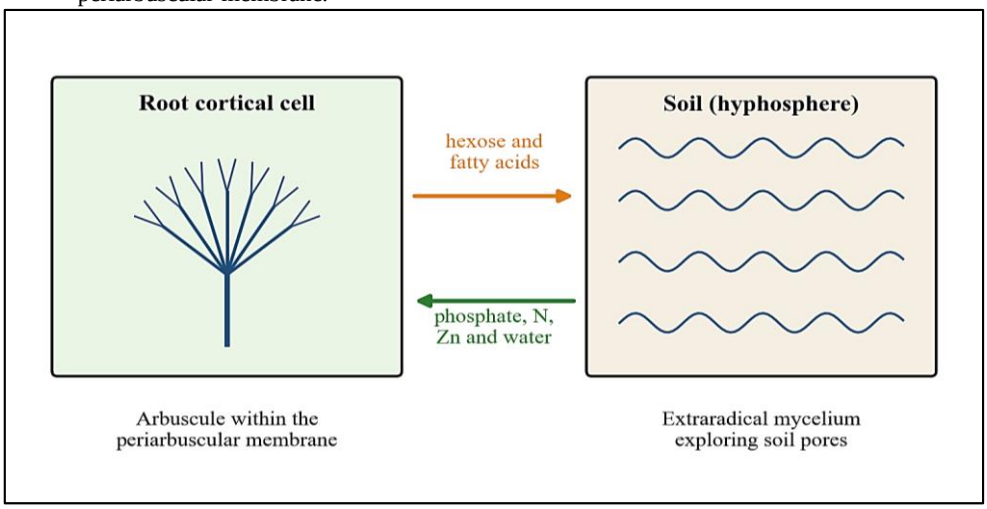
3. ESTABLISHING THE ARBUSCULAR SYMBIOSIS

Colonisation begins before contact. Roots release strigolactones into the rhizosphere, and these stimulate hyphal branching in germinating fungal spores. The fungus replies with lipochitooligosaccharides and short-chain chitin oligomers, collectively termed Myc factors, which the plant perceives through lysin-motif receptor kinases .[6] Signal perception activates a pathway involving nuclear calcium spiking, a calcium and calmodulin-dependent kinase and its transcriptional target, and this same pathway is required for nodulation in legumes. Its conservation indicates that the machinery for accommodating nitrogen-fixing bacteria was recruited from a much older fungal programme .[7]

After contact the host builds a transient structure, the prepenetration apparatus, which guides the hypha through the epidermal cell. In the cortex the hypha branches repeatedly and the plant plasma membrane invaginates around it to form the periarbuscular membrane. This membrane carries a distinct set of transporters, including mycorrhiza-specific phosphate transporters whose loss abolishes the nutritional benefit without preventing colonisation. Figure 2 shows the arrangement schematically. The interface is short lived; individual arbuscules degenerate within days and are replaced, so the exchange surface is continually renewed .[4]

What the fungus receives in return was long assumed to be sugar alone. Two studies published together showed otherwise. Isotope labelling and analysis of mutants demonstrated that the fungal lipids are synthesised by the plant and transferred across the interface, and that a plant fatty-acid biosynthetic pathway is induced specifically in arbuscule-containing cells .[8,9]Since glomeromycotinan genomes lack the genes for cytosolic fatty-acid synthesis, this dependence explains why these fungi cannot be cultured away from a host and why the association has never reverted to a free-living state.

Fig 2: Bidirectional exchange at the arbuscular interface. Carbon moves to the fungus as hexose and as fatty acids, while phosphate and other mineral nutrients move to the plant across the periarbuscular membrane.[8,9]



4. STABILITY OF THE EXCHANGE

An exchange of this kind invites cheating, since a fungus that took carbon without delivering phosphate would gain at the expense of its host. Experiments in which single root systems were offered fungal partners of differing quality showed that plants direct more carbon to the partner supplying more phosphate, and that the fungus reciprocally allocates more phosphate to the more generous host. The reward operates at the level of individual root regions rather than the whole plant, which is what allows discrimination among partners sharing one mycelium .[10]

The benefit is nevertheless conditional. A meta-analysis of inoculation experiments found responses ranging from strong growth promotion to net depression, with the outcome depending on soil phosphorus availability, on the identity of both partners and on light availability to the host.[11] Plants growing in phosphorus-rich soil suppress the symbiosis through systemic signalling, which is the expected response when the carbon cost exceeds the nutritional return. The association is therefore best described as occupying a continuum between mutualism and parasitism whose position shifts with conditions.[12]

5. CONSEQUENCES FOR COMMUNITIES AND SOILS

Mycorrhizal fungi influence plant coexistence as well as plant nutrition.[13] In microcosms, increasing the number of arbuscular fungal species increased plant diversity and productivity, because different fungi favoured

different hosts.[14] In the field, soil biota accumulate around established plants and generate feedbacks that may be negative, which promotes turnover, or positive, which promotes dominance. The direction depends on mycorrhizal type: ectomycorrhizal tree seedlings tend to experience positive feedback, whereas arbuscular species more often experience negative feedback mediated by accumulating pathogens, and this difference predicts patterns of recruitment in temperate forests.[15]

The most consequential ecosystem effect concerns carbon and nitrogen. Under the framework proposed by Phillips and colleagues, stands dominated by arbuscular trees cycle nutrients rapidly through inorganic pools, while ectomycorrhizal stands rely on organic nitrogen mobilised by their fungal partners and accumulate litter that decomposes slowly.[16] Because those fungi compete with free-living decomposers for nitrogen, they slow decomposition and increase the quantity of carbon retained in soil; a global synthesis found consistently higher soil carbon under ectomycorrhizal vegetation after accounting for climate and productivity.[17] Fungal hyphae also bind soil particles into aggregates, which protects organic matter physically and improves water relations.[18]

These effects are large enough to appear in global maps. Analyses of forest inventory data covering more than a million plots show that the distribution of tree symbioses follows the climatic controls on decomposition, with ectomycorrhizal dominance where cold or dry conditions slow the release of nutrients from litter and arbuscular dominance where decomposition is rapid.[19] Global maps of mycorrhizal vegetation constructed on this basis link the distribution of symbiosis types to terrestrial carbon stocks.[20]

6. RESPONSES TO ENVIRONMENTAL CHANGE

Rising atmospheric carbon dioxide increases plant growth only where nutrient supply permits, and mycorrhizal type is a strong predictor of which stands respond. A synthesis of free-air enrichment and chamber experiments found that plants with ectomycorrhizal partners showed substantial biomass gains under low nitrogen availability, while arbuscular plants under the same conditions showed little response.[21] The proposed explanation is that ectomycorrhizal fungi can mobilise organic nitrogen and so relieve the constraint, whereas arbuscular fungi cannot.

Nitrogen deposition acts in the opposite direction. Chronic addition reduces colonisation, shifts fungal community composition and lowers the diversity of ectomycorrhizal fungi, effects observed consistently across long-term experiments in temperate forests. Warming shifts the geographic ranges of both partners, and because fungal and plant ranges need not move together, novel combinations are likely. Conversion of land to intensive agriculture reduces arbuscular fungal abundance and diversity through tillage, phosphate fertilisation and the loss of continuous host cover.

The practical response has been to sell fungal inoculants, and here the evidence is uneven. Field trials show benefits that are largest in degraded or phosphorus-poor soils and often negligible in fertilised systems where native fungal communities are already present and plant demand is low.[11] Management that maintains host cover through cover cropping and reduces soil disturbance tends to support the native community more reliably than inoculation does.

Drought interacts with the symbiosis in both directions. Colonised plants often maintain higher water potential during dry spells, since hyphae reach water held in pores that roots cannot enter and improve the contact between soil and root surface. Severe drying nevertheless damages the extraradical mycelium, and repeated cycles of drying and rewetting reduce the length of hyphae in soil. Whether the net effect of a drier climate is protective or damaging therefore depends on the severity of the episodes rather than on annual averages, which is one reason field results have been inconsistent.

7. NETWORKS BETWEEN PLANTS AND THE LIMITS OF THE EVIDENCE

A single fungal mycelium can colonise the roots of several plants, which raises the possibility that resources move between individuals through the fungus. Isotope labelling in a Douglas fir and paper birch system reported a net transfer of carbon between trees of different species, with the direction depending on shading.[22] A later field experiment labelled the canopies of mature spruce trees and detected the label in neighbouring trees of other species, indicating transfer at ecologically meaningful scales in an old forest.[23]

These results have been extended in popular accounts into claims that trees deliberately support their offspring and warn neighbours of attack through a fungal network. A systematic review of the primary literature examined the evidence for three such claims and found it much weaker than the secondary literature suggests: field evidence that networks are widespread in forests is limited, the effect of networks on seedling performance is inconsistent in sign, and no peer-reviewed study has demonstrated defence signalling between adult trees through a network in the field. The review also documented citation practices in which unsupported claims accumulated support through repetition.[24]

The reasonable position is that transfer occurs, that its magnitude relative to a tree's carbon budget is generally small, and that transfer through a shared mycelium is difficult to separate from transfer through soil solution or through decomposing material. Distinguishing these routes requires labelling designs that few studies have implemented. The topic is a useful reminder that a mechanism demonstrated under controlled conditions

does not establish an ecological function, and that the strength of a claim should be judged from the primary measurements rather than from how often it has been repeated.

8. CONCLUSION

Mycorrhizal associations are the normal condition of terrestrial plants, not an accessory to it. The molecular events that establish the arbuscular symbiosis are now traced from the first diffusible signal to the transporters that carry phosphate and lipids across the interface, and the mechanisms that keep the exchange stable are partly understood. At larger scales the type of association predicts how nitrogen is cycled, how much carbon soils retain and how vegetation responds to enrichment of the atmosphere. The main gaps lie between these scales. Fungal community composition is still poorly represented in ecosystem models, the fate of the carbon delivered to fungi is quantified for few systems, and the consequences of reassembling plant and fungal communities under a changing climate remain largely unexamined. Progress on these points would improve predictions of soil carbon storage more than further work on the signalling pathway alone.

REFERENCES

1. Remy W, Taylor TN, Hass H, Kerp H. 1994. Four hundred-million-year-old vesicular arbuscular mycorrhizae. Proc Natl Acad Sci USA. 91(25):11841–3.
2. Field KJ, Pressel S, Duckett JG, Rimington WR, Bidartondo MI. 2015. Symbiotic options for the conquest of land. Trends Ecol Evol. 30(8):477–86.
3. Brundrett MC, Tedersoo L. 2018. Evolutionary history of mycorrhizal symbioses and global host plant diversity. New Phytol. 220(4):1108–15.
4. Genre A, Lanfranco L, Perotto S, Bonfante P. 2020. Unique and common traits in mycorrhizal symbioses. Nat Rev Microbiol. 18(11):649–60.
5. Smith SE, Read DJ. 2008. Mycorrhizal symbiosis. 3rd ed. London: Academic Press.
6. Maillet F, Poinsot V, André O, et al. 2011. Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. Nature. 469(7328):58–63.
7. Gutjahr C, Parniske M. 2013. Cell and developmental biology of arbuscular mycorrhiza symbiosis. Annu Rev Cell Dev Biol. 29:593–617.
8. Jiang Y, Wang W, Xie Q, et al. 2017. Plants transfer lipids to sustain colonization by mutualistic mycorrhizal and parasitic fungi. Science. 356(6343):1172–5.
9. Luginbuehl LH, Menard GN, Kurup S, et al. 2017. Fatty acids in arbuscular mycorrhizal fungi are synthesized by the host plant. Science. 356(6343):1175–8.
10. Kiers ET, Duhamel M, Beesetty Y, et al. 2011. Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. Science. 333(6044):880–2.
11. Hoeksema JD, Chaudhary VB, Gehring CA, et al. 2010. A meta-analysis of context-dependency in plant response to inoculation with mycorrhizal fungi. Ecol Lett. 13(3):394–407.
12. Johnson NC, Graham JH, Smith FA. 1997. Functioning of mycorrhizal associations along the mutualism-parasitism continuum. New Phytol. 135(4):575–85.
13. Tedersoo L, Bahram M, Zobel M. 2020. How mycorrhizal associations drive plant population and community biology. Science. 367(6480).
14. van der Heijden MGA, Klironomos JN, Ursic M, et al. 1998. Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. Nature. 396(6706):69–72.
15. Bennett JA, Maherali H, Reinhart KO, Lekberg Y, Hart MM, Klironomos J. 2017. Plant-soil feedbacks and mycorrhizal type influence temperate forest population dynamics. Science. 355(6321):181–4.
16. Phillips RP, Brzostek E, Midgley MG. 2013. The mycorrhizal-associated nutrient economy: a new framework for predicting carbon-nutrient couplings in temperate forests. New Phytol. 199(1):41–51.
17. Averill C, Turner BL, Finzi AC. 2014. Mycorrhiza-mediated competition between plants and decomposers drives soil carbon storage. Nature. 505(7484):543–5.
18. Rillig MC, Mummey DL. 2006. Mycorrhizas and soil structure. New Phytol. 171(1):41–53.
19. Steidinger BS, Crowther TW, Liang J, et al. 2019. Climatic controls of decomposition drive the global biogeography of forest-tree symbioses. Nature. 569(7756):404–8.
20. Soudzilovskaia NA, van Bodegom PM, Terrer C, et al. 2019. Global mycorrhizal plant distribution linked to terrestrial carbon stocks. Nat Commun. 10:5077.
21. Terrer C, Vicca S, Hungate BA, Phillips RP, Prentice IC. 2016. Mycorrhizal association as a primary control of the CO2 fertilization effect. Science. 353(6294):72–4.
22. Simard SW, Perry DA, Jones MD, Myrold DD, Durall DM, Molina R. 1997. Net transfer of carbon between ectomycorrhizal tree species in the field. Nature. 388(6642):579–82.
23. Klein T, Siegwolf RTW, Körner C. 2016. Belowground carbon trade among tall trees in a temperate forest. Science. 352(6283):342–4.
24. Karst J, Jones MD, Hoeksema JD. 2023. Positive citation bias and overinterpreted results lead to misinformation about common mycorrhizal networks in forests. Nat Ecol Evol. 7(4):501–11.