Modeling the Photoisomerization of Retinal

Jonathan Pyle

18 March 1998

Abstract

The *cis* to *trans* photoisomerization of retinal, a common biomolecular photon detector, is modeled using the Su, Schrieffer, and Heeger model of conjugated polyenes and the extended Hubbard model of electronic interactions. This attempt improves on previous attempts by shedding the mean-field approximation, and by treating photoisomerization as a current acting on electrons in the p_z orbitals of carbon chains rather than an artificial HOMO-LUMO electron displacement within a molecular orbital framework. The model reproduces *ab initio* and experimental data well, including the energy of the photoexcited state, and with finer tuning may reproduce the 11-*cis* to all-*trans* isomerization of retinal, and may help explain the electronic mechanism of this peculiarly fast and efficient process.

Contents

1 Introduction to retinal

1.1 Biological context

The retina of the eye is covered with rod cells and cone cells. The rod cells are responsible for night vision, and the cone cells for color vision. There are about 3×10^6 cone cells and about 1×10^8 rod cells on the human retina (Fig. 1a). The outer segment of the rod cell contains a stack of about 1×10^3 disc-shaped membrane sacs, each about 160 A thick, in which the protein rhodopsin makes up 80% of the protein content (Fig. 1b,c). Rhodopsin consists of the protein opsin bound to a molecule called retinal.

Retinal, which changes shape when hit by a photon, is the first step in the process of vision. It is present in cone cells and rod cells, the only difference being the proteins to which it is bound, which influence rhodopsin's absorption spectrum. Although the opsin proteins depend on cell type and animal species, retinal appears to be a universal photon detector in animal vision. Rhodopsin's absorption spectrum peaks around 500 nm, correlating well with the solar spectrum. (Nathans 1992; Stryer 1995).

1.2 Chemical structure

The retinal molecule has a backbone of 15 carbon atoms, forming a ring of six carbons at one end, with a zigzag tail of nine carbon atoms terminated by an oxygen atom (Fig. 2a). It is synthesized from all-*trans*-retinol, or Vitamin A, which differs from retinal only in that it has CH_2OH at the tip of the tail (Fig. 2). The tail end of retinal binds to the protein through a molecule of H_2N attached to a lysine group, $(CH_2)_4$, which is attached to the opsin protein. With the introduction of a proton, retinal loses its oxygen, H_2N loses a hydrogen and an electron, and a molecule of H_2O forms, leaving a positively charged nitrogen in the oxygen's place on the retinal chain, bound to opsin via the lysine group (Fig. 3). This nitrogen is significant because its positive charge affects retinal's electronics.

Retinal's carbons have been numbered arbitrarily. Carbons 5, 9 and 13 each have an attached "methyl group," which consists of one carbon, bonded to the backbone, and three

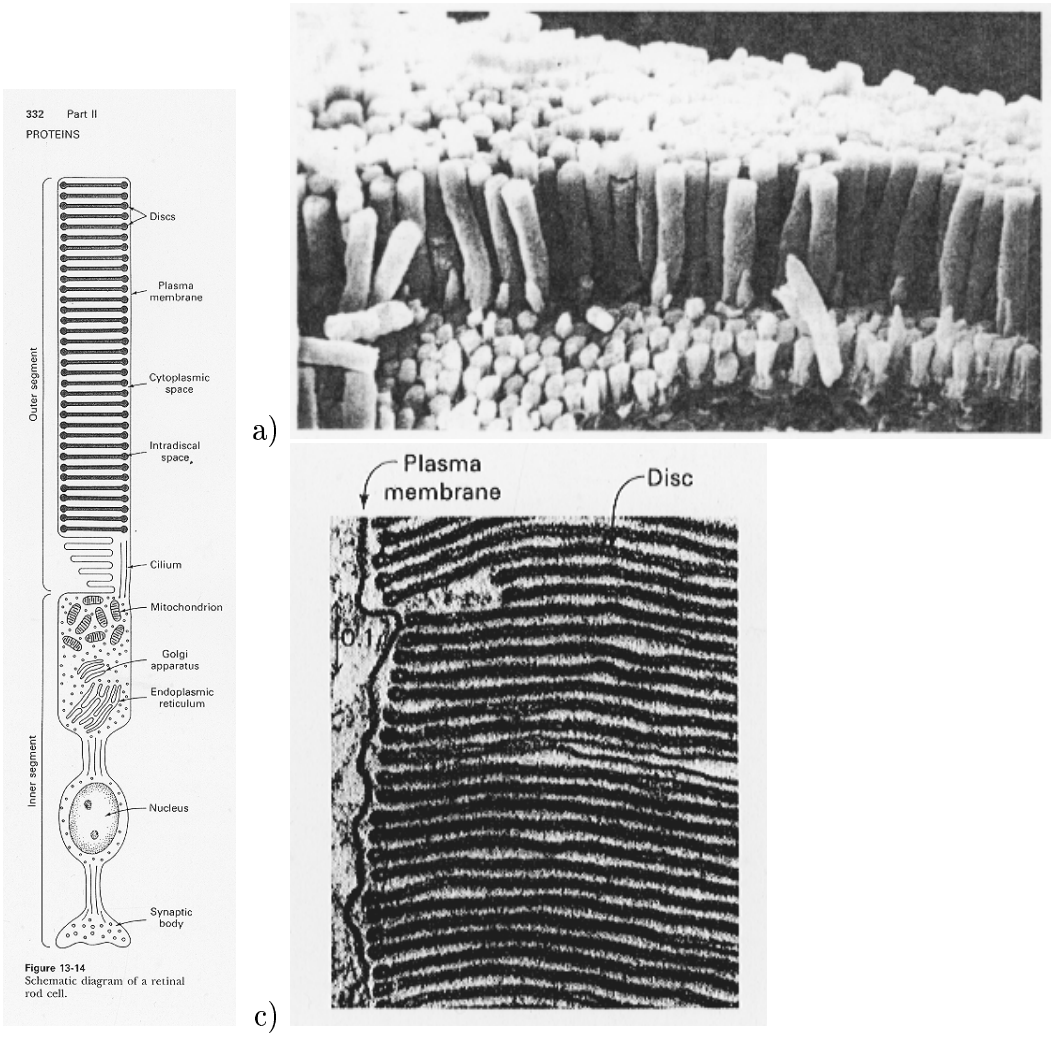


Figure 1: a) Scanning electron micrograph of retina, showing rod cells and the smaller cone cells. b) One rod cell, containing stack of membrane sacs. The rhodopsin protein resides in the membrane. c) Cross section of rod cell, showing flat membrane sacs. The rhodopsin protein is embedded in the membrane (Stryer 1995).

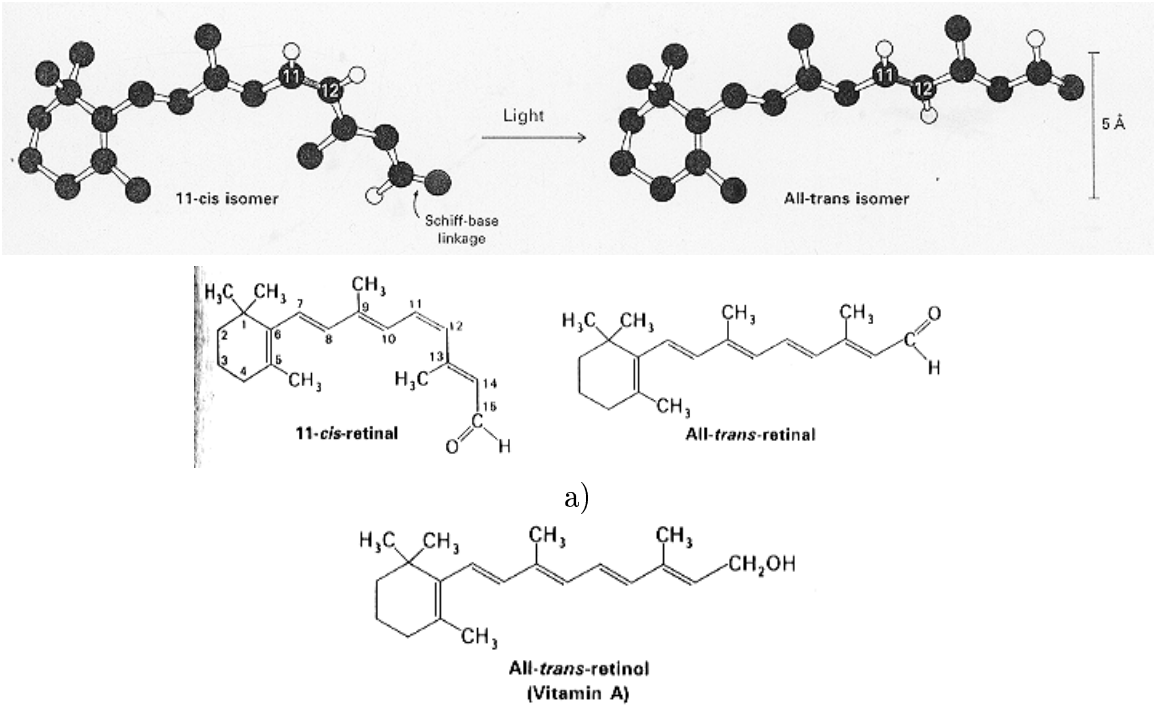


Figure 2: a) The molecule 11-*cis*-retinal, not tied to any protein. In the ball-and-stick picture, carbons are shown in black; hydrogens are not shown. In the chemical diagram, hydrogens are not shown but are implied. At the tail end of bare retinal is an oxygen atom. b) Vitamin A, from which retinal derives (Stryer 1995).

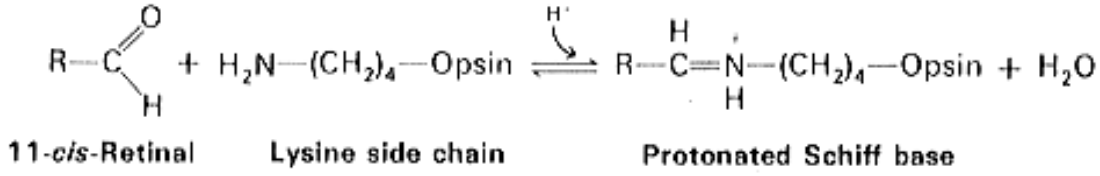


Figure 3: 11-*cis*-retinal attaches to the rhodopsin protein via a lysine group and forms a protonated Schiff base, or a positive charge at the tip of retinal's tail (Stryer 1995).

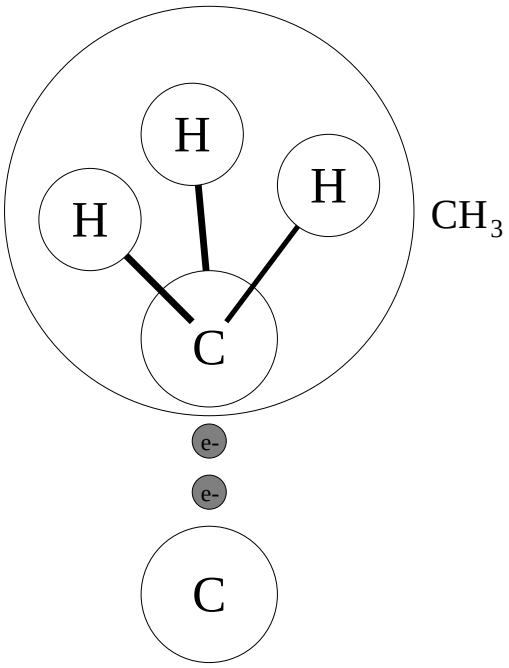


Figure 4: A methyl group. In retinal, methyls are bonded through their carbons to retinal's carbons 5, 9, and 13.

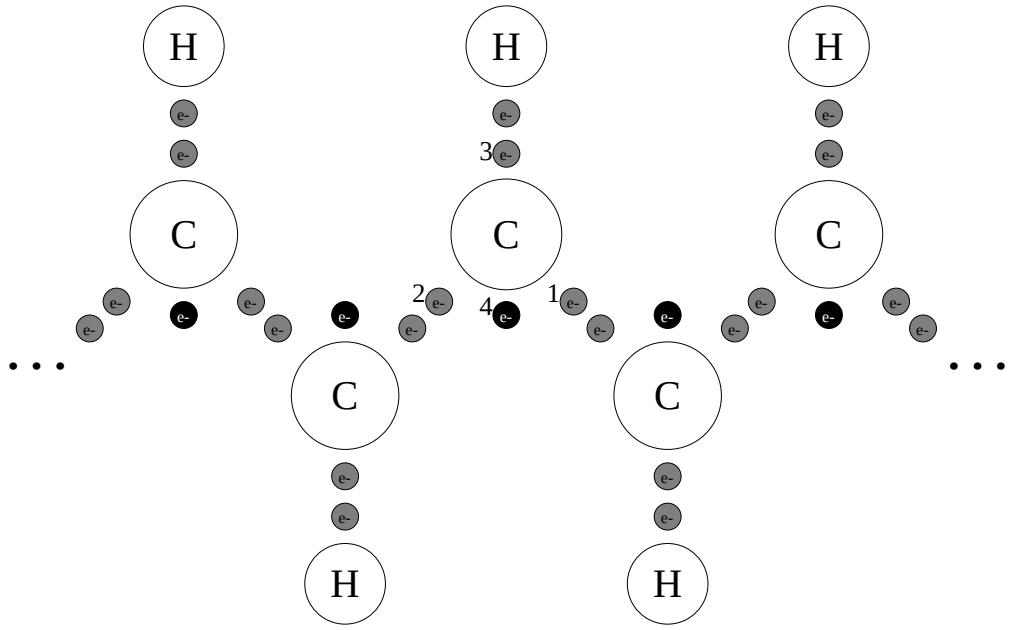


Figure 5: Each carbon in the carbon chain is bonded to a hydrogen. (The hydrogens are omitted from chemical diagrams). Each carbon has two valence electrons caught up in bonds to the left and right neighboring carbons. The third electron is in the bond with the hydrogen, and the fourth (shown dark) inhabits the p_z orbital and can jump from site to site, as if the carbon chain were a kind of one-dimensional metal.

hydrogens (Fig. 4). The other carbons have an attached hydrogen instead.

The retinal molecule can take many forms, or "isomers." The surrounding protein constrains the molecule so that it assumes a bent form, called 11-*cis,* because the bond at carbon number 11 is a "cis" bond, rather than a "trans" bond. If the tail of the molecule were rotated 180° around the 11–12 bond, then the molecule would be straight. This straight configuration is called "all-trans" retinal. When the molecule changes shape, without changing its atomic content, it is said to "isomerize."

1.3 Electronics

Each carbon in the chain has four valence electrons (Fig. 5). One of the electrons is caught up in a covalent bond with the neighboring carbon to the left. Another is caught up in a covalent bond with the neighbor on the right. The third is caught up in a covalent bond with the hydrogen atom or methyl group. The three peripheral atoms repel each other, so that they form a planar configuration around the center carbon, with bonds 120° apart. The fourth moves up and down through the plane, occasionally exchanging places with a corresponding electron from a neighboring carbon. In this way the carbon chain is like a metal.

Since the atoms are bound together into a molecule, the electrons inhabit not pure electronic orbitals like s and p (Fig. 6a), but "hybrid orbitals," which combine elements of s and p. When the s and p orbitals of neighboring carbons overlap, the overlapping s orbitals form a σ bond and the exchanging of electrons between p orbitals forms a π bond. The orbitals of each carbon in the chain are of the type known as sp^2p_z (Fig. 6b,c). If two atoms are bonded by σ bonds but not strongly by π bonds, their bond is called "single." When π bonding is significant, however, the atoms will be closer together and their bond is called "double."

The potential wells of double bonds are deeper than those of single bonds (Fig. 7). Therefore, a chain of carbon-hydrogen pairs will form as many double bonds as possible.

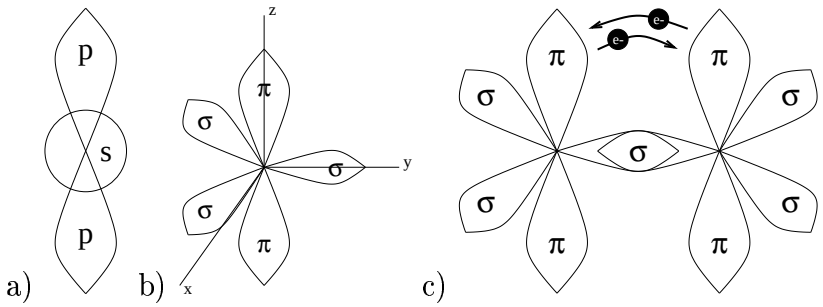


Figure 6: a) Standard s and p orbitals. b) Hybrid orbitals σ and π, in the sp^2p_z configuration. c) Bonds form when the orbitals are close enough to exchange electrons and form a wavefunction of their own.

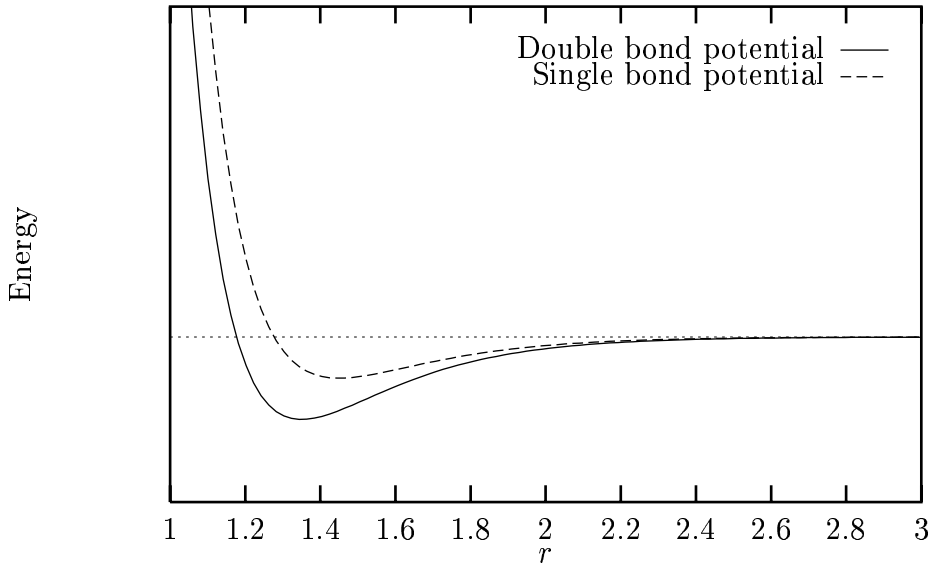


Figure 7: The potential well of the double bond is deeper than that of the single bond, even though the double bond brings the nucleons closer together. (schematic diagram)

The configuration with the lowest energy turns out to consist of alternating single and double bonds (Fig. 8). When long bonds and short bonds alternate, the system is said to be "dimerized."

Bonding by π orbitals creates torsional stiffness, because neighboring p_z orbitals overlap the most—and lower the total energy the most—when they are parallel, so that twisting away from the parallel orientation would cause the energy to increase. When they are perpendicular, the overlap is negligible.

The bend in 11-*cis*-retinal brings the methyl group at carbon 13 in close contact with the

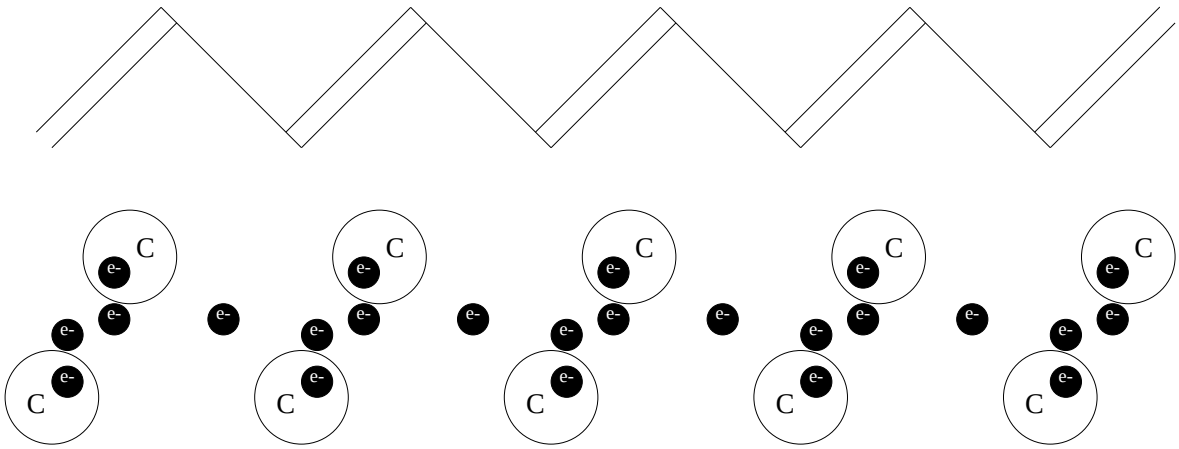


Figure 8: The polyacetylene molecule, in chemical and schematic notation. The ground state of a chain of carbons has alternating single and double bonds. Double bonds are physically shorter than single bonds. In a double bond, electrons hop back and forth between the carbons more often than in a single bond.

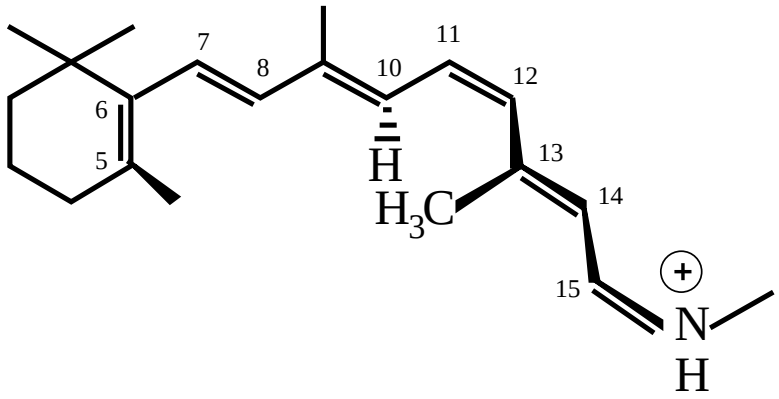


Figure 9: The 11-*cis* configuration, showing the close proximity of the hydrogen at carbon 10 and the methyl at carbon 13.

hydrogen from carbon 10 (Fig. 9). The interaction potential between the methyl and the hydrogen is a strong antibonding one; the particles repulse one another other. The 11–12 bond is double, however, and therefore torsionally stiff, so the molecule is stuck in the bent formation even though it could avoid the methyl–hydrogen discomfort and have a lower total energy if it assumed the all-*trans* form. Experimental analysis of retinal has found a twist of about 10° around the 11–12 bond in the ground state of 11-*cis*-retinal (Aalberts 1997).

1.4 Behavior

When 11-*cis* retinal absorbs a photon, it will isomerize to the all-*trans* form, with a quantum yield of 67%. If 100 photons are absorbed by rhodopsin proteins, 67 all-*trans* retinal molecules will be formed. The transformation of rhodopsin, which holds 11-*cis* retinal, into an intermediate bathorhodopsin, a form of rhodopsin in which retinal's 11 bond is on its way to becoming *trans*, takes place in 200 fs (Wang 1994). This is the fastest known photochemical reaction (Vos 1996).

Retinal's efficiency and speed could be related. Experiments have shown that the motion of 11-*cis* retinal molecules in solution is vibrationally coherent during the 2 ps after the impact of a 35 fs pulse of light. This result suggests that the transition is "barrierless," because if a barrier inhibited a photoexcited retinal molecule from isomerizing, then the unisomerized molecules, those that could not pass the energy barrier, would exhibit different modes than the isomerized molecules that did, and the vibrational modes would not be observed to be coherent.

Traditionally, photoisomerization was thought to occur when a photon created an electronic particle-hole pair in retinal, bringing the molecule to an excited state, where it would vibrate and critically damp out, riding its energy surface down to the ground state of the all-*trans* "photoproduct" through an energy transition known as "internal conversion." However, such a scenario would not exhibit vibrational coherence. Furthermore, the transformation occurs so quickly it is unlikely the vibrations have time to relax. The vibrational coherence suggests that a nonstationary wave packet travels directly from the 11-cis excited state to the all-*trans* ground state, through Landau-Zener tunneling process (Fig. 10) whose probability (P_{LZ}) depends only on the velocity of the nuclei:

$$P_{LZ} \propto \exp -\frac{1}{v}$$

In such a process, then, quantum yield would depend primarily on nuclear velocity (v), which is related to the speed of the transformation. There is supporting evidence for this

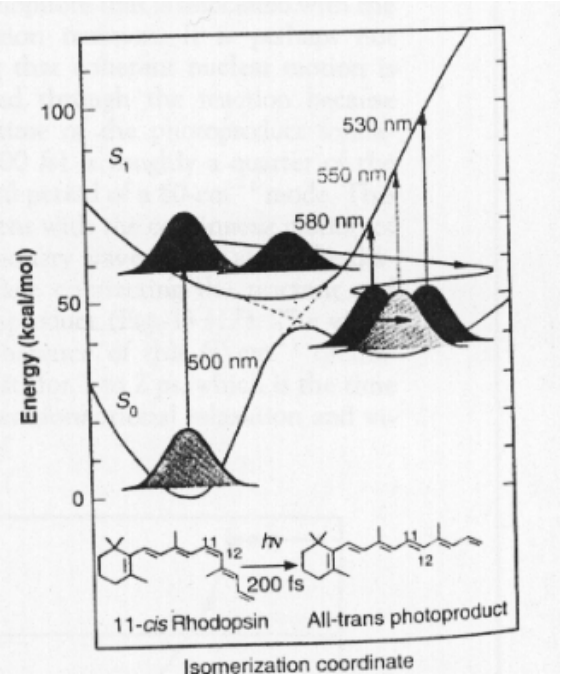


Figure 10: The photoproduct is vibrationally coherent because the excited state tunnels into the all-*trans* form of the ground state. The dashed lines show the tunneling path (Wang 1994).

relationship between speed and efficiency in the case of isorhodopsin, which has a slower isomerization speed (600 fs) and a lower quantum yield (22%) (Wang 424).

It is unknown exactly how this transformation proceeds. However, it is likely that the absorption of the photon reduces the torsional stiffness in the 11–12 bond, allowing the methyl–hydrogen repulsion to twist and unbend 11-*cis*-retinal into all-*trans*-retinal. To investigate this process, we first need a theoretical model of the dimerized carbon chain.

2 Modeling

2.1 Polyacetylene as a model of retinal

The backbone of retinal is much like the molecule polyacetylene, $(CH)_x$. Polyacetylene is a simple molecule which can be synthesized with any number of CH groups. It has been studied experimentally and theoretically for decades, so it is fairly well understood. Therefore,

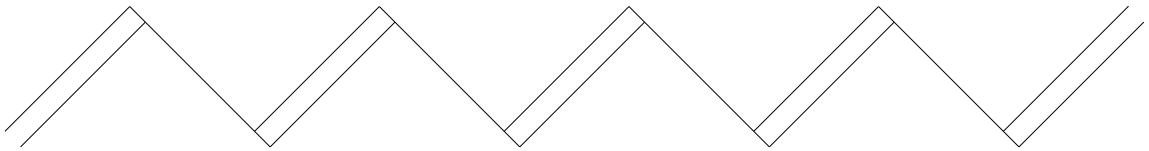


Figure 11: Polyacetylene, CH_x, is a well-studied that is similar to the backbone of retinal.

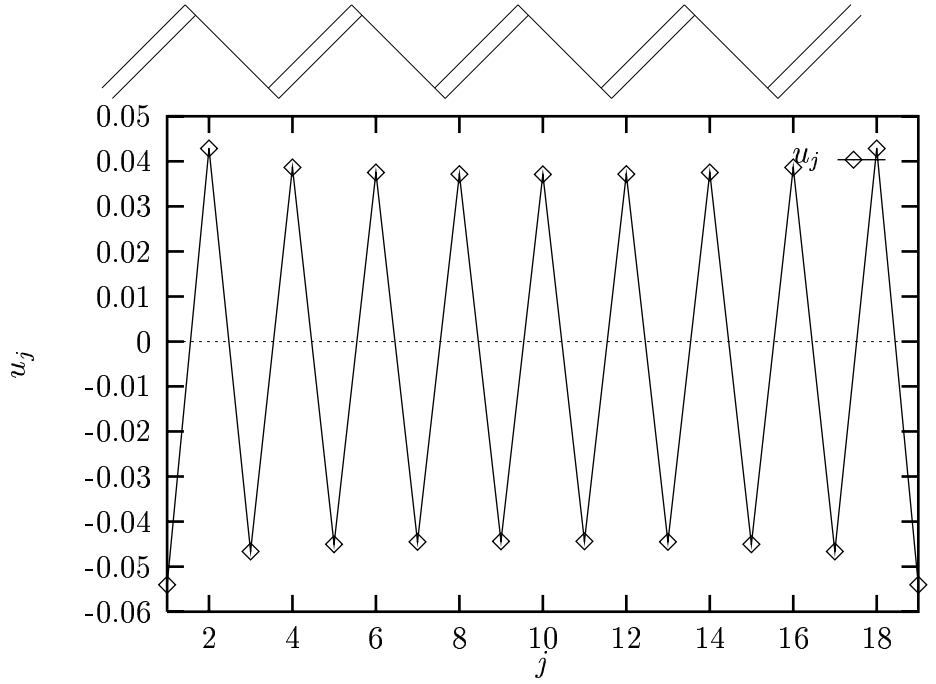
conclusions about the structure and dynamics of polyacetylene may lead to understanding of retinal. The validity of such an analogy is strengthened by observations that retinal's ring and its lysine tail end are stationary during photoisomerization.

Retinal has significant differences from polyacetylene with its two methyl groups in place of hydrogens off of carbons 9 and 13; however, these differences are manifest primarily in their effect on the classical moments of inertia, and not on the electronic structure of the molecule, since all the valence electons on the methyl's carbon are tied up in σ bonds, within an sp^3 hybrid orbital, rather than an sp^2p_z orbital that would affect the electronics of the backbone.

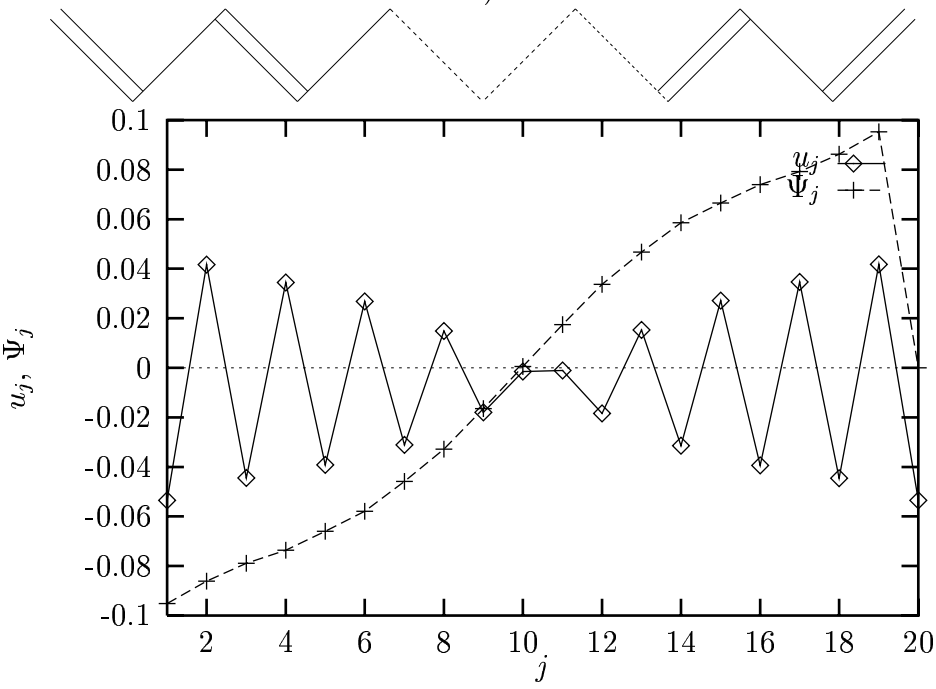
Studies of polyacetylene have shown that the dynamics of carbon chains can be explained not as a disorderly mess of phonons, but as a zoology of species of stable, self-localized non-linear excitations in bond-length patterns which propogate throughout the chain on timescales much longer than typical phonon modes. Topical excitations are important for the study of retinal's photoisomerization because they demonstrate the effect of added energy, as from a photon, on the bond lengths. These well-studied topical excitations include "solitons," "polarons," and "breathers."

2.2 Solitons, polarons, and breathers

Solitons arise spontaneously in the ground state of some polyacetylene chains. It turns out that the lowest energy configuration of all finite chains has double bonds at both ends of the chain. If the chain has an odd number of bonds, the bond alternation pattern is an unambiguous pattern of alternating double and single bonds: $u_n = (-1)^n u_0$ (Fig. 12a) where u_j is the difference between the bond length $|\mathbf{r}_{j+1} - \mathbf{r}_j|$ and the proper bond length



a)



b)

Figure 12: Whether bond number is odd or even, double bonds form at the two ends of the chain. a) When the number of bonds is odd, the bond alternation is consistent throughout the chain. b) When the number of bonds is even, the bond alternation pattern must shift from double-single to single-double in order to have double bonds at both ends of the chain. The spot at the middle, where $u_j \approx 0$, is called a soliton. The graph Ψ_j crosses zero at the middle, indicating with a linear approximation the precise location of the soliton.

a_0:

$$u_j = |\mathbf{r}_{j+1} - \mathbf{r}_j| - a_0$$

If the chain has an even number of bonds, this alternation pattern would give a single bond at the end. But the ground state of an odd-bond chain actually forms in such a way that there is a gradual transition from a pattern of $u_n = (-1)^n u_0$ for low n to a pattern of $u_n = (-1)^{n+1} u_0$ for high n, so that double bonds appear on both ends of the chain. Toward the middle region, the alternation magnitude gradually decreases to zero, and in the very middle of the chain the alternation magnitude is zero. A soliton is said to exist at that point (Fig. 12b).

Like solitary waves, solitons have a width. The pattern of $|u_j|$ fits an inverted Gaussian curve, with a width proportional to the length of the chain. (In a chain as short as retinal, that width is around $2a$; in a chain of 100, the width is around $10a$). The pattern of $\Psi_j = (-1)^j u_j$ fits a $\pm b \tanh(c(j' - j'_0))$ curve, where b and c are constants, and j' is a continuous variable (Fig. 12b). The soliton can be pinpointed in continuous space at the location j_0 where the tanh curve crosses zero. In these simulations, Ψ_j is defined in relative terms, as $\Psi_j = (-1)^j(u_j + 1 - u_j)$. Under this definition Ψ_j is a smooth function, and it exposes the solitons inherent in formations which distort the proper average bond length.

The solitons are interesting because they behave like particles. As time evolves, the motion of the solitons is continuous. Solitons will "bounce" of the boundaries of chains, and can annihilate with one another.

These so-called "solitons" qualify as the type of solitons that physicists and mathematicians know because they arise from nonlinear phenomena, exhibit particle-like motion, and do not disperse. However, they do not move through one another the way proper solitons do. They carry charge (but can also be neutral), mass (they are quite light—just a few electron masses), and spin (charged solitons have spin 0, while neutral ones have spin 1/2) (Su 1980; Heeger 1988). They are perhaps better known as "kinks."

Solitons appear in even-bond chains, but can arise from other means. For instance, if

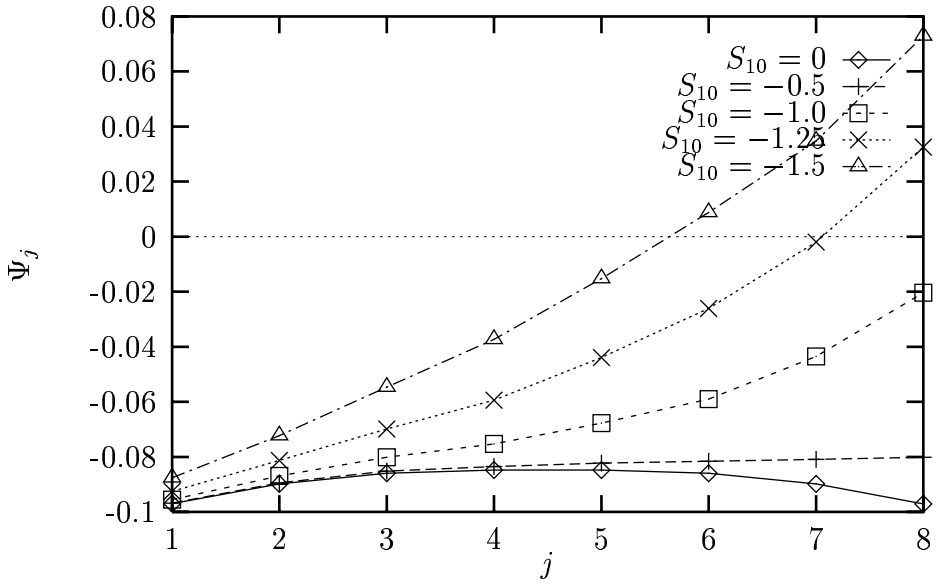


Figure 13: In an odd-bond chain, the dimerization pattern is perfect. But as a negative charge is added, the dimerization pattern on the right hand side is reversed, and a soliton gradually moves in from the right. Solitons exist at the roots of the the Ψ_j function. (The simulation algorithm will be explained in section 2.6.)

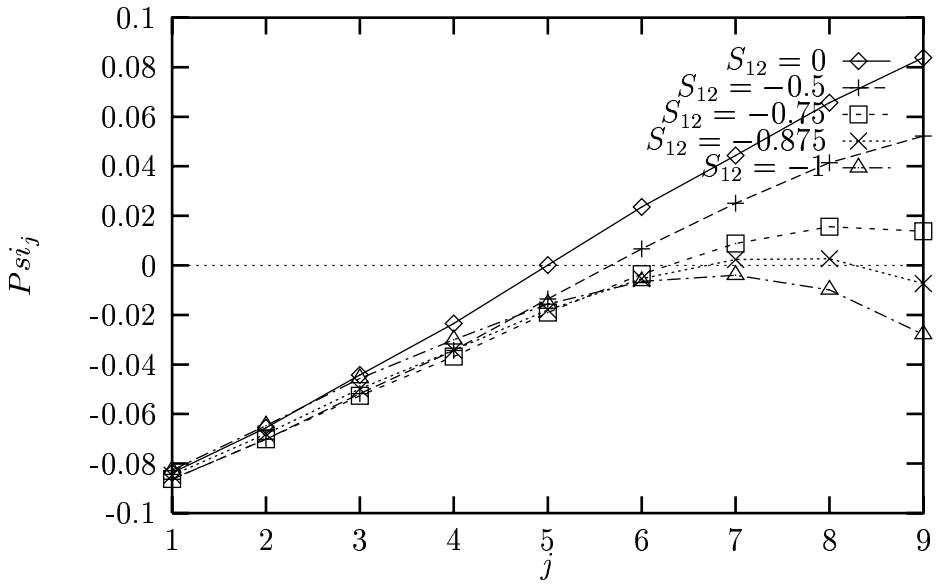


Figure 14: In a chain of even bond number, the stationary soliton in the middle of the chain can be removed by the addition of a negative charge a the right hand end of the chain.

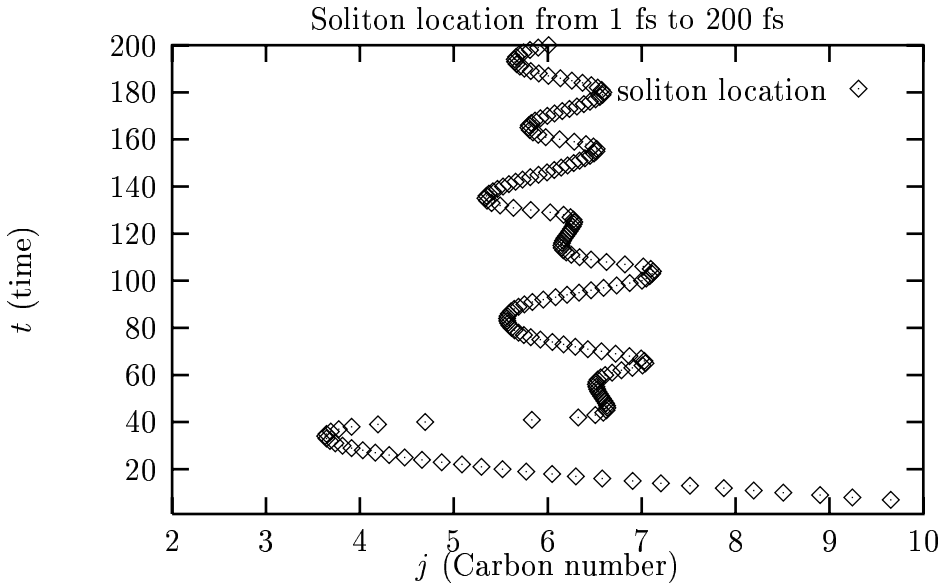


Figure 15: At time zero, the 11 bonds were arranged as if a 12th double bond existed. When system was allowed to evolve, a soliton appeared at the end, moved toward the left, and gradually converged around the middle.

a strong positive charge is placed near a chain, it will increase the electron charge density nearby, creating an electron shield that will repel other electrons, thereby pulling up charge density at a site where it was low, and pulling down charge density at the neighboring site where it was high. This causes a soliton to emerge at the end of the chain, and move in more the greater the charge (Fig. 13).

Conversely, if a soliton already exists at the center, as in an even-bond chain, increasing the charge at the end will "erase" the soliton (Fig. 14). If the final carbon in an odd-bond chain is instantaneously plucked off, a soliton will form at that endpoint, and travel toward the other end, "bounce" off of that end, and then make its way toward the center (Fig. 15).

Or, if an electron is moved from the valence band to the conduction band by photon absorption, then a soliton and an "antisoliton," a pair of oppositely charged solitons, form at the center and move away from each other toward the ends, bounce off the walls, and cancel each other out (Fig. 16). The energy of the solitons is not lost except through damping. Therefore, even if solitons cancel each other out, and the Ψ_j function no longer crosses zero, their energy is still stored in the motion of the carbons. Therefore, it is likely

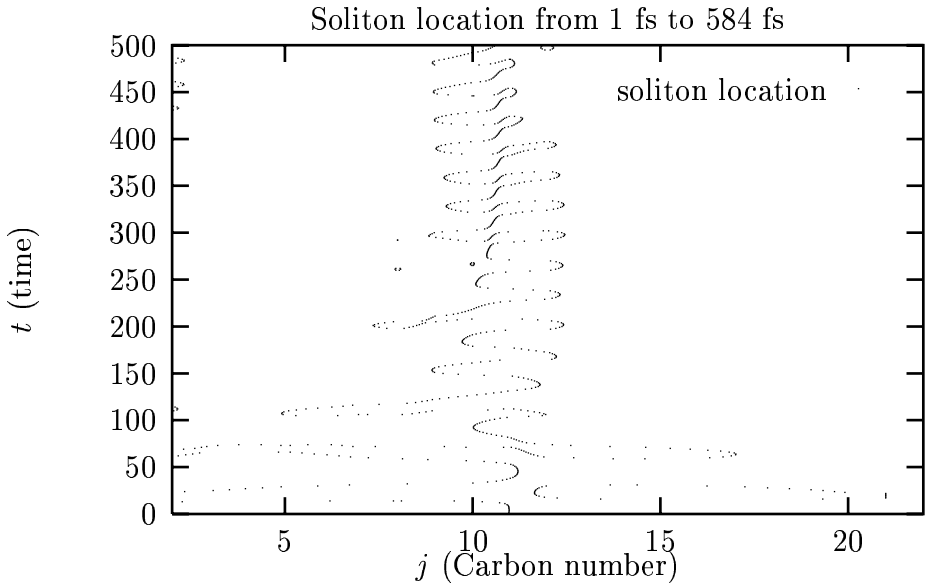


Figure 16: An even-bond system was in its ground state when it was hit with a photon at time $t = 0$, introducing energy that resulted in the creation and subsequent annihilation of soliton-antisoliton pairs.

that soliton-antisoliton creation follows soon after annihilation (Fig. 17).

Other forms of excitation have been named. The polaron, for example, is like a soliton-antisoliton pair in which Ψ_j never crosses zero (Fig 19). When a soliton annihilates with an antisoliton, it "turns into" a polaron. Of course, such an annihilation is really a soliton's Ψ_j function moving away from the x axis, causing the zero crossings to become closer together; the polaron's creation is merely the opposite (Guinea 1984; Heeger 1988).

If the energy added to the system with the promotion of electron is greater than the energy required to create a soliton-antisoliton pair, then the extra energy might be stored in a "breather," which is defined as a "spatially localized dynamically bound state" (Heeger 1988) (Fig. 18). Breathers arise in the wake of solitons. Solitons have a certain threshold velocity (which is near the "speed of sound" in the center of mass frame), below which they are bound together, and above which they shed energy, leaving behind a breather that is stationary (in the same frame). (Guinea 1984; Su 1980).

It is likely that retinal, like polyacetylene, exhibits localized excitations such as solitons, polarons, and breathers. Such excitations in the bond lengths, fueled by the energy of an

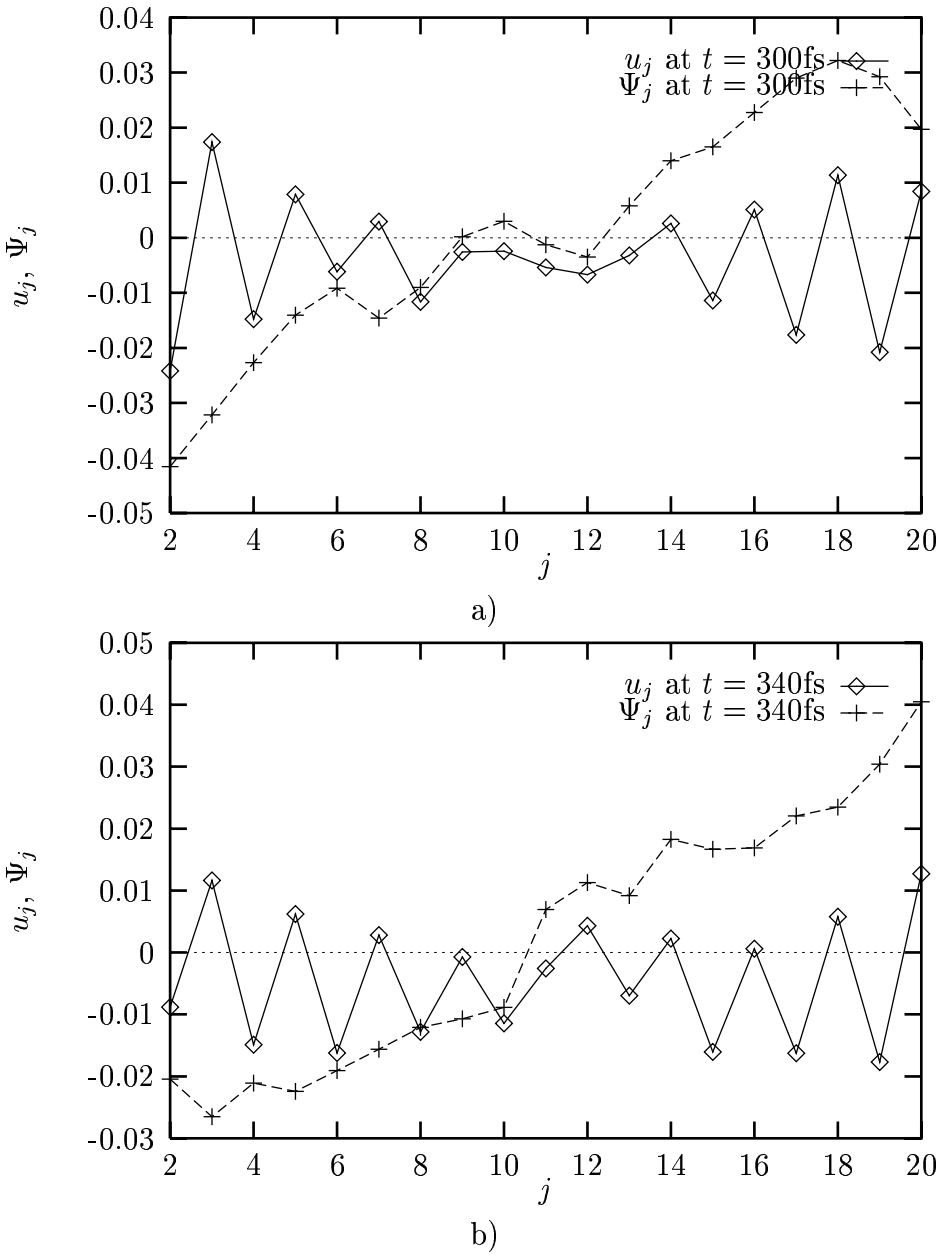


Figure 17: a) At time $t = 300$fs, the system in Figure 16 has a soliton/anti-soliton pair in addition to the original soliton. b) Just 40fs later, the soliton/anti-soliton pair has annihilated, leaving the original soliton.

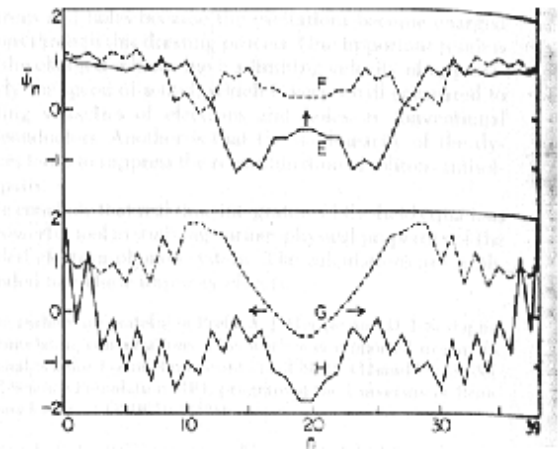


Figure 18: A breather in polyacetylene, formed in the wake of a soliton/anti-soliton pair (Su 1980).

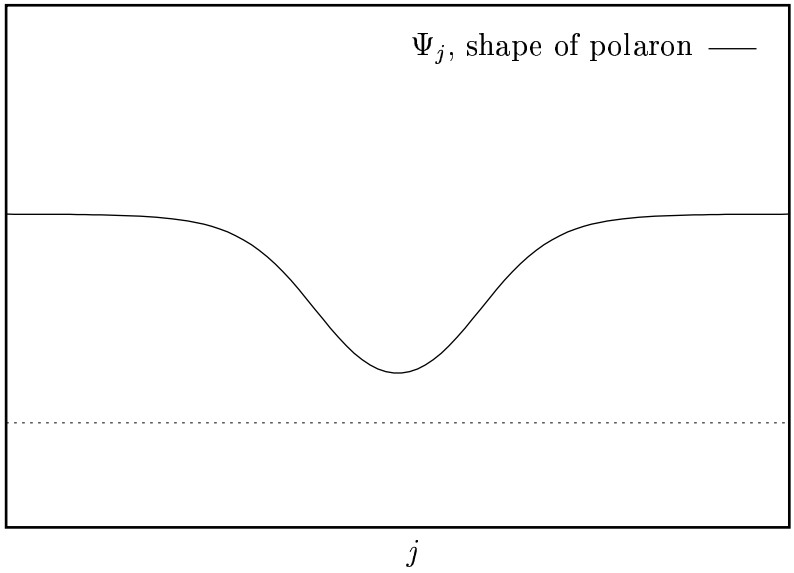


Figure 19: The polaron is a "dip" in the Ψ_j surface. Like a soliton-antisoliton pair, it holds phonon energy. This figure shows a theoretical polaron.

absorbed photon, may cause the strong double bond at the 11–12 bond to change into a weak single bond, allowing the steric repulsion between the methyl and the hydrogen to force 11-*cis*-retinal into the all-*trans* isomer.

2.3 The Su, Schrieffer, and Heeger analytic model of polyacetylene (SSH)

It is possible to model retinal and its behaviors, since there are empirical data with which a model can be verified. *Ab initio* calculations give a good picture of retinal's ground state that comport with experimental evidence (Vreven 1997). Using NMR spectroscopy on retinal in rhodopsin, experimentalists can measure a carbon's proximity to other carbons using an NMR "magic angle spinning" technique. Using NMR spectroscopy with ^{13}C labeling, they can measure the charge density at each carbon in retinal, within the rhodopsin protein (Aalberts 1997). Retinal has been crystalized and its bond lengths measured in a frozen state (Gilardi 1972; Hamanaka 1972). Retinal can be synthesized from from ground-up carrots and placed in solution, so its photochemical properties can be analyzed (Wang 1994).

Unfortunately, the empirical evidence is limited. *Ab initio* calculations give good predictions of the ground state, but cannot calculate the dynamics of the full retinal molecule in the photoexcitated state. Results from experiments on crystalline retinal or retinal in solution are not immediately transferable to retinal within rhodopsin. Crystallization deforms all-*trans* retinal so that it is rotated around the 10–11 bond in addition to the rotation around the 11–12 bond that makes 11-*cis* retinal (Gilardi 1972).

The predominant model for the dynamics of polyacetylene, developed by Su, Schrieffer, and Heeger, is transferable to retinal. Their model ("SSH") models the σ bonds and the 120° bond separation with classical springs, and the electronic with the "tight-binding model." (Ashcroft 1976).

The model Hamiltonian is not deduced from first principles, but rather constructed to fit the needs of the computational task. For that reason, it treats the electronic behavior

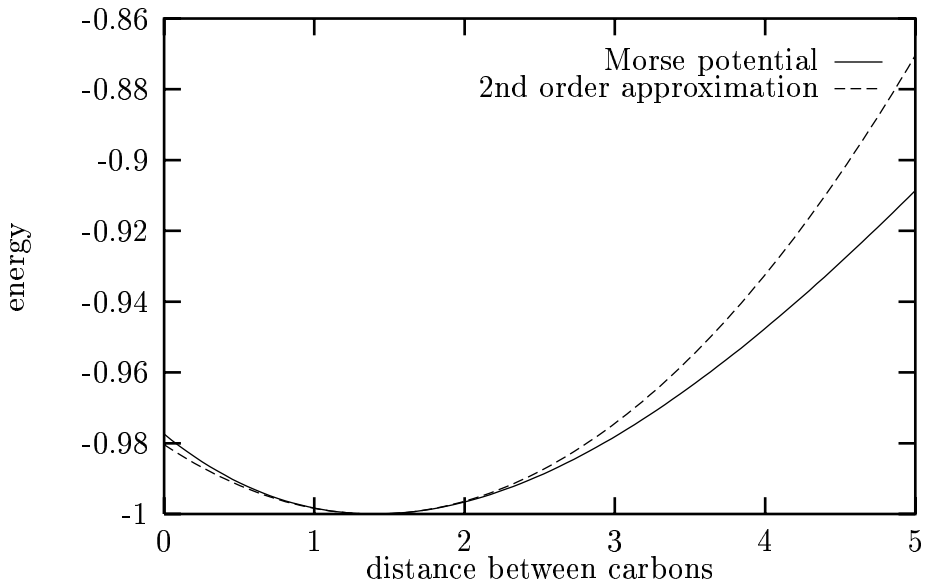


Figure 20: The quadratic approximation is a good approximation of the well of a Morse potential, when changes in distance are small.

discretely in one dimension, using the probability of finding electrons at a "site," one of the carbons along the chain, rather than calculating all of the electron clouds in three dimensions. The model uses a crude spring approximation for the sigma bonds in order to make the setting of parameters more flexible. Nevertheless, it predicts many variables well.

The SSH model treats simply the sp^2 portion of carbon's sp^2p_z hybrid orbital structure. The potential well of carbon-carbon bonding is steeply repulsive at short distances and mildly attractive at distances beyond the average bond length a_0. For small oscillations, however, a u^2 approximation suffices, since the second order approximation of the Morse potential

$$D(e^{-2k(d-d_0)} - 2e^{-k(d-d_0)})$$

is

$$2D(-\frac{1}{2} + \frac{1}{2}k^2(d-d_0)^2)$$

where d is the distance between the particles, and d_0 is the location of the bottom of the well. Since the bond length in polyacetylene has been observed to vary during dimerization by no more than 6% of its average value, the second order approximation is good as long as

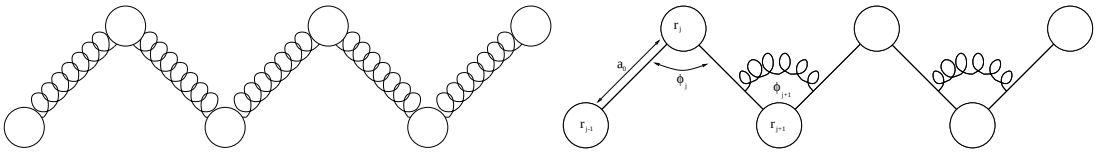


Figure 21: The σ bonds in the SSH model are treated simply as springs.

$k < 0.1$ in the Morse approximation (Fig. 20). Therefore, the σ bonds can be treated as springs (Fig. 21):

$$\hat{H} = \sum_{j=1}^{n-1} \frac{1}{2} K_\sigma u_j^2 + \cdots$$

where $u_j = |\mathbf{r}_{j+1} - \mathbf{r}_j| - a_0$, and the hydrogens are included only implicitly, in the maintenance of 120° angles between the bonds, modeled also with springs:

$$\hat{H} = \cdots + \sum_{j=1}^{n-1} \frac{1}{2} K_{\text{bend}} (\phi_j - 120°)^2 + \cdots$$

where ϕ_j is the angle between the σ bonds connecting carbon j to carbons $j+1$ and $j-1$. By the law of cosines,

$$\phi = \arccos \frac{|\mathbf{r}_j - \mathbf{r}_{j-1}|^2 + |\mathbf{r}_{j+1} - \mathbf{r}_j|^2 - |\mathbf{r}_{j+1} - \mathbf{r}j - 2|^2}{2|\mathbf{r}_j - \mathbf{r}_{j-1}||\mathbf{r}_{j+1} - \mathbf{r}_j|}$$

The average bond length a_0 is the location of the minimum of the σ bonding potential energy surface. Therefore, the purpose of these terms is not only to model the σ bonds, but more generally to ensure that the configuration of the carbons mirrors observations. The constants K_σ and K_{bend} result in calculable normal modes for the system, and so can be chosen to fit available spectroscopic evidence.

The model treats the π electrons in the p_z orbitals with the "tight-binding model." This quantum mechanical model is based on the fact that there is a finite probability that electrons in neighboring p_z orbitals can exchange places with one another. The closer the carbon atoms are together, the more the orbitals overlap, the greater the probability. The uncertainty Δx in the electron's location increases the more the electron moves along the chain, thereby decreasing the uncertainty in its momentum:

$$\Delta x \Delta p \quad = \quad \text{const}$$

$$\Delta p \;\propto\; \Delta x^{-1}$$

Since the electrons are confined to the molecule, however, the expectation value of the momentum is zero, and therefore the momentum is proportional to its uncertainty. Therefore, the electronic energy will decrease the closer together the carbons get, and the more closely parallel the p_z orbitals become.

$$E = \frac{p^2}{2m} \propto \frac{1}{\Delta x^2}$$

Therefore, the molecule can lower its total energy by delocalizing its p_z electrons through hopping.

Since the probability of nearest neighbor hops is much greater than that of next-nearest neighbor hops, we can consider only nearest neighbor hopping. By a theorem of solid state physics known as "Peierl's instability," the configuration in which the carbons are equally spaced is unstable, and electron-phonon coupling causes the ground state bond lengths to alternate:

$$u_n = (-1)^n u_0 \,.$$

Since the index n can be assigned arbitrarily, the physics is invariant under an $n \to n+1$ transformation, so the resulting state is two-fold degenerate (Kittel 1986). Soliton creation collapses this degeneracy.

The physics of Peierl's instability is implemented by the electron-phonon coupling term,

$$\hat{H} = \cdots - t \sum_{j=1,s}^{n-1} (c^\dagger_{j,s} c_{j+1,s} + c^\dagger_{j+1,s} c_{j,s}) + \cdots$$

where the c_{js} are standard creation and annihilation operators. The operator product, $c^\dagger_{j+1} c_j$, acts on a state by destroying an electron at site $j+1$ and creating an electron at site j, and its expectation value is the propensity for an electron to travel from site $j+1$ to site j. In accordance with the Pauli exclusion principle, the $c^\dagger_{j,s}$ operator returns zero when it is applied to a wavefunction in which an electron of spin s already exists at site j, just as the

$c_{j,s}$ returns zero where one does not exist.

$$\begin{aligned} c\,|n\rangle &= \sqrt{n}\,|n-1\rangle \\ c^\dagger\,|n\rangle &= \sqrt{n+1}\,|n+1\rangle \end{aligned}$$

where $|n\rangle$ is a state where the site has n electrons on it. Thus, the expectation value of the electron density operator n_{js} is

$$\begin{aligned} \langle\Psi\,|n_{js}|\,\Psi\rangle &= \left\langle \Psi \left| c^\dagger_{j,s} c_{j,s} \right| \Psi \right\rangle \\ &= \left\langle \Psi \left| c^\dagger_{j,s}\sqrt{n} \right| \Psi'(n' = n_{js} - 1) \right\rangle \\ &= \left\langle \Psi \left| \sqrt{n-1+1}\sqrt{n} \right| \Psi \right\rangle \\ &= n \end{aligned}$$

and returns the expected number of electrons of spin s at site j.

Since the sign of t is always positive, the term has a negative sign. The greater the hopping probability, the more "delocalized" the electrons will be, and the less energetic they will be. The factor t is defined as

$$t = (t_0 - \alpha u_j)\cos\theta$$

The t_0 term sets a constant hopping probability that is increased if the bond is short ($u_j < 0$), or double, and decreased if long, or single. The angle θ refers to the angle between the p_z orbitals on sites j and $j+1$, so that the hopping probability goes to zero if they are oriented perpendicularly to one another. More specifically, θ_j is the angle between the plane formed by points $\mathbf{r}_j$, $\mathbf{r}_{j+1}$, and $\mathbf{r}_{j+2}$, and the plane formed by points $\mathbf{r}_{j-1}$, $\mathbf{r}_j$, and $\mathbf{r}_{j+1}$, where $\mathbf{r}_j$ is the location of carbon j:

$$\cos\theta = \frac{\mathbf{w}_j \times \mathbf{w}_{j+1}}{|\mathbf{w}_j \times \mathbf{w}_{j+1}|}$$

where

$$\mathbf{w}_j = \frac{(\mathbf{r}_j - \mathbf{r}_{j-1}) \times (\mathbf{r}_{j+1} - \mathbf{r}_j)}{|(\mathbf{r}_j - \mathbf{r}_{j-1}) \times (\mathbf{r}_{j+1} - \mathbf{r}_j)|}$$

The $\cos\theta$ factor, which is always positive, will bring the ground state into the correct planar form ($\theta_j = 0$ for all j), and will introduce a force of torsional stiffness to the molecule, by creating an energy barrier against the *cis-trans* isomerization.

The α term tends to shrink the bond lengths to an unrealistic extent, requiring an additional term,

$$\hat{H} = \cdots + \sum_{j=1}^{n-1} -\Gamma u_j + \cdots$$

to counterstretch the bonds. The value of Γ is set so that the average bond length is a_0.

We can model the effect of the Peierl's distortion on the energy bands. Neglecting the $cos\theta$ contribution, the electron-phonon coupling is

$$H(u) = -\sum_{n,s}[t_0 + (-1)^n 2\alpha u](c^\dagger_{n+1,s}c_{n,s} + c^\dagger_{n,s}c_{n+1,s}) + 2NKu^2$$

which can be diagonalized by the Bloch operators

$$c_{ks-} = (N^{-1/2})\sum_{ns}\exp(-ikna)c_{ns}$$

$$c_{ks+} = -i(N^{-1/2})\sum_{ns}\exp(-ikna)(-1)^n c_{ns}$$

for valence and conduction bands, respectively.

The Hamiltonian can then be rewritten

$$H(u) = \sum[\epsilon_k(c^\dagger_{ks+}c_{ks+} - c^\dagger_{ks-}c_{ks-}) + \Delta_k(c^\dagger_{ks+}c_{ks-} - c^\dagger_{ks-}c_{ks+}) + 2NKu^2]$$

where $\Delta_k = 4\alpha u_0 \sin(ka)$ and $\epsilon_k = 2t_0\cos(ka)$.

Then, by making the transformation

$$a_{ks-} = \alpha_k c_{ks-} - \beta_k c_{ks+}$$

$$a_{ks+} = \alpha_k c_{ks-} + \beta_k c_{ks+}$$

and requiring

$$|\alpha|^2 + |\beta_k|^2 = 1$$

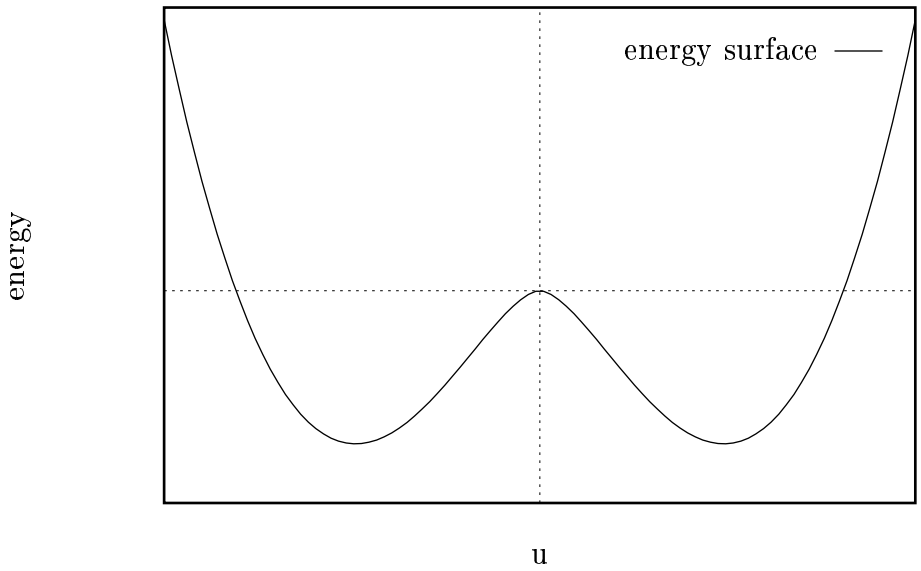


Figure 22: The energy surface created by the tight-binding model creates an unstable equilibrium at $u = 0$, pushing the u_j to the bottom of the wells, which are located at $u = \pm u_0$. This is Peierl's instability.

to satisfy Fermi anticommutation relations, and by eliminating terms that cross + and − operators, we get

$$H = \sum E_k(n_{ks+} - nks-) + 2NKu^2$$

where $n_{ks+} = c^\dagger_{ks+}c_{ks+}$ and $n_{ks-} = c^\dagger_{ks-}c_{ks-}$ are the number of conduction and valence electrons, respectively, and

$$E_k = \sqrt{\epsilon_k^2 + \Delta_k^2}\,.$$

Now, when $u_0 = 0$, the conduction and valence bands intersect, so that an electron can move from one to another. However, a non-zero u_0 introduces a gap between the two bands.

The ground state energy can calculated:

$$E_0(u)/N = -4t_0/\pi - (2t_0/\pi)[\ln(4/|z|) - 1/2]z^2 + Kt_0^2z^2/2\alpha^2$$

where $x = 2\alpha u/t_0$ (Fig 22). The ground state energy thus has a maximum at $u = 0$, and minima at $u = \pm u_0$ (Su 1980).

The tight-binding model assumes that hopping probability is linear with u_j. But this is clearly false: hopping probability depends on the overlap integral of two electron clouds,

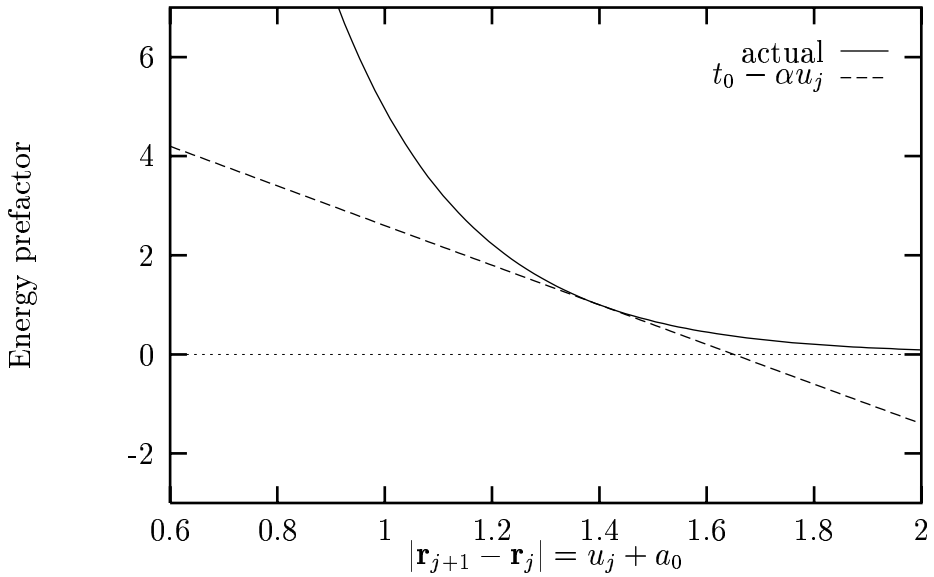


Figure 23: The linear approximation to the hopping prefactor is good for $t_0 = 1$, $\alpha = 4.0$, $u_0 < 0.05$. In this case, $a_0 = 1.4$.

and since electron clouds decay exponentially with distance, so should hopping probability. In fact, the tight-binding prefactor, $t_0 - \alpha u_j$, is merely a first-order approximation to a more realistic exponential function, $t_0 \exp(-\frac{\alpha}{t_0} u_j)$ (Fig. 23). With typical values $t_0 = 1.0$, $\alpha = 4.0$, and a variation $\Delta u_j = \pm 0.05$, there is a reasonable 2.25% maximum deviation from the exponential. The linear approximation works especially well at the value $t_0 = 2.5$ initially used by Su, Schrieffer, and Heeger (error of 0.25%). However, in our studies we lowered t_0 to as low as 0.4, for which the error reaches an unreasonable 17.5%. Perhaps this is one place where the model could be improved.

The model is crude in other ways. The force of planarity does not belong exclusively in the tight-binding term, since the planarity of the molecule depends on nuclear repulsion as well as electronic preference. The Γ term appears completely arbitrary. It represents nothing physical, but only acts as a corrective to the tight-binding force. However, models are positivist approximations of reality, and their terms need not always represent identifiable physical phenomena.

This limited set of terms was chosen by Su, Schrieffer, and Heeger in the early 1980's in

part so that the full quantum mechanical calculations could be carried out in a reasonable amount of time. By necessity they applied the mean-field approximation, because including coulomb interactions would have complicated the model impractically. They devoted much language to justifying their disregard of interactions (Heeger 1988).

The SSH model can be extended with the "Hubbard model," which models same-site Coulomb interactions simply:

$$\hat{H} = \cdots + \sum_{j=1}^{n} U_j n_{j,\uparrow} n_{j,\downarrow} + \cdots$$

where $n_{j,s} = c_{j,s}^{\dagger} c_{j,s}$. This term increases the energy if electrons of different spin congregate around the same carbon atom, creating a negative incentive for the formation of doubly occupied ionic levels. The positive coefficient U_j is a constant which is usually treated as the same for all carbon atoms, but which may be different on other types of atoms. When U_j is the same for all j, the charge density remains uniform across j.

The "extended Hubbard model" includes a term for intersite coulomb repulsion:

$$\hat{H} = \cdots + \sum_{j=1}^{n-1} V_j (n_{j,\uparrow} + n_{j,\downarrow})(n_{j+1,\uparrow} + n_{j+1,\downarrow}) + \cdots$$

The term favors an electronic configuration in which electrons congregate at every other site. The positive coefficients V_j, like the U_j, are usually the same for all j. Nonzero V_j values across j will dimerize the charge density ρ_j, alternately reducing and increasing electron density from site to site (Fig. 24).

To model the individual characteristics of molecules like retinal, terms creating specific charge densities can be added:

$$\hat{H} = \cdots + \sum_{j=1,s}^{n} S_j n_{j,s} + \cdots$$

A negative S_j models a buildup of electron density at site j, and a positive S_j will repel electrons from that site. To model a negative ion near the chain, one can compute the potential W_j due to that charge at each of the carbon locations $\mathbf{r}_j$, and then set $S_j = eW_j$.

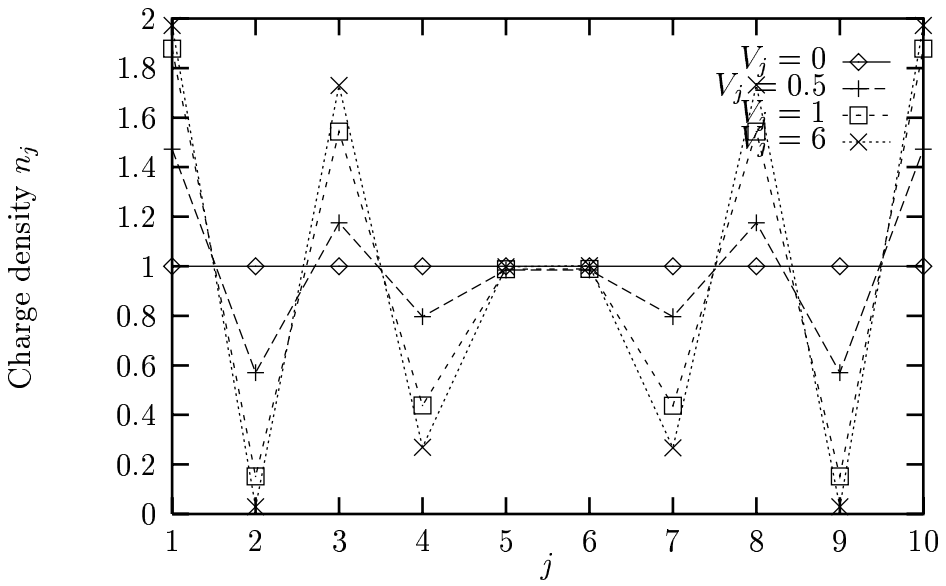


Figure 24: When the V_j are uniformly increased, they introduce an alternation pattern into the charge density pattern.

At close distances, unbonded atoms repel one another with a "steric force." This close-range force can be modeled with a Lennard-Jones potential:

$$\hat{H} = \cdots + \frac{A_{\text{steric}}}{r^{12}} + \cdots$$

The precise magnitude of the steric repulsion is not well known, so A_{steric} can be chosen so that the relative atomic coordinates agree with experiment (Haile 1992).

For better accuracy, the steric repulsion can be modeled with the antibonding term of a Morse potential, $De^{-2a(r-r_0)}$, where D and a are constants, r is the distance of the interaction, and r_0 is the natural distance between the two atoms (Morse 1929).

2.4 Implications of the Hubbard model

The Hubbard model does not offer easy analytical solutions to systems with more than two sites. But the two-site case illustrates how the Hubbard model affects energy eigenvalues (Ashcroft 1976). For example, consider the problem of one electron which can be bound to one of two protons, or sites. There are two electronic orbital levels, $|R\rangle$ and $|R'\rangle$, one for each proton, which are taken to be orthogonal, $\langle R|R'\rangle = 0$. If the two protons are far apart,

there will be a hydrogen atom and a proton. If they are close together, there will be H_2^+, in which the electron has a probability of being in the orbital of each proton. The tight-binding model can be expressed in terms that are off-diagonal in $|R\rangle$ and $|R'\rangle$:

$$\langle R\,|H|\,R\rangle' = \langle R'\,|H|\,R\rangle = -t$$

Meanwhile, we know that the binding energy of an electron bound to a proton E_e, so that

$$\langle R'\,|H|\,R'\rangle = \langle R\,|H|\,R\rangle = E_e$$

Therefore, in $|R\rangle$ and $|R'\rangle$ space, the Hamiltonian is

$$\hat{H} = \begin{bmatrix} E_e & -t \\ -t & E_e \end{bmatrix}$$

with eigenvalues $E_e \pm t$ and corresponding eigenvalues

$$\frac{1}{\sqrt{2}}(|R\rangle \mp |R'\rangle)$$

Thus, the introduction of the hopping parameter splits the energy spectrum in two.

To extend this system to two electrons, the Hartree approximation (Ashcroft 1976), which treats electrons as independent, can be applied to the Hubbard model. The ground state energy of the one-electron system is $E_e - t$, as shown. If the two electrons could inhabit the same orbital without interaction, the ground state energy of the two-electron system would be simply twice that energy, or $2(E_e - t)$. We can introduce interaction into our model as a perturbation to the Hartree approximation, by assuming that there is an additive energy cost of U whenever two electrons inhabit the same site. The total energy is then $2(E_e - t) + pU$, where p is the probability of two electrons inhabiting the same site, and since there are four ways to distribute two pebbles in two buckets, two of which involve both pebbles in the same bucket, the probability p is $1/2$. (The basis states are $|RR\rangle$, $|RR'\rangle$, $|R'R\rangle$, $|R'R'\rangle$.) Therefore, the ground state energy within the Hartree approximation is $E = 2(E_e - t) + \frac{1}{2}U$ where U is the coefficient for the same-site interaction energy term.

But electrons are indistinguishable, so the full set of states of the two site, two electron system properly has a basis set of three states, not four: two electrons on site 1 (Φ_1), two electrons on site 2 (Φ_2), and one electron on each site (Φ_0). The Hamiltonian in the (Φ_0, Φ_1, Φ_2) basis is

$$\hat{H} = \begin{bmatrix} 2E_e & -\sqrt{2}t & -\sqrt{2}t \\ -\sqrt{2}t & 2E_e + U & 0 \\ -\sqrt{2}t & 0 & 2E_e + U \end{bmatrix}$$

This Hamiltonian shows that the energy is just $2E_e$ when there is one electron on each site, and that the energy goes up by U when both electrons are on the same site. The matrix element is zero for transitions between the two doubly occupied states, because a probability amplitude for a simultaneous two electron hopping is not defined in this model. The matrix element for a hopping between Φ_0 and a doubly occupied state is $-\sqrt{2}t$ rather than $-t$ because it is the probability amplitude sum of two equally possible events, both of which have amplitude $-t$. For example:

$$\begin{aligned} \langle \Phi_0 | \Phi_1 \rangle &= \frac{1}{\sqrt{2}} \left(\langle RR' | + \langle R'R | \right) \hat{H} \, | RR \rangle \\ &= \frac{1}{\sqrt{2}} \left\langle RR' \left| \hat{H} \right| RR \right\rangle + \frac{1}{\sqrt{2}} \left\langle R'R \left| \hat{H} \right| RR \right\rangle \\ &= -\frac{2}{\sqrt{2}} t = -\sqrt{2}t \end{aligned}$$

The ground state of this system is proportional to

$$\frac{1}{\sqrt{2}}\Phi_0 + \left(\sqrt{1 + \left(\frac{U}{4t} \right)^2} - \frac{U}{4t} \right) \frac{1}{2}(\Phi_1 + \Phi_2)$$

with eigenvalue

$$E = 2E_e + \frac{1}{2}U - \sqrt{4t^2 + \frac{1}{4}U^2}$$

When U becomes greater than about $2t$, the effect of the interactions on the energy becomes small, and the energy tends toward $2E_e$ (Fig. 25). In this basis, the probability of finding two electrons on the same site is found to decrease exponentially with U. The "Heitler-London approximation," which takes the ground state as Φ_0, where each site is single occupied, is

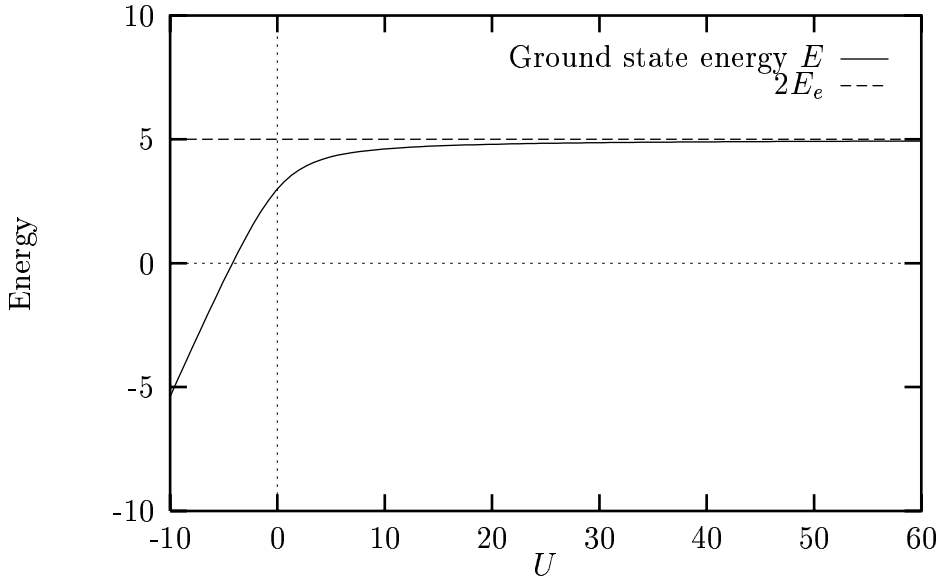


Figure 25: Analytic solution of energy as a function of U in the two-site, two-electron problem, where $E_e = 2.5$ and $t = 1$. Note that the energy stays near $2E_e$ when U is greater than $\sim 2t$.

thus correct for large U/t. However, in systems that undergo Peierl's distortion, the mixing of the Φ_0 with the Φ_1 and Φ_2 states becomes important, and such an approximation is invalid.

Su, Schrieffer, and Heeger argued at length that U is negligible in polyacetylene, and achieved good agreement with experiment without taking electron interactions into account. However, they underestimated the energy gap between the ground and the photoexcited state. Although errors in energy could be due to a number of factors, it is possible that their explanation for the insignificance of interactions was flawed. The analytic two-site, two-electron case suggests that the energy gap would be increased if U were taken into account. We have demonstrated in our own calculations that U causes the energy gap to increase (Fig. 27).

In the multi-carbon chain, therefore, increasing U should tend to cut down on electron traffic from site to site. On average, of course, there is only one electron per site, but this means that some of the time there are no electrons there, and at other times there are two electrons there. For the electrons to delocalize, they need to be able to cohabit the same site,

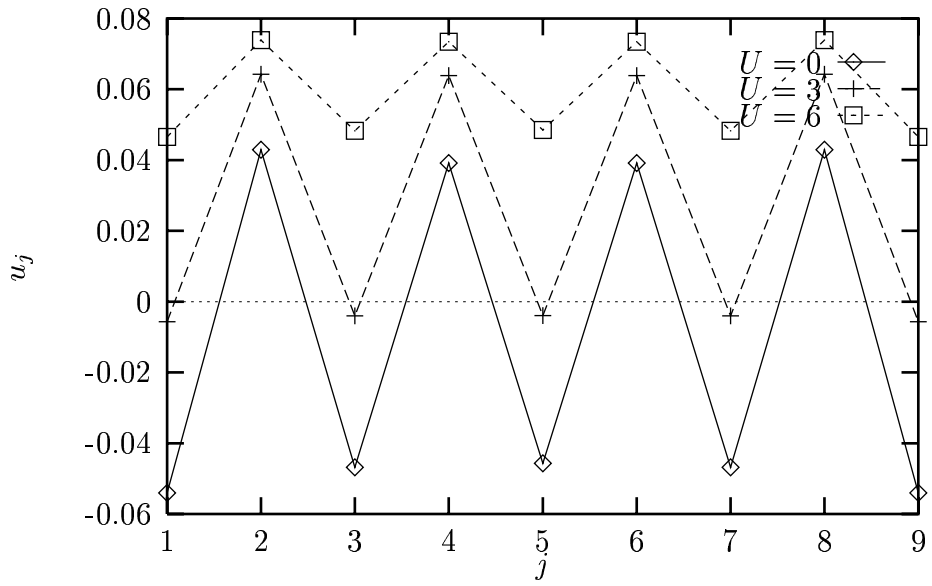


Figure 26: In polyacetylene, As U increases the single bonds and the double bonds enlongate, because the Coulomb interaction within each site inhibits hopping.

since electron traffic without occasional cohabitations, in a system of n sites and n electrons, is possible only in a continuous ring of sites. Electrons will be less likely to hop to places where they will face the energy cost of cohabitation. Therefore, the intrasite interaction potential U should have the effect of weakening all of the π bonding. Single bonds, which have some π character, should elongate somewhat, and double bonds should elongate even more. This was verified in simulation (Fig. 26).

The introduction of interactions thus has a major effect on the SSH tight-binding model. If U is increased to a non-negligible value, then Γ must be reduced—all the bonds must be shrunk in order that the bond lengths oscillate around the observed value a_0. The variable α must also be increased, so that the value u_0, where $u_j = (-1)^n u_0$ is brought to its proper value. After these adjustments have been performed, the quantum component of the energy eigenvalue is greater, although the tight-binding model works as before. Thus, adjusting the variable U allows for more flexible modeling of the energy.

The site-to-site interaction potential V causes alternations in the charge density pattern, causing electrons to bunch up every other site, thereby counteracting U's effect of distributing

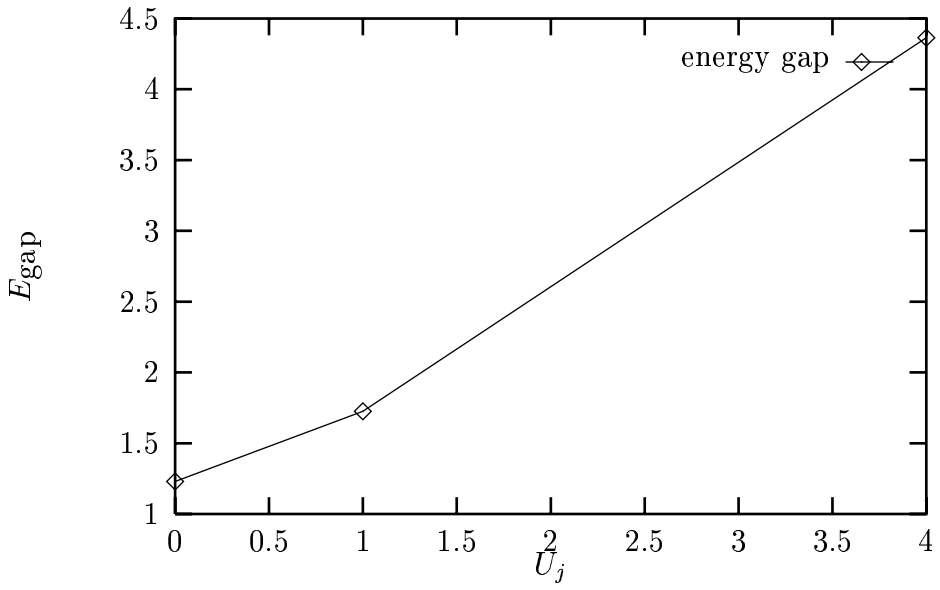


Figure 27: Energy as a function of U, in a polyacetylene chain of 10 carbons and nine bonds, where α and Γ were adjusted to yield the proper dimerization pattern.

the electrons equally across the sites. It increases electron density around double bonds, and decreases it for single bonds. In practice, the electrons tend to congregate at one site, in the middle of the chain, because the energy due to the V term of a chain with consistent dimerization is more than that of a chain with a single short bond at the middle.

However, the two-site model does not consider the spin of the electrons. The two-site model with one up electron and one down electron is more appropriate to studying conjugated chains, where there is an average of one electron per site, and where the population of electrons can be assumed to contain up and down electrons in equal numbers.

When spin matters, the two site model has four possible states:

	site 1	site 2
1	$\uparrow\downarrow$	
2	$\uparrow$	$\downarrow$
3	$\downarrow$	$\uparrow$
4		$\uparrow\downarrow$

In a basis with this ordering, The Hamiltonian thus becomes

$$\hat{H} = \begin{bmatrix} U & -t & -t & 0 \\ -t & V & 0 & -t \\ -t & 0 & V & -t \\ 0 & -t & -t & U \end{bmatrix}$$

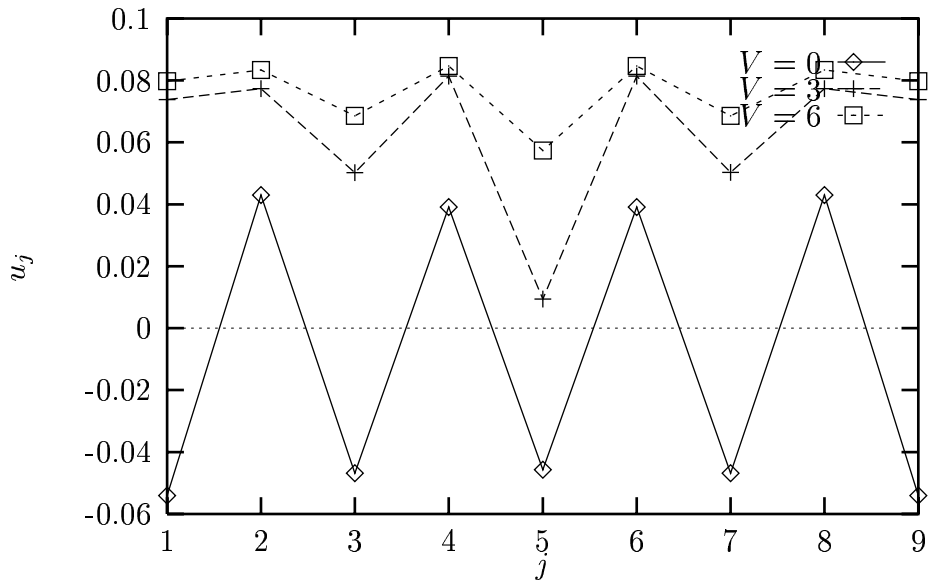


Figure 28: As V increases, the single bonds and the double bonds enlongate, but the bond at the center of the chain stays relatively short.

2.5 Modeling photoexcitation

The photoexcitation operator can be modeled in this basis. Photoexcitation causes an electron-hole pair to be created; that is, by moving an electron from the valence band to the conduction band. Quantum mechanically, this involves an electromagnetic potential $\mathbf{A}$. To first order in $\mathbf{A}$, the photoexcited state evolves as

$$|\Psi(t)\rangle = e^{-iHt}\left[1 + \frac{i}{c}\int_0^t \mathbf{j}(t')\mathbf{A}(t')\,dt'\right]|\Psi(0)\rangle$$

where $\mathbf{j}$ is the electric current in the chain and c is the speed of light (Su 1980). This photoinduced current can be understood formally as the operator

$$\sum_{j=1}^{N-1}(\epsilon\cdot\hat{\mathbf{u}})\hat{J} = \sum_{j=1}^{N-1}(\epsilon\cdot\hat{\mathbf{u}})(c_{j,s}^{\dagger}c_{j+1,s} - c_{j+1,s}^{\dagger}c_{j,s})$$

where ϵ is the polarization vector of the photon, and

$$\hat{\mathbf{u}} = \frac{\mathbf{r}_{j+1} - \mathbf{r}_j}{|\mathbf{r}_{j+1} - \mathbf{r}_j|}$$

The potential $\mathbf{A}$ and its corresponding polarization vector ϵ can be assumed to be uniform across j, since the molecules in question are on the order of tens of angstroms in size, whereas the wavelength of light for photoexcitation is hundreds of nanometers.

In the spin-conscious two-site problem, the unit current operator $\hat{J}$ can be written

$$\hat{J} = \begin{bmatrix} 0 & -1 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

Some computational models keep track of electron states by population numbers within each band. In such a system, photoexcitation can be modeled as the artificial, instantaneous moving of an electron from the valence band to the conduction band. However, when only ground expectation values are computable, and no wavefunctions are known, photoexcitation must be modeled as an operator $\hat{P}$ that has the effect of making $\hat{P}\,|G\rangle = |E\rangle$, where $|G\rangle$ is the ground state and $|E\rangle$ is the optically excited state.

Such an operator can be found within the photon transition rate

$$R_{i\to f} = C\,|\langle f\,|A\cdot p|\,i\rangle|^2\,\delta(E_f - E_i - w)$$

(where $h = 1$) which gives the probability of a transition between the state i and the state f, involving either absorption or emission. The product $A\cdot p$ can be rewritten as the current operator $\hat{J}$, using Mahan's formalism. The absorption spectrum $I(\omega) = \sum_f r_{i\to f}$ can be broken down into its moments. For example,

$$\begin{aligned}
\langle\omega\rangle = \frac{\int I(\omega)\omega\,d\omega}{\int I(\omega)\,d\omega} &= \frac{\sum_f |\left\langle f\left|\hat{J}\right|i\right\rangle|^2(E_f - E_i)}{\sum_f |\left\langle f\left|\hat{J}\right|i\right\rangle|^2} \\
&= \frac{\sum_f \left\langle i\left|\hat{J}\right|f\right\rangle\left\langle f\left|\hat{J}\right|i\right\rangle(E_f - E_i)}{\sum_f |\left\langle f\left|\hat{J}\right|i\right\rangle|^2} \\
&= \frac{\sum_f \left\langle i\left|\hat{J}^\dagger\hat{H}\right|f\right\rangle\left\langle f\left|\hat{J}\right|i\right\rangle - \left\langle i\left|\hat{J}^\dagger\right|f\right\rangle\left\langle f\left|\hat{J}\hat{H}\right|i\right\rangle}{\sum_f |\left\langle f\left|\hat{J}\right|i\right\rangle|^2} \\
&= \frac{\left\langle i\left|\hat{J}^\dagger\hat{H}\hat{J}\right|i\right\rangle - \left\langle i\left|\hat{H}\right|i\right\rangle}{\langle i|i\rangle}
\end{aligned}$$

It is true by definition that in any transition the energy gap between the photoexcited state ($|E\rangle$) and the ground state ($|G\rangle$) should be the energy of the incoming photon: $E_E - E_G = \omega$.

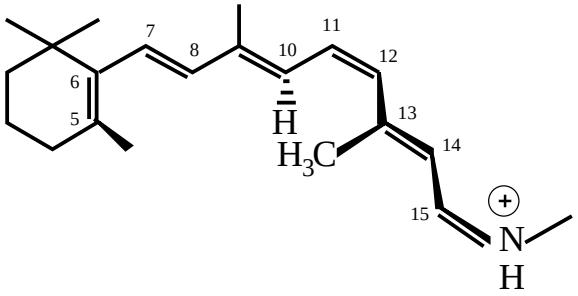


Figure 29: The retinal molecule. Carbons 5 through 16 are modeled as a polyacetylene chain of length 12, and the ring is neglected.

The same must therefore hold for the expectation values of these quantities:

$$\left\langle E \left| \hat{H} \right| E \right\rangle - \left\langle G \left| \hat{H} \right| G \right\rangle = \langle \omega \rangle$$

Therefore, it follows that $\hat{J} \left| i \right\rangle = \left| E \right\rangle$, since the moment must reduce to

$$\langle \omega \rangle >= \left\langle i \left| \hat{J}^{\dagger} \hat{H} \hat{J} \right| i \right\rangle = \left\langle E \left| \hat{H} \right| E \right\rangle - \left\langle G \left| \hat{H} \right| G \right\rangle = E_E - E_G = E_{\text{gap}}$$

where $E_{\text{gap}}$is the name for the energy gap between the photoexcited state and the ground state. Higher order moments are also possible. The moments of the quantum energy gap can construct the full absorption spectrum, and vice versa.

2.6 Extending the polyacetylene model to retinal

Within this theory, the problem of retinal's isomerization can be outlined. It can be imagined as a polyacetylene chain of 12 carbons, where the last carbon is not really a carbon but a nitrogen atom (Fig 29). The fact of the nitrogen atom is significant only for its mass, since, like a carbon, it contributes one π electron to the chain. The nitrogen atom at site 16 is a good place to end the chain because it has no other possible double bonds, only single bonds to the lysine group and to a hydrogen. Similarly, carbon 5 is a good place to end the chain, since its bond to carbon 4 can never be double. The positive charge on the nitrogen can be modeled by setting S_{16} in order that the ground state configuration comports with experiment. This potential should cause a soliton to be created around site 13 in the ground state (Aalberts 1997).

The effect of the methyl from site 13 can be modeled by including in the Hamiltonian the antibonding steric potentials between the methyl and hydrogens on neighboring sites. These terms will lead the Hamiltonian to recognize that the all-*trans* configuration has a lower energy than the 11-*cis* configuration. However, the $\cos\theta$ factor in the tight-binding term causes torsional stiffness in the π bonds, creating an energy barrier in the ground state between the *cis* and *trans* formations.

Ab initio calculations have shown that photoisomerization takes place in three sequential steps: a photon interacts with an electron in retinal; the tail of retinal stretches; and then the tail rotates around the 11–12 bond in to the *trans* configuration (Fig. 30). This is consistent with a hypothesis that photoisomerization causes the 11–12 bond to change from a short double bond to a long single bond, thereby decreasing the torsional stiffness so that there is no barrier separating the *cis* configuration from the *trans*. This is also consistent with the experimental finding that the vibrational modes of retinal molecules in solution are coherent during the time of isomerization: the molecule does not vibrate down into the all-*trans* state from an excited state, but rather moves there through a barrierless internal conversion. We further hypothesize that the stretching is caused by the passing of a soliton past the 11–12 region. (Garavelli 1997; Wang 1994).

The full theory of conjugated polyenes until now has not been used to investigate this hypothesis. We incorporate the full theory into our model, considering electron interactions (the U and V terms, the photoexcitation operator ($\hat{P}$), mechanical and electronic ground states, charge doping, soliton dynamics, and vibrational modes.

We include electronic interactions even though Su, Schrieffer, and Heeger concluded that electron interactions were not important. We think it is possible that the increase due to the U term in the energy gap between the ground state and the photoexcited state is necessary for the ground state to reach the photoexcited state, from which it can tunnel into the all-*trans* ground state. Furthermore, we have the technology to perform quantum mechanical calculations with a density matrix method which can handle electron interactions

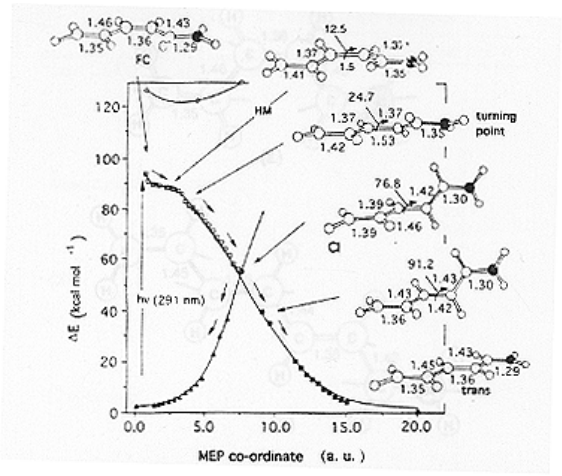


Figure 30: Garavelli et al. performed an *ab initio* calculation of the dynamics of retinal's photoisomerization, and found it to stretch first and then twist. Note that the bond lengths change between the FC and HM states, even though the molecule has not traveled far down its minimum energy path (MEP). Only after the central bond is long does the molecule twist around it (Garavelli 1997).

in a reasonable amount of time, which Su, Schrieffer, and Heeger did not. The density matrix method (DMRG) gives quick answers because it yields expectation values without calculating all of the wave functions (White 1992; White 1993; du Croo de Jongh 1998).

3 Computing

We consider the full Hamiltonian

$$\begin{aligned}
\hat{H} &= \sum_{j=1}^{n-1} \frac{1}{2} K_\sigma u_j^2 + \sum_{j=1}^{n-1} \frac{1}{2} K_{\text{bend}} (\phi_j - 120^\circ)^2 + \frac{A_{\text{steric}}}{(r_{\text{H10-Meth13}})^{12}} \\
&- \sum_{j=1,s}^{n-1} (t - \alpha u_j) cos\theta (c_{j,s}^\dagger c_{j+1,s} + c_{j+1,s}^\dagger c_{j,s}) + \sum_{j=1}^{n-1} -\Gamma u_j \\
&+ \sum_{j=1}^{n} U_j n_{j,\uparrow} n_{j,\downarrow} + \sum_{j=1}^{n-1} V_j (n_{j,\uparrow} + n_{j,\downarrow})(n_{j+1,\uparrow} + n_{j+1,\downarrow}) \\
&+ \sum_{j=1,s}^{n} S_j n_{j,s} + \sum_{j=1}^{n} \frac{1}{2} m v_j^2
\end{aligned}$$

where all of the terms except the obvious kinetic energy term $\sum_{j=1}^{n} \frac{1}{2} m v_j^2$ have been defined above.

The potential energy terms of the Hamiltonian are functions only of coordinates $\mathbf{r}_j$ and operators $c^\dagger_{j,s}$ and $c_{j,s}$. Given the coordinates $\mathbf{r}_j$, the DMRG method yields $\left\langle \Phi \left| \hat{O} \right| \Phi \right\rangle$ for any operator $\hat{O}$, including $\hat{H}$ and $\hat{J}^\dagger \hat{H} \hat{J}$. Thus, the energy $E = \left\langle \Phi \left| \hat{H} \right| \Phi \right\rangle$ of the ground state or the photoexcited state is easily obtainable by substituting expectation values of operators in place of the Hamiltonian's operators. The energy of any configuration of $\mathbf{r}_j$ is computable.

Since x, y, z are real parameters of $\hat{H}$, and since the eigenfunctions $\Psi(x, y, z)$ are normalized, it is true by the Hellman-Feynman theorem that

$$\frac{dE}{dx} = \frac{d}{dx} \left\langle \Psi(x,y,z) \left| \hat{H}(x,y,z) \right| \Psi(x,y,z) \right\rangle = \left\langle \Psi(x,y,z) \left| \frac{d}{dx} \hat{H}(x,y,z) \right| \Psi(x,y,z) \right\rangle$$

and the same for y and z. Since the dependence of $\hat{H}$ on the coordinates is simple, computing the forces on the carbons, $F = -\nabla E = -\left\langle \Psi \left| \nabla \hat{H} \right| \Psi \right\rangle$ is a simple matter of plugging in expectation values for the operators in analytic derivatives of the Hamiltonian. For example, the force due to the "hopping term" is just

$$\begin{aligned} F_{\text{hop}} &= -\left\langle \Psi \left| \nabla \hat{H}_{\text{hop}} \right| \Psi \right\rangle \\ &= \sum_{j=1,s}^{n-1} \nabla \left(t - \alpha \cos\theta (u_{j+1} - u_j) \right) (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n) \left\langle \Psi \left| (c^\dagger_{j,s} c_{j+1,s} + c^\dagger_{j+1,s} c_{j,s}) \right| \Psi \right\rangle \end{aligned}$$

where $\cos\theta$ and u_j are simple functions of the coordinates.

3.1 Finding the ground state

Finding the spatial ground state is a simple matter of minimizing the potential energy E with respect to the x_j, y_j, and z_j. A conjugate gradient method of numerical minimization was used. This method is more sophisticated than simple methods in which the coordinates are merely stepped by their derivatives until the deriviatives are near zero. Instead, the function is followed along the line of its gradient in coordinate space until it reaches a minimum. Then the coordinates are reassigned to this point on the the line, and the gradient is computed again. The algorithm stops when the coordinates change very little with each step (Press 1992, sec. 10.6). This method finds local minima, not global minima.

3.2 Time evolution

With knowledge of the forces, it is possible to compute the trajectory of the coordinates in time, and to watch how the system evolves.

Time evolution is a simple matter of integrating the carbon coordinates with their time derivatives, and integrating the time derivatives using the coordinates' second time derivative, which we calculate as the force divided by the carbon mass. (The mass of the electrons is neglected.) There are n carbons, each of which has a location and a velocity, and the carbons are modeled in three dimensions. Evolving the system through time, therefore, requires the integration of $6n$ variables.

We perform this integration using the Bulirsch-Stoer method with Richardson extrapolation (Press 1992, section 16.4). The method, generally, is to perform the integration of a function $f(x)$ (in our case a component of location or velocity) along x (time) in a series of non-infinitesimal steps of stepsize H. Within each step of size H, it integrates between x and $x+H$ with the modified midpoint method many times, each time with a successively smaller stepsize h, from $h = H/2$ to $h = H/4$, to $h = H/6$, and so on, until a suitably accurate value of $f(x + H)$ can be extrapolated. The value H can be considered the mega-stepsize, and the value h the mini-stepsize. The extrapolation method considers $f(x + H)$ as a function of stepsize h, and uses rational function extrapolation to guess at the value of $f(x + H)$ at $h = 0$. The number of integrations depends on the sensitivity of $f(x + H)$ to various stepsizes. When extra $f(x + H)$ datapoints at successively smaller values of h cease to change appreciably the fitted rational function, the routine sets $f(x + H)$ to the extrapolated value $f(x + H)$ at $h = 0$, and moves forward to integrate from $x + H$ to $x + 2H$.

The mega-stepsize H is chosen adaptively. If H is too small, then the extrapolation method will converge on an answer too quickly, which is accurate but inefficient. If H is too large, then the extrapolation method might not be able to converge at all. The algorithm adapts by esimating a new mega-stepsize $H' = \alpha H$ after the completion of an integration

from t to $t + H$. The extrapolation algorithm's sequence of successive mini-stepsizes is determined by $h = H/n_j$ where

$$n_j = 2, 4, 6, 8, \ldots, 2j$$

It happens that the error ϵ_j in the fitted function at step j goes as $\epsilon_j \approx H^{2j+1}$. Therefore, if the previous series of integrations (under mega-stepsize H) converged at step $j = k$, the next stepsize H' can be assigned a value

$$H' = H \left(\frac{\epsilon_j}{\epsilon} \right)^{1/(2k+1)}$$

where ϵ is the threshold error below which the extrapolation is said to have converged. A more complicated estimate considers the ideal j at which to converge, called q, can be estimated by looking at the previous sequence of integrations and setting q to the j at which the sequence of integrations yielded the least error with the least amount of computation. If the extrapolation converges before the integration with the ideal mini-stepsize H/n_q, then the next mega-stepsize H can be expanded by a factor that aims at achieving convergence at mini-stepsize H/n_q. If the extrapolation looks like it is going to fail to converge before mini-stepsize H/n_{q+1}, the extrapolation routine can be restarted with a shrunk mega-stepsize H. This integration routine is carefully controlled and highly accurate.

Since the full quantum mechanical time evolution is impossible to calculate in a reasonable amount of time, we must perform the classical time evolution in short steps, while recalculating the quantum eigenvalues with larger steps. The time evolution method calculates new expectation values of the quantum mechanical operators each time a new mega-step H is considered, which in the model occurs every few femtoseconds. In reality, of course, the electronic expectation values change continuously, not discretely. But the discrete approximation is accurate by the standard of the Frank-Condon principle, which holds that electronic transitions, which occur on the timescale of femtoseconds, are so fast compared to vibrational motion, which occurs on the timescale of picoseconds, that the nuclei have

the same locations and velocities in the new state as they had in the old state (Turro 1960). A timescale of electronic change of ~ 1 fs is small compared to the fastest observed phonon motion, which in the model spectrum has a period of about 25 fs (Fig. 31).

The process of photoexcitation can be modeled by first taking the system to its ground state, then by allowing the system to evolve from some arbitrary time $t = 0$ during which time the expectation values of the Hamiltonian are computed for the photoexcited state:

$$\left\langle \hat{O} \right\rangle = \left\langle \Phi \left| \hat{O} \right| \Phi \right\rangle \quad , \quad t < 0$$
$$\left\langle \hat{O}_{\mathrm{ex}} \right\rangle = \left\langle \Phi \left| \hat{J}^{\dagger} \hat{O} \hat{J} \right| \Phi \right\rangle \quad , \quad t \geq 0$$

It is impossible to know, quantum mechanically, exactly when the photon interacts with the electron; there is uncertainty in both time and energy. But the immediate imposition of the excited state is justified by the Franck-Condon principle. Since the excited state must occur sometime, and because it forms quickly relative to the motion of the coordinates, the model system can realistically photoexcite in one step of the time evolution algorithm.

3.3 Setting the parameters

The Hamiltonian depends on several coefficients that should be assigned the values that make the computed ground state correlate well with experimental data.

To set the variables t and α, one can use experimental measurements and trustworthy theoretical calculations of the lengths of the double and single bonds. A study of hexatriene (polyacetylene of length 6) found that the single bonds were 1.47 A, and the double bonds were 1.354 A (Negri 1989). Although hexatriene is certainly different from retinal, the bond lengths are similar to those from ab initio calculations of retinal's tail: 1.45 A for single bonds, and 1.35 A for double bonds (Gilardi 1972). In both cases, the average bond length is 1.4 ± 0.01 A. The stretching parameter Γ can be set so that the average bond length conforms to experiment.

A force field study of hexatriene found that the stretch constant k (for the spring model, where $F = -ku$) of a double bond is 8.50 mdyn/ A, or 53 eV/ A^2, and for a single bond, 5.42 mdyn/ A, or 34 eV/ A^2. This should lead to a vibrational frequency of $\omega = \sqrt{k/m}$, where m is the mass of a carbon. The K_σ coefficient can be set so that Fourier analysis of the coordinates during a time evolution reveals such a frequency. Using values of the bending constants (for carbon-carbon-carbon bending along the chain) from the same study, which amounted to 8.68 eV, the $K_{\rm bend}$ coefficient can be set. In Fig. 31, for example, the peak in u_{12} around 22 fs suggests the central frequency due to the K_σ term. The peak in ϕ_{12} around 50 fs suggests the central frequency of the $K_{\rm bend}$ term. The two separate peaks in θ_{12} suggest the vibrational modes resulting from the different torsional stiffnesses of single and double bonds. Since the vibrational modes of polyacetylene can be calculated analytically, (Aalberts 1998; Inagaki 1975), we can use Fourier analysis to verify the primary optical and acoustical modes, in order to check that the time evolution algorithm proceeds as expected. Also in this way, the behaviors of different molecules can be compared. For example, the vibrational modes reveal the different character of *cis* configurations and *trans* configurations (Fig. 32).

Also in this way, the behaviors of different molecules can be compared. The vibrational modes also reveal the different character of *cis* configurations and *trans* configurations (Fig. 32).

A study of hexatriene found that the torsion constant k (for the torsional spring model, where the torque is $\tau = k\theta$) is 0.13 mdyn A/ $\rm rad^2$, or 0.81 eV for two carbons connected with a double bond at the center of a chain, and 0.03 mdyn A/ $\rm rad^2$, or 0.18 eV for the case of a single bond. The energy expression for the tight-binding term in the Hamiltonian is

$$(\alpha u_j - t)cos\theta(c^\dagger_{j,s}c_{j+1,s} + c^\dagger_{j+1,s}c_{j,s})$$

which is proportional to $\cos\theta \approx 1 - \frac{\theta^2}{2}$. Thus, θ enters the energy expression in a $-\frac{1}{2}k\theta^2$

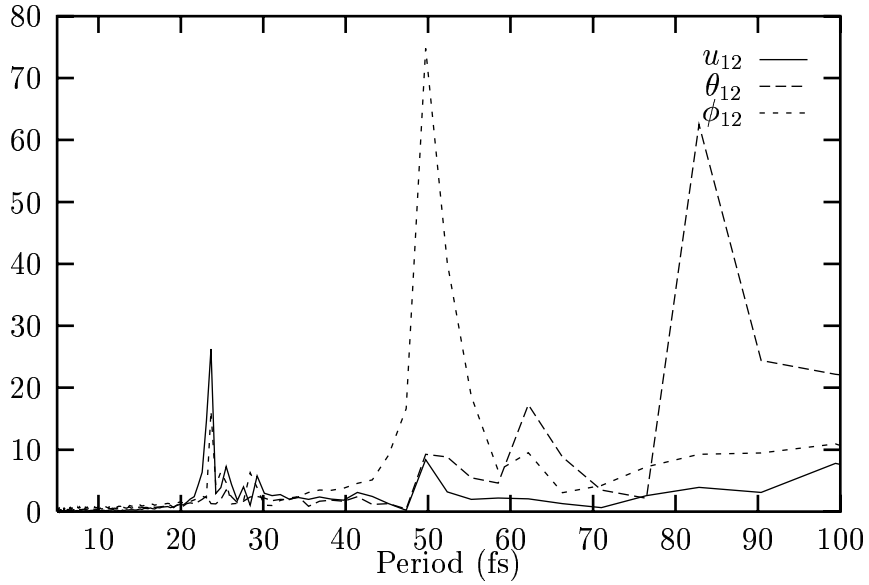


Figure 31: Fourier transform of a time evolution of polyacetylene in its excited state. Discrete peaks represent vibrational modes, which can be checked with the spectrum observed in experiments.

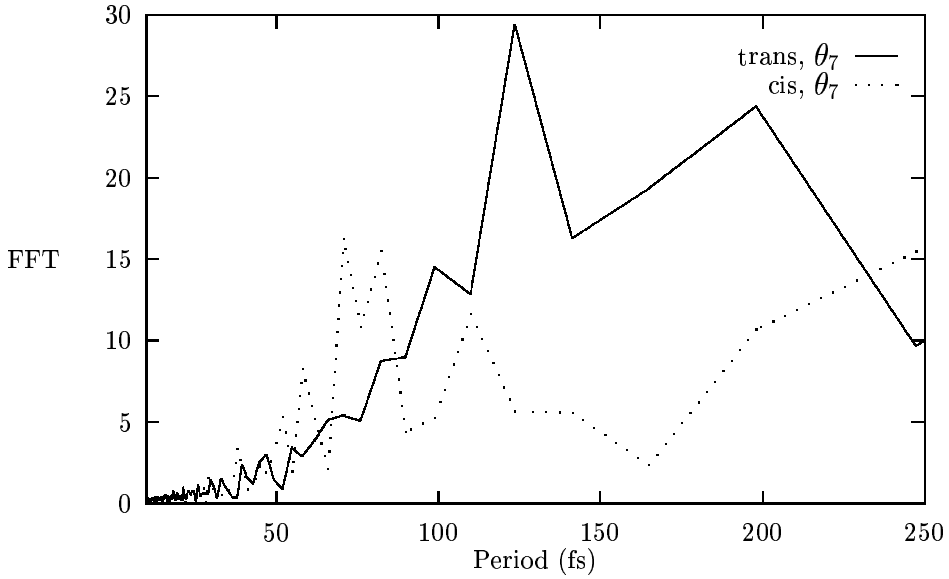


Figure 32: Two time evolutions of retinal were performed, in which both evolutions used the same parameters, but one was in the *cis* isomer and the other in the *trans*. The atoms of the molecules were given random, slight displacements, and the system was allowed to evolve. The two systems exhibited very different vibrational modes.

form, which is that of a standard harmonic oscillator, where the torsion constant k is

$$k = (t - \alpha u_j)(c^{\dagger}_{j,s} c_{j+1,s} + c^{\dagger}_{j+1,s} c_{j,s})$$

Thus, the variables t and α can be set along with any other variables that affect bond length in order that the expression k matches the values of 0.18 eV and 0.81 eV for single and double bonds j, respectively.

Hexatriene calculations show that in the transition from the ground state to the lowest triplet excited state, the torsional constants of the double and single bonds switch places: the constants of double bonds drop to the level of ground state single bonds, and the constants of single bonds rise to within 75% of that of ground state double bonds. Constants for double bonds at the ends of the chain do not fall as low as those for double bonds in the middle of the chain, but they are still less than the constants of former single bonds. This is not true for higher excited states (Negri 1989).

Retinal's isomerization may occur when photoexcitation takes the system to a higher energy state in which the double bonds and single bonds around the center of the tail are inverted, as in hexatriene's lowest triplet excited state. This inversion will turn the 11–12 double bond into a single bond, a process that is aided by the steric repulsion between the methyl group off carbon 13 and the hydrogen off carbon 10. The steric force will cause the 11–12 bond to quickly (within hundreds of fs) twist into a 90° position. Relaxation to the planar all-*trans* state may take picoseconds, since the half-untwisted state kills the torsional force. Vibrational motion might be required to push the torsional coordinate down the energy hill to the *trans* configuration, and such a transition would require time. The truth of this scenario is bolstered by the finding of intermediates in the photoproduct of photoexcited retinal that do not form until picoseconds later. Nevertheless, a twist of the 11–12 bond to the 90° position is a sufficient mechanism to change the chemical properties of the rhodopsin protein, setting off the process of generating a nerve signal.

Experiments using NMR spectroscopy can measure the charge density at each site along

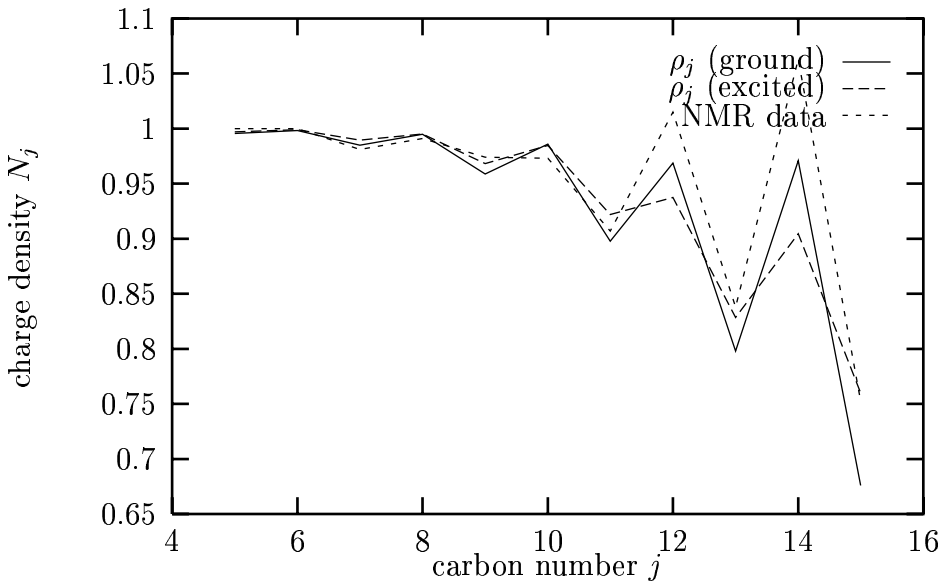


Figure 33: Measurements of charge density give good agreement with experimental data.

retinal's chain in rhodopsin. Using semiempirical methods, and applying the "zero differential overlap" approximation, theorists have tried to reproduce such charge density patterns by positioning an ion at the end of the chain, where the positively charged NH_2 group is located, as well as a "counterion" near carbon 13 (Han 1993). The existence of such a counterion is certainly a possibility. However, it is also possible that the electron interactions, especially U and V, affect the charge density pattern in ways similar to those of a counterion. The U and V should thus be set according to NMR measurements of retinal's charge density in solution. Since they also affect the dimerization pattern, they also need to be set in order to comply with ab initio bond length predictions (Aalberts 1997, refs. 13, 19.) (for *ab initio* measurements on retinal, see Garavelli 1997) (for calculations on hexatriene, see Negri 1989). We found that values of $U_j \approx 1$ and $V_j \approx 0.3$ gave good agreement with the charge density data (Fig. 33). We also experimented with altering the S_j electron density terms in order to model the effect of charges near retinal, such as might come from retinal's protein pocket.

The distances between carbons can be measured with NMR spectroscopy. The distance between the methyl carbon from site 13 and carbons 10 and 11 has been measured. The magnitude of the steric repulsion A_{steric} can be set so that these distances are correct. In

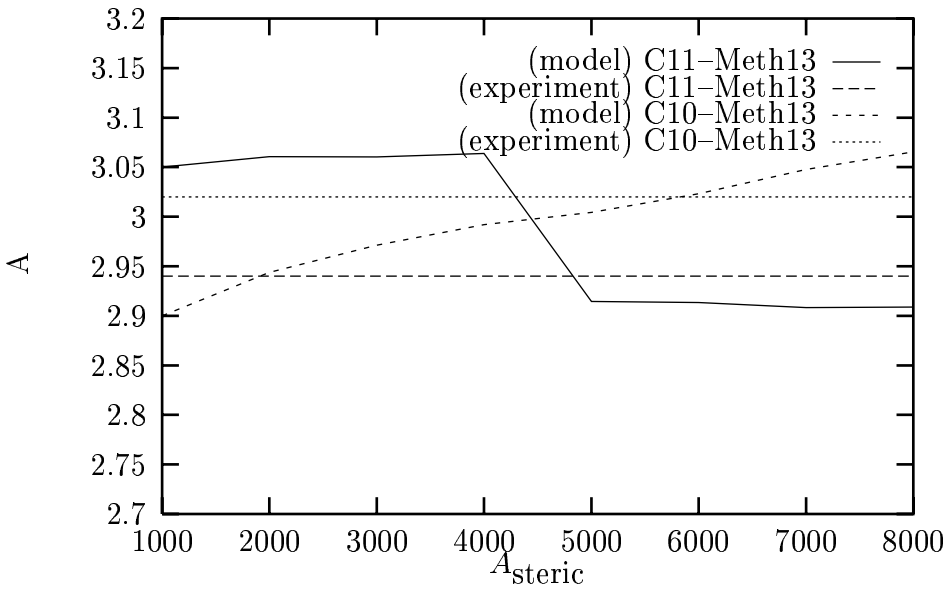


Figure 34: The steric force parameter can be set according to the experimentally measured carbon-carbon distances. A value of 6000 was found to be realistic.

Fig. 34, for instance, the value of $A_{\text{steric}} = 6000$ was found to result in reasonably good agreement with experiment.

Setting these parameters accurately is a grand challenge, since each measurable quantity is affected by several parameters, and each parameter affects several measurable quantities. Since the goal of the modeling is to explain photoisomerization, the model can match experiment relatively loosely, perhaps with a 10% accuracy in oscillation times and relative distances—but one can only guess the accuracy requirement. However, it seems likely that a molecule with a 67% quantum efficiency at room temperature would not have stringent requirements (Wang 1994)

Of great importance is the energy difference between the ground state and optically excited state, E_{gap}. It is a particularly difficult variable to tune, since every parameter in the Hamiltonian affects the energy. However, the great failing of models like SSH which did not take interactions into account was their underestimate of E_{gap}. Since increasing U increases the energy gap, we think that the inclusion of interactions is what is needed to model photoisomerization properly.

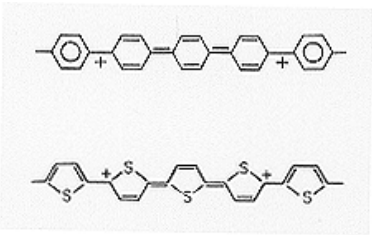


Figure 35: Molecules such as these give off light in the visual spectrum in response to currents, and could be used in future computer displays. Our simulator is capable of modeling such molecules, and can predict the wavelength of the emitted photon as a function of the molecule's structure and conditions (Heeger 1988).

The $E_{\text{gap}}$is especially important for practical applications. Conjugated polymers have important optical properties not only in their absorption, as in vision, but also in their emission. It has been shown that the absorption of a photon produces an electronic current in the conjugated chain. The mathematics is time symmetric, however, so that an induced current causes a photon to be emitted. The value E_{gap}determines the wavelength of the emitted photon. Conducting polymers are the mechanism behind low-power "Indi-glo" lights in wristwatches and nightlights, which would be highly desirable for computer monitors, as they require no heat-producing incandescent backlights. A color monitor would be possible if the molecules at a particular point on the screen could be deformed by charges or by a piezo so that they emit (or absorb, on a non-glowing screen) at a specific frequency. The problem of engineering such applications is to set $E_{\text{gap}}$to a specific value; and that is the same problem that our model attempts to solve (Fig. 35).

Solitonic excitation will be a very important area of study. It is thought that photoexcitation proceeds when the soliton at carbon 13 collides with a soliton that emerges around site 5 as a result of the energy introduced by the current. When the soliton passes by the 11 bond, it switches the double bond into a single bond, which is less torsionally stiff than the single bond. Although soliton-antisoliton pairs may continue to be produced, and annihilated, they will stay away from the 11–12 bond for several hundred femtoseconds. Noninteracting mod-

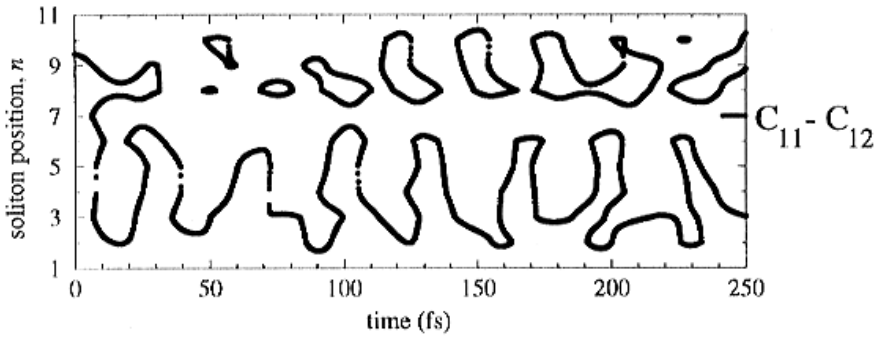


Figure 36: Upon photoexcitation, soliton/anti-soliton pairs are created in retinal. A soliton crosses the 11–12 bond, turning it from a double bond into single bond. The solitons continue to propagate, but they do not recross the 11–12 region (Aalberts 1997).

els have demonstrated this, but yield the incorrect energy gap (Aalberts 1997). (In Figure 36, a soliton changed the 11–12 bond from a double bond into a single bond, but the system did not isomerize because the excitation energy was not sufficient to overcome the torsional energy barrier.) In the excited state, the hopping expectation value is reduced. Therefore, the torsional spring constant can be tuned so that it goes negative upon photoexcitation, on the 11–12 bond. This will make the bond very weak, allowing the molecule to isomerize without vibration.

Our simulation has not yet reproduced retinal's photoinduced isomerization. However, it is likely that the model has sufficient flexibiliy to show photoisomerization, if only the parameters were set correctly. In one simulation retinal isomerized around the 10–11 bond, a transformation not observed in nature, instead of the 11-12 bond (Figs. 37, 38). The failure of this simulation suggests that the classical model of retinal's forces must be modeled with greater complexity, since retinal's tail has so many degrees of freedom.

4 Conclusion

More simulations must be performed, and the parameters must be tuned more carefully, but the preliminary results demonstrate a promising model of retinal photoisomerization.

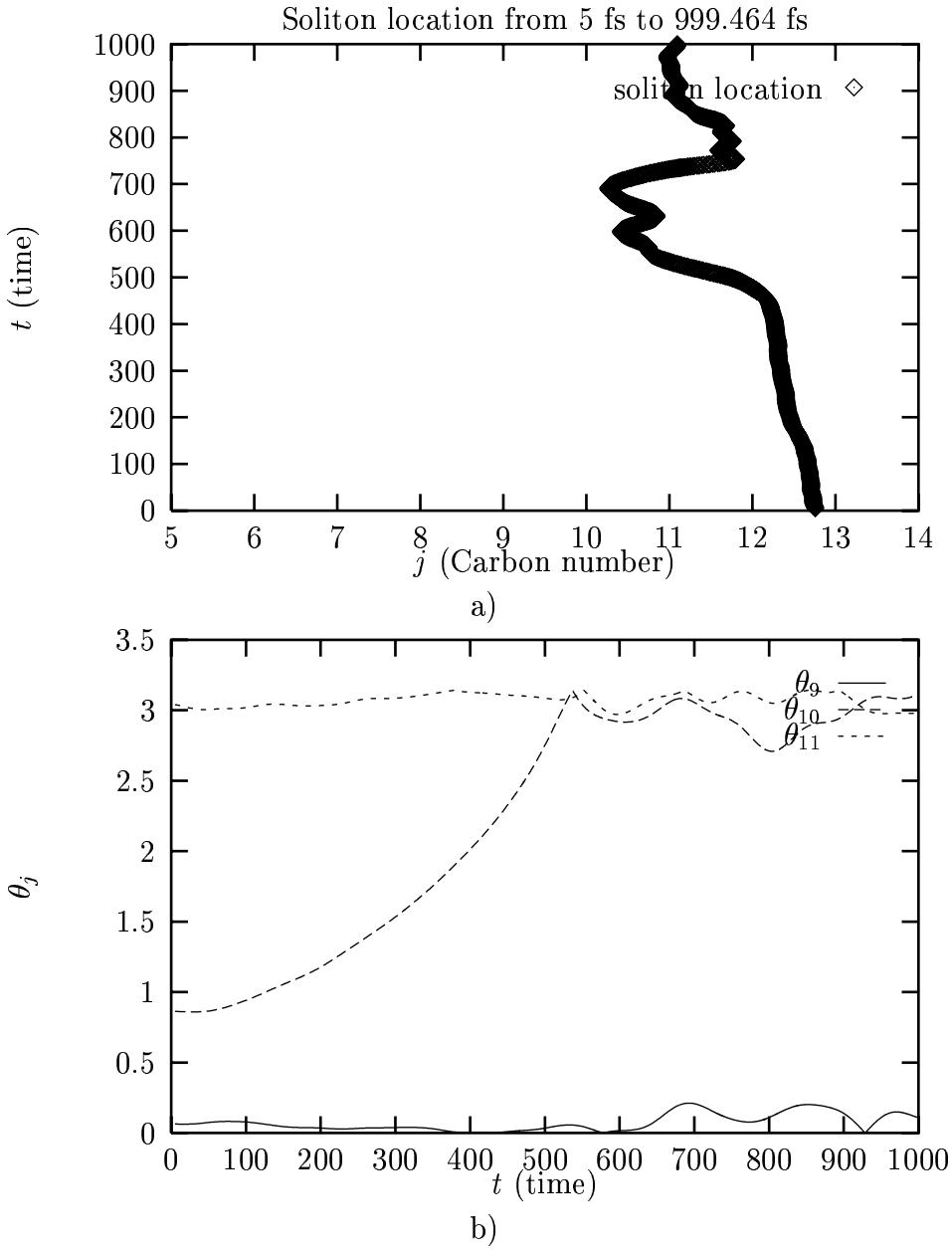
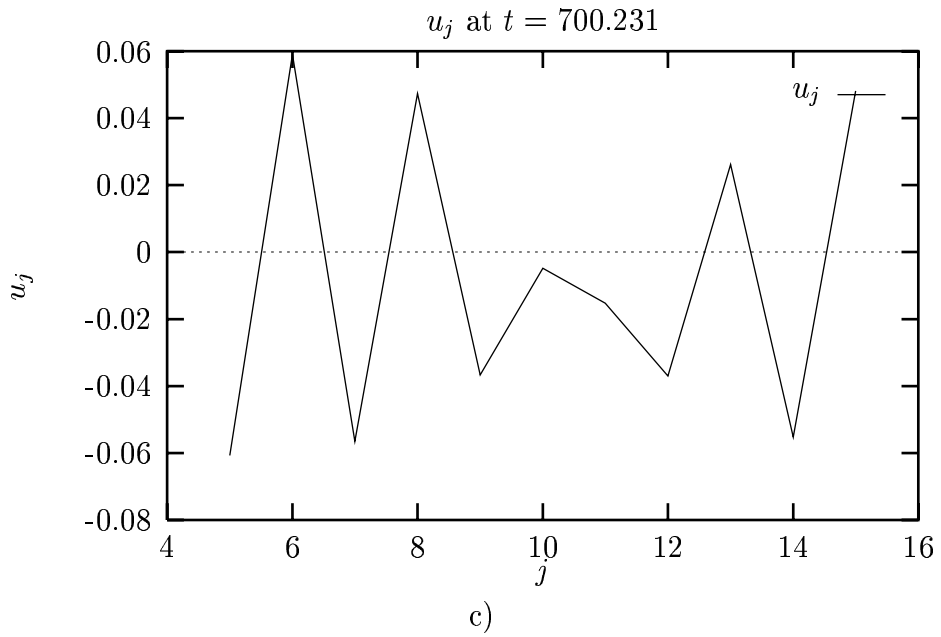


Figure 37: The 11-*cis*-retinal molecule at time $t = 0$ was in its ground state, but then was "hit with a photon." a) The soliton around site 13 traveled leftward, then bounced back, weakening the 10–11 single bond. b) The 10–11 bond is so weak that the molecule, which is already somewhat twisted there, twists all the way around it.



c)

Figure 38: By $t = 700$fs in the simulation of Fig. 37, the double bond at 11–12 was weakened, but sufficiently for isomerization around 11–12 to occur.

References

Aalberts, D. P., F. L. J. Vos, and W. van Sarloos, 1997, Pure and Appl. Chem. 69, 2099.

Aalberts, D. P., F. L. J. Vos, and W. van Sarloos, 1998, unpublished.

Ashcroft, N. W. and N. D. Mermin, 1976, *Solid State Physics* (Saunders College, Philadelphia), 334-37, 685, 689-91.

Atkins, P. W., 1990, *Physical Chemistry* (Freeman, New York), 473.

du Croo de Jongh, M. S. L., and J. M. J. van Leeuwen, 1998, Phys. Rev. B 57, 8494.

Heeger, A. J., S. Kivelson, J. R. Schrieffer, and W. P. Su, 1988, Rev. Mod. Phys. 60, 781.

Garavelli, M., P. Celani, F. Bernardi, M. A. Robb, M. Olivucci, 1997, J. Am. Chem. Soc. 119, 6891.

Hamanaka, T., T. Mitsui, T. Ashida, and M. Kakudo, 1972, Acta Cryst. B28, 214.

Gilardi, R. D., I. L. Karle, and J. Karle, 1972, Acta Cryst. B28, 2605.

Haile, J. M., 1992, *Molecular dynamics simulation* (Wiley, New York), 189.

Han, M., B. S. DeDecker, S. O. Smith, 1993, Biophysical Journal 65, 899.

Guinea, F., 1984, Phys. Rev. B. 30, 1884.

Kittel, C., 1986, *Introduction to Solid State Physics* (Wiley, New York), 284-85.

Morse, P. M., 1929, Phys. Rev. 34, 57.

Nathans, J., 1992, Biochemistry 31, 4923.

Negri, F., G. Orlandi, A. Brouwer, F. W. Langkilde, and R. Wilbrandt, 1989, J. Chem. Phys. 90, 5944.

Press, W. H. et al., 1992, *Numerical Recipes in C* (Cambridge, New York), 420-25; 724-732.

Stryer, L., 1995, *Biochemistry* (W. H. Freeman, New York).

Su, W. P., J. R. Schrieffer, and A. J. Heeger, 1979, Phys. Rev. B. 22, 2099.

Su, W. P. and J. R. Schrieffer, 1980, Proc. Natl. Acad. Sci. USA 77, 5626.

Turro, N. J., 1965, *Molecular Photochemistry* (W. A. Benjamin, New York), 178.

Wang, Q., R. W. Schoenlein, L. A. Peteanu, R. A. Mathies, and C. V. Shank, 1994, Science 266, 422.

White, S. R., 1992, Phys. Rev. Lett. 69, 2863.

White, S. R., 1993, Phys. Rev. B 48, 10345.

Vos, F. L. J., D. P. Aalberts, and W. van Sarloos, 1996, Phys. Rev. B. 53, R5986.

Vos, F. L. J., D. P. Aalberts, and W. van Sarloos, 1996, Phys. Rev. B. 53, 14922.

Vreven, T., F. Bernardi, M. Garavelli, M. Olivucci, M. A. Robb, and H. B. Schlegel, 1997, J. Am. Chem. Soc. 119, 12687.