Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Spin-Selective Transport in Closed Molecular Backbones: A Geometry-Based Quantum Model and Physics-Informed Machine-Learning Study

Shreyash Krishna
Indian Institute of Technology Roorkee
Mayank Goswami
mayankfph@iitr.ac.in

Indian Institute of Technology Roorkee https://orcid.org/0000-0001-8829-9980

Research Article

Keywords: spin-selective transport, molecular geometry, spin–orbit coupling, non-equilibrium Green's function, Büttiker probe, machine learning, CISS

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11710946/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Spin-Selective Transport in Closed Molecular Backbones: A Geometry-Based Quantum Model and Physics-Informed Machine-Learning Study

Shreyash Krishna[1] and Mayank Goswami [1, 2, #]

[1] Department of Physics, IIT Roorkee

[2]*Mehta Family School of Data Science and Artificial Intelligence*, IIT Roorkee,

Roorkee, Uttarakhand-247667, India.

[#]mayank.goswami@ph.iitr.ac.in

Abstract—We study how molecular geometry, spin–orbit coupling, and electrode placement affect spin-dependent electron transport in closed molecular backbones. We compare planar loops, nonplanar unknots, and trefoil knots using a 60-site spinful tight-binding model. Local curvature and binormal direction enter an effective spin–orbit hopping term. Non-equilibrium Green's functions provide drain-resolved transmission and spin polarization for coherent multi-terminal devices and for devices with charge-conserving Büttiker voltage probes. The simulation dataset contains 25,920 drain-resolved observations across 180 independent geometry groups. Among 25,712 usable observations with transmission $T \geq 0.01$, 202 reach polarization magnitude $|\mathbf{P}| \geq 0.90$. All 202 occur in coherent two- or three-drain configurations; the largest value is 99.9708% at $T = 0.1368$. Strong polarization also appears in planar loops, so it cannot by itself establish a chirality-specific effect. We further construct 185 physics-informed features and train a gradient-boosting model to rank high-polarization candidates. On 1,836 active candidates from held-out geometries, its top-10% shortlist contains 63 of 191 high-response cases, giving 34.24% precision and a 3.29 × enrichment over random selection. These results support simulation-based screening of spin-selective molecular transport while identifying the need for stricter chirality and robustness tests.

Index Terms— spin-selective transport, molecular geometry, spin–orbit coupling, non-equilibrium Green's function, Büttiker probe, machine learning, CISS.

I. INTRODUCTION

Electron transport through a molecule can depend on the electron's spin. A prominent example is chirality-induced spin selectivity (CISS), which has been observed in several chiral molecular systems [1], [2]. The origin and generality of strong CISS signals remain active questions [3]. In particular, a large polarization measured at one output terminal does not necessarily show that molecular handedness is responsible: spin–orbit coupling (SOC), quantum interference, and the arrangement of multiple electrodes can also affect the outgoing spin distribution.

Here, we use an effective numerical model to examine these effects in closed molecular backbones. The three geometry families are planar loops, nonplanar unknots, and trefoil knots. A planar loop serves as an achiral comparison. A nonplanar unknot bends out of the plane without being topologically knotted, whereas a trefoil has nontrivial knot topology. We use the same quantum-transport equations for every family.

We also ask a computational question: can inexpensive geometric and contact descriptors help identify strong spin-selective responses without solving the full transport problem for each new candidate? We generate transport labels using the non-equilibrium Green's-function (NEGF) method and use physics-informed features to train a ranking model.

Our contributions are threefold. First, we implement and numerically validate a geometry-dependent, spinful transport model with multiple physical drains and optional voltage probes. Second, we analyze the distribution of very high polarization across the three geometry families and transport regimes. Third, we test a machine-learning (ML) screen on geometry groups excluded from model fitting and selection. The study is numerical and model based; it does not claim a microscopic description of a specific molecule, a measurement of optical rotation, or experimental proof of CISS.

II. GEOMETRIC AND QUANTUM-TRANSPORT MODEL

A. Closed-backbone geometry

Each backbone is represented by a smooth, closed spatial curve $\mathbf{r}(u)$, resampled into $N = 60$ points at approximately uniform arc-length spacing Δs. We denote the unit tangent, principal normal, binormal, and curvature by $\mathbf{T}$, $\mathbf{N}$, $\mathbf{B}$, and κ, respectively. The central geometric quantity is

$$\mathbf{q}(s) = \kappa(s)\mathbf{B}(s) = \frac{\mathbf{r}'(s) \times \mathbf{r}''(s)}{|\mathbf{r}'(s)|^3}.$$

Primes denote derivatives with respect to a curve parameter; the formula is unchanged under regular reparameterization with the consistent derivative convention. Unlike a separately normalized binormal, $\mathbf{q}$ stays well behaved when the curvature approaches zero. In the numerical implementation, derivatives are estimated from neighboring uniformly spaced sites. Its units are nm^{-1}.

Fig. 1 shows representative backbones. Each mirror pair is generated by a reflection of the three-dimensional coordinates. The family labels describe shape and topology; they are not assumed to determine the transport response on their own.

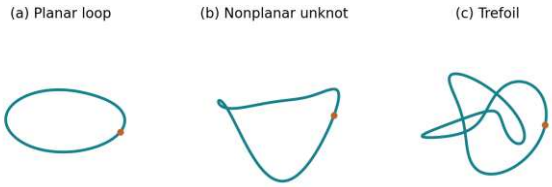


Fig. 1. Representative simulated closed backbones: (a) planar loop, (b) nonplanar unknot, and (c) trefoil knot. Curves show the effective one-dimensional backbone, not atomic structures. The marker indicates the beginning of the numerical site ordering.

B. Spinful tight-binding Hamiltonian

Each site carries one orbital with spin states $|\uparrow\rangle$ and $|\downarrow\rangle$, so the Hamiltonian is 120×120. We use

$$\begin{aligned} H &= \sum_i \varepsilon_i |i\rangle\langle i| \otimes I_2 \\ &+ \sum\langle ij\rangle |i\rangle\langle j| \otimes K_{ij} \\ &+ \sum\langle ij\rangle |j\rangle\langle i| \otimes K_{ij}\dagger \quad (2) \\ K_{ij} &= tI_2 + i\,\mathbf{d}_{ij}\cdot\boldsymbol{\sigma}, \quad \mathbf{d}_{ij} = \lambda_g(\mathbf{q}_i + \mathbf{q}_j). \end{aligned}$$

Here $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ are Pauli matrices, t is spin-independent hopping, and λ_g controls the geometry-dependent SOC strength. The backbone includes the closing bond from site $N-1$ to site 0. The numerical examples use $t = 0.3$ eV and zero onsite energy unless otherwise stated.

Equation (3) is a *chosen effective lattice discretization*. It is not an ab initio SOC calculation and should not be treated as an exact atomistic interaction for a particular chemical species. Real hopping and SOC coefficients, together with conjugate reverse hoppings, make H Hermitian and preserve time-reversal symmetry in the absence of magnetic terms.

C. Electrodes, Green's functions, and polarization

The source and drains are nonmagnetic, spin-degenerate reservoirs coupled locally to distinct backbone sites. In the wide-band limit, the

retarded self-energy of terminal a is $\Sigma_a^r = -i\Gamma_a/2$. The retarded Green's function is [4], [5]

$$G^r(E) = \left[(E + i\eta)I - H - \sum_a \Sigma_a^r\right]^{-1},$$

where $\eta > 0$ is a numerical regulator and $G^a = (G^r)^\dagger$. With local spin-independent contact broadenings γ_a, the transmission from terminal b to a is

$$T_{ab} = \mathrm{Tr}(\Gamma_a G^r \Gamma_b G^a) = \gamma_a \gamma_b \mathrm{Tr}\left(G_{ab} G_{ab}^\dagger\right).$$

Here G_{ab} is the relevant 2×2 spin block. This is the local-contact form of the NEGF transmission relation [5], [6].

For an unpolarized source, the outgoing spin-transmission matrix for drain D, without probes, is

$$M_D = \gamma_D \gamma_S G_{DS} G_{DS}^\dagger.$$

We calculate drain transmission and spin components as

$$T_D = \mathrm{Tr} M_D, \quad S_{D,k} = \mathrm{Tr}(\sigma_k M_D), \quad P_{D,k} = \frac{S_{D,k}}{T_D},$$

with $|\mathbf{P}_D| = \sqrt{P_{D,x}^2 + P_{D,y}^2 + P_{D,z}^2}$. Polarization is undefined numerically if the drain transmission is essentially zero; this issue is handled by an explicit usability check.

For the single-orbital-channel, nonmagnetic, coherent two-terminal setup studied here, time-reversal symmetry constrains the unpolarized incident current to have zero net transmitted spin polarization. A coherent device with multiple drains can nevertheless route different spin components to different outputs. Our analysis therefore reports *individual-drain* polarization rather than assuming that polarization at one drain measures chirality.

D. Floating voltage probes

We additionally consider phenomenological Büttiker probes [7], [8]. Each probe has a self-energy of the same local wide-band form, but its voltage is adjusted until its *net charge current* vanishes. For reciprocal spin-summed transmissions, the linear-response probe equations are

$$\sum_{b \neq p} T_{pb}\left(V_p - V_b\right) = 0.$$

The resulting linear system is solved for all floating voltages at once. The probe reinjection is unpolarized in our implementation, so the probes can relax spin as well as disrupt coherent phase relations. For one biased source at V_S and grounded drains, Eq. (6) becomes

$$M_D = \gamma_D \gamma_S G_{DS} G_{DS}^\dagger + \sum_p \frac{V_p}{V_S} \gamma_D \gamma_p G_{Dp} G_{Dp}^\dagger \quad (9)$$

This restores charge-conserving reinjection; adding an imaginary probe self-energy without the reinjection term would instead represent an artificial sink. These probes are an effective environmental model, not a microscopic phonon calculation.

III. NUMERICAL STUDY AND DATASET CONSTRUCTION

A. Sampling and transport calculation

The numerical study contains 180 independent geometry groups: 60 planar loops, 60 nonplanar unknots, and 60 trefoils. Each group includes mirror-related realizations. Simulations vary electron energy, geometry-dependent SOC, contact layout, source/drain broadening, the number of physical drains (one to three), and the optional probe regime. The sampled energy and SOC ranges are approximately $-0.6 \leq E \leq 0.6$ eV and $0.015 \leq \lambda_g \leq 0.230$ eV nm. The main regimes include coherent transport, six probes, and eighteen probes; contacts are spin-degenerate.

The complete simulation campaign contains 11,424 jobs, including main calculations and additional validation cases. The main analysis comprises 6,480 jobs and yields 25,920 drain-resolved observations after accounting for both mirror realizations and multiple drains. A job specifies a transport configuration, whereas an ML observation represents one drain response. The analysis uses checksum-verified simulation outputs; the accompanying notebook does not regenerate the full production campaign.

To reduce computational cost, we write the Green's-function problem as $AX = B$, where $A = (E + i\eta)I - H - \Sigma^r$. Dense LU factorization is performed once per fixed A, and B selects only the two spin columns of each contacted site. Hence the full 120×120 inverse need not be formed. Transmission coefficients are then evaluated using the 2×2 terminal blocks, followed by the floating-probe voltage solve where applicable.

B. Validation and safeguards

The notebook checks Hamiltonian Hermiticity, mirror covariance, coherent two-terminal and zero-SOC controls, current conservation for voltage probes, and the physical bound $|\mathbf{P}| \leq 1$. A two-site spin-degenerate benchmark gives $T = 0.72$ and matches its analytical solution. A Fisher–Lee scattering calculation independently agrees with the NEGF approach; the recorded unitarity error is 8.04×10^{-10}. Changing η over a tested range gives stable transport outputs. These checks establish internal numerical consistency, not experimental correctness or general convergence for every possible parameter point.

C. High-polarization descriptive analysis

We call an observation usable if its polarization is defined, its reported values are finite, and $T \geq 0.01$. The main dataset has 25,712 usable observations under this criterion. We report the number with $|\mathbf{P}| \geq 0.90$, and distinguish geometry family, number of drains, and probe regime. This descriptive analysis includes both mirror-related rows, so the count of observations should not be interpreted as a count of independent molecules or experimental measurements.

IV. PHYSICS-INFORMED MACHINE LEARNING

A. Inputs, target, and split strategy

Each raw ML row contains 427 predictors: seven operating parameters; 180 curvature-binormal components (three per site); 60 entries each for source coupling, drain coupling, probe coupling; and a 60-entry target-drain indicator. The deterministic function build_physics_features maps these inputs to 185 physically motivated descriptors. They include local geometry near the source and target, source-centered Fourier coefficients of the geometry, two directional path averages, contact-weighted descriptors, and explicit geometry–SOC products. These features use only known inputs; no NEGF target label enters feature construction.

A group-disjoint 70/15/15 split assigns 126 geometries to training, 27 to validation, and 27 to testing. Mirror copies and operating-point variants of any geometry stay in the same split. The exported single-reflection data contain 9,072 training, 1,944 validation, and 1,944 test observations. Coherent single-drain cases are excluded from active screening because their spin polarization is symmetry constrained. The final training-eligible set contains 8,508 rows after label-usability and $T \geq 0.01$ requirements.

Rather than attempting to reproduce the full spin vector, the final model ranks candidates according to predicted polarization magnitude. The training target is

$$z = \ln\left(1 + \frac{|\mathbf{P}|}{P_{\mathrm{scale}}}\right),$$

where P_{scale} is the median polarization magnitude among eligible training rows. A HistGradientBoostingRegressor [9], [10] uses 160 boosting iterations, at most 15 leaf nodes per tree, learning rate 0.05, minimum 25 samples per leaf, and L_2 regularization 1.0.

We define a high-response candidate by $T \geq 0.01$ and $|\mathbf{P}| \geq P_{90}$, where $P_{90} = 0.3993$ is the 90th percentile of eligible *training* magnitudes. Test labels are not used to set this cutoff or to train the model. The candidate rankings are saved and fixed before the held-out labels are opened for the final evaluation.

B. Ranking metrics

We select the top 10% of active test candidates by predicted score. If k candidates are selected and h satisfy the high-response criterion, precision and enrichment (lift) are

$$\text{Precision@10\%} = \frac{h}{k}$$
$$Lift@10\% = (h/k)/(N_+/Ntest) \quad (11)$$

where N_+ is the number of high-response cases among all active test candidates. We additionally report ROC area under the curve (ROC-AUC) and average precision. This is a *screening* task: the model provides a useful ordering, not a validated substitute for every transport observable.

V. RESULTS

A. Near-complete drain polarization occurs in multiple families

Among 25,712 usable main observations, 202 (0.79%) reach $|\mathbf{P}| \geq 0.90$. They span 75 geometry groups. By geometry family, 72 observations come from nonplanar unknots, 68 from planar loops, and 62 from trefoils. All 202 occur in *coherent* configurations with two or three physical drains. No observation reaches this threshold in the sampled six- or eighteen-probe regimes, nor in the coherent single-drain regime.

TABLE I — High-polarization frequency in coherent multi-drain configurations (|P|≥0.90, T≥0.01). Counts include mirror-related observations.

Geometry family	Two drains	Three drains
Nonplanar unknot	8/916 (0.87%)	64/1408 (4.55%)
Planar loop	12/946 (1.27%)	56/1424 (3.93%)
Trefoil	18/938 (1.92%)	44/1404 (3.13%)
All families	38/2800 (1.36%)	164/4236 (3.87%)

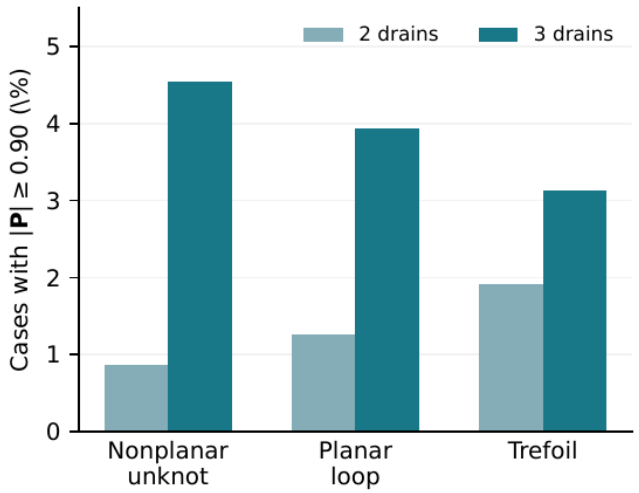


Fig. 2. Observed frequency of high polarization in the coherent two- and three-drain subsets. The percentages use the number of usable observations in each category, not the size of the complete dataset.

The largest observed polarization is 99.9708% at $T = 0.136801$ for a nonplanar unknot with source site 6 and target drain site 35, in a coherent three-drain configuration. A trefoil example reaches 99.9148% at $T = 0.558361$. The latter illustrates that very high spin selectivity can coexist with appreciable drain transmission within the model. These are sampled operating points, not demonstrated broad or disorder-tolerant operating regions.

The larger three-drain occurrence rate in Table I is consistent with enhanced opportunities for drain-resolved spin routing, although the comparison is observational: contact arrangements and sampled conditions differ between configurations. The zero count in probe-assisted regimes similarly does not prove that environmental dephasing always suppresses spin polarization.

B. The planar control limits the chirality claim

High polarization is not restricted to trefoils or nonplanar backbones. In particular, 68 high-polarization observations occur in the planar-loop family, which is achiral by construction. Thus, large $|\mathbf{P}|$ at one output is not enough to identify a chirality-induced effect. Spin-dependent multi-terminal routing provides a competing explanation within this model. A chirality-specific analysis would need to isolate a response that changes under a controlled handedness comparison and cannot be explained by electrode geometry alone.

C. Held-out ML screening

On the 27 held-out geometry groups, the active test set has 1,836 candidates. Of these, 191 meet the training-defined high-response criterion. The model selects $k = 184$ top-ranked candidates, of which 63 are true high-response cases. This yields

$$\text{Precision@10\%} = \frac{63}{184} = 34.24\%$$
$$\text{Lift@10\%} = 3.29 \quad (12)$$

The random-selection prevalence is $191/1836 = 10.40\%$. Recall at the selection cutoff is $63/191 = 32.98\%$. ROC-AUC is 0.859 and average precision is 0.354. Fig. 3 compares shortlist precision with the random baseline.

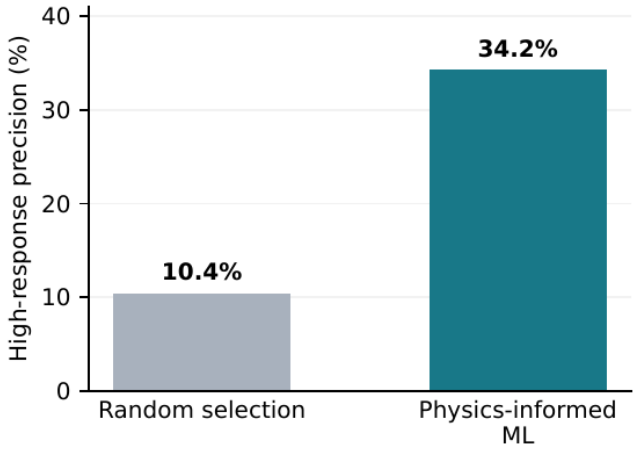


Fig. 3. High-response fraction in the top-10% ML shortlist compared with random selection on the same active held-out test population.

On the separate validation set, the ML screen achieves 69 hits among 184 selections (37.50% precision), whereas ranking by SOC strength alone gives 14/184 (7.61%). This simple baseline indicates that information beyond the SOC coefficient is useful. It does not replace comparisons against stronger physics-based heuristics or alternative ML models.

VI. DISCUSSION AND LIMITATIONS

The results establish that the chosen geometric-SOC Hamiltonian permits strong spin filtering into individual drains. Their presence across planar loops, nonplanar unknots, and trefoils shows that topological knotting is not a necessary condition for high drain-resolved polarization in the sampled multi-terminal setup. Care is needed in using the term CISS: this study tests transport in geometries motivated by chirality, but does not isolate a uniquely handedness-driven signal.

The voltage probes provide a controlled way to study phenomenological environmental coupling while conserving charge.

They do not resolve a microscopic mechanism for dephasing or spin relaxation. Likewise, the SOC coefficient is an effective parameter, not one derived from an atomistic electronic-structure calculation. The calculations assume nonmagnetic local contacts, the wide-band approximation, a zero-temperature single-energy linear-response treatment, and idealized closed backbones. No transport experiment, disorder ensemble, thermal averaging, or systematic robustness map around the 99.97% operating point is presented here.

High-polarization counts include related operating points and mirror configurations, so they should not be interpreted as statistically independent measurements. The study also uses a restricted set of geometry families and contact layouts; extrapolation outside them is unverified. The ML result supports prioritizing candidates within the simulated distribution, not reliable prediction of all quantum transport observables or experimental CISS values.

Two saved model evaluations use the same test candidates but give slightly different rankings: the reported evaluation selects 63/184 high-response cases (3.29× enrichment), compared with 64/184 (3.34×) in an earlier model evaluation. The cause of this small numerical difference has not been isolated. Both prediction files, model settings, software environments, and checksums should be retained for reproducibility.

VII. CONCLUSION

We developed a consistent numerical workflow linking closed-backbone geometry to spin-dependent tight-binding transport and ML-assisted screening. In the simulation dataset, 202 of 25,712 usable drain-resolved observations exceed 90% spin polarization; all occur in coherent multi-terminal configurations. Nearly complete polarization appears with finite transmission, but also in an achiral planar control, demonstrating why strong spin selectivity alone cannot establish a chirality-specific mechanism. A physics-informed gradient-boosting model improves high-response candidate selection by 3.29× relative to random selection on geometry-held-out test cases.

The immediate next steps are to examine robustness around strong-response operating points, perform controlled mirror and contact comparisons, and test stronger screening baselines. The present work is best understood as a validated effective-model and computational-screening study, rather than a complete explanation or experimental confirmation of CISS.

DATA AND CODE AVAILABILITY

The analysis uses preserved simulation outputs, an executed research notebook, and saved model predictions and evaluation records with integrity checks. The accompanying notebook verifies the full simulation dataset before analysis. Its optional demonstration generator follows a different experiment schedule and is not the source of the results reported here. Data and code may be supplied with the academic project submission; no public repository or permanent identifier is claimed.

Video: https://youtu.be/0QTnhiuBAXo

REFERENCES

[1] B. Göhler *et al.*, “Spin selectivity in electron transmission through self-assembled monolayers of double-stranded DNA,” *Science*, vol. 331, no. 6019, pp. 894–897, 2011, doi: 10.1126/science.1199339.

[2] R. Naaman and D. H. Waldeck, “Spintronics and chirality: Spin selectivity in electron transport through chiral molecules,” *Annu. Rev. Phys. Chem.*, vol. 66, pp. 263–281, 2015, doi: 10.1146/annurev-physchem-040214-121554.

[3] F. Evers *et al.*, “Theory of chirality induced spin selectivity: Progress and challenges,” *Adv. Mater.*, vol. 34, Art. no. 2106629, 2022, doi: 10.1002/adma.202106629.

[4] S. Datta, *Electronic Transport in Mesoscopic Systems*. Cambridge, U.K.: Cambridge Univ. Press, 1995.

[5] C. Caroli, R. Combescot, P. Nozières, and D. Saint-James, “Direct calculation of the tunneling current,” *J. Phys. C: Solid State Phys.*, vol. 4, pp. 916–929, 1971, doi: 10.1088/0022-3719/4/8/018.

[6] D. S. Fisher and P. A. Lee, “Relation between conductivity and transmission matrix,” *Phys. Rev. B*, vol. 23, pp. 6851–6854, 1981, doi: 10.1103/PhysRevB.23.6851.

[7] M. Büttiker, “Four-terminal phase-coherent conductance,” *Phys. Rev. Lett.*, vol. 57, pp. 1761–1764, 1986, doi: 10.1103/PhysRevLett.57.1761.

[8] J. L. D’Amato and H. M. Pastawski, “Conductance of a disordered linear chain including inelastic scattering events,” *Phys. Rev. B*, vol. 41, pp. 7411–7420, 1990, doi: 10.1103/PhysRevB.41.7411.

[9] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, 2001, doi: 10.1214/aos/1013203451.

[10] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.