LONI Institute - LATech

Highlights of research projects being conducted by LI scientists at LA Tech.

High Throughput High Energy Physics Data Processing on LONI

Z. Dick Greenwood

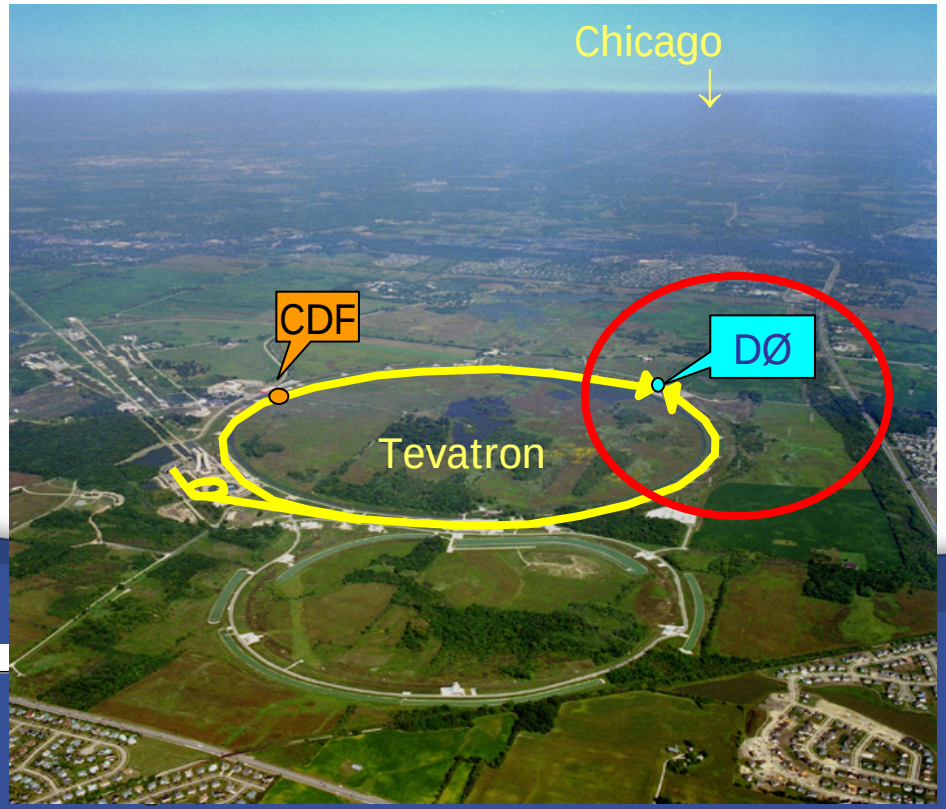
Fermilab Tevatron and LHC at CERN

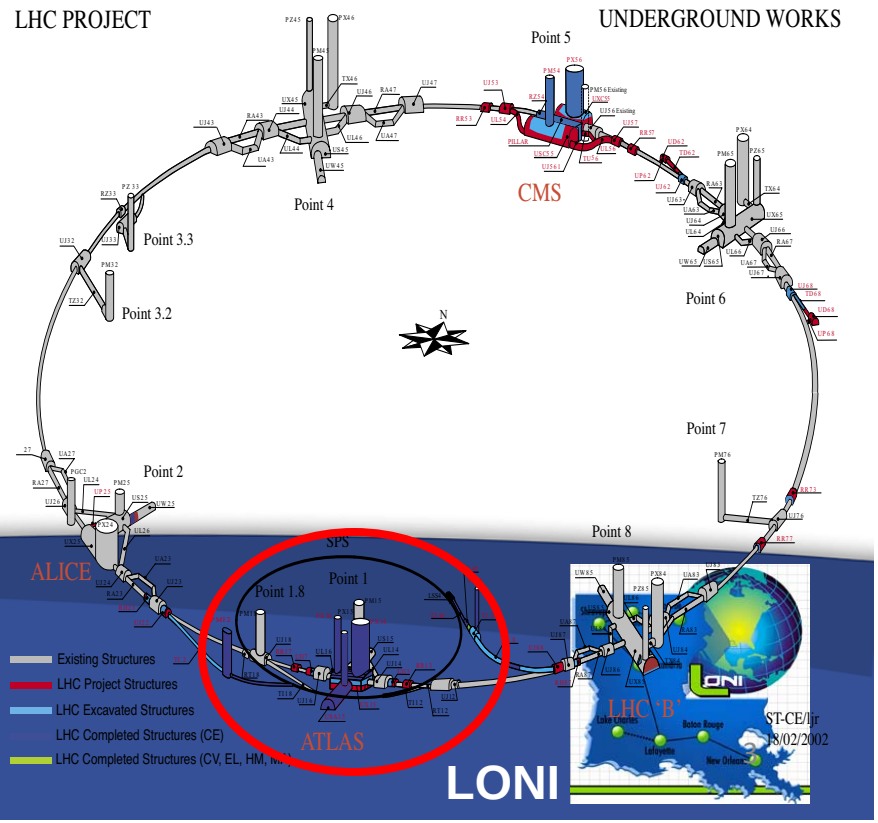
- Present world's Highest Energy proton-anti-proton collider
  - 4km circumference
  - E_{cm}=1.96 TeV (=6.3x10^{-7}J/p ➔ 13M Joules on 10^{-4}m^2)

  ⇒ Equivalent to the kinetic energy of a 20t truck at a speed 130km/hr (81mi/hr)

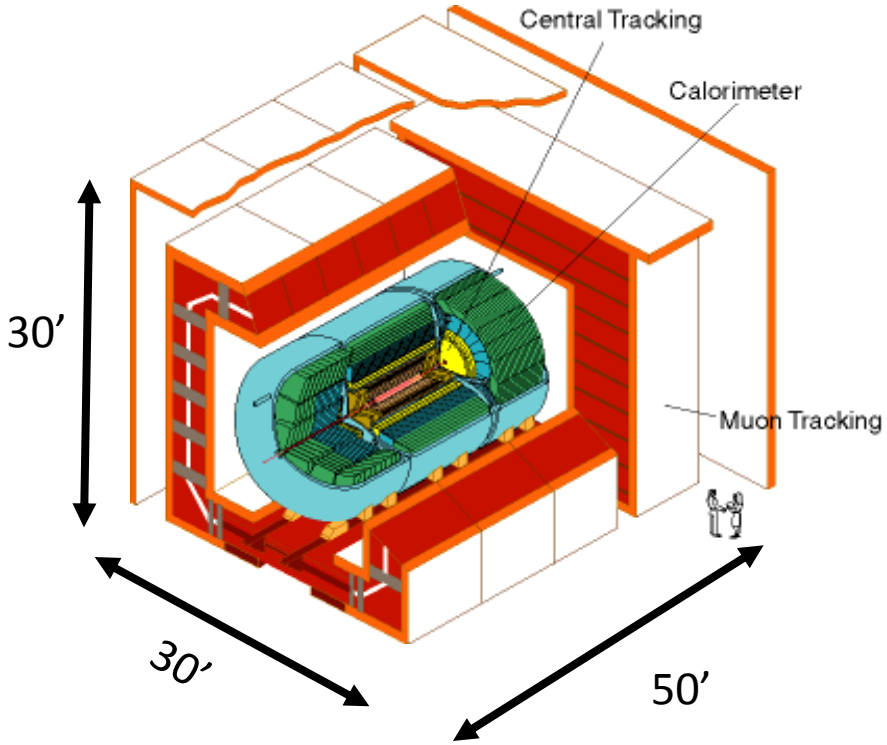
- World's Highest Energy proton-proton collider in 2008
  - 27km circumference
  - E_{cm}=14 TeV (=44x10^{-7}J/p➔ 1000M Joules on 10^{-4}m^2)

  ⇒ Equivalent to the kinetic energy of a 20t truck at a speed 1140km/hr (711mi/hr)



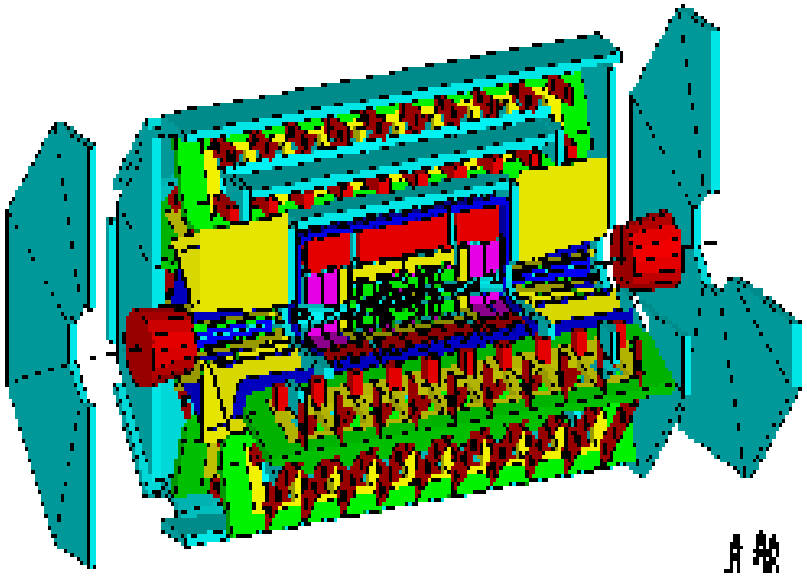


DØ Detector



- Weighs 5000 tons and 5 story tall
- Can inspect 3,000,000 collisions/second
- Records approximately 10,000,000 bytes/second
- Records 0.5×10^{15} (500,000,000,000,000) bytes per year (0.5 PetaBytes).

ATLAS Detector



- Weighs 10000 tons and 10 story tall
- Can inspect 1,000,000,000 collisions/second
- Records approximately 300,000,000 bytes/second
- Will record 1.5×10^{15} (1,500,000,000,000,000) bytes each year (1.5 PetaByte).

Open Science Grid Locations
(Courtesy Open Science Grid)

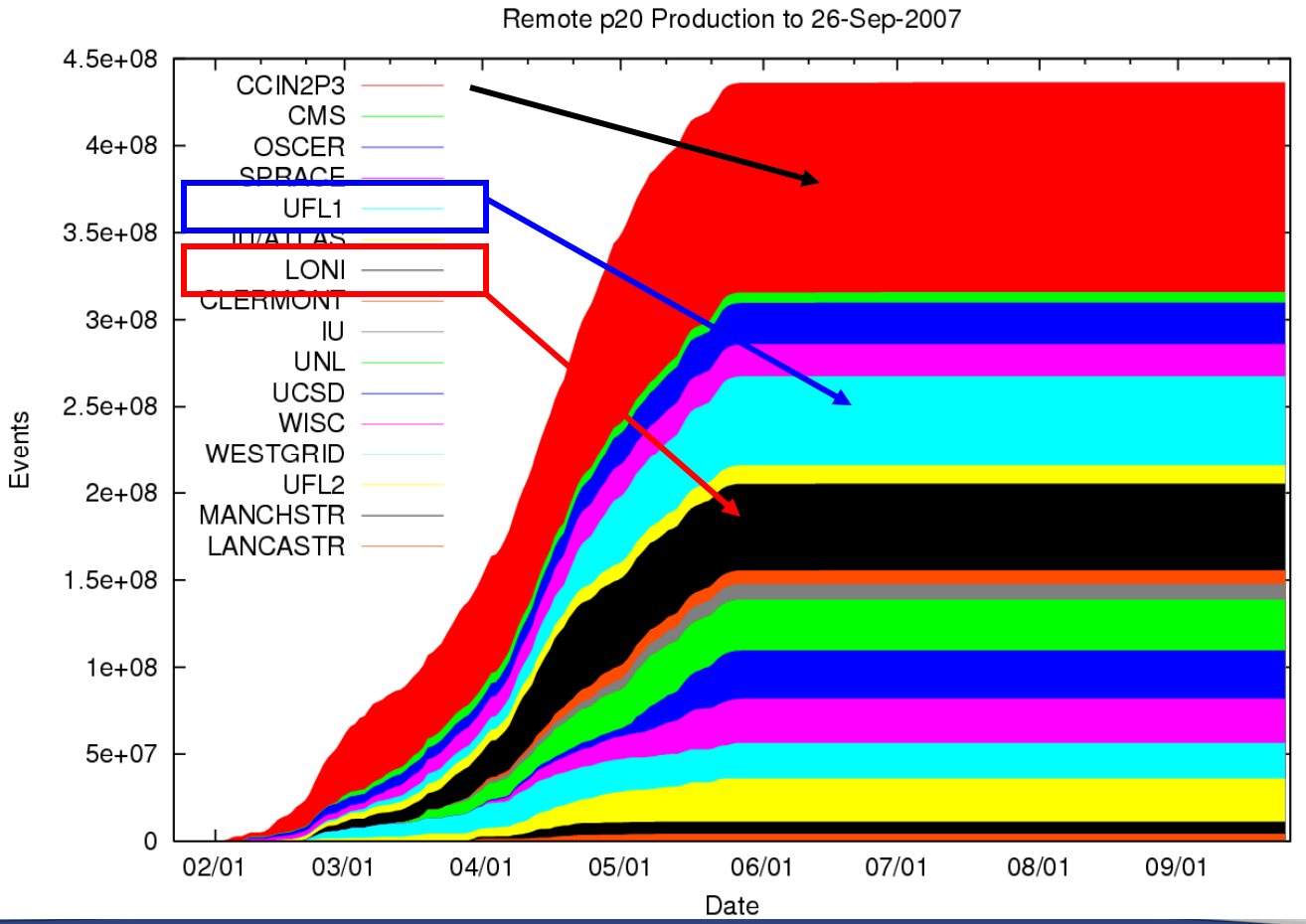
Louisiana Tech University

LONI

Dzero Reprocessing, 2007

Of the 455 million events processed remotely, LONI processed 47,755,321 (more than 10%), and was one of the top two performing OSG sites (tied with the University of Florida)!

Included in the OSG resources were >200 LONI CPUs managed by the group.

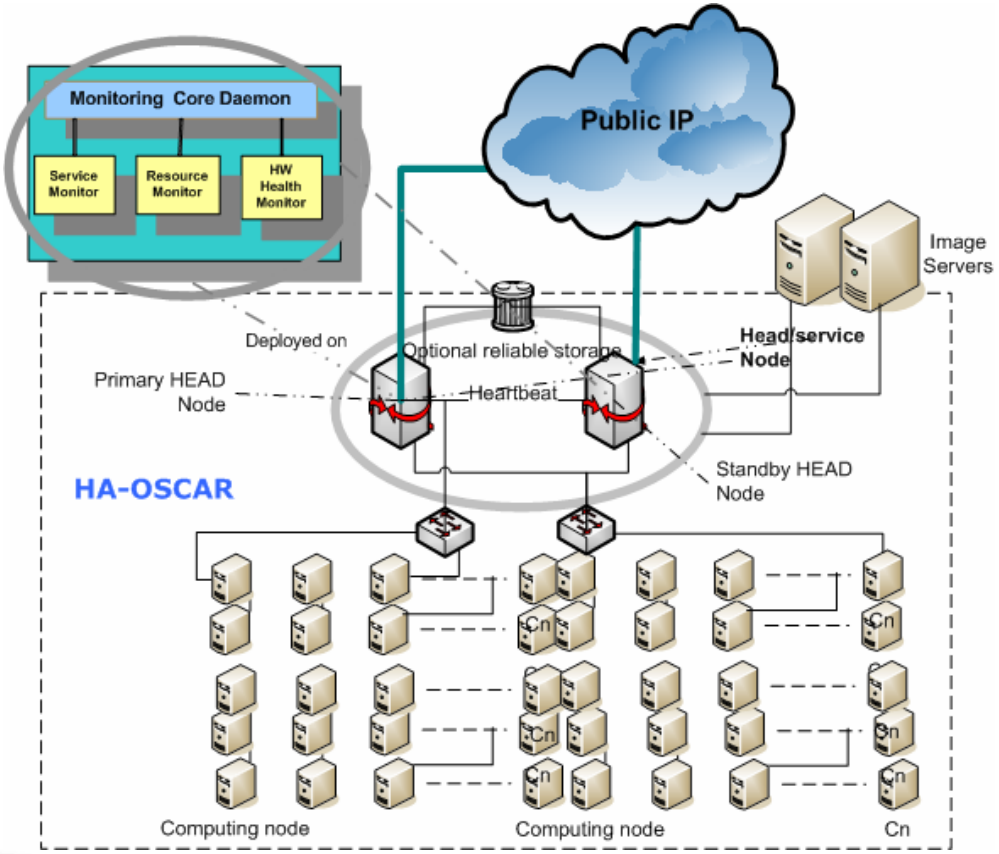


Towards Non-stop Computing in HPC environments

Box Leangsuksun, PhD

http://hpci.latech.edu

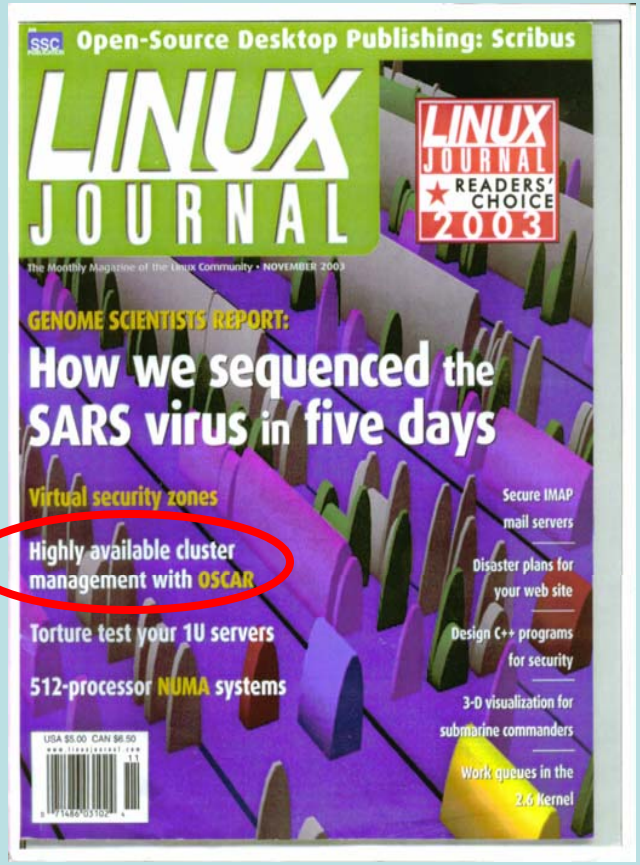
HA-OSCAR: Self-monitoring/healing cluster



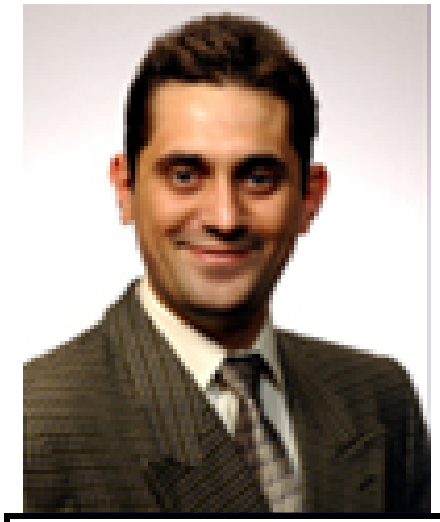
- Production-quality Linux-cluster project
- HA and HPC clustering techniques to enable critical computing infrastructure via self-healing and self-configuration cluster system
- HA-enabled HPC and enterprise Services.
- Self-healing with 3-5 sec automatic failover time
- The first known field-grade open source HA Beowulf cluster release
- Support transparent recovery for MPI and Job Queue Fault Tolerance
- Funded by DOE office of Science

HA-OSCAR Magazine Covers.

web site: http://xcr.cenit.latech.edu/ha-oscar



Simulation of Apoptotic Pathway in Cells



Andrei Paun

SIMULATION OF APOPTOTIC PATHWAY IN HIV-1 INFECTED CELLS

Andrei Paun

A major obstacle in obtaining a cure for the HIV virus is the infection of the latent T cells of our immune system. The major cause for the depletion of the T cells in infected persons is the cell suicide (apoptosis).

We are creating a new simulation method for the apoptotic pathway in HIV-1 infected cells called “Membrane Systems” simulation.

The Membrane System simulation method is an improvement over current simulation methods for studying the behavior of cells – nearly as accurate as solving ODE’s, yet as fast as stochastic methods.

Progress in this area will lead to clinical advances in the treatment of HIV infections.

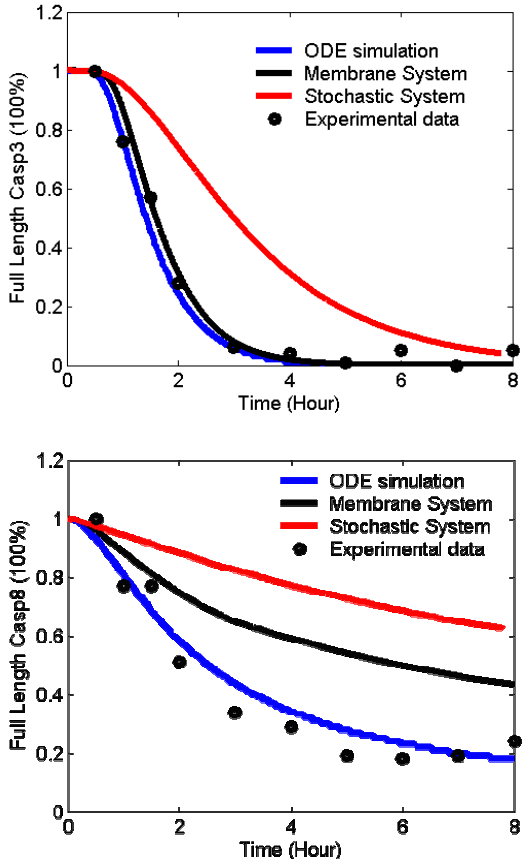


Figure 1. Comparison between the experimental data, ODE, Stochastic System (Gillespie’s algorithm) and Membrane System simulations of a model system. All the 3 distinct simulation techniques aforementioned use the exact same model of the caspase mediated apoptosis.

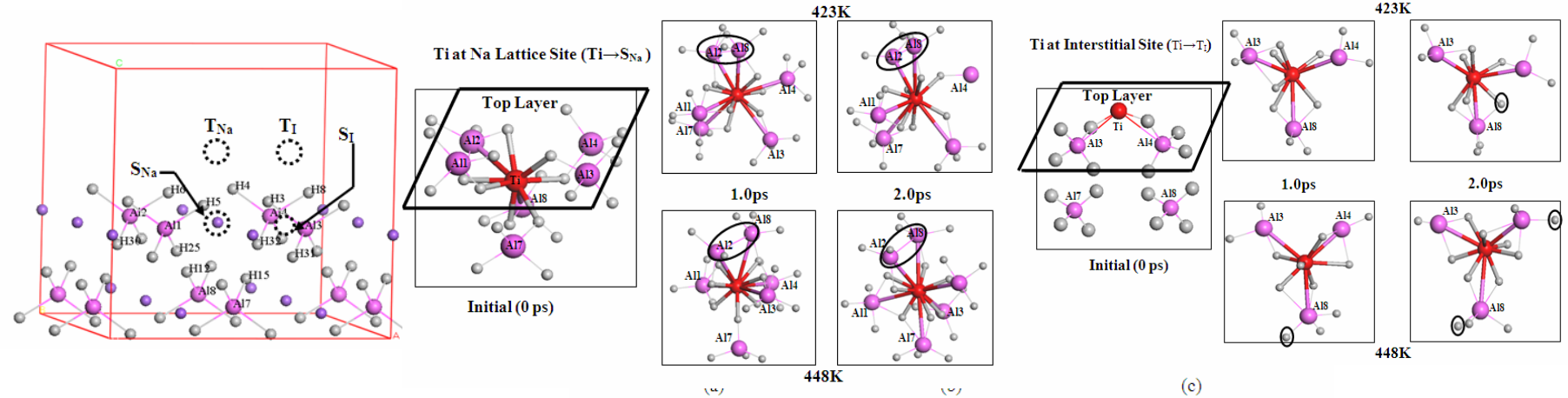
Computational Chemistry for Hydrogen Storage and Fuel Cells

Daniela Mainardi

COMPLEX METAL HYDRIDES FOR ON-BOARD HYDROGEN STORAGE

Daniela Mainardi

Role of Titanium Catalyst in Ab/desorption of Complex Metal Hydrides



TDOS and PDOS plots from optimised (a) pristine, (b) Ti→S_{Na} and (c) Ti→T_I models.

1. Calculated substitution energies of Ti-doped (001) $NaAlH_4$ surfaces have shown almost equal probability of substitution at both lattice and interstitial sites
2. $TiAl_2H_7$ and $TiAl_2H_2$ complexes are observed after geometry optimization of doped-surface models
3. Results from DFT-MD simulations have shown the existence of the observed Ti-Al-H complexes with time at both temperatures (423 and 448K) as well as increased association of Al and H with the complexes as time evolves

G. K. P. Dathara and D. S. Mainardi, Molecular Simulation, 2008

Supported by DOE Basic Energy Sciences DOE-BES/ DE-FG02-05ER46246

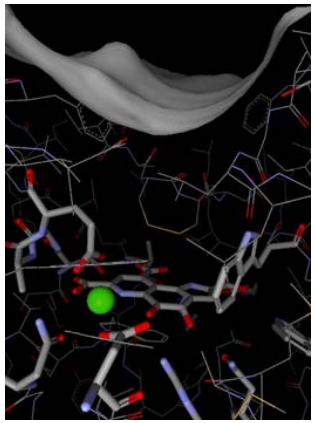
Louisiana Tech University

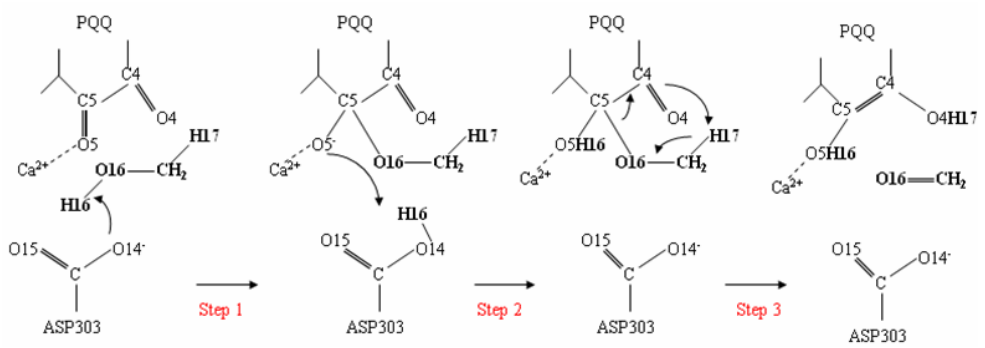
LONI

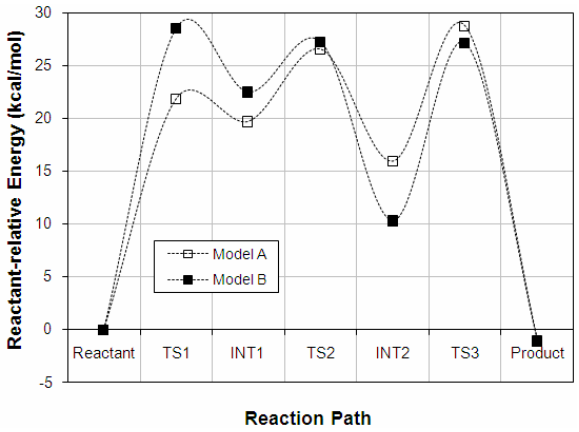
METHANOL OXIDATION MECHANISMS BY METHANOL DEHYDROGENASE ENZYME

Daniela Mainardi

Computational studies on Addition-Elimination and Hydride Transfer Mechanisms







1. The results obtained from the gas-phase calculations demonstrate that the oxidation of methanol can proceed through the addition-elimination mechanism

2. The rate-determining step (step 3, cleavage of Cmet-H17) for MDH Models A and model B shows energy barriers of 13 and 17 kcal/mol respectively

3. The presence of an activator (ammonia) in solvent and/or protein environment could further stabilize the intermediates and transition states, thereby modifying the energy barriers for a smoother catalytic process

Supported by DOE Basic Energy Sciences DOE-BES/ DE-FG02-05ER46246

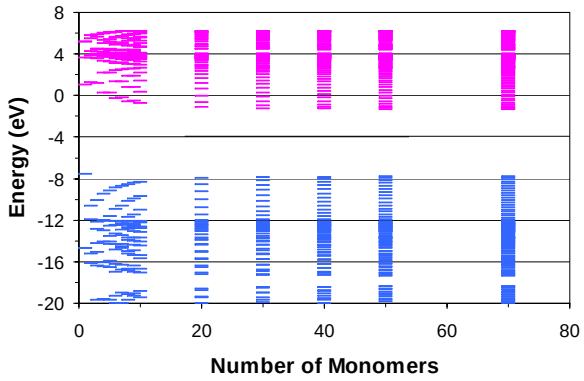
N. B. Idupulapati and D. S. Mainardi, Molecular Simulation, 2008

Multiscale Modeling for Materials Science

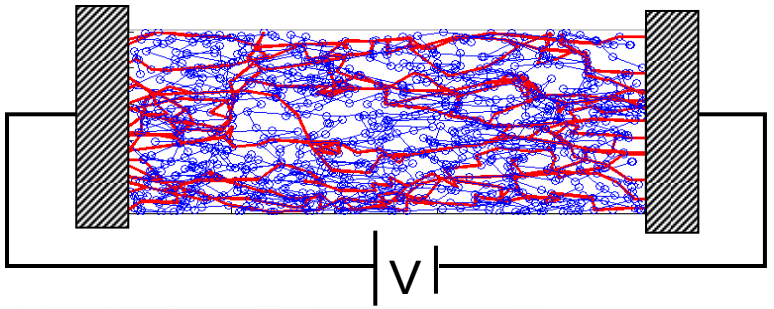
Pedro Derosa

Electron Transport in Polymers

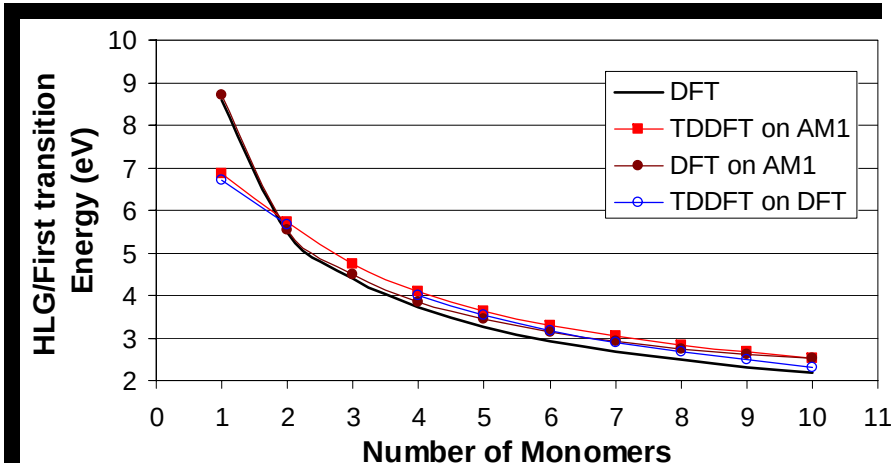
- Use quantum mechanical-based calculations to study electronic and geometrical structure of polymers.
  - Geometry at a Semi-empirical level (AM1, MNDO, PM3)
  - Electronic properties at a Density Functional Theory level (B3PW91/6-311++G**)

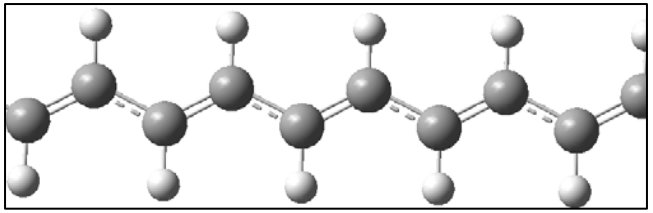


- Charge transport in realistic systems is calculated with a Monte-Carlo approach using parameters obtained from quantum mechanics



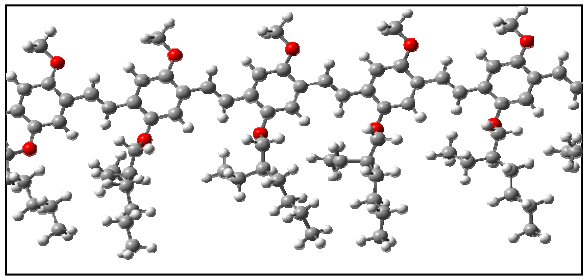
From monomers to polymers





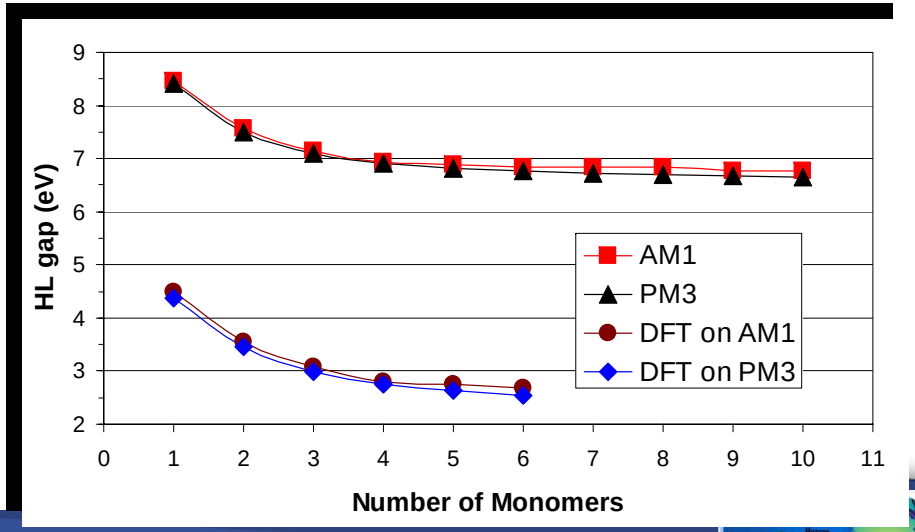
Exp. band gap 1.5 -1.8eV

Asymptotic limit Sim. 1.52-1.73eV



Exp. band gap 2.3 -2.45 eV

Asymptotic limit Sim. 2.29-2.32 eV

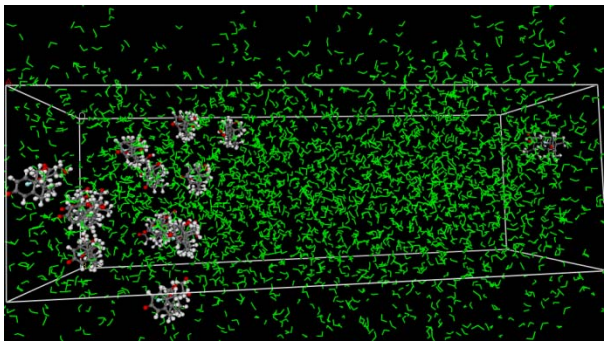


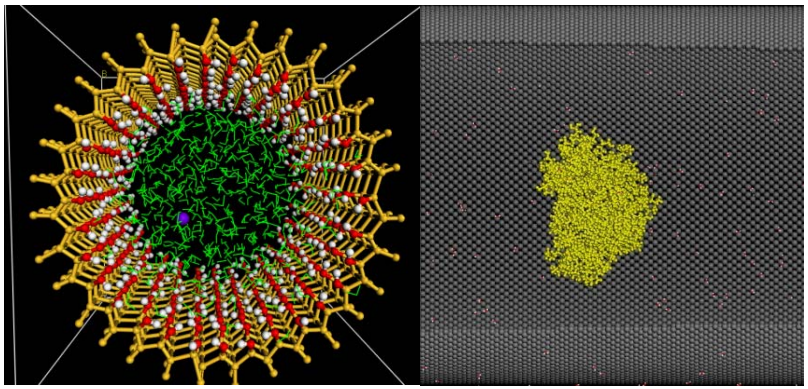
Louisiana Tech University

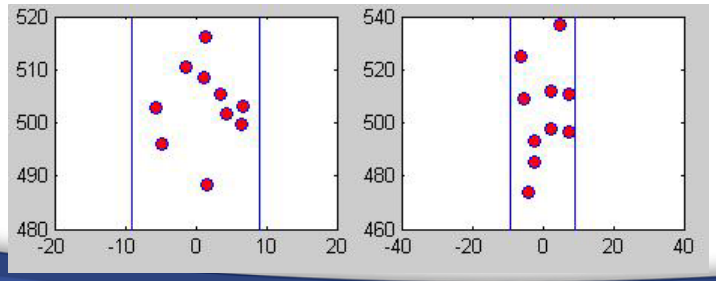
LONI

Computational Study of Diffusion in nanostructures

- Modeling of nanoscale processes require modeling at a variety of space and time scales.
- The development and implementation of reliable multiscale models is the challenge to be faced in the next few years.







- Combination of computational tools working at different scales a valid approach under implementation.
- Atomic quantum mechanics simulations, semiclassical and quantum molecular dynamics and coarse grained Monte Carlo are among the tools being employed.
- Applications include diffusion of species in nanostructure and modeling of structural, mechanical and electronic properties of nanocomposites.

Data Mining and Bioinformatics

Sumeet Dua

Overview

Data Mining

- Dimensionality reduction, data shrinking
- Unsupervised and supervised classification
- Associative theory
- Automated Target Recognition and Tracking

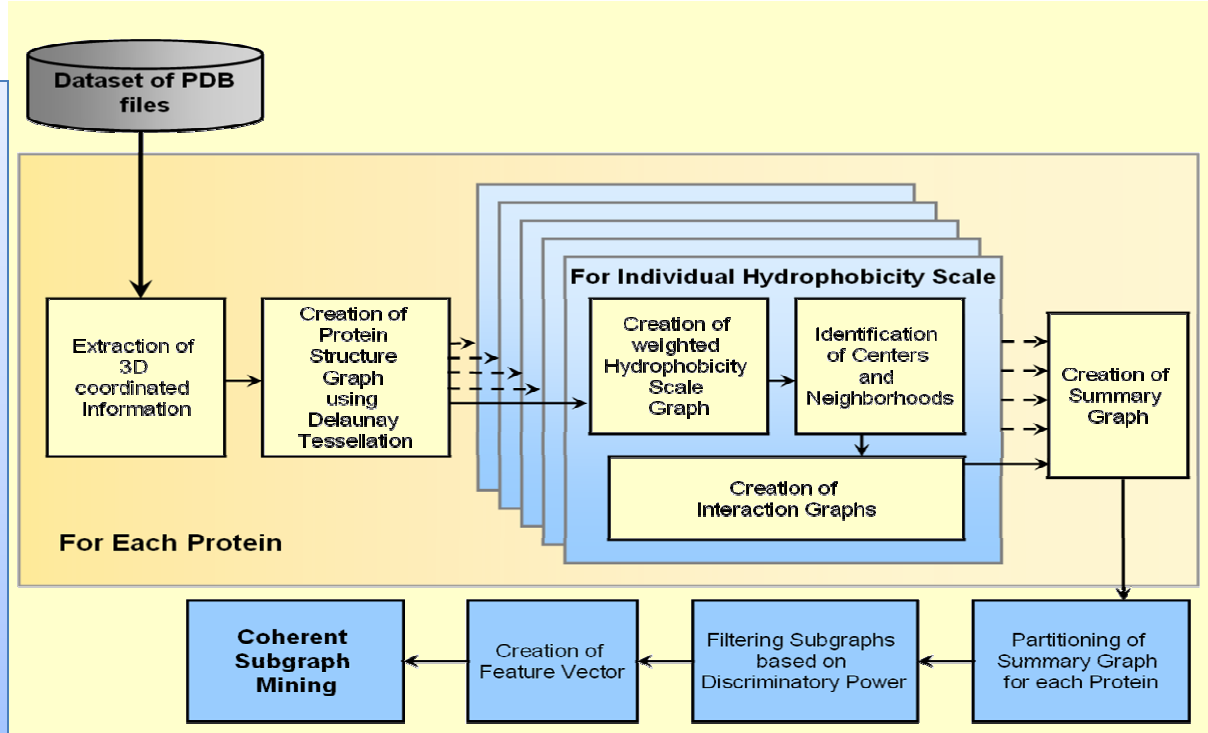
Bioinformatics

- Protein data mining
  - Problems with stereochemical property integration, fold prediction, structural classification, functional annotation, multi-domain proteins
- Microarray gene expression mining
  - Problems with marker gene selection, DR
- Biomedical Image Mining: Image indexing and Feature Modeling

Information fusion: Integration of protein stereochemical properties for classification

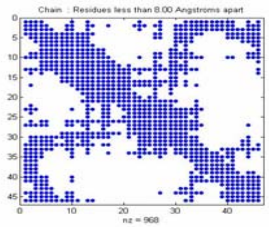
- Protein sequence based tools are not sensitive enough to discover similarity between proteins because of the exponential growth in diversity of sequences.
- We have developed a Graph Theory based Data Mining Framework to extract and isolate protein structural features that sustain invariance in evolutionary proteins.

We have hypothesized that proteins of the same homology contain conserved hydrophobic residues that exhibit analogous residue interaction patterns in the folded state.



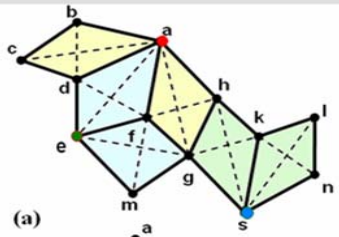
Methodology

(a)

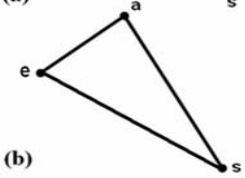


(b)

Creation of Protein Structure Graphs $G_{Hydn}(P)$
Example: SCOP Class Small Protein: 1CRN(A)
(a) Protein Structure Graph using Delaunay Tessellations with distance set at 8 Angstroms
(b) The Adjacency Matrix of the Protein Structure Graph



(a)



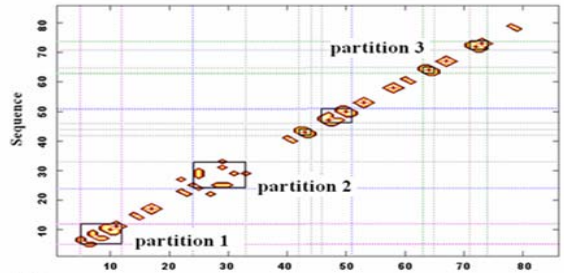
(b)

t_1	a=10	b=7	c=8	d=5
t_2	a=10	e=12	f=9	d=5
t_3	a=10	f=9	g=3	h=6
t_4	e=12	f=9	g=3	m=7
t_5	s=9	g=3	h=6	k=4
t_6	s=9	k=4	l=3	n=5

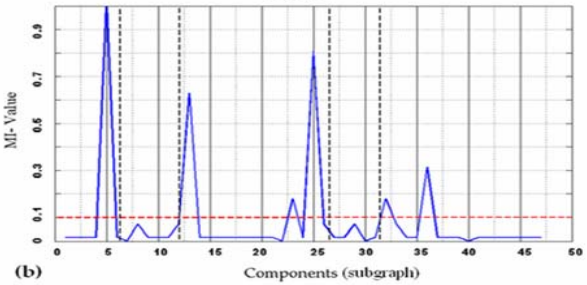
$C(t_1)$=a $C(t_2)$=e $C(t_3)$=a $C(t_4)$=e
$C(t_5)$=s $C(t_6)$=s

$H(a)=\{t_1,t_3\}$ $H(e)=\{t_4,t_2\}$
$H(s)=\{t_5,t_6\}$

Identification of centers and neighborhoods.
a) Representation of the central residue "C" uniquely colored as "e" (green), "a" (red), "s" (blue), and their respective neighborhoods "H" (identified here by the shades of tetrahedra). The table shows each tetrahedron (t), its respective residues, and its respective weights.
b) The resultant Interaction Graph (IG), where a proximity edge is drawn between centers if they share vertices in common.

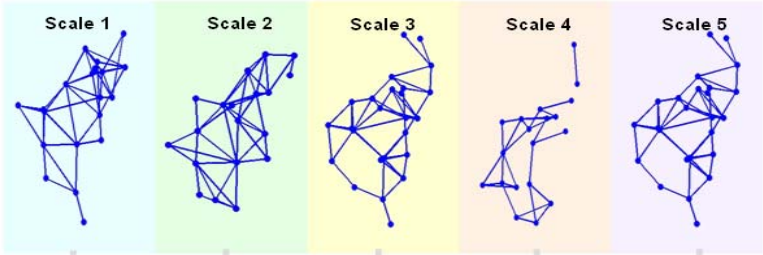


(a)



(b)

Summary Graph of protein 1CRN(A)
a) Identification of connected components
b) Filtering of connected components using Mutual Information



With five different hydrophobicity scales, we obtain a set of five Interaction Graphs *(IG)* representing protein 1CRN(A). The vertices for each *IG* have a common vertex set *V* ($G_{Hydn}(P)$), but possess different edge sets.

Protein Mining *(snapshot of results)*

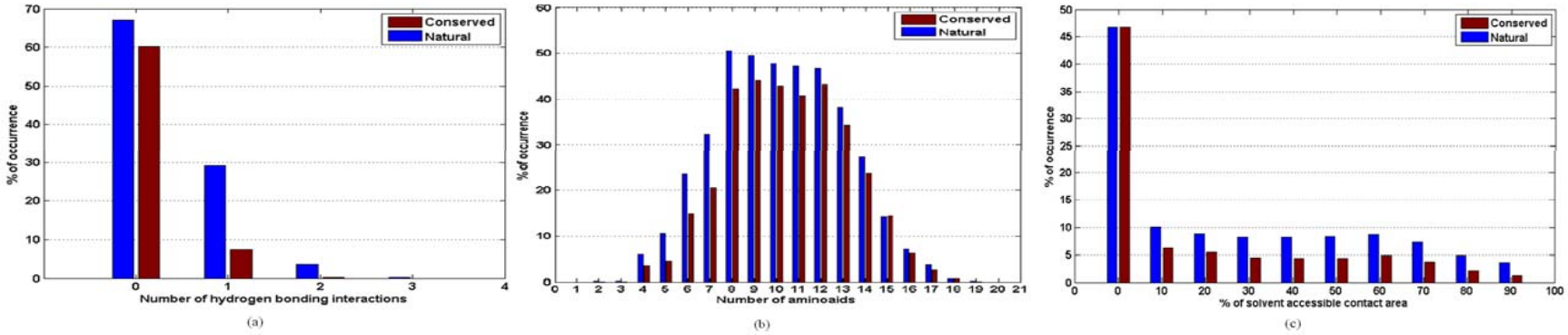
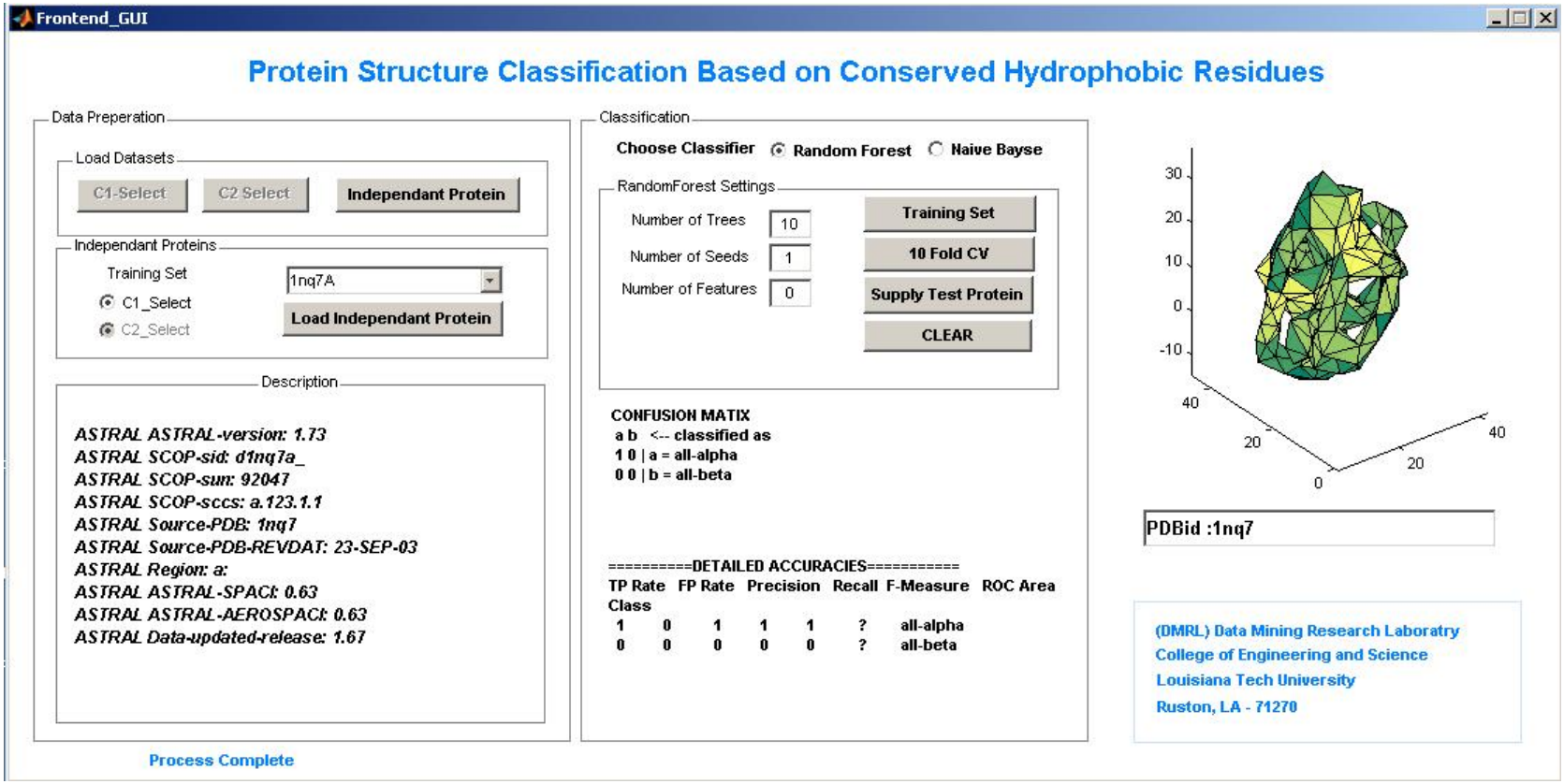


Fig. Composition of amino acids in conserved residues of the summary graphs compared with the entire protein representative set. On the Y-axis is the percentage of amino acids and on the X-axis: a. hydrogen bonding interactions, b. Ooi number in an 8 Å radius around the amino acid and c. solvent accessible contact area as a percentage of residue accessibility.

- Ref.: P. Chowriappa, S. Dua, J. Kanno and H. Thompson, "*Protein Structure Classification Based on Conserved Hydrophobic Residues*", to appear in the *IEEE/ACM Transactions on Computational Biology and Bioinformatics.*
- Ref.: S. Dua, P. Chowriappa and R. Rajagopalan, "*Spectral Coherence Feature Extraction from Stereochemical Scales for Protein Classification*", *under review for IEEE/ACM Transactions on Computational Biology and Bioinformatics.*

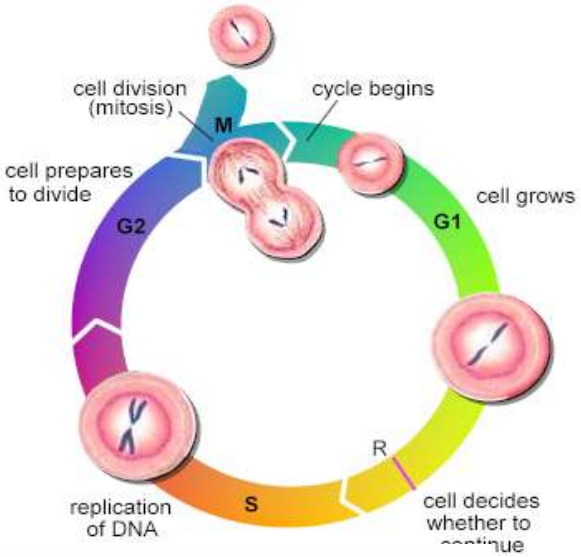
Tool Features



- Provides for the identification of conserved regions within proteins of the same family
- Integration of five physico-chemical properties
- Classification using Random Forest and Naïve Bayes classifier
- Provides for classification of independent proteins into specific classes

Information fusion: Gene Ranking through fusion of Synchronization Experiments

- The *cell cycle,* or *cell-division cycle*, is the series of events that take place in a cell leading to its replication.
- The cell-division cycle is one of the most fundamental processes of life, allowing cells to multiply and faithfully pass on their genetic information to future generations.
- The first critical task in understanding such cyclic systems is to identify the genes that are periodically expressed during the cell cycle – focus of our work.



Our Approach

A. DATA PREPROCESSING
A.1 Interpolation
A.2 Denoising
A.3 Intersecting Gene Sets
A.4 Normalization

→ **B. DATA ALIGNMENT** →

C. INFORMATION FUSION
cdc15 ⊕ cdc28

↓

D. FEATURE EXTRACTION
D.1 Principal Component Analysis
D.2 Skewness Moment
D.3 Kurtosis Moment

←

E. CLASSIFICATION USING GENETIC ALGORITHMS
E.1 Training: Top-ranked (300, 500, 800) genes
E.2 Testing: Published benchmark (B1, B2 and B3: known cyclic genes)

Gene ranking *(snapshot of results)*

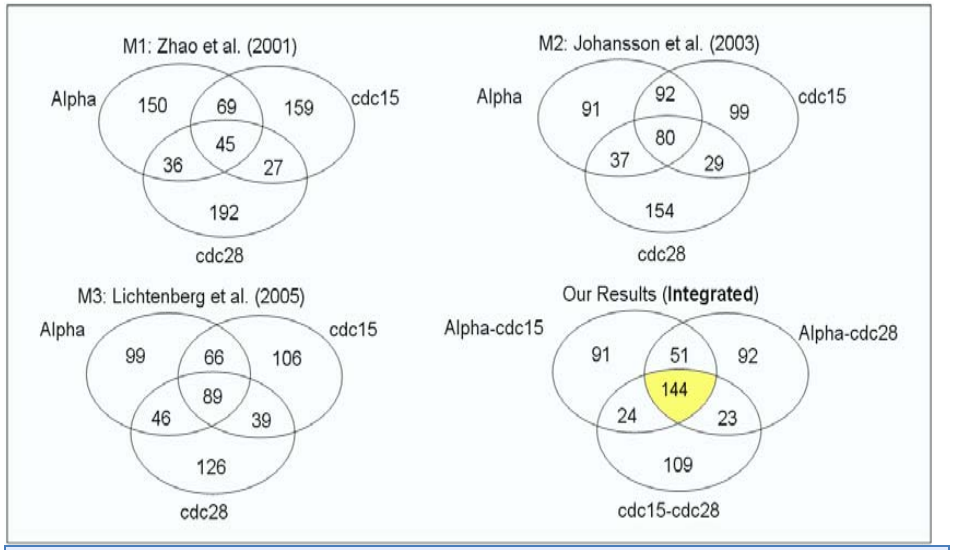


Fig. Agreement across experiments. Venn diagram based on the top 300 genes from each experiment are shown for the methods that provide ranked lists for the individual and integrated experiments.

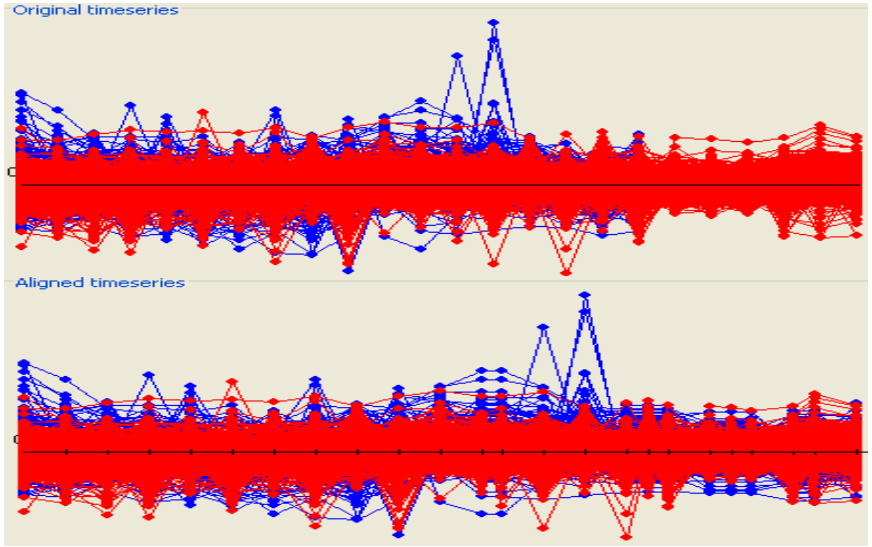
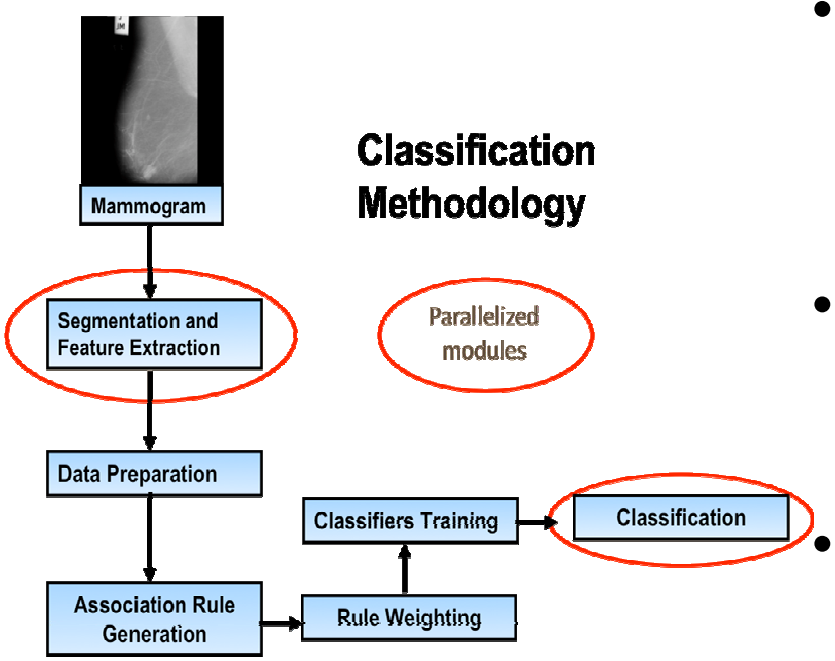


Fig. Data alignment for alpha and cdc15 datasets.

References: A. Alex, S. Dua, P. Chowriappa, "Gene Ranking through the Integration of Synchronization Experiments", to appear in the *Proceedings of 2008 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (IEEE-CIBCB08).*

S. Dua, P. Chowriappa and A. E. Alex; "Ranking through Integration of Protein-similarity for Identification of Cell-cyclic Genes", to appear in the *Proceedings of the Biotechnology and Bioinformatics Symposium (BIOT-2008).*

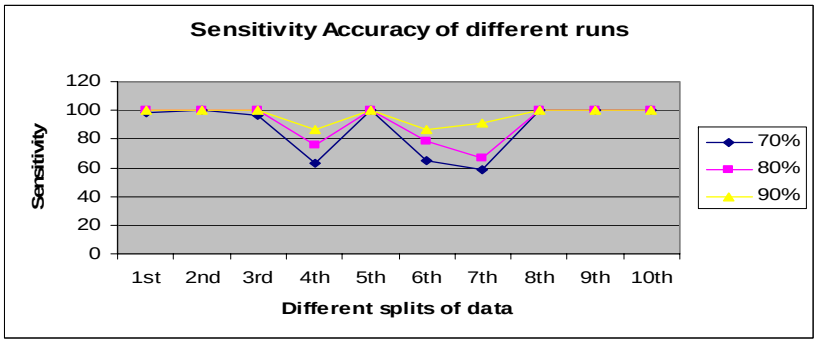
Medical Image Classification using Weighted Association Rules



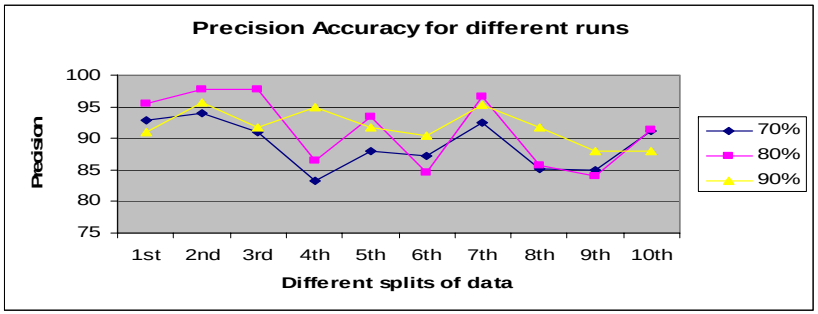
- We have developed a novel method for the classification of medical images (mammograms) using a unique weighted association rule based classifier.
- Isomorphic association rules are derived between various texture components extracted from segments of images,
- These discriminatory rules are then used for the classification through exploitation of their intra- and inter-class dependencies.

- Rigorous experimentation has been performed to evaluate the rules' efficacy under different classification scenarios.
- The algorithm delivers accuracies as high as 89%, which far surpasses the accuracy rates of other rule based classification techniques.

Mammogram classification *(snapshot of results)*



(a)



(b)

The change of Precision (a) and Recall (b) with different percentages of training versus testing data.

Reported Classes

True Classes

	Normal	Benign	Malign
Normal	22	0	0
Benign	1	5	0
Malign	1	0	3

The confusion matrix for three classes considered for classification. The number indicates the number of cases reported.

Reference: S. Dua, V. Jain, H.W. Thompson, "Patient Classification using Association Mining of Clinical Images", Proc. *Of the 5th IEEE International Symposium on Biomedical Imaging (ISBI '08).*

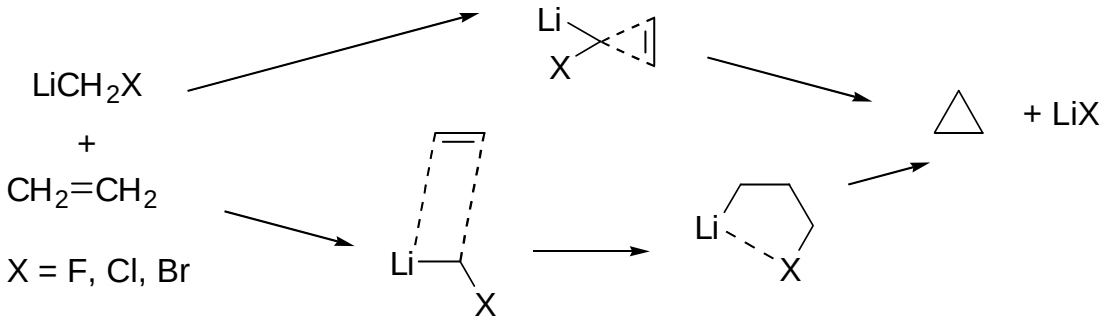
- S. Dua, H. Singh, H.W. Thompson, "Associative Classification of Mammograms using weighted Rules based Classification", under review for *Expert Systems and Applications Journal (*Elsevier).

Organolithium Chemistry

Ramu Ramachandran

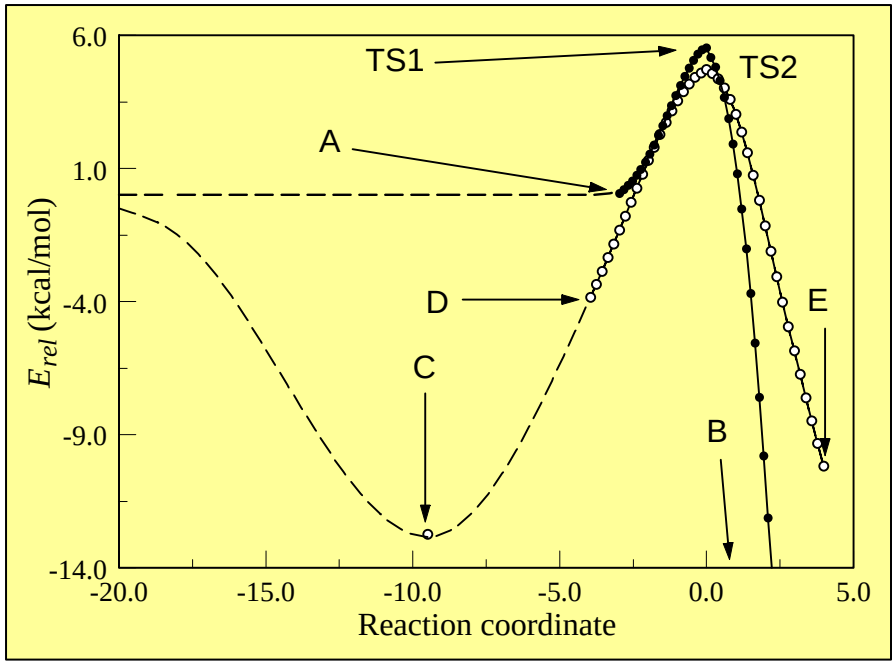
CYCLOPROPANATION REACTIONS OF HALOMETHYLLITHIUM CARBENOIDS

Ramu Ramachandran (with Larry Pratt at Fisk University)



- Experimental evidence suggests that the direct mechanism is favored. But the evidence is not conclusive.
- An open question since the direct mechanism was proposed in 1958, and the stepwise mechanism was suggested in 1962.
- Computational studies, until recently, indicated that the step-wise mechanism is slightly favored, contradicting most experimental findings.
- However, until recently, such studies have ignored the strong tendency of organolithium species to form aggregates.

Reactions of the gas phase dimers – comparison to monomers

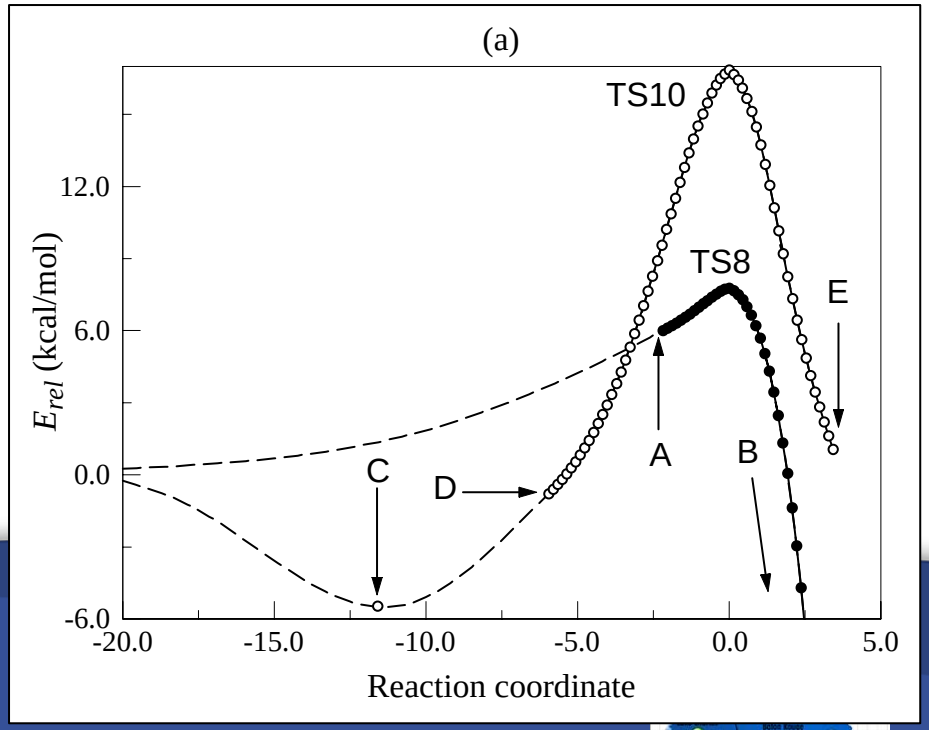


Monomers:

TS1 = direct

TS2 = stepwise

No energetic preference for either path

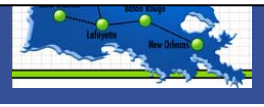


Dimer 6:

TS8 = direct

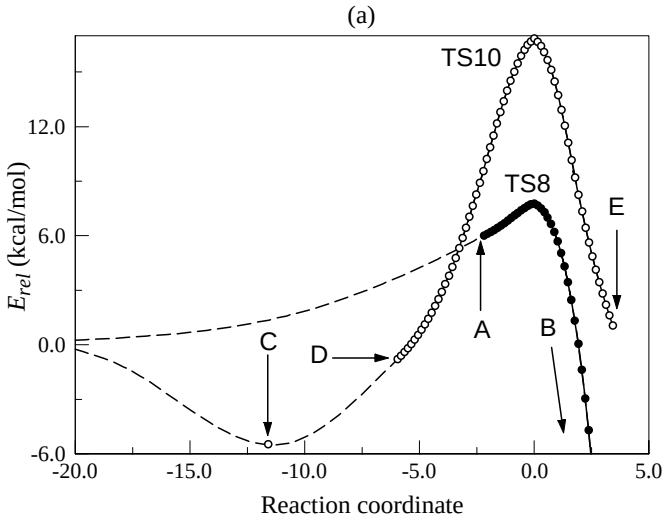
TS10 = stepwise

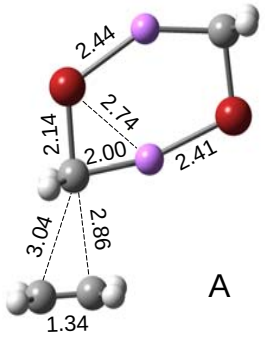
Clear energetic preference for the direct path

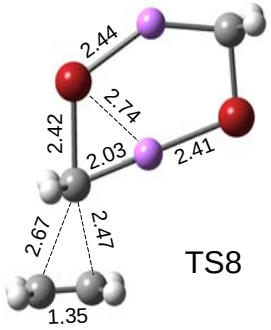


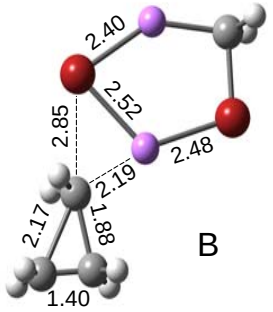
LONI

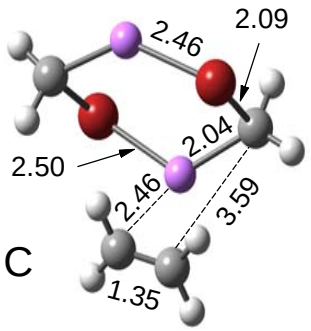
Dimer 6 – structures on the reaction path

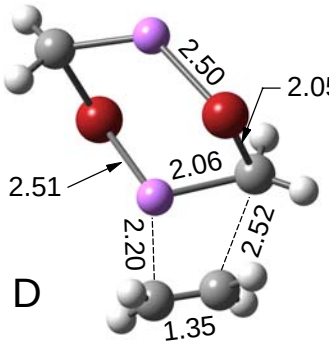


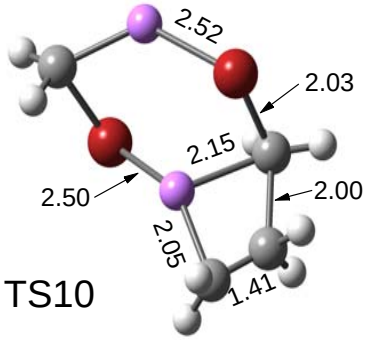


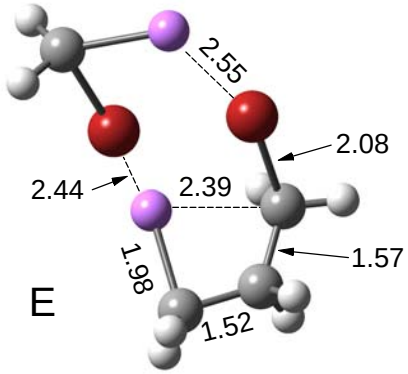








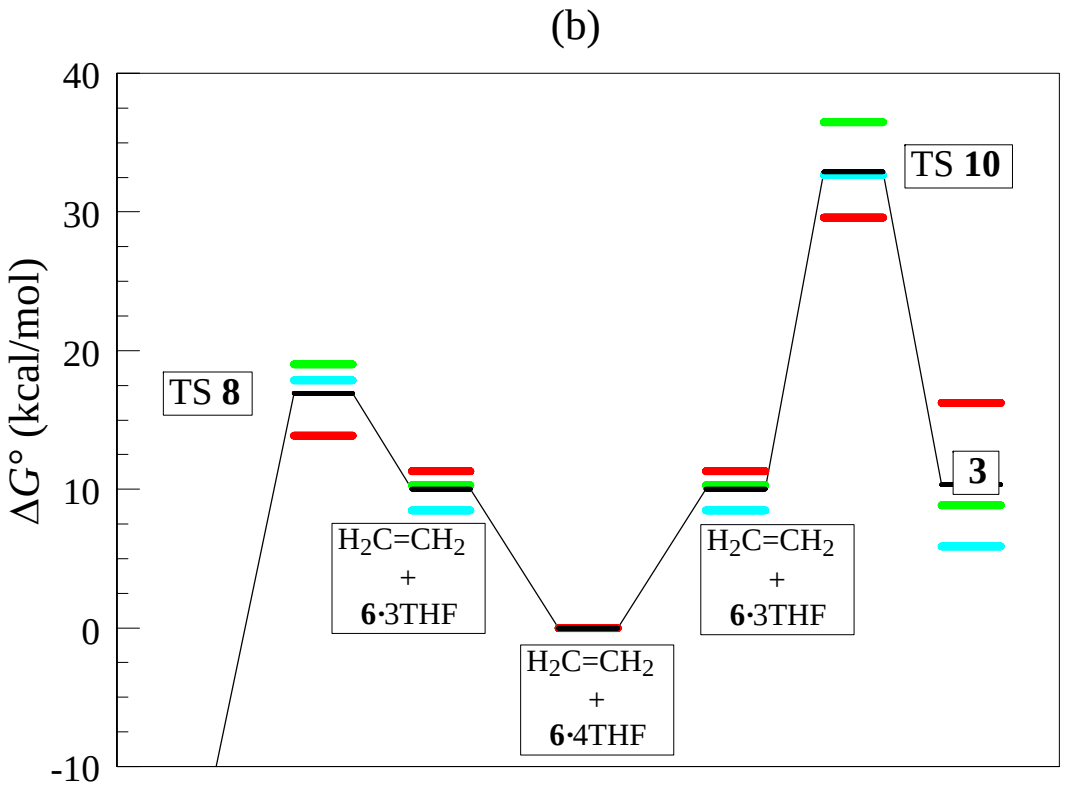




Louisiana Tech University

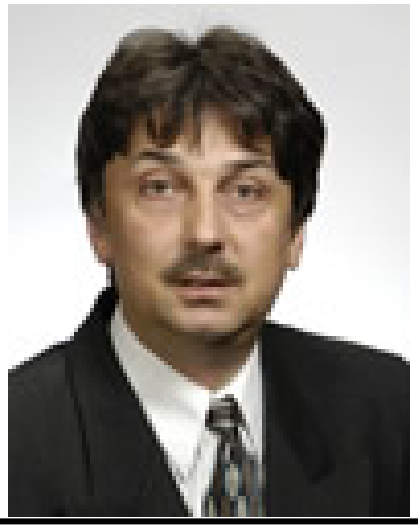
LONI

THF-solvated dimer 6 – free energy profiles



- Direct mechanism is energetically more favorable in THF solution.
- Similar preference also reported in DME by Philip and coworkers.

Finite-Difference Time-Domain Method for Simulations



Neven Simicevic

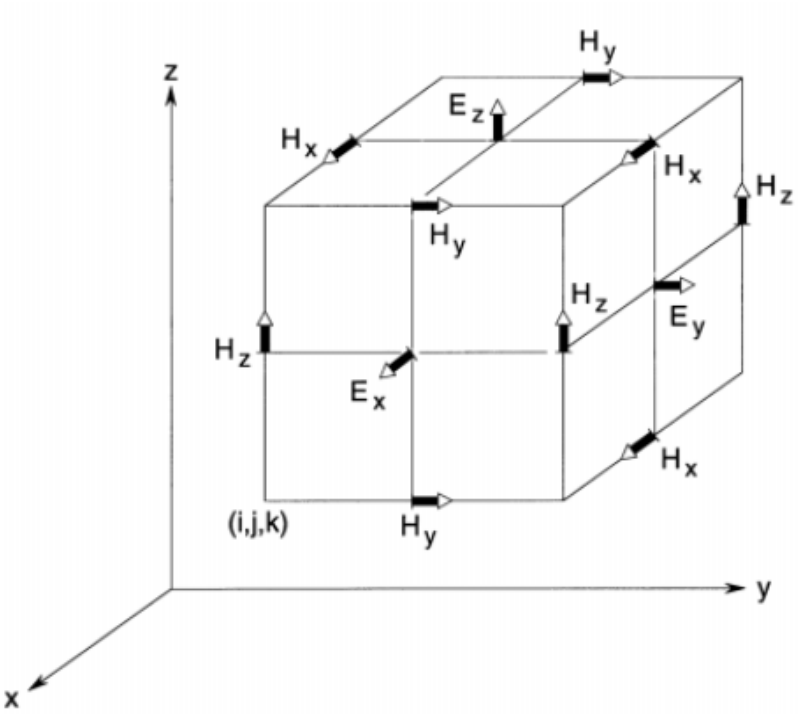
Basis of the Method: Discretize space as well as time

$$\nabla \times \mathbf{H} = \frac{1}{c}\mathbf{J_c} + \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t}$$

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}$$

$$H_{k+1/2}^{n+1/2} = H_{k+1/2}^{n-1/2} + \frac{c\triangle t}{\triangle z}(E_{k+1}^{n} - E_{k}^{n})$$

$$E_{k}^{n+1} = E_{k}^{n} + \frac{c\triangle t}{\epsilon_r(k)\triangle z}(H_{k+1/2}^{n+1/2} - H_{k-1/2}^{n+1/2})$$



Why there is never a big enough computer

$$\Delta x \simeq \frac{v}{10\, f_{max}}$$

$$\Delta t \leq \frac{1}{c\sqrt{(\Delta x)^{-2} + (\Delta y)^{-2} + (\Delta z)^{-2}}}$$

The computation cannot be applied to an arbitrarily large geometry and an arbitrarily short pulses, independently. The slide borrowed from acoustic FDTD shows that the computation time can easily become unreachable.

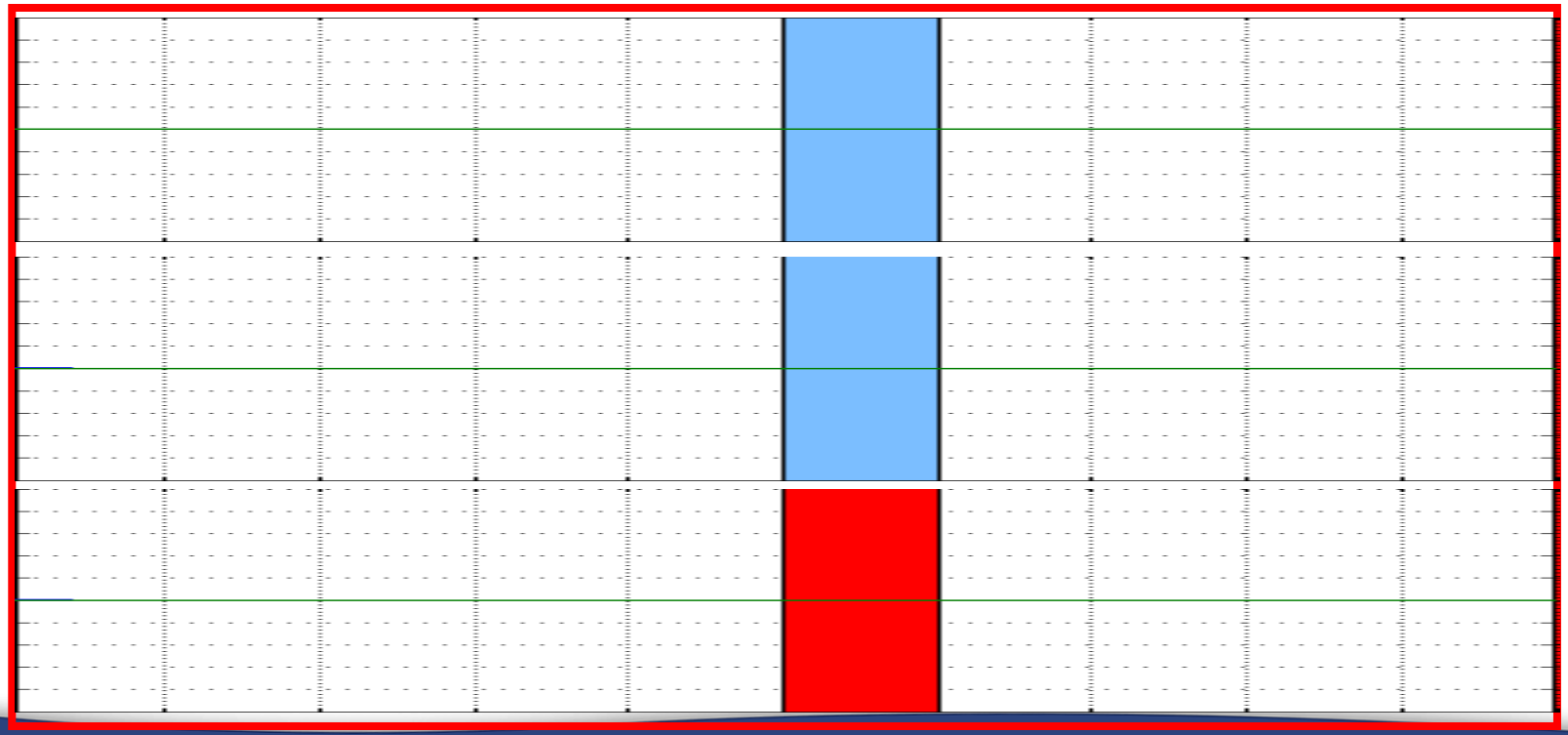
STATE-OF-THE-ART FDTD, 2

	1 kHz	2 kHz	4 kHz	8 kHz
5000 m³	10 GB 1 day	80 GB 16 days	640 GB 256 days	4.8 TB 11 yrs
40000 m³	80 GB 2 days	640 GB 32 days	4.8 TB 512 days	
160000 m³	640 GB 4 days	4.8 TB 64 days		300 TB 44 yrs

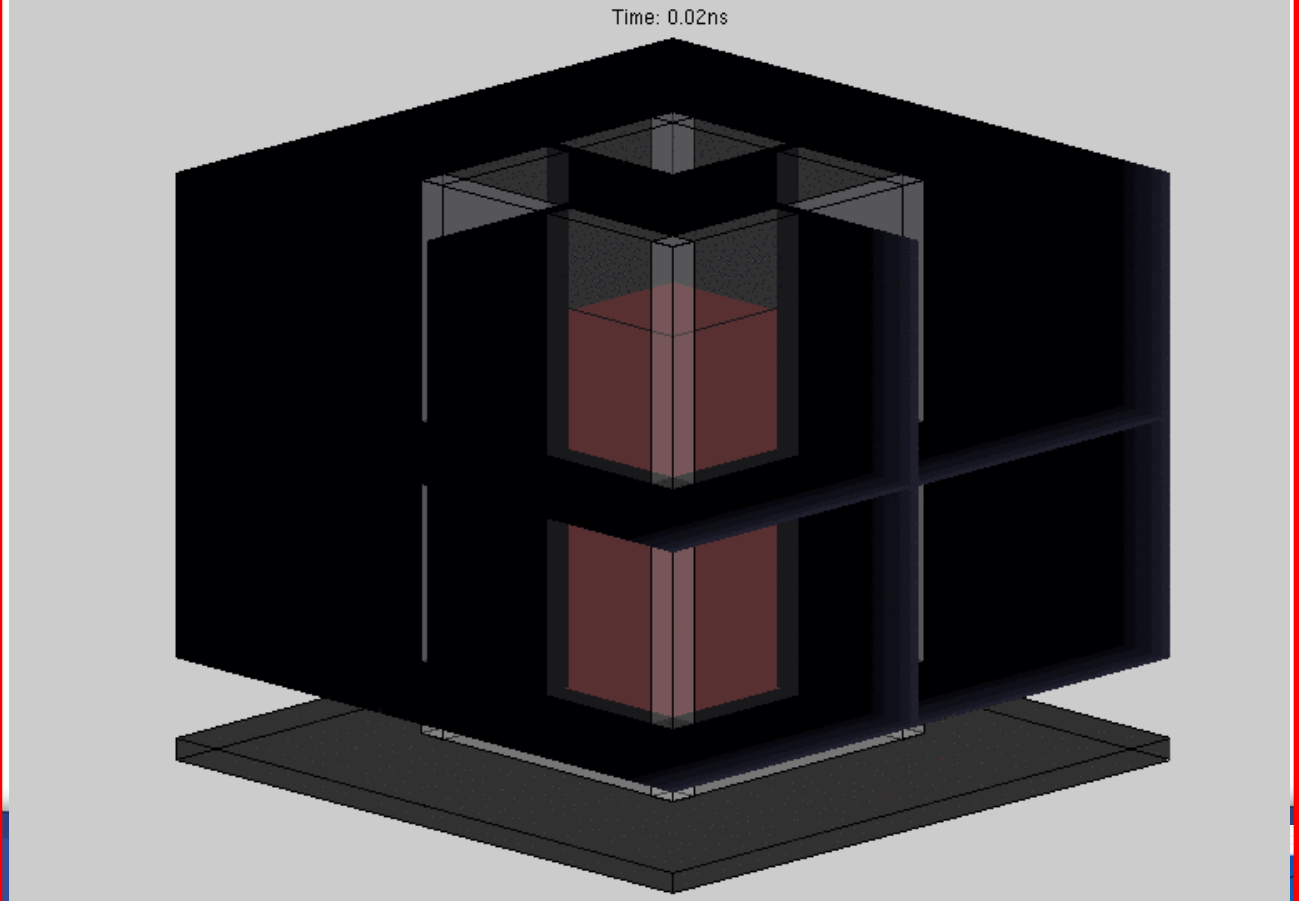
U. Peter Svensson – Norwegian University of Science and Technology

Proof of principle: EM pulse interaction with material

Interaction of 1-dimensional Gaussian pulse with pure water, water having conductivity of the blood, and blood. Part of the pulse energy is reflected, part is absorbed, and part of it is re-radiated into the space.

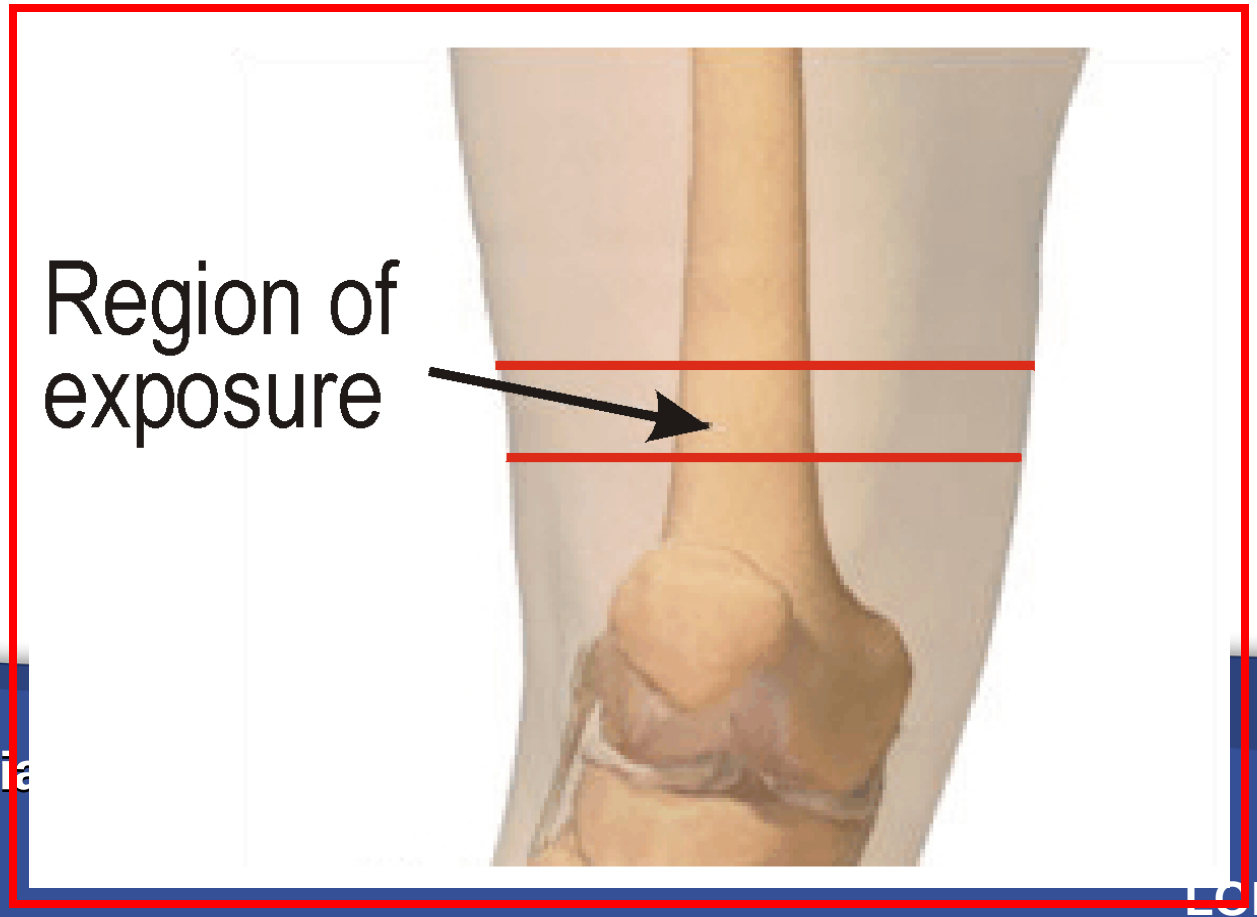


On a larger scale



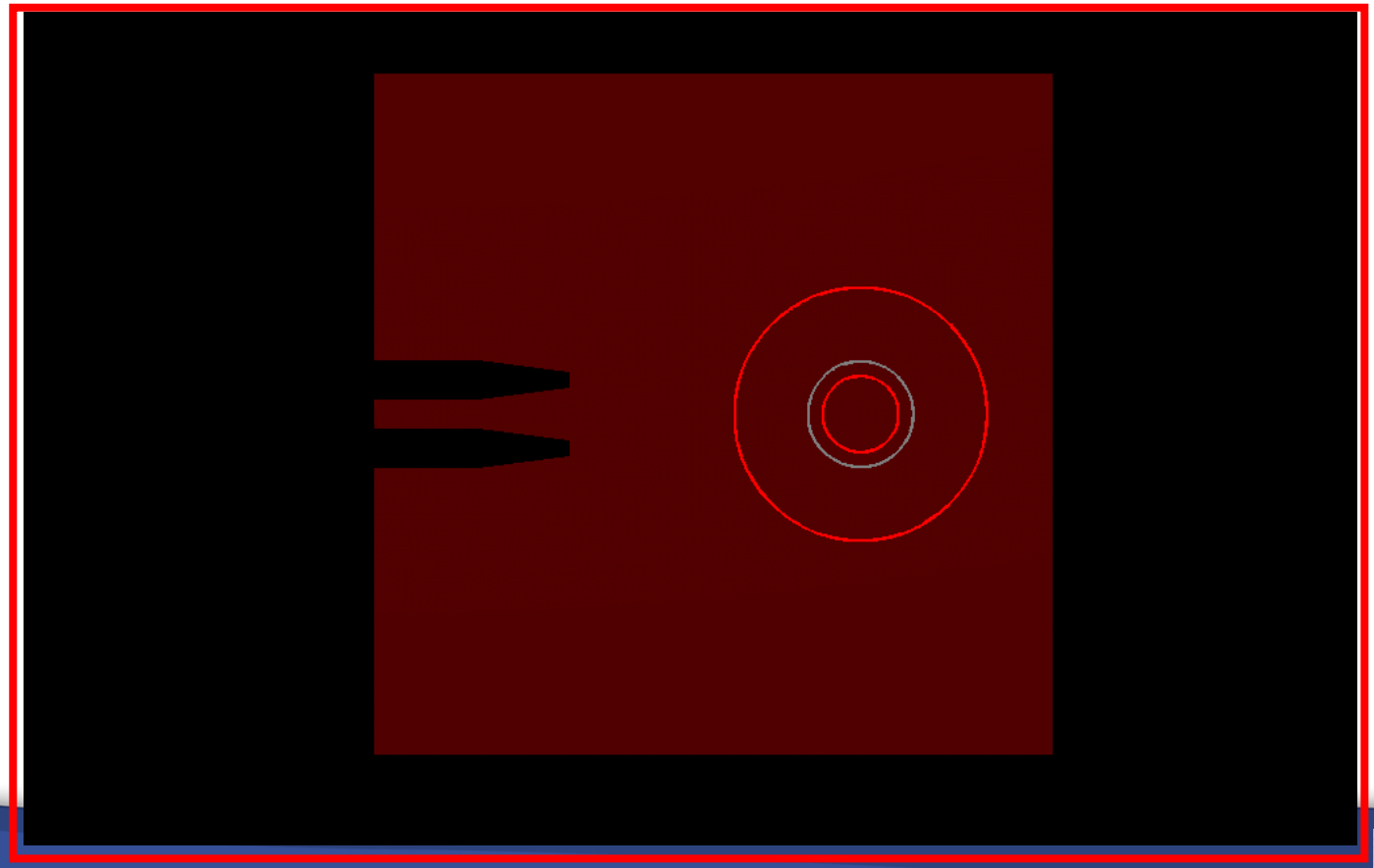
Application in Medicine

Numerical experiment: partial body exposure to UWB Gaussian pulse.



Exposure to UWB pulse

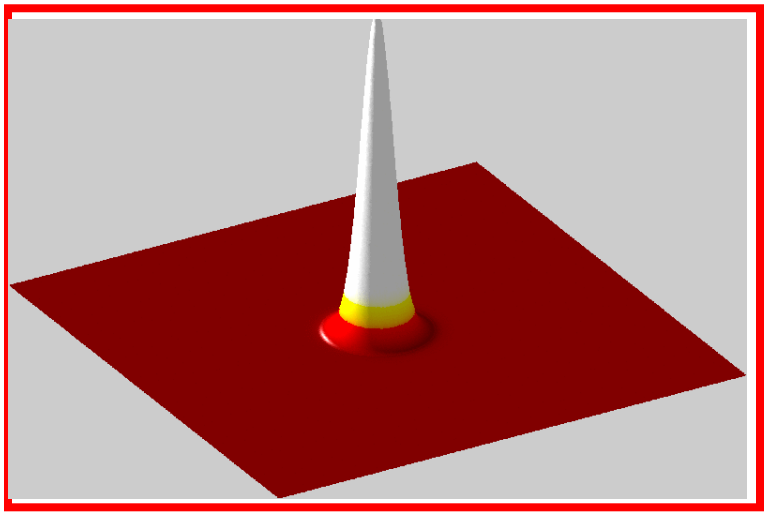
Propagation of the UWB pulse through exposed sample.

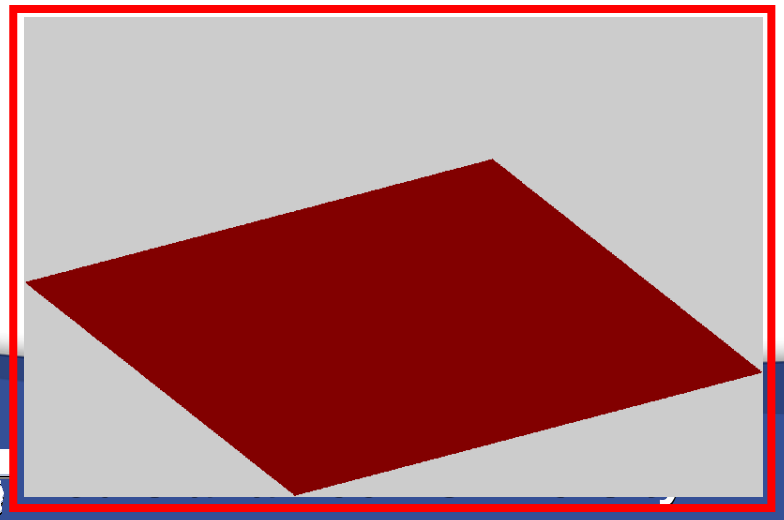


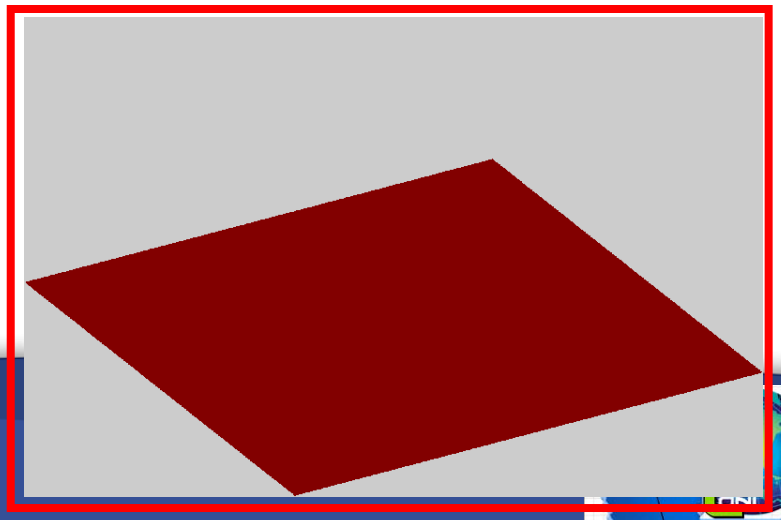
More: FDTD for Quantum Mechanics

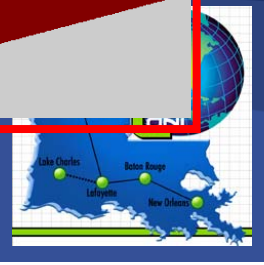
$$\begin{aligned}\Psi_1^{n+1/2}(I,J,K) &= \frac{2-C^n(I,J,K)}{C^n(I,J,K)}\Psi_1^{n-1/2}(I,J,K) - \frac{c\Delta t}{2\Delta x C^n(I,J,K)}\left[\Psi_3^n(I,J,K+1) - \Psi_3^n(I,J,K-1)\right. \\ &+ \left.\Psi_4^n(I+1,J,K) - \Psi_4^n(I-1,J,K) - i(\Psi_4^n(I,J+1,K) - \Psi_4^n(I,J-1,K))\right] \\ &+ i\frac{e\Delta t}{\hbar C^n(I,J,K)}\left[A_1^n(I,J,K)\Psi_4^n(I,J,K) - iA_2^n(I,J,K)\Psi_4^n(I,J,K) + A_3^n(I,J,K)\Psi_3^n(I,J,K)\right]\end{aligned} \tag{11}$$

FDTD solutions of the Dirac equation: Relativistic treatment of electron motion.



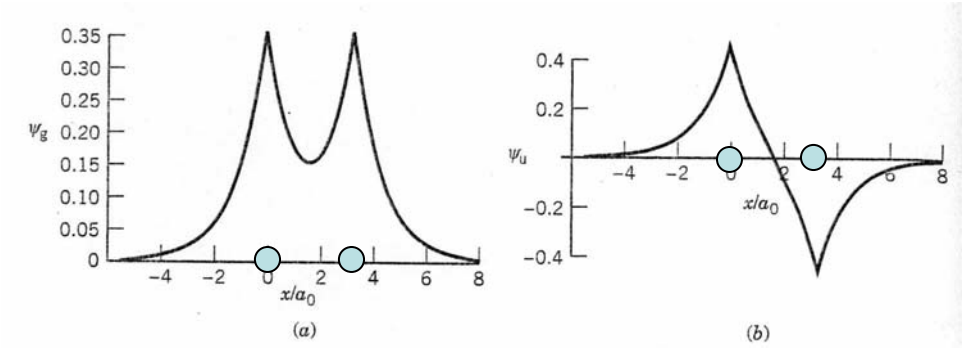




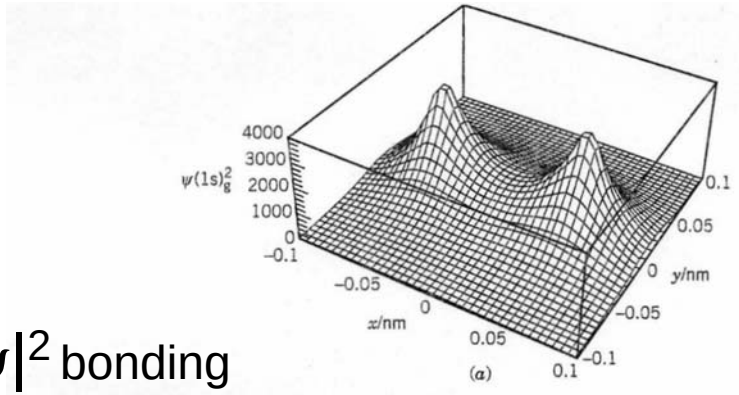


LONI

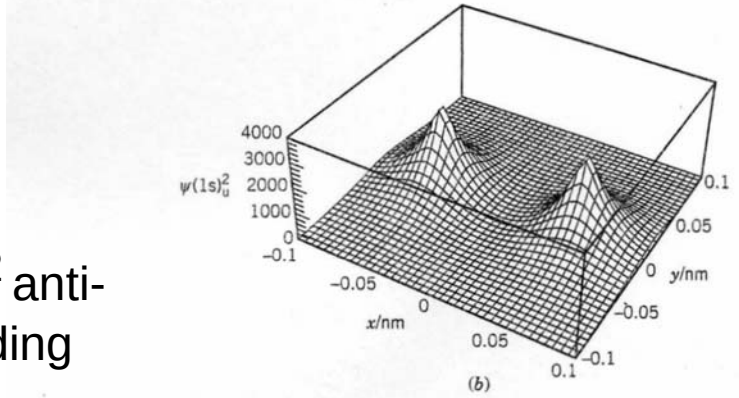
Non-relativistic model of the simplest chemical bond: H_2^+



$|\psi|^2$ bonding



$|\psi|^2$ anti-bonding



Plot of the electron density in the xy plane for the hydrogen molecule ion H_2^+ with (a) a 1s sigma bonding orbital and (b) a 1s sigma antibonding orbital. Note the buildup of electron density between the nuclei with the bonding orbital. Also note that with the antibonding orbital the electron density is zero along a line perpendicular to the line between the nuclei and halfway between the nuclei. The internuclear axis is along x.

Bonding and anti-bonding molecular orbitals.

Static.

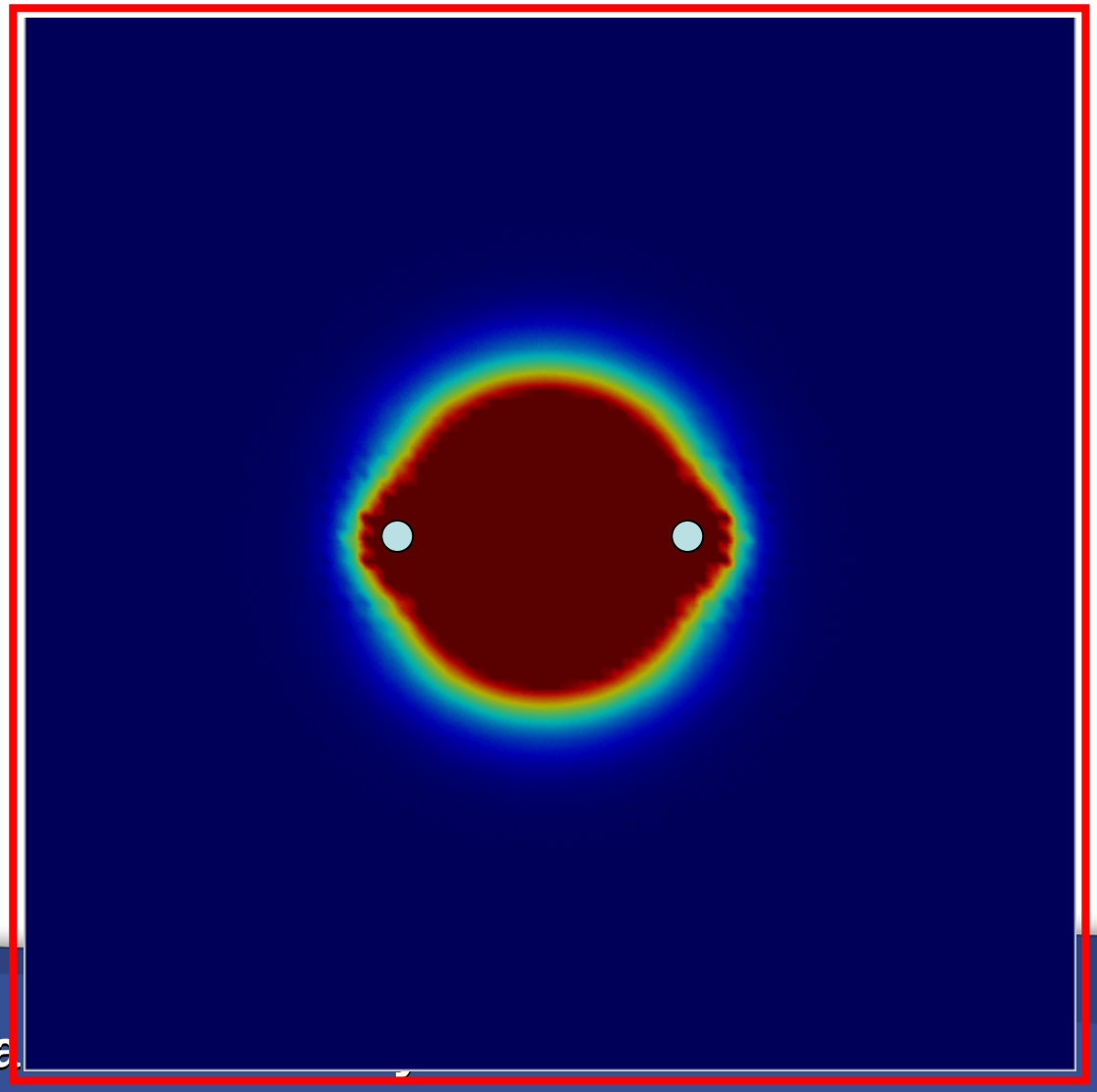
Basis of almost all of quantum chemistry.

But … what is an electron in such a chemical bond really doing, as a function of time?

- You are among the first 50 - 200 people in the world to see this …

- The first was Neven, of course.
- The second was me, and I was here in Baton Rouge, at the Cook Hotel at the time!
- The third was probably Les Guice.

The electron at work, in the simplest chemical bond



Thank You!


Shreveport
Monroe
Ruston
Alexandria
LONI
Lake Charles
Baton Rouge
Lafayette
New Orleans