The GIT Molecular Simulations Discussion Group presents,

Normal Mode Analysis
of
Macromolecules
--Expanded Version--

Jason Mears

Russell Hanson

June 8, 2004

Normal mode math

$$|\nabla^2 E - \lambda \mathbf{M}| = \mathbf{0} \tag{1}$$

$\nabla^2 E$:= the Hessian, matrix of second derivatives w.r.t. time
$\mathbf{M}$:= diagonal mass matrix of macromolecule

$$|\mathbf{M}^{-1/2}(\nabla^2 E)\mathbf{M}^{-1/2} - \lambda \mathbf{I}| = \mathbf{0} \tag{2}$$

λ := eigenvalues proportional to the squares of the vibrational frequencies

$$[M]\{\ddot{x}\} + [K]\{x\} = \{0\} \tag{3}$$

$[M]$:= mass matrix of macromolecule
$[K]$:= stiffness matrix; second derivatives of potential energy of molecule
$\{x\}$:= displacement vectors of all atoms from their equil. positions
$\{\ddot{x}\}$:= second derivatives w.r.t. time
Let $\{x\} = \{\chi \sin(\omega t)\}$; χ are normal mode variables, ω are circular frequency variables.

$$\{[K] - \omega^2[M]\}\{\chi\} = 0 \tag{4}$$

Solving Eq. (4) yields natural frequencies and corresponding normal mode vectors. The harmonic dynamics of macromolecular system are fully described thus.

Approximate potential energy function by harmonic modes around minimum energy conformation. By diagonalizing the Hessian matrix of mass-weighted second derivatives of the potential energy arrive at analytical solution to equations of motion.

Eigenvectors are the normal modes; eigenvalues are the squares of the associated frequencies.

(M.H. Hao and S. Harvey, 1992, Biopolymers)

Harmonic approximation

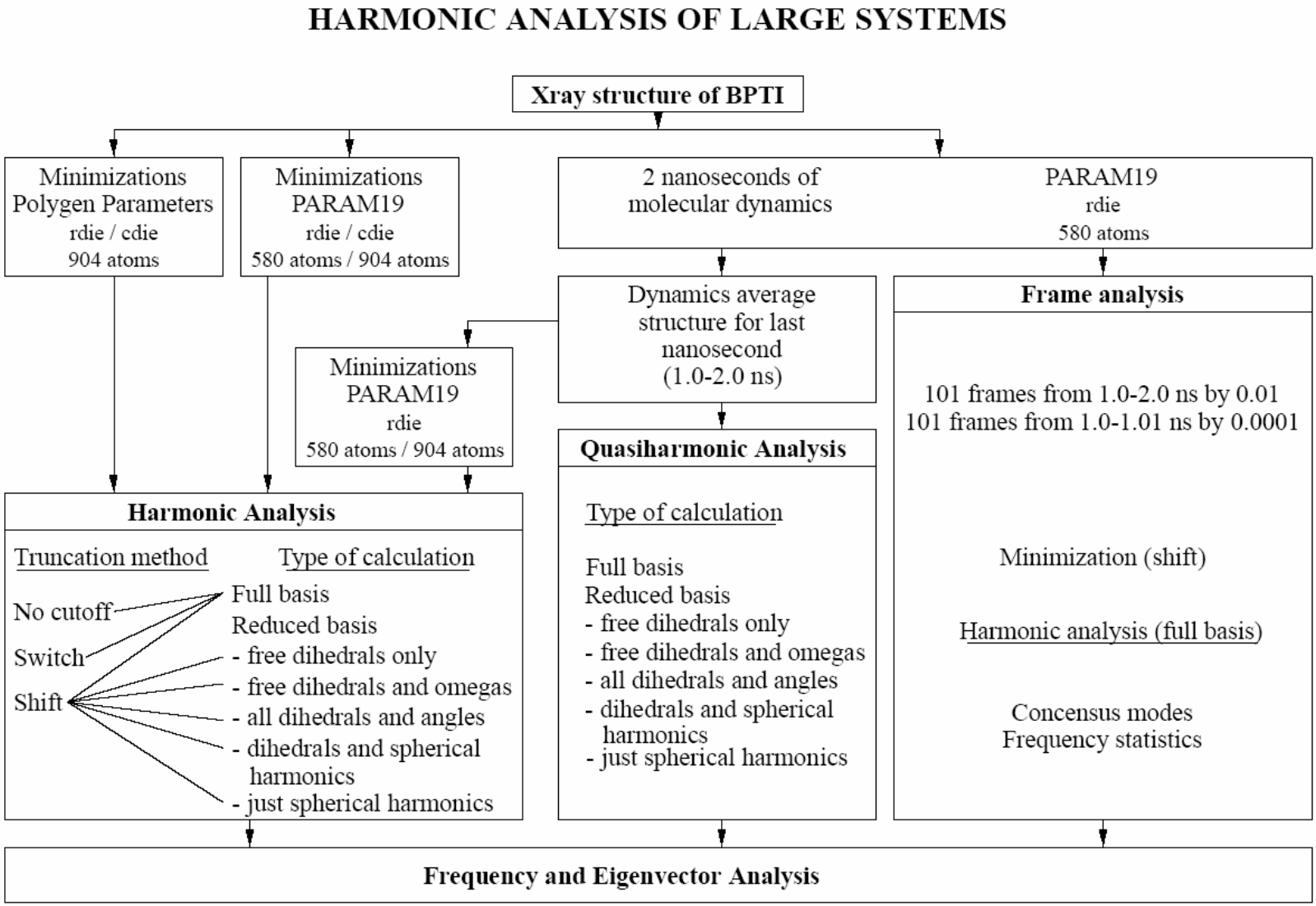
- Def: **Harmonic Approximation:** Assumes that potential energy function can be approximated as sum of quadratic terms in displacements.
- Coefficients of these terms are:
  - 1) Force constant matrix &
  - 2) Atomic masses.
- Matrix equation of molecular vibrational modes

(B. Brooks & M. Karplus, J Comp Chem, 1995, Harmonic Analysis of Large Systems. I. Methodology)

Energy function contributions, etc.

- Types of contributions to energy functions found in Hessian matrix:
  - 1) Diagonal interaction (atom with self)
  - 2) Close interactions (atom connected by bond or angle term)
  - 3) Long-range interactions (VdW and EM)
  - 4) Additional close-range interactions with assoc. long-range term (1-4 dihedrals and H-bonding)
  - 5) Zero interactions (atom pairs beyond long-range cutoffs)
- For any pair interaction, up to 9 contributions to the second derivative matrix *(dx_i,dy_i,dz_i* with any of *dx_j, dy_j, dz_j)*, however, can calculate all 9 from two magnitudes and two angles defining direction. $\partial E/\partial r$ and $\partial^2 E/\partial r^2$
- Matrix-specific methods
  - Gram-Schmidt, tridiagonalization, etc.

Large systems techniques



(Janežič, 2002, Cell. Mol. Biol. Lett.)

Memory requirements

- Matrix storage size: set of basis vectors, size M x $3N$, N is number of atoms, M is number of modes.

- `1eul.pdb calcium ATPase:`
- `(50 modes) x 3(7672 atoms)`
  `= 1 150 800 elements`
- `Double-precision floating point uses 8 bytes`
- `8 * 276 192 = 9 206 400 bytes`

(Hinsen, 1998, Proteins)

http://dasher.wustl.edu/tinker/

Tinker's vibrate.f – Hessian memory requirements

```
 Current Atom :      2      Total Atoms :    642
 Current Required Hessian Storage :        11535
 Maximum Allowed Hessian Storage :         10000
 Minimum Significant Hessian Value :     0.00000

 Current Atom :     18      Total Atoms :    642
 Current Required Hessian Storage :       102519
 Maximum Allowed Hessian Storage :        100000
 Minimum Significant Hessian Value :     0.00000

 Current Atom :    207      Total Atoms :    642
 Current Required Hessian Storage :      1002915
 Maximum Allowed Hessian Storage :       1000000
 Minimum Significant Hessian Value :     0.00000

>>> 1002915/207
4845
>>> 642 * 5500
3531000

 HESSIAN  --  Increase MAXHESS and/or HESSCUT
```

Do you love statically-allocated Fortran programs as much as I do?

Tinker's vibrate.f – heme.xyz

```
heme.xyz 73 atoms

 Enter Cartesian Coordinate File Name :  heme.xyz

 Atoms with an Unusual Number of Attached Atoms :

 Type           Atom Name      Atom Type       Expected    Found

 Valence          1-FE          107              6          4
 HESSIAN  --     23871 Elements   100.00 % Off-diag H    0.48 % Storage

 Eigenvalues of the Hessian Matrix :

    1    0.000     2    0.000     3    0.000     4    0.000     5    0.000
    6    0.000     7    0.092     8    0.150     9    0.344    10    0.412
   11    0.528    12    0.590    13    0.808    14    0.883    15    0.965
   16    1.109    17    1.401    18    1.764    19    2.403    20    2.667

 Vibrational Frequencies (cm-1) :

    1    0.057I    2    0.000     3    0.001     4    0.001     5    0.011
    6    0.030     7   12.819     8   22.509     9   27.896    10   30.731
   11   31.492    12   36.431    13   39.879    14   42.884    15   52.237
   16   55.964    17   63.268    18   79.309    19   80.520    20   90.178
```

Tinker's vibrate.f – peptide.xyz

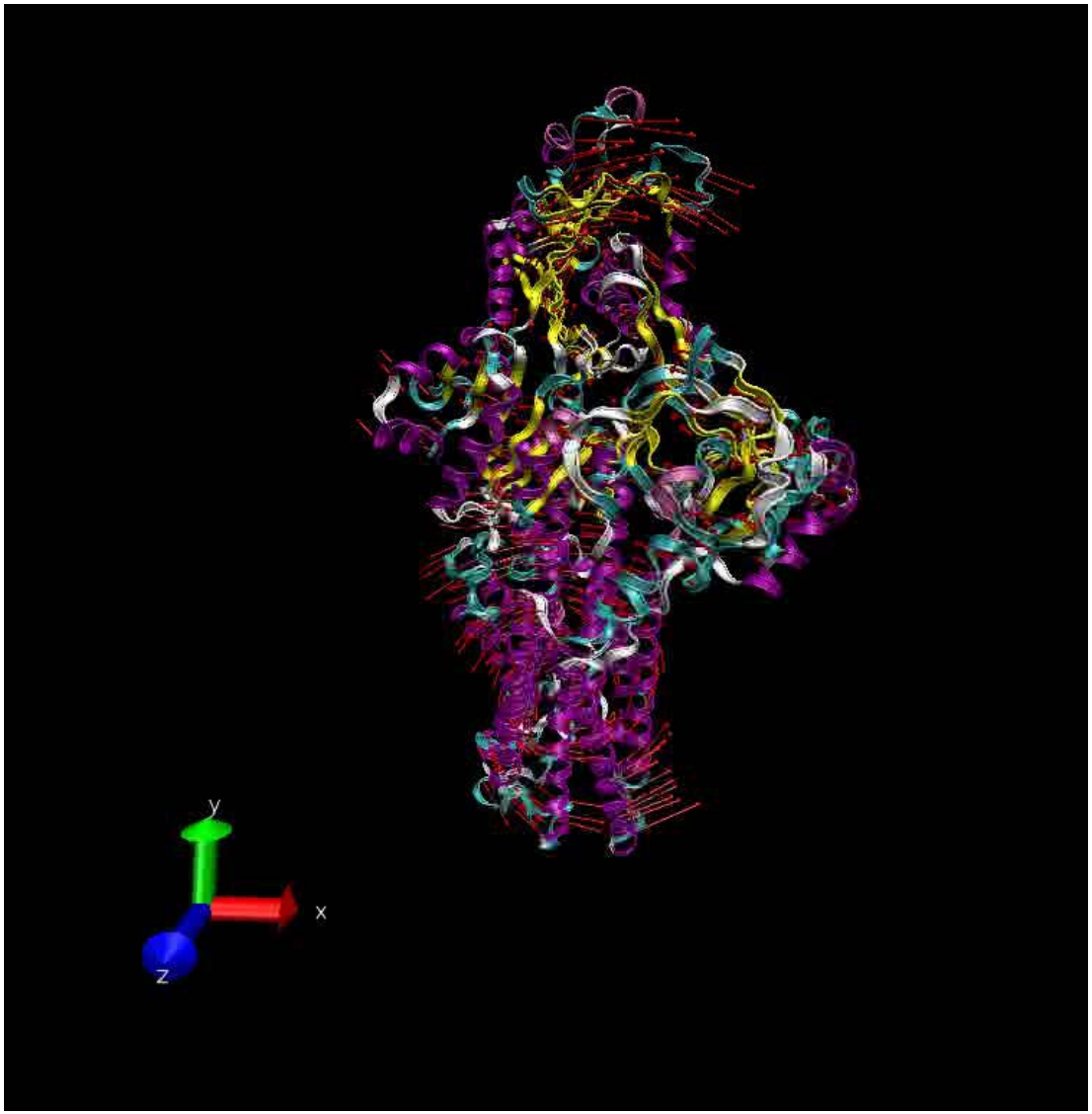
```
peptide.xyz 328 atoms:

 Enter Cartesian Coordinate File Name :  peptide.xyz
 HESSIAN  --     483636 Elements   100.00 % Off-diag H    9.67 % Storage

 Eigenvalues of the Hessian Matrix :

    1*********     2*********     3*********     4*********     5*********
    6*********     7*********     8*********     9-3506.064    10-2387.767
   11-2234.117    12-2199.782    13-1511.355    14 -896.670    15 -398.115
   16 -297.187    17 -277.633    18 -260.510    19  -94.945    20  -57.980

 Vibrational Frequencies (cm-1) :

    188500.966I    288451.359I    375889.056I    470912.131I    541313.948I
    625565.421I    715218.648I    812210.608I    9 4070.718I   10 3906.274I
   11 3444.554I   12 2427.541I   13 2308.384I   14 1562.948I   15 1020.599I
   16  663.011I   17  653.237I   18  587.842I   19  570.628I   20  307.442I
```

Interpretation of low and high frequencies

- 1) Global domain motions have no energy contribution from internal degrees of freedom of the domains because there is no deformation.
- 2) Long-range interactions between domains are weaker than short-range interactions between neighboring atoms.
- -> High-frequency modes are localized motions involving few atoms
- -> Low-frequency modes represent global movements of large domains

(Hinsen, 1998, Proteins)

1EUL.pdb Ca^{2+}ATPase



Tinker's vibrate.f - implementation

- 0) establish potential force field parameters
- 1) calculate the Hessian matrix of second derivatives
- 2) store upper triangle of the Hessian in "matrix"
- 3) perform diagonalization to get Hessian eigenvalues
- 4) store upper triangle of the mass-weighted Hessian matrix
- 5) diagonalize to get vibrational frequencies and normal modes
- 6) form Cartesian coordinate displacements from normal modes
- 7) print the vibrational frequency and normal mode

Subspace methods and free energy and entropic effects

- Constrain degrees of freedom: backbone dihedral angles ϕ and ψ; Fourier basis space
- Basis vectors of the subspace are not coordinates but coordinate differentials—each basis vector describes a direction in 3*N*-dimensional coordinate space. Basis vector regarded as set of atomic displacement vectors.

$$\mathbf{d}_I = \mathbf{D}(\mathbf{R}_I), \qquad (1)$$

where $\mathbf{R}_I$ is the position of atom i and $\mathbf{d}_I$ is its displacement vector. Obviously, there is more than one vector field $\mathbf{D}(\mathbf{r})$ corresponding to a given set of displacement vectors $\mathbf{d}_I$, although the inverse relation is unique. Because the vector field $\mathbf{D}(\mathbf{r})$ has no

A precise specification of this normal mode subspace basis is given by the vector fields

$$\mathbf{B}_\alpha^{ijk}(\mathbf{r}) = w(x, k_i^{(x)})\, w(y, k_j^{(y)})\, w(z, k_k^{(z)})\, \mathbf{e}_\alpha, \qquad (2)$$

where $\mathbf{e}_\alpha$, $\alpha = x, y, z$ is a unit vector along one of the three Cartesian axes and

$$w(x, k) = \begin{cases} \sin(kx) \text{ for } k < 0 \\ \cos(kx) \text{ for } k \geq 0. \end{cases} \qquad (3)$$

- Comparison to free energy results:

"Moreover, the fact that the simplified protein model is able to reproduce the low-frequency modes of large proteins rather well explains why normal mode analysis, despite its exploration of only a single local energy minimum of the configurational space of the system, can make meaningful predictions for the system in its real physiological environment. **Such environments have temperatures at which entropic effects are not negligible, and hence the relevance of studying minima of potential energy is questionable.** Instead, the free energy as a function of slow variables should be analyzed. As explained in this article, the simplified protein model can in fact be regarded as a crude approximation to the free energy as a function of residue positions. **Because such a model produces essentially the same low-frequency motions as an atomic model with a potential energy surface, it can be concluded that the neglect of entropic effects in standard normal mode analysis has no important consequences as far as domain motions are concerned.**" (Hinsen, 1998, Proteins)

Local minima of potential energy

"The implication of these observations for the energy landscape of proteins is that the multiple local minima of the potential energy in the subspace of low-frequency motions and the corresponding smoothed-out minima of the free energy profile must have similar shape. This shape is essentially determined by the condition that deformations should be limited to small regions and/or regions with a low atom density, because a low atom density implies a lower energetic cost of deformations." (Hinsen, 1998, Proteins)

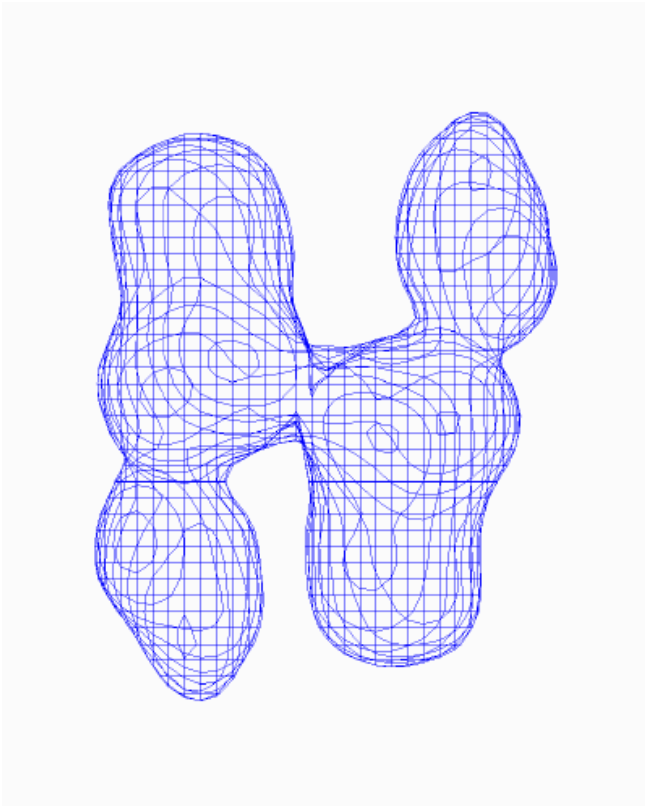
Quantized elastic deformational model (QEDM)

- Allows calculate normal modes based on low-resolution (20--30 Å) cryo-EM density maps without atomic coordinates or amino acid sequence.
- Ma: “The success of the initial study of QEDM-assisted refinement procedure demonstrates the potential of improving the resolution of the final reconstruction in single-particle cryo-EM by dividing the particle images into more homogeneous particle subsets in terms of molecular conformations.”
- Substructure synthesis method (SSM): determine modes at very long length scales; determine substructure modes; link substructures together; enforce geometric compatibility at interfaces of neighboring structures; eigenvalue problem on smaller substructures
- As harmonic approximation, presumably cannot overcome energy barrier separating two states—how reveal trajectories of motion?

Jianpeng Ma (2004). New Advances in Normal Mode Analysis of Supermolecular Complexes and Applications to Structural Refinement. *Curr. Prot. Pept. Sci.* **5:** 119-123

Quantized elastic deformational model (QEDM) (p2)

- Fatty Acid Synthase (FAS) – application of QEDM to a 19 Å cryo-EM density map, revealing deformational modes.
- Using multi-copy x-ray crystallographic refinement, a simultaneous multi-reference refinement in presence of structural variations in cryo-EM images was performed using QEDM-predicted conformers. (Brink & Ma, Structure, 2004)



Simplified elastic network model

"What is the predictive power of the method? In an attempt to characterize more fully the lowest frequency normal modes, the following approach has been followed: **each residue is scanned in turn and its associated mass is increased by a factor of 100 compared to the other residues**. Then, the shift in the frequency of each of the ten lowest modes is recorded. In this manner, residues whose mass contributes most to these low frequency modes are highlighted and a residue-by-residue "signature" is being built for each of the ten lowest frequency normal modes. In general, we find that for hinge motions of loosely connected domains, the residues that matter most are the ones at the tip of the distal domains."

(Delarue, 2002, Simplified Normal Mode Analysis of Conformational Transitions in DNA-dependent Polymerases: the Elastic Network Model J Mol Biol)

Some timings with simplified model

TABLE VII. CPU Time Required for the Calculation of 50 Normal Modes on a HP Workstation, Using RTB With One Residue per Block, or the Standard Method Available in CHARMM for Large Matrices, Namely, DIMB†

	Standard method		RTB method			
Protein	Matrix size	DIMB (min)	Projection (min)	Matrix size	DIAGQ* (min)	Total (min)
HIV-1 protease	2,766	30	0.6	594	0.9	1.5
Triglyceride lipase	7,491	515	3.6	1,590	14	17.6

†Note that DIMB yield exact normal modes.
*The projected matrix was diagonalized with the DIAGQ routine found in the VIBRAN module of CHARMM.

F. Tama et al. (2000). Building blocks approach for determining low-frequency normal modes in macromolecules. Proteins 41, 1–7.

Langevin modes in macromolecules background

- Langevin eqns solved in terms of Langevin modes; Newton's eqns solved in terms of normal modes

$$m\frac{\mathrm{d}\nu}{\mathrm{d}t} = -\zeta\nu + \delta F(t)$$

- Langevin modes: Recall the Langevin eqns.

G. Lamm & A. Szabo. Langevin modes of macromolecules. J. Chem. Phys. 85, 7334-7348 (1986)

$$m_i\ddot{q}_i + \sum_{j=1}^{3N}\zeta_{ij}\dot{q}_j + \sum_{j=1}^{3N}V''_{ij}(q_j - q_j^0) = R_i(t), \quad i = 1,\ldots,3N$$

$$\langle R_i(t)\rangle = 0$$

$$\langle R_i(t)R_j(t')\rangle = 2\beta^{-1}\zeta_{ij}\delta(t-t')$$

The random forces $R_i(t)$ satisfy the averages above and m_i denote the particle masses and ζ represents the friction matrix.

Langevin modes in macromolecules

In Sec. IV we found that the full and reduced distribution functions can be expressed in terms of the eigenvalues (λ) of **A**, a $3N \times 6N$ matrix **L** formed from the eigenvectors of **A**, and the $6N \times 3N$ transpose of this matrix ($\mathbf{L}^T$). In the next section we will see that this can also be done for the correlation functions. We now show that **L** specifies the free modes of vibration of a system of damped harmonic oscillators in the absence of stochastic forces and designate such modes Langevin modes.

The equation of motion for a system of N three-dimensional damped oscillators is

$$\dot{\boldsymbol{\alpha}} = \frac{d}{dt}\begin{bmatrix}\mathbf{s}(t)\\ \dot{\mathbf{s}}(t)\end{bmatrix} = \mathbf{A}\boldsymbol{\alpha}\,, \tag{6.1}$$

where **A** is defined in Eq. (2.6) and **s**, the mass-weighted displacement, is defined in Eq. (4.9). It follows from Eq.

$$\mathbf{A} = \begin{pmatrix} 0 & \mathbf{1} \\ -\mathbf{F} & -\boldsymbol{\gamma}\end{pmatrix}, \quad \mathbf{B} = \beta^{-1}\begin{pmatrix}0 & 0\\ 0 & \boldsymbol{\gamma}\end{pmatrix}, \tag{2.6}$$

with **1** denoting the $3N \times 3N$ unit matrix and

$$\gamma_{ij} = (m_i m_j)^{-1/2}\zeta_{ij}\,, \tag{2.7a}$$

$$F_{ij} = (m_i\, m_j)^{-1/2}V''_{ij}\,. \tag{2.7b}$$

The full distribution function $p(\boldsymbol{\alpha}t\,|\boldsymbol{\alpha}_0)$ describes the evolution of the system in phase space and incorporates the coupling between the coordinate and the velocity degrees of freedom. The reduced distribution function involving only the coordinates may be found by integrating the full distribution function over all final velocities and an equilibrium (Boltzmann) distribution of initial velocities. Defining the mass-weighted displacement coordinates

$$s_i = m_i^{1/2}\Delta q_i, \quad i = 1,\ldots,3N\,, \tag{4.9}$$

Langevin modes results compared to normal modes

TABLE V. Comparison of the gas phase normal mode frequencies of butane with the Langevin eigenvalues of Eq. (3.1) for low solvent viscosity ($\eta = 0.2$ cp). The hydrodynamic radii are in the ratio 2:0:0:2 (with units 0.77 Å). See the text for details.

	Gas phase frequencies[a] (cm^{-1})	Langevin eigenvalues (cm^{-1}) $\eta = 0.2$ cp		
		Exact	Zeroth-order[b]	Second-order[c]
	1046	$-16.32 \pm 1044.12\,i$	$-16.47 \pm 1045.71\,i$	$-16.29 \pm 1044.13\,i$
	1004	$-1.35 \pm 1003.77\,i$	$-1.37 \pm 1003.95\,i$	$-1.35 \pm 1003.77\,i$
	903	$-29.72 \pm 900.23\,i$	$-29.87 \pm 902.35\,i$	$-29.72 \pm 900.24\,i$
	437	$-51.85 \pm 434.34\,i$	$-51.81 \pm 434.14\,i$	$-51.88 \pm 434.33\,i$
	406	$-29.12 \pm 400.39\,i$	$-29.87 \pm 405.07\,i$	$-29.12 \pm 400.42\,i$
	120	$-20.12 \pm 103.09\,i$	$-29.87 \pm 116.19\,i$	$-21.52 \pm 102.19\,i$
T	0	−83.52	−64.04	−80.75
		−65.53	−64.57	−65.69
	0	−64.32	−64.19	−64.11
		0.00	0.00	0.00
	0	0.00	0.00	0.00
		0.00	0.00	0.00
R	0	−123.77	−111.70	−125.22
		−108.49	−107.50	−108.53
	0	$-0.09\,i$	−12.04	1.49
		$0.09\,i$	0.00	0.00
	0	0.00	0.00	0.00
		0.00	0.00	0.00

[a] *T* and *R* indicate the frequencies of the three translationally and three rotationally symmetric gas phase normal modes, respectively.
[b] Calculated according to Eq. (8.7a). Both the zeroth- and first-order eigenvalues are identical (see the text).
[c] Calculated according to Eq. (8.16).

NB. As solvent viscosity increases accuracy decreases (data not shown).