E5-2001-225

E.A.Ayrjan, I.Pokorný, I.V.Puzynin, J.Skřivánek

SOLVENT ACCESSIBLE SURFACE AREA CALCULATION USING TRANSFORMATION INTO THE PLANE

1 Introduction

A spherical probe is allowed one to roll on the outside while maintaining contact with the van der Waals surface. The *accessible surface* is a continuous sheet defined by the locus of the centre of the probe. The accessible surface area of the macromolecule is crucial for computing the effect of protein solvation. The free energy of protein-solvent interaction can be approximately derived from the solvent-accessible surface areas A_i of atoms by the relation

$$E_{hyd} = \sum_i \sigma_i A_i, \tag{1}$$

where σ_i is an empirical solvation parameter depending on the atom type.

Except the top point, there exists a natural continuous correspondence between the points of the sphere and the points of the plane.

2 Parametrization of the Sphere

Let (x_i, y_i, z_i) be Cartesian coordinates of the centre of the i-th sphere of the accessible surface of a molecule and r_i be the radius of this sphere. So,

$$(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 = r_i^2 \tag{2}$$

for every its point (x, y, z). Equations

$$x = x_i + \frac{4r_i^2 t}{t^2+s^2+4r_i^2}, \; y = y_i + \frac{4r_i^2 s}{t^2+s^2+4r_i^2}, \; z = z_i + r_i - \frac{8r_i^3}{t^2+s^2+4r_i^2} \tag{3}$$

describe a relation between the points of the plane $(t, s) \in \mathbf{R}^2$ and the points of the sphere, except the point $(x_i, y_i, z_i + r_i)$. Indeed, it is a projection of points of the i-th sphere from the top point of this sphere onto the plane. The reverse correspondence is

$$\begin{aligned} t &= -2r_i \tfrac{x-x_i}{z-z_i-r_i} \\ s &= -2r_i \tfrac{y-y_i}{z-z_i-r_i}. \end{aligned} \tag{4}$$

The part of the i-th sphere which is not covered by the j-th sphere, satisfies (2) and the following inequality

$$(x - x_j)^2 + (y - y_j)^2 + (z - z_j)^2 \geq r_j^2. \tag{5}$$

We only need a relation

$$a_j^i(t^2 + s^2) + b_j^i t + c_j^i s + d_j^i \geq 0 \tag{6}$$

to describe the corresponding area in the plane where

$$\begin{aligned} a_j^i &= (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i + r_i - z_j)^2 - r_j^2 \\ b_j^i &= 8r_i^2(x_i - x_j) \\ c_j^i &= 8r_i^2(y_i - y_j) \\ d_j^i &= 4r_i^2 \left[(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - r_i - z_j)^2 - r_j^2\right]. \end{aligned} \tag{7}$$

Part of the surface area A_i parametrized by (3) can be computed by the following form

$$A_i = \iint_{\Omega_i} \left\| \left(\frac{\partial x}{\partial t}, \frac{\partial y}{\partial t}, \frac{\partial z}{\partial t}\right) \times \left(\frac{\partial x}{\partial s}, \frac{\partial y}{\partial s}, \frac{\partial z}{\partial s}\right) \right\| dt\, ds. \tag{8}$$

It can be easily shown that

$$A_i = \iint_{\Omega_i} \frac{16r_i^4 dt\, ds}{(t^2 + s^2 + 4r_i^2)^2} \tag{9}$$

where

$$\Omega_i = \{(t,s) \in \mathbf{R}^2; \quad a^i_j(t^2+s^2) + b^i_j t + c^i_j s + d^i_j \geq 0 \quad \text{for all} \quad j \neq i\}. \tag{10}$$

3 Calculation of the Area

Inequality (6) presents a subset of $\mathbf{R}^2$ that can be only of the next type:

a) empty set
b) singleton
c) interior of a circle ($a^i_j < 0$)
d) half plane ($a^i_j = 0$)
e) exterior of a circle ($a^i_j > 0$)
f) all plane $\mathbf{R}^2$.

So, by (10) the region Ω_i is an intersection of such sets (see fig. 1). There is a possibility to exclude lines and circles with big radius to avoid inaccuracy in computation. We can turn all molecular body by matrix

$$\begin{pmatrix} \cos\varphi_0 \sin\gamma_0 & -\sin\varphi_0 & \cos\varphi_0\cos\gamma_0 \\ \sin\varphi_0\sin\gamma_0 & \cos\varphi_0 & \sin\varphi_0\cos\gamma_0 \\ -\cos\gamma_0 & 0 & \sin\gamma_0 \end{pmatrix} \tag{11}$$

to come an appropriate point $(x_i + r_i\cos\varphi_0\cos\gamma_0,\ y_i + r_i\sin\varphi_0\cos\gamma_0,\ z_i + r_i\sin\gamma_0)$ of the i-th sphere in the top.

Suppose, Ω_i is bounded and of nonzero measure, N_i is the set of order numbers of the spheres which intersect the i-th sphere and Λ^i_j is the number of arcs which generate the boundary of Ω_i and descend from the j-th sphere. Greenes theorem allows one to express the integral (9) as the sum

$$A_i = \sum_{j\in N_i} \sum_{\lambda=1}^{\Lambda^i_j} \int_{C^i_{j,\lambda}} 2r_i^2 \frac{tds - sdt}{t^2+s^2+4r_i^2}, \tag{12}$$

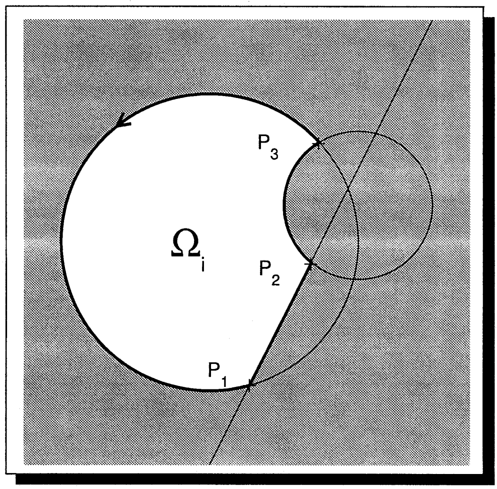

Figure 1: Region Ω_i with three vertices

where all arcs $C^i_{j,\lambda}$ together design the boundary of Ω_i and are orientated positively with respect to Ω_i. The image of each arc $C^i_{j,\lambda}$ is part of a circle or a line.

If $C^i_{j,\lambda}$ is the circle arc

$$\begin{aligned} t &= -\frac{b^i_j}{2a^i_j} + \sqrt{\frac{{b^i_j}^2+{c^i_j}^2-4a^i_j d^i_j}{4{a^i_j}^2}}\cos\varphi \\ s &= -\frac{c^i_j}{2a^i_j} + \sqrt{\frac{{b^i_j}^2+{c^i_j}^2-4a^i_j d^i_j}{4{a^i_j}^2}}\sin\varphi \end{aligned} \qquad \text{for} \quad \varphi \in \langle \alpha^i_{j,\lambda}; \beta^i_{j,\lambda}\rangle, \tag{13}$$

we get (dropping the upper index i) the next integral $I^i_{j,\lambda}$ in the form

$$\int_{C^i_{j,\lambda}} 2r_i^2 \frac{tds-sdt}{t^2+s^2+4r_i^2} = r_i^2 \cdot \Big[(\alpha_{j,\lambda}- \\ -\beta_{j,\lambda})\cdot \operatorname{sign}(a_j) + \frac{d_j+4r_i^2 a_j}{V_j}\left(\pi - 2\arctan\frac{U_j}{2a_j^2 V_j \sin\frac{\beta_{j,\lambda}-\alpha_{j,\lambda}}{2}}\right)\Big], \tag{14}$$

where

$$U_j = |a_j|(b_j^2 + c_j^2 - 2a_j d_j + 8r_i^2 a_j^2)\cos\frac{\beta_{j,\lambda}-\alpha_{j,\lambda}}{2} - \\ -a_j\sqrt{b_j^2+c_j^2-4a_j d_j}\,(b_j\cos\frac{\alpha_{j,\lambda}+\beta_{j,\lambda}}{2} + c_j \sin\frac{\alpha_{j,\lambda}+\beta_{j,\lambda}}{2}) \tag{15}$$

and

$$V_j = \sqrt{(4r_i^2 a_j - d_j)^2 + 4r_i^2(b_j^2 + c_j^2)}. \tag{16}$$

If the arc C_j^i is a full circle and $\beta_j - \alpha_j = 2\pi$, we can write I_j^i as

$$\int\limits_{C_j^i} 2r_i^2 \frac{tds - sdt}{t^2 + s^2 + 4r_i^2} = 2r_i^2\pi \cdot \left(-\text{sign}(a_j) + \frac{d_j + 4r_i^2 a_j}{V_j}\right). \tag{17}$$

One can easily derive from (14) the relation we will need below

$$\lim_{r\to\infty} \int\limits_{C(r)} 2r_i^2 \frac{tds - sdt}{t^2 + s^2 + 4r_i^2} = 2r_i^2\gamma\ , \tag{18}$$

where $C(r)$ is a positively orientated circle part with the fixed centre point accordant to the radius r and the angle γ.

If $C_{j,\lambda}^i$ is the line segment

$$\begin{aligned} t &= t_0 + c_j \cdot \varphi \\ s &= s_0 - b_j \cdot \varphi \end{aligned} \qquad \text{for} \quad \varphi \in \langle \alpha_{j,\lambda}; \beta_{j,\lambda} \rangle, \tag{19}$$

the curvilinear integral $I_{j,\lambda}^i$ is

$$\int\limits_{C_{j,\lambda}^i} 2r_i^2 \frac{tds - sdt}{t^2 + s^2 + 4r_i^2} =$$

$$= \frac{2r_i^2 d_j}{\sqrt{d_j^2 + 4r_i^2(b_j^2 + c_j^2)}} \cdot \left(\arctan \frac{(b_j^2 + c_j^2)\beta_{j,\lambda} + c_j t_0 - b_j s_0}{\sqrt{d_j^2 + 4r_i^2(b_j^2 + c_j^2)}} - \right. \tag{20}$$

$$\left. - \arctan \frac{(b_j^2 + c_j^2)\alpha_{j,\lambda} + c_j t_0 - b_j s_0}{\sqrt{d_j^2 + 4r_i^2(b_j^2 + c_j^2)}}\right).$$

Substitution $a_j = 0$ in (17) or $\beta_j = \infty$ and $\alpha_j = -\infty$ in (20) brings the integral I_j^i

$$\int\limits_{C_j^i} 2r_i^2 \frac{tds - sdt}{t^2 + s^2 + 4r_i^2} = \frac{2r_i^2\pi d_j}{\sqrt{d_j^2 + 4r_i^2(b_j^2 + c_j^2)}} \tag{21}$$

over the full line C_j^i.

An unbounded area Ω_i forms the whole plane except some rings (fig. 2) or an angle of size γ without exempt bounded part (fig. 3). So, by (12) and (18) the surface area is

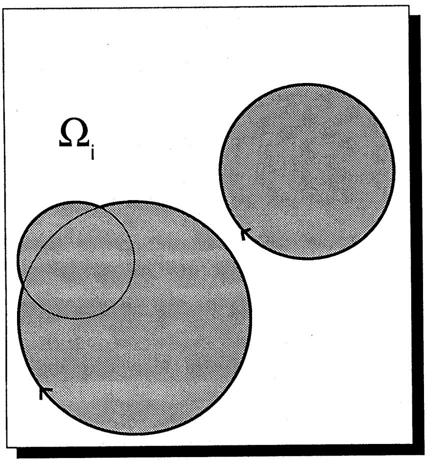

Figure 2: Full angle around

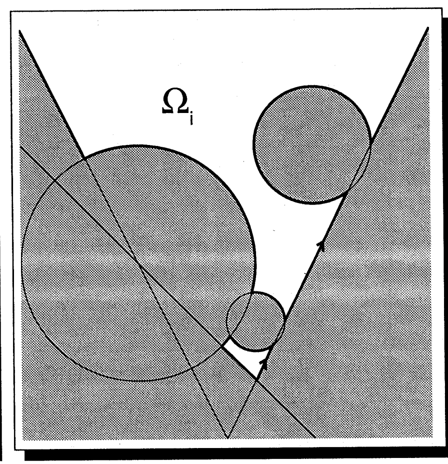

Figure 3: Restricted angle

$$A_i = \chi(\Omega_i) + \sum_{j \in N_i} \sum_{\lambda=1}^{\Lambda_j^i} \int_{C_{j,\lambda}^i} 2r_i^2 \frac{tds - sdt}{t^2 + s^2 + 4r_i^2}, \tag{22}$$

where

$$\chi(\Omega_i) = \begin{cases} 0, & \Omega_i \text{ is bounded} \\ 4\pi r_i^2, & \Omega_i \text{ is all plane except several rings} \\ 2r_i^2\gamma, & \Omega_i \text{ is an angle of size } \gamma \text{ with some} \\ & \text{picked bounded part.} \end{cases} \tag{23}$$

4 Implementation in ACCAR

The system ACCAR was originally created for the MATLAB environment. Every spherical part A_i of the accessible surface area is calculated particularly in several cycles. In the first cycle all spheres which do not achieve the i-th sphere are removed from the list of meaningful spheres. If some sphere in full cover the i-th sphere, partial calculation is stopped with result $A_i = 0$.

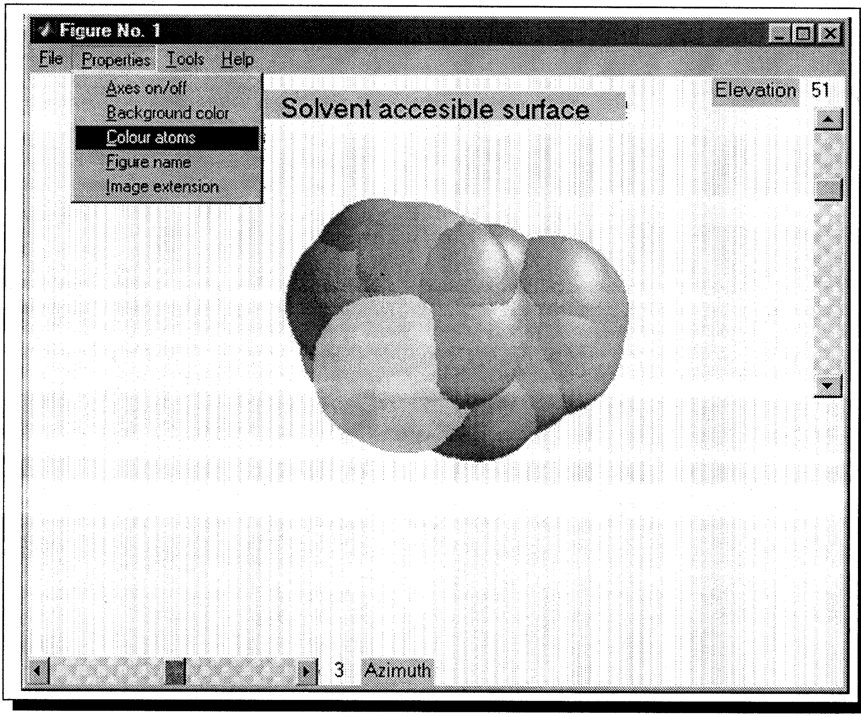

Figure 4: Visualization in ACCAR

The further cycle compares each pair of relevant regions (6). The region which is a superset of another is removed. If an intersection of two regions is of measure zero, we have $A_i = 0$ again. In this cycle the vertices are calculated. Let the vertices $[t_{j,k,1}; s_{j,k,1}]$, $[t_{j,k,2}, s_{j,k,2}]$ come from the intersection of borders of the regions j and k. If this borders are circles, then

$$t_{j,k,\nu} = \frac{-2(b'_j-b'_k)(d'_j-d'_k)-(b'_kc'_j-b'_jc'_k)(c'_j-c'_k)\pm(c'_j-c'_k)\sqrt{D_{j,k}}}{2\left((b'_j-b'_k)^2+(c'_j-c'_k)^2\right)}$$

$$s_{j,k,\nu} = \frac{-2(c'_j-c'_k)(d'_j-d'_k)-(c'_kb'_j-c'_jb'_k)(b'_j-b'_k)\mp(b'_j-b'_k)\sqrt{D_{j,k}}}{2\left((b'_j-b'_k)^2+(c'_j-c'_k)^2\right)} ,$$

where $D_{j,k} = 4(b'_kd'_j - b'_jd'_k)(b'_j - b'_k)+$

$$+4(c'_kd'_j - c'_jd'_k)(c'_j - c'_k) - 4(d'_j - d'_k)^2 + (c'_kb'_j - c'_jb'_k)^2 \tag{24}$$

for $b'_j = \frac{b_j}{a_j}$, $c'_j = \frac{c_j}{a_j}$, $d'_j = \frac{d_j}{a_j}$, $b'_k = \frac{b_k}{a_k}$, $c'_k = \frac{c_k}{a_k}$, $d'_k = \frac{d_k}{a_k}$ and $\nu \in \{1;2\}$. If the j-th region is bounded by circle and the k-th one-by line,

$$t_{j,k,\nu} = \frac{-2b_kd_k-c_k(b'_jc_k-b_kc'_j)\pm c_k\sqrt{D_{j,k}}}{2\left(b_k^2+c_k^2)^2\right)}$$

$$s_{j,k,\nu} = \frac{-2c_kd_k+b_k(b'_jc_k-b_kc'_j)\mp b_k\sqrt{D_{j,k}}}{2\left(b_k^2+c_k^2)^2\right)} , \tag{25}$$

where $D_{j,k} = 4(b'_jd_k - b_kd'_j)b_k+$

$$+4(c'_jd_k - c_kd'_j)c_k - 4d_k^2 + (c_kb'_j - c'_jb_k)^2$$

for $b'_j = \frac{b_j}{a_j}$, $c'_j = \frac{c_j}{a_j}$, $d'_j = \frac{d_j}{a_j}$ and $\nu \in \{1;2\}$. If $[t_{j,k}; s_{j,k}]$ is an initial point or an endpoint of an arc on the j-th circle then

$$\tan\gamma = \frac{2a^i_j \cdot s_{j,k} + c^i_j}{2a^i_j \cdot t_{j,k} + b^i_j} , \tag{26}$$

where $\gamma = \alpha^i_j$ or $\gamma = \beta^i_j$ respectively. The following implicit equation is preferred in the ACCAR for computation of the gradient

$$\begin{aligned}(a_jb_k - a_kb_j)^2 + (a_jc_k - a_kc_j)^2 + (b_j^2 + c_j^2 - 4a_jd_j)\cdot a_k^2 - \\ -(b_k^2 + c_k^2 - 4a_kd_k)\cdot a_j^2 + 2\sqrt{b_j^2 + c_j^2 - 4a_jd_j}\cdot a_k \cdot \\ \cdot \operatorname{sign} a_j \cdot \left((a_jb_k - a_kb_j)\cos\gamma + (a_jc_k - a_kc_j)\sin\gamma\right).\end{aligned} \tag{27}$$

The intersection of two lines is evidently

$$\begin{aligned} t_{j,k} &= \frac{c_j d_k - c_k d_j}{b_j c_k - b_k c_j} \\ s_{j,k} &= \frac{d_j b_k - d_k b_j}{b_j c_k - b_k c_j}. \end{aligned} \tag{28}$$

Subsequently three lists are built up:

a) the list of circles (lines) with no vertex,
b) the list of vertices with record of order numbers of pertinent circles (lines),
c) the list of circles (lines) with vertices.

The procedure is completed by integration over the relevant circles, lines and arcs by (17), (21), (14) and (20) and summarized by (22).

5 Calculation of the Gradient

The calculation of the gradient of the surface area dependent upon a position of the centre of j-th sphere bears on derivation of I_j^i in (17), (21), (14) and (20). If the j-th circle on the border of Ω_i contains no vertex, then

$$\frac{\partial A_i}{\partial(x_j, y_j, z_j)} = \frac{\partial I_j^i}{\partial(a_j^i, b_j^i, c_j^i, d_j^i)} \cdot \frac{\partial(a_j^i, b_j^i, c_j^i, d_j^i)}{\partial(x_j, y_j, z_j)}, \tag{29}$$

where

$$\frac{\partial(a_j^i, b_j^i, c_j^i, d_j^i)}{\partial(x_j, y_j, z_j)} = \begin{pmatrix} 2(x_j - x_i) & 2(y_j - y_i) & 2(z_j - z_i - r_i) \\ -8r_i^2 & 0 & 0 \\ 0 & -8r_i^2 & 0 \\ 8r_i^2(x_j - x_i) & 8r_i^2(y_j - y_i) & 8r_i^2(z_j - z_i + r_i) \end{pmatrix} \tag{30}$$

by (7) and

$$\frac{\partial I_j^i}{\partial(a_j^i, b_j^i, c_j^i, d_j^i)} =$$

$$= \frac{8r_i^4\pi}{\left((4r_i^2 a_j^i - d_j^i)^2 + 4r_i^2(b_j^{i\,2} + c_j^{i\,2})\right)^{\frac{3}{2}}} \cdot \Big(4r_i^2(b_j^{i\,2} + c_j^{i\,2} - 2a_j^i d_j^i) + \tag{31}$$

$$+2d_j^{i\,2}, -b_j^i(d_j^i + 4r_i^2 a_j^i), -c_j^i(d_j^i + 4r_i^2 a_j^i), b_j^{i\,2} + c_j^{i\,2} - 2a_j^i d_j^i + 8r_i^2 a_j^{i\,2}\Big)$$

by (17).

Let the j-th circle on the border of Ω_i consist of Λ_j^i arcs (see 12). Then

$$\frac{\partial A_i}{\partial(a_j^i, b_j^i, c_j^i, d_j^i)} = \sum_{\lambda=1}^{\Lambda_j^i} \frac{\partial(I_{j,\lambda}^i + I_{k_\lambda,\mu_\lambda}^i + I_{l_\lambda,\nu_\lambda}^i)}{\partial(a_j^i, b_j^i, c_j^i, d_j^i)}\,, \tag{32}$$

where $C_{k_\lambda,\mu_\lambda}^i$ is part of the k_λ-th circle on the border of Ω_i that takes up $C_{j,\lambda}^i$ in the initial point and $C_{l_\lambda,\nu_\lambda}^i$ is part of the l_λ-th circle that takes up $C_{j,\lambda}^i$ in the endpoint. There was used (27) in differentiation of the right side of (32).

References

[1] R. Fraczkiewicz, W. Braun *Exact and Efficient Analytical Calculation of the Accessible Surface Areas and Their Gradients for Macromolecules* Journal of Computational Chemistry, Vol. 19, No. 3, pp. 319-333 (1998).

[2] B. von Freyberg, W. Braun *Minimization of Empirical Energy Functions in Proteins Including Hydrophobic Surface Area Effects* Journal of Computational Chemistry, Vol. 14, No. 5, pp. 510-521 (1993).

[3] T. J. Richmond *Solvent Accessible Surface Area and Excluded Volume in Proteins* Journal of Molecular Biology, No. 178, pp. 63-89 (1984).

Received by Publishing Department
on October 23, 2001.

Айрян Э.А. и др. E5-2001-225
Расчет области, достижимой растворителем, путем преобразования в плоскость

Традиционные методы аналитического вычисления достижимой поверхностной области основаны на теореме Гаусса–Боннэ. Эта поверхность составлена из частей сфер, ограниченных дугами. Мы преобразовали проблему вычисления части поверхностной области на криволинейный интеграл на плоскости. Уравнения

$$x=x_i+\frac{4r_i^2 t}{t^2+s^2+4r_i^2},\quad y=y_i+\frac{4r_i^2 s}{t^2+s^2+4r_i^2},\quad z=z_i+r_i-\frac{8r_i^3}{t^2+s^2+4r_i^2}$$

описывают отношение между точками поверхности $(t,s)\in \boldsymbol{R}^2$ и точками сферы, кроме точки (x_i, y_i, z_i+r_i). Площадь незакрытой части поверхности i-й сферы может быть выражена как

$$A_i=\sum_{j\in N_i}\sum_{\lambda=1}^{\Lambda_j^i}\int_{C_{j,\lambda}^i 2} 2r_i^2\frac{tds-sdt}{t^2+s^2+4r_i^2}.$$

Поскольку сферические круги преобразуются в плоскости на круги или линии, рассматриваются только интегралы по частям кругов и линий.

Работа выполнена в Лаборатории информационных технологий ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна, 2001

Ayrjan E.A. et al. E5-2001-225
Solvent Accessible Surface Area Calculation Using Transformation into the Plane

Traditional routines for analytical calculation of the accessible surface area are based on the global Gauss–Bonnet theorem. This surface is composed from parts of spheres bounded by circle arcs. We transformed the problem of calculation of the part of surface area onto curvilinear integral in the plane. Equations

$$x=x_i+\frac{4r_i^2 t}{t^2+s^2+4r_i^2},\quad y=y_i+\frac{4r_i^2 s}{t^2+s^2+4r_i^2},\quad z=z_i+r_i-\frac{8r_i^3}{t^2+s^2+4r_i^2}$$

describe relation between points of the plane $(t,s)\in \boldsymbol{R}^2$ and points of the sphere, except the point (x_i, y_i, z_i+r_i). The surface area of the exposed part of ith sphere can be expressed as

$$A_i=\sum_{j\in N_i}\sum_{\lambda=1}^{\Lambda_j^i}\int_{C_{j,\lambda}^i 2} 2r_i^2\frac{tds-sdt}{t^2+s^2+4r_i^2}.$$

As spherical circles are mapped by this transformation onto circles or lines, only integrals along parts of circles and lines are considered.

The investigation has been performed at the Laboratory of Information Technologies, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna, 2001

Макет Т.Е.Попеко

Подписано в печать 28.11.2001
Формат 60 × 90/16. Офсетная печать. Уч.-изд. л. 0,96
Тираж 325. Заказ 52979. Цена 1 р. 15 к.

Издательский отдел Объединенного института ядерных исследований
Дубна Московской области