Matter tells space how to curve
Space tells matter how to move

The boundary of a boundary is zero

John Wheeler

"a gravitationally completely collapsed object"

A black hole has no hair, Mass without Mass,
Law without Law, Magic without Magic

No phenomenon is a physical phenomenon until it is an observed phenomenon.

Time is defined so that motion looks simple.

We live on an island surrounded by a sea of ignorance. As our island of knowledge grows, so does the shore of our ignorance.

If you haven't found something strange during the day, it hasn't been much of a day.

“Time is what prevents everything from happening at once”

Behind it all is surely an idea so simple, so beautiful, that when we grasp it - in a decade, a century, or a millennium - we will all say to each other, how could it have been otherwise? How could we have been so stupid?

In any field, find the strangest thing and then explore it.

What’s new?

In contrast to other “popular” books on gravitation which attempt no more than a broad panoramic survey, Wheeler’s **A Journey into Gravity and Spacetime** seeks depth. Wheeler’s goal is a thorough explanation of how gravitation works, how “mass grips spacetime, telling it how to curve” and how “spacetime grips mass, telling it how to move.” To provide such an explaation without the help of tensors or differential forms is a daunting task, but Wheeler approaches this challenge with his characteristic zeal and joyful enthusiasm. From his lectures at Princeton 25 years ago, I still remember the intensity and passion that permeated his explanations, and I was gald to see that these explanations permeate this book. For Wheeler, an explanation is a battle of ideas, to be won by a skillful, spirited attack on several fronts, aided by a battery of clever, multicolored diagrams. He has the unique ability to breathe life and excitement into even the dullest of topics---who else could lend excitement to the Bianchi identity, which appears in the guise of “the boundary of a boundary is zero”? ...... At the heart of the book he presents us with a remarkable statement linking the curvature of spacetime to the distribution of matter: For any (small) cube, the sum of the moments of rotation of the geodesics forming the edges equals 8π times the amount of “momenergy” enclosed in the cube. This marvelously simple and concise formulation of Einstein’s equation is analogous to Gauss’ law for electricity. It is an adaptation of the mathematical treatment of differential forms, given in chapter 15 of Misner, Thorne, and Wheeler’s **Gravitation** textbook. I had read that chapter years ago, but it made little impression on me, because there the physics is camoflaged by a thick layer of Cartan calculus. In **A Journey into Gravity and Spacetime** the physics is laid bare, and the surprising simplicity of Einstein’s gravitational equation stands revealed.

--Hans C. Ohanian

Most difficult for a layman to understand is how spacetime acts on masive objects, but the author explains it brilliantly in the next chapter, taught via the concept of "momenergy". This entity is a 4-vector, and the author uses it to show how its creation in a spacetime region can be written as the sum of 8 terms, reflecting the fact that the "boundary" of a four-dimensional block in spacetime consists of eight three-dimensional cubes. That the contents of these cubes sum to zero is the famous "boundary of a boundary is zero", which is discussed in the next chapter. This chapter is one of the best explanations ever given (at this level) of the physics behind spacetime curvature and massive objects. The actual mathematical quantification of curvature is detailed in chapters 8 and 9, using elementary mathematics. The author discusses nicely the famous Scharwzschild geometry.

--Lee D. Carlson

A JOURNEY INTO GRAVITY AND SPACETIME

John Archibald Wheeler

**SCIENTIFIC
AMERICAN
LIBRARY**

A Division of HPHLP
New York

heat! Energy available in the form of the relative motion of the two objects before they collide and stick goes into energy of internal excitation—or heat—after the two combine. The amalgamated system has more energy than the two objects would if they were tamely juxtaposed. Therefore it has more mass. For the same reason, a hot object has to weigh more than the same object does when cold.

Never yet, however, has anyone succeeded in weighing heat. In 1787, long before Einstein, Count Rumford made valiant attempts to detect and measure the difference in weight between the same barrel of water when hot and when cold; he was unsuccessful. He concluded that "all attempts to discover any effect of heat upon the apparent weight of bodies will be fruitless." Since Einstein, we know that there must be such a difference and can easily figure it. However, the increase in the internal energy of agitation—and therefore in mass—is too small even for today's best instruments to detect. Yet detect it we someday can and must! What a challenge to mankind's ingenuity!

Conservation of Momenergy in Many Collisions

Never has the Pacific seen a man-initiated event that released a greater energy than the Mike nuclear explosion of November 1952. The bomb set off on the Eniwetok Atoll released an energy equivalent to ten million tons of TNT, a thousandfold times the output of the atomic bomb dropped on Hiroshima in August 1945. In the first 100 picoseconds, the first 10^{-7} second, after the uranium, deuterium, tritium, and lithium had completed their reaction, the energy set free resided for the most part right there. It took the form of heat, heat of two kinds. One was heat of agitation—particles zigzagging wildly back and forth between encounters with other particles. The other carrier of heat energy was radiation—particles of light, quanta of electromagnetic radiation, photons. Mike's interior at that instant, at twenty million degrees Celsius or so, was like a barrel of star interior suddenly brought down to Earth.

In that hot mass, during that 10^{-7} second, billions upon billions of photons, electrons and nuclei traded momenergy, each of them billions upon billions of times. Collisions unbelievable in number! Could we have followed one of them? Could we have compared the sum of the momenergy of any two participants before colliding with the sum after their collision? Perhaps. But could we have done that for all the collisions of each participant, and for all the participants? That's beyond the power of any human, any machine, any instrument.

That impossible bookkeeping enterprise spacetime itself nevertheless accomplishes easily, quietly, successfully. It conserves momenergy as rigorously at the wholesale level during billions and billions of collisions as it does at the retail level during one collision.

How can we picture most vividly this macroscopic conservation of momenergy? Not on any one mass, not even on any colliding pair of masses, do we any longer turn our attention, but on the totality of momenergy carried by all the masses in some standard volume of space (cubic centimeter, cubic meter, or cubic lightyear) and in like standard volumes east, west, north, south, up, and down from it. We do not segment space alone in our bookkeeping. Time, too, we partition into standard segments whose dimensions we measure off by light-travel time (centimeter, meter, or lightyear). Thus we end up with spacetime itself partitioned into blocks. The exact dimensions don't matter. For definiteness, however, picture the block on which we fix our attention as one unit "long," or nearly so, in each space dimension and one unit "high" in the time dimension.

A wild melee of collisions occupies this typical spacetime block. Billions of battles go on, encounters in which one particle gains momenergy, another loses it. But never, in this or any other spacetime block, is any energy or momentum ever created or destroyed. From this single simple *never* statement follows all that need be said, all that can be said, about conservation of momenergy. Does "no creation or destruction of momenergy" in a given region during a given time mean that there's just as much momenergy in a specified region at the end of a specified time interval as there was at its beginning? Not at all! Zillions of particles and photons may flow—and typically do flow—in or out across the boundaries of the region during the time in question. So "no creation or destruction" means—and demands—a more imaginative formulation. Has the bank created any money in my account during the month, or destroyed any, or behaved as it should? To test this point I work out a single simple number. I call it my measure of "creation," although I suspect Mrs. Mykietyn at the bank wouldn't like the name! "Creation" is the amount in the account at the end of the month, minus the amount in it at the beginning, plus the checks paid out in those 30 days, minus the deposits made. Thus formed out of these four numbers, my creation index, were it positive, would indicate creation of money in my account; were it negative, destruction. But always it comes out zero, as the bank and I require. The grip of spacetime on momenergy likewise demands—but also, as we will see in the next chapter, *automatically brings about*—a zero value for an analogous index of creation of momenergy:

$$
\begin{aligned}
&\begin{pmatrix}\text{creation of momenergy}\\ \text{in a specified spacetime region}\end{pmatrix}\\
&\quad= \begin{pmatrix}\text{momenergy in specified}\\ \text{region of space at end}\\ \text{of specified time}\end{pmatrix} - \begin{pmatrix}\text{momenergy in specified}\\ \text{region of space at beginning}\\ \text{of specified time}\end{pmatrix}\\
&\quad+ \begin{pmatrix}\text{flow of momenergy out of}\\ \text{left-hand face of region}\\ \text{during specified time}\end{pmatrix} + \begin{pmatrix}\text{flow of momenergy out of}\\ \text{right-hand face of region}\\ \text{during specified time}\end{pmatrix}\\
&\quad+ \begin{pmatrix}\text{flow of momenergy out of}\\ \text{front face of region}\\ \text{during specified time}\end{pmatrix} + \begin{pmatrix}\text{flow of momenergy out of}\\ \text{back space of region}\\ \text{during specified time}\end{pmatrix}\\
&\quad+ \begin{pmatrix}\text{flow of momenergy out of}\\ \text{bottom face of region}\\ \text{during specified time}\end{pmatrix} + \begin{pmatrix}\text{flow of momenergy out of}\\ \text{top face of region}\\ \text{during specified time}\end{pmatrix}
\end{aligned}
$$

In brief, when any change in the content of momenergy in the region under watch is added to the net outflow of momenergy, the result must be zero.

If the change is positive, then outflow is negative. To say that the net outflow is negative is a fancy but nevertheless useful way to state that the net inflow is positive. That inflow is exactly what brings about the increase of momenergy in the region under consideration. If the change is negative and the momenergy has decreased, then the net outflow is positive. That outflow is what brings about the decrease in momenergy in the specified spacetime block. No "creation" or "destruction" actually occurs. Momenergy is not created or destroyed, it is simply transferred.

An important principle teaches us something new every time we learn to state it in a new way. So here. No creation or destruction of momenergy in a four-dimensional spacetime volume, we now discover, translates itself into information about the *content* of momenergy in eight three-dimensional volumes, eight boundary "faces," eight 3-cubes.

Just as a three-dimensional cube is defined and in a sense "surrounded" by its six two-dimensional faces, so a four-dimensional block of spacetime is defined and "surrounded" by eight three-dimensional cubes. Each of these cubes corresponds to one of the eight terms in the expression above. For example, let's consider two of those cubes: cube 1 defined by the first term, that for the momenergy in the unit region of space at the end of the specified time, and cube 4, defined by the fourth term, that for the flow of momenergy out across the right-hand face of the original 4-cube in the specified time. Cube 4 differs from cube 1 only in the directions of its three dimensions. Cube 1 has three dimensions: left-right, front-back, down-up. For cube 4, however, the dimensions are front-back, down-up, and past-future. Timelike though one of those dimensions is, the cube it helps to define is no less a cube. This cube, like the first, has a content of mass-in-motion for inclusion in our count of momenergy. The amount of momenergy in cube 1 is the amount in the unit region of space at the end of the specified time, while cube 4 contains all the momenergy that has flowed out through the right-hand face of the original 4-cube in the specified time.

The grip of spacetime holds firmly onto the content of momenergy in every one of the 3-cubes that surround a 4-cube. That grip demands a zero *sum* for the contents of all those 3-blocks that bound that four-dimensional spacetime region. In that way the grip of spacetime, there and likewise anywhere, forever bars any creation—or destruction—of momenergy anywhere in spacetime.

A grip on momenergy so vigilant! Does spacetime maintain it by eagle-eyed vision, by clever machinery, by an all-reaching spy system? No, behind all that apparent vigilance lies a linkup of momentum-and-energy to spacetime so simple and so clever, so beautiful and so spare, that it does not count as "machinery" at all! The magic of that grip is now about to show itself in the principle that "the boundary of a boundary is zero!"

The Einstein field equations [edit]

Main article: Einstein field equations

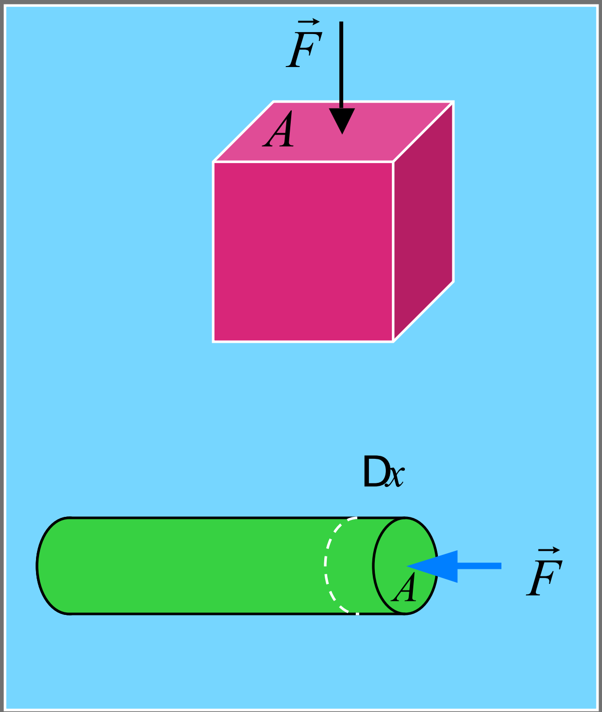
In general relativity, the stress tensor is studied in the context of the Einstein field equations which are often written as

$$R_{\mu\nu} - \frac{1}{2}R\,g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu},$$

where $R_{\mu\nu}$ is the Ricci tensor, R is the Ricci scalar (the tensor contraction of the Ricci tensor), and G is the universal gravitational constant.

Stress Tensor

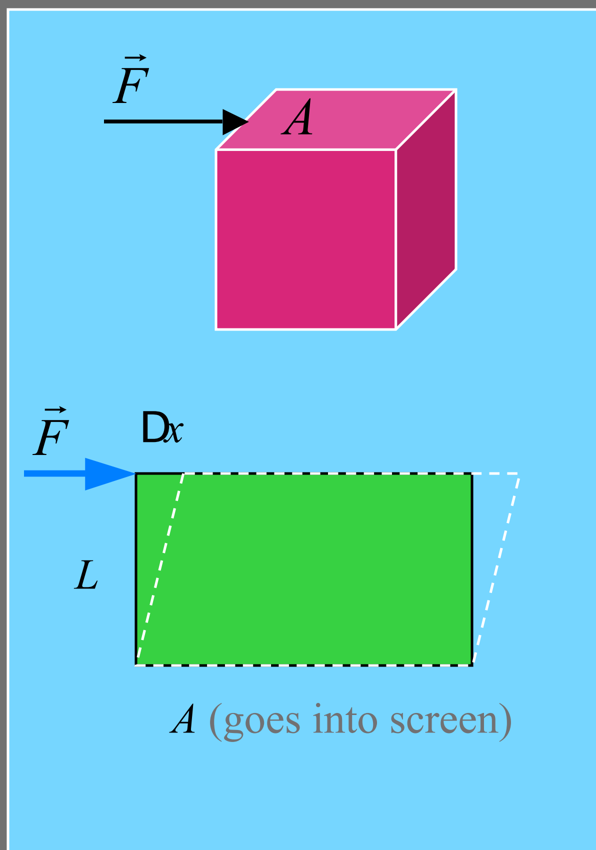
Normal Stress



- A stress measures the surface force per unit area.
  - Elastic for small changes
- A normal stress acts normal to a surface.
  - Compression or tension

$$\vec{t} = \frac{\vec{F}}{A}$$

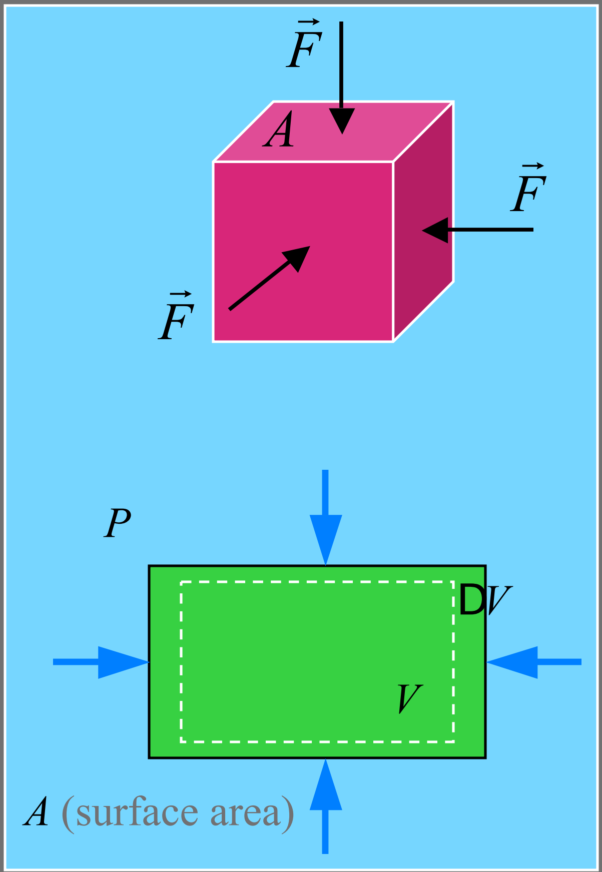
Shear Stress



- A shear stress acts parallel to a surface.
  - Also elastic for small changes
- Ideal fluids at rest have no shear stress.
  - Solids
  - Viscous fluids

$$\vec{t} = \frac{\vec{F}}{A}$$

Volume Stress



- Fluids exert a force in all directions.
  - Same force in all directions
- The force compared to the area is the pressure.

$$P = \frac{F}{A}$$

Symmetric Form

- The stress tensor includes normal and shear stresses.
  - Diagonal normal
  - Off-diagonal shear
- An ideal fluid has only pressure.
  - Normal stress
  - Isotropic
- A viscous fluid includes shear.
  - Symmetric
  - 6 component tensor

$$\mathbf{T} = \begin{bmatrix} \sigma_1 & \tau_{12} & \tau_{13} \\ \tau_{21} & \sigma_2 & \tau_{23} \\ \tau_{31} & \tau_{32} & \sigma_3 \end{bmatrix}$$

$$T_{ij} = P\delta_{ij}$$

$$T_{ij} = T_{ji} \qquad \mathbf{T} = \begin{bmatrix} \sigma_1 & \tau_{12} & \tau_{13} \\ \tau_{12} & \sigma_2 & \tau_{23} \\ \tau_{13} & \tau_{23} & \sigma_3 \end{bmatrix}$$

Force Density

$$\vec{F} = \iint_S \mathbf{T} \cdot d\vec{S}$$

$$\vec{F} = -\iint_S \mathbf{T} \cdot \hat{n} dS$$

$$\vec{F} = -\iiint_V \nabla \cdot \mathbf{T} dV$$

$$\vec{f}_S = -\nabla \cdot \mathbf{P}$$

- The total force is found by integration.
  - Closed volume with Gauss' law
  - Outward unit vectors
- A force density due to stress can be defined from the tensor.
  - Due to differences in stress as a function of position

next

$$T^{\alpha\beta} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}$$

Maxwell stress tensor

From Physics wiki

< Classical electrodynamics

The tensor

$$T_{ij} \equiv \epsilon_0 \left(E_i E_j - \frac{1}{2}\delta_{ij} E^2 \right) + \frac{1}{\mu_0} \left(B_i B_j - \frac{1}{2}\delta_{ij} B^2 \right).$$

SI units

In free space in SI units, the electromagnetic stress-energy tensor is (in flat space-times)

$$T^{\mu\nu} = \frac{1}{\mu_0}[F^{\mu\alpha}F^{\nu}{}_{\alpha} - \frac{1}{4}\eta^{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}]\,.$$

where $F^{\mu\nu}$ is the electromagnetic tensor. Note: The tensor $T^{\mu\nu}$ is a symmetric tensor.
And in explicit matrix form:

$$T^{\mu\nu} = \begin{bmatrix} \frac{1}{2}(\epsilon_0 E^2 + \frac{1}{\mu_0}B^2) & S_x/c & S_y/c & S_z/c \\ S_x/c & -\sigma_{xx} & -\sigma_{xy} & -\sigma_{xz} \\ S_y/c & -\sigma_{yx} & -\sigma_{yy} & -\sigma_{yz} \\ S_z/c & -\sigma_{zx} & -\sigma_{zy} & -\sigma_{zz} \end{bmatrix},$$

with

Poynting vector $\vec{S} = \frac{1}{\mu_0}\vec{E} \times \vec{B}$,

Electromagnetic field tensor $F_{\mu\nu}$,

Minkowski metric tensor $\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$, and

Maxwell stress tensor $\sigma_{ij} = \epsilon_0 E_i E_j + \frac{1}{\mu_0}B_i B_j - \frac{1}{2}\left(\epsilon_0 E^2 + \frac{1}{\mu_0}B^2\right)\delta_{ij}$.

Note that $c^2 = \frac{1}{\epsilon_0\mu_0}$ where *c* is light speed.

Energy, Matter, and the Curvature of Spacetime

We've gotten a sneak preview of Einstein's equations before: **G=8πT**. The **G** on the left stands for the different numbers in the Einstein tensor. But, the Einstein tensor represents the geometry of spacetime, so this is what the left side really represents. We also know that the curvature of spacetime is caused by matter, so the **T** on the right must represent matter.

Just like **G**, the symbol **T** stands for a set of numbers:

$$T_{xx}\,,\ T_{xy}\,,\ T_{xz}\,,\ T_{xt}\,,\ T_{yy}\,,\ T_{yz}\,,\ T_{yt}\,,\ T_{zz}\,,\ T_{zt}\,,\ \text{and}\ T_{tt}\,.$$

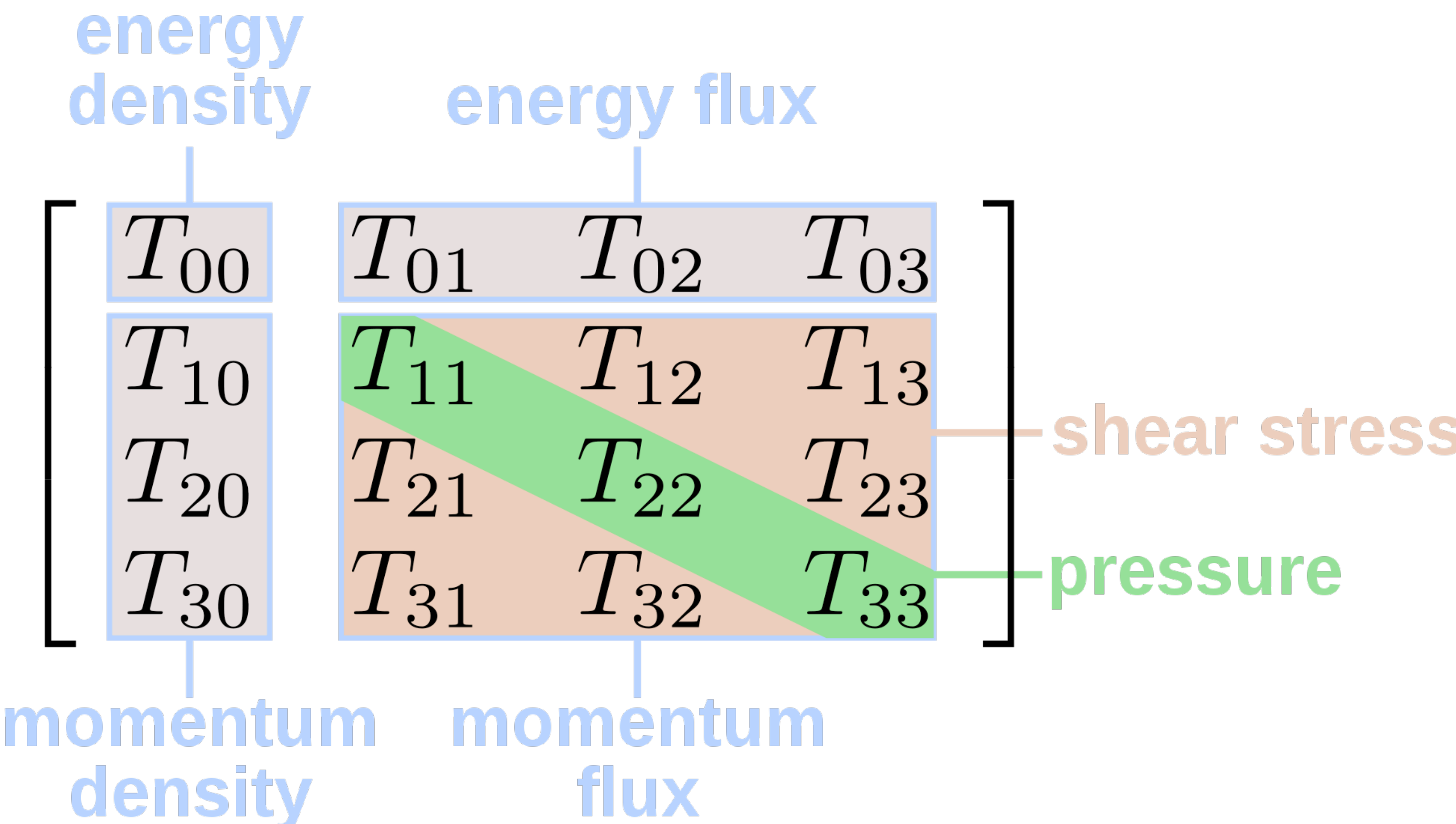
These numbers measure different things about matter. Together, they make up the Stress-Energy Tensor. Each component of this tensor has a slightly different physical interpretation:

Pieces of the Stress-Energy Tensor

T_{tt}	Measures how much mass there is at a point—how much density
T_{xt} , T_{yt} and T_{zt}	Measures how fast the matter is moving—its momentum
T_{xx} , T_{yy} and T_{zz}	Measures the pressure in each of the three directions
T_{xy} , T_{xz} and T_{yz}	Measures the stresses in the matter

As we see from the table, things like stress, pressure, and momentum come into Einstein's equations. That is, stress, pressure, and momentum all have some effect on the warping of spacetime. This is related to Einstein's most famous equation, $\mathbf{E=mc^2}$, which says that energy has mass.

Warped spacetime affects how matter moves by changing its geodesics. On the other hand, Einstein's equations show us how matter—and its movement and pressures—affect the shape of spacetime. Thus, Einstein solved the fundamental problem in Physics—in principle. Of course, solving something in principle is very different from solving in practice. Finding real solutions has proven to be very difficult. Often, it is a job best left to computers.

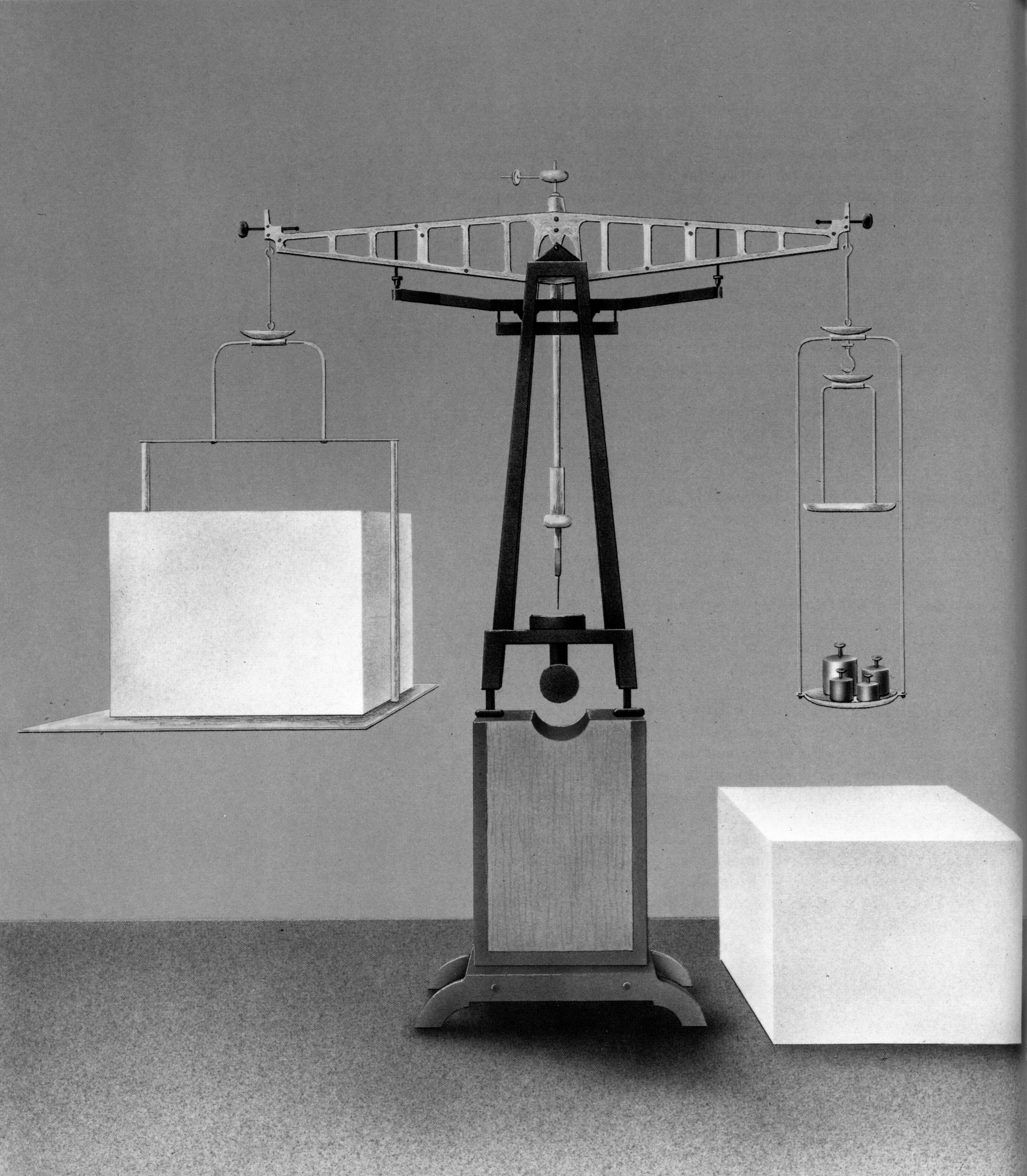

energy density
energy flux
T_{00} T_{01} T_{02} T_{03}
T_{10} T_{11} T_{12} T_{13}
T_{20} T_{21} T_{22} T_{23}
T_{30} T_{31} T_{32} T_{33}
shear stress
pressure
momentum density
momentum flux

7 *The Boundary of a Boundary: Where the Action Is!*

Leave us now you must —
So you signal to us, Minerva;
But, bringer of wisdom,
Before you go,
Reveal, we beg,
The magic central idea of gravity.

We would venture forth
On many an exploration of our own —
From the harmony of the planets
To the power of the black hole,
And from the clout of a gravity wave
To the dynamics of the cosmos —
But confidently so
Only with your talisman,
Your magic key to it all,
In our hand.

At the boundaries of the box the mass inside grips spacetime outside. Spacetime in turn grips the boundaries and drags the box down upon the scales.



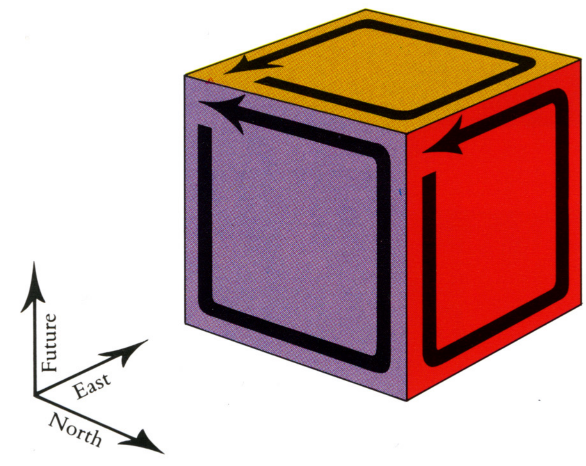
In desperation, we turn to Minerva, the goddess of wisdom, for the key to the magic grip of gravity. Mysteriously she says, "The secret of the grip lies in the boundary of the boundary," and vanishes.

From her enigmatic smile, we know her words somehow divulge the very heart of the mystery. But how can we translate her cryptic message into anything loud and clear?

"The boundary of the boundary," Minerva said. Then we'll start by remembering all that we know about a boundary.

A directed or "oriented" line—a one-dimensional *manifold*—has for its boundary the starting point and the terminal point, both zero-dimensional. Convention counts the end point as positive, as "payoff"; the starting point as negative, as "debt incurred" to start that line. A two-dimensional manifold, a bit of oriented surface cut out of a sheet of paper by a single circuit of the scissors, has for its boundary the directed line—the one-dimensional manifold—traced out by the scissors. And that directed line itself? What is its boundary? Zero! Zero because whatever the point at which we consider that line to have started, that is also the point at which the line terminates. The debt incurred at the starting point annihilates, consumes, eats up the payoff at its end point. Otherwise stated, the *zero*-dimensional boundary of the *one*-dimensional boundary of a *two*-dimensional region is **zero.**

When we change the last sentence a bit, we get another true statement: the *one*-dimensional boundary of the *two*-dimensional boundary of a *three*-dimensional region is **zero.** Really true? Yes, as a friendly dia-



The one-dimensional boundary of the two-dimensional boundary of a three-dimensional space (here a cube) is zero.

gram reminds us. The interior of a cube can be thought of as oriented. That is, each face of the cube inherits from the interior an orientation, a swirl, a direction of circulation as indicated by the arrow that runs around the periphery of that face. All six of the two-dimensional faces together constitute the boundary of the cube.

The one-dimensional boundary of one face of the cube is the directed line indicated by the arrow. Not its start—wherever *that* is—and not its end, because the zero-dimensional start and end cancel. No, the line itself is the boundary. Nothing zero about it, regarded in and by itself. But pick out a segment of that line. Oops! Next to it, on the adjacent face, runs a counter-directed line! Those two line segments annul, kill, annihilate each other. And so it goes at every edge. Result? Total washout of the one-dimensional boundary. The one-dimensional boundary of the two dimensional boundary of a three-dimensional region totals to zero.

Automatic Conservation of Momenergy

Minerva's words, "the boundary of the boundary," have a rhythmic beat. They run through our minds, enchant us, lead us on. Can it be that the *two*-dimensional boundary of the *three*-dimensional boundary of a *four*-dimensional region is likewise self-canceling, self-annihilating, zero?

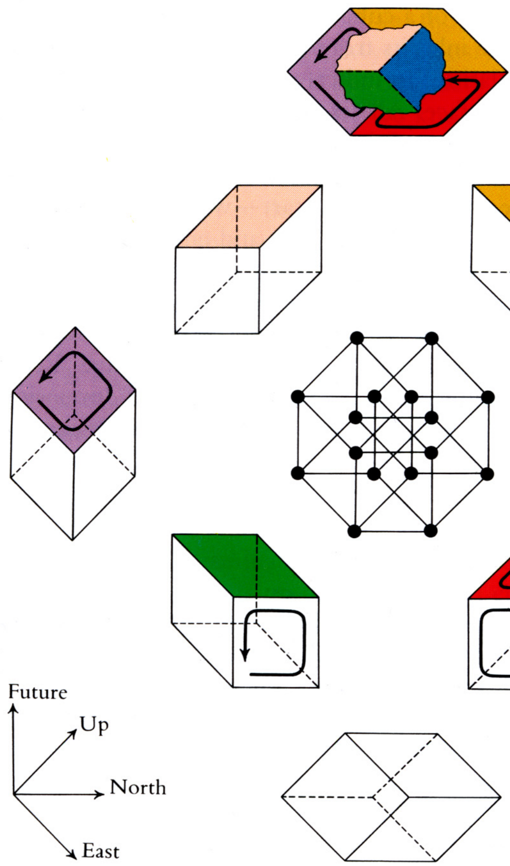
The phrase "four-dimensional region" sets up vibrations in our memory. Suddenly we remember that in no four-dimensional region anywhere does the grip of spacetime ever permit any change in momenergy whatsoever. No creation or destruction of momenergy in any block of four-dimensional spacetime. And this demand, we discovered in Chapter 6, translates into a rule of regal reach: *content of momenergy totaled for all the 3-cubes that bound any spacetime region must sum to zero.*

Amid the dark mystery of how the grip of gravity works, all at once a flicker of light begins to gleam. Boundary! Our regal rule puts right before our eyes a boundary—the 3-cubes. But where is the *boundary* of the boundary that Minerva mentioned?

Let's turn back, then, and visualize the three-dimensional boundary of a four-dimensional region. It presents to view eight pieces, eight 3-cubes. In the mind's eye one of those 3-cubes begins to gleam all over its two-dimensional faces, gleam like a luminous box of Tiffany glass. Its top glows gold; its front face,red; its right face, blue. Equally vivid colors shine forth on the opposite three faces.

We muse on that many-colored cube, and as we do, our sixth sense picks up the glint of an arriving idea. The glowing faces of the cube start

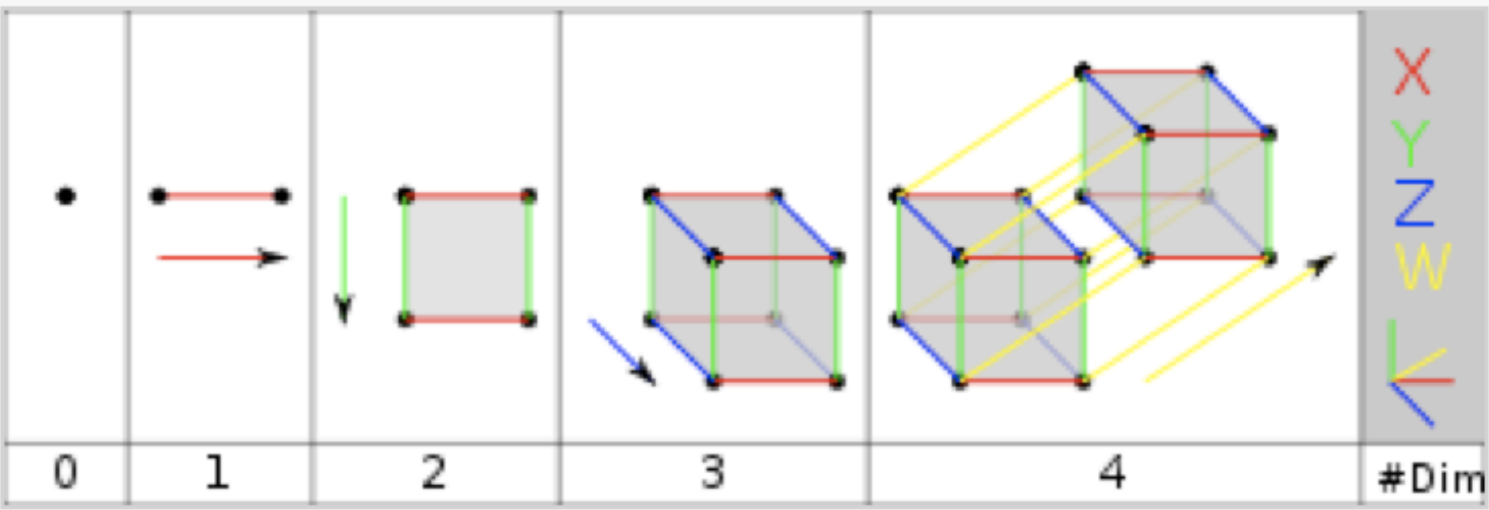
The 4-D block of spacetime depicted at the center is bounded by eight 3-cubes, shown here as exploded off that block. Each 3-cube has six 2-faces. The content of momenergy—of mass-in-motion—inside each 3-cube expresses itself on its faces. Each face is endowed with an "orientation" or sense of swirl represented here—for a few sample faces—by one-dimensional arrows around its perimeter. Faces with matching colors butt up against each other. Two faces that abut have opposite orientations, opposite senses of swirl, and thus make contributions equal in magnitude but opposite in sign to the audit of momenergy creation in the 4-D block during the stretch of time from "start" (bottom of diagram) to "end" (top of diagram).



to pulse. We are getting hot on the trail. Does spacetime curvature on the *faces* of the cube grip the content of momenergy inside the cube? In turn, does the content of momenergy inside the cube grip the spacetime curvature on the *faces* of the cube? As if in answer to this reciprocal query, the cube of many colors flashes faster.

Suddenly the light of *the* great idea floods our consciousness. Before we even fully grasp what it is, we know in our bones that the glittering central mechanism of gravity at last lies here, exposed. Our eyes, blinded by the brightness, bit by bit adapt. Outlines become visible before we discern any details.

The two-dimensional boundary of the cube—that's where the action is! That is where the content of momenergy is revealed, and ulti-



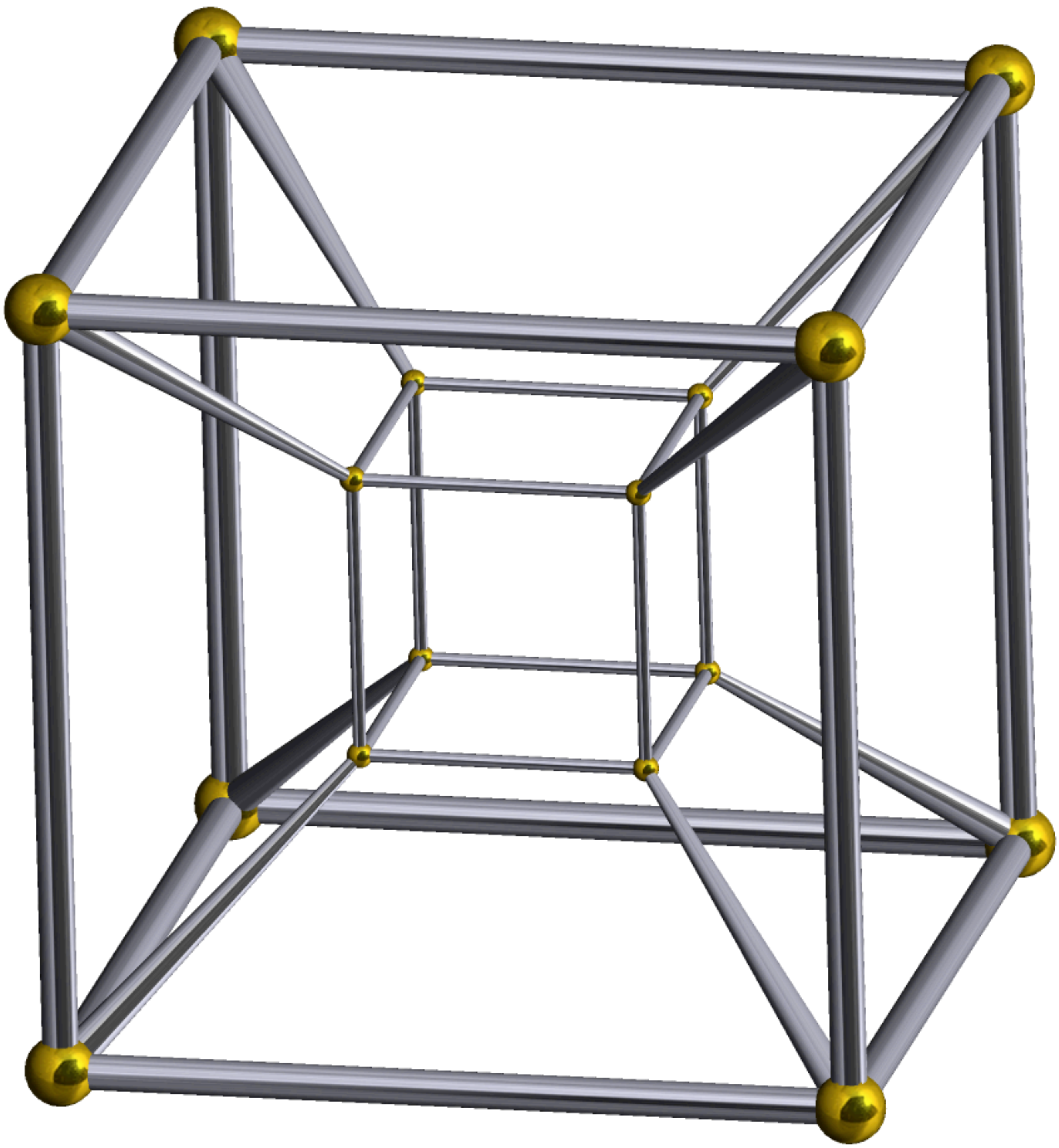
A diagram showing how to create a tesseract from a point.

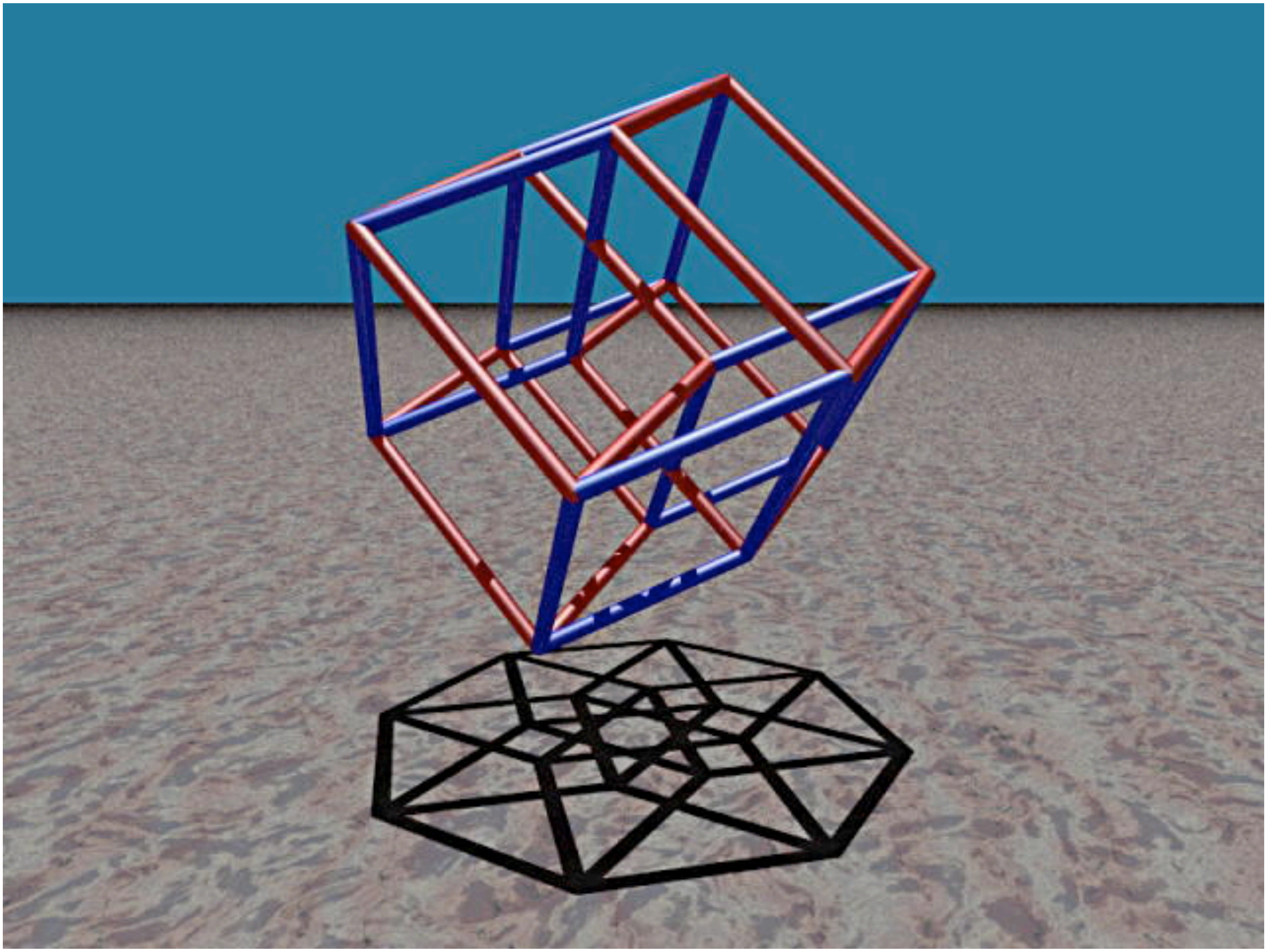
Hypercube elements $E_{m,n}$

			m	0	1	2	3	4	5	6	7	8	9	10
n	γ_n	n-cube	Names Schläfli symbol Coxeter-Dynkin	Vertices	Edges	Faces	Cells	*4*-faces	*5*-faces	*6*-faces	*7*-faces	*8*-faces	*9*-faces	*10*-faces
0	γ_0	0-cube	Point -	1										
1	γ_1	1-cube	Line segment {}	2	1									
2	γ_2	2-cube	Square **Tetragon** {4}	4	4	1								
3	γ_3	3-cube	Cube **Hexahedron** {4,3}	8	12	6	1							
4	γ_4	4-cube	Tesseract **Octachoron** {4,3,3}	16	32	24	8	1						
5	γ_5	5-cube	Penteract **Decateron** {4,3,3,3}	32	80	80	40	10	1					
6	γ_6	6-cube	Hexeract **Dodecapeton** {4,3,3,3,3}	64	192	240	160	60	12	1				
7	γ_7	7-cube	Hepteract **Tetradeca-7-tope** {4,3,3,3,3,3}	128	448	672	560	280	84	14	1			
8	γ_8	8-cube	Octeract **Hexadeca-8-tope** {4,3,3,3,3,3,3}	256	1024	1792	1792	1120	448	112	16	1		
9	γ_9	9-cube	Enneract **Octadeca-9-tope** {4,3,3,3,3,3,3,3}	512	2304	4608	5376	4032	2016	672	144	18	1	
10	γ_{10}	10-cube	Dekeract **icosa-10-tope** {4,3,3,3,3,3,3,3,3}	1024	5120	11520	15360	13440	8064	3360	960	180	20	1

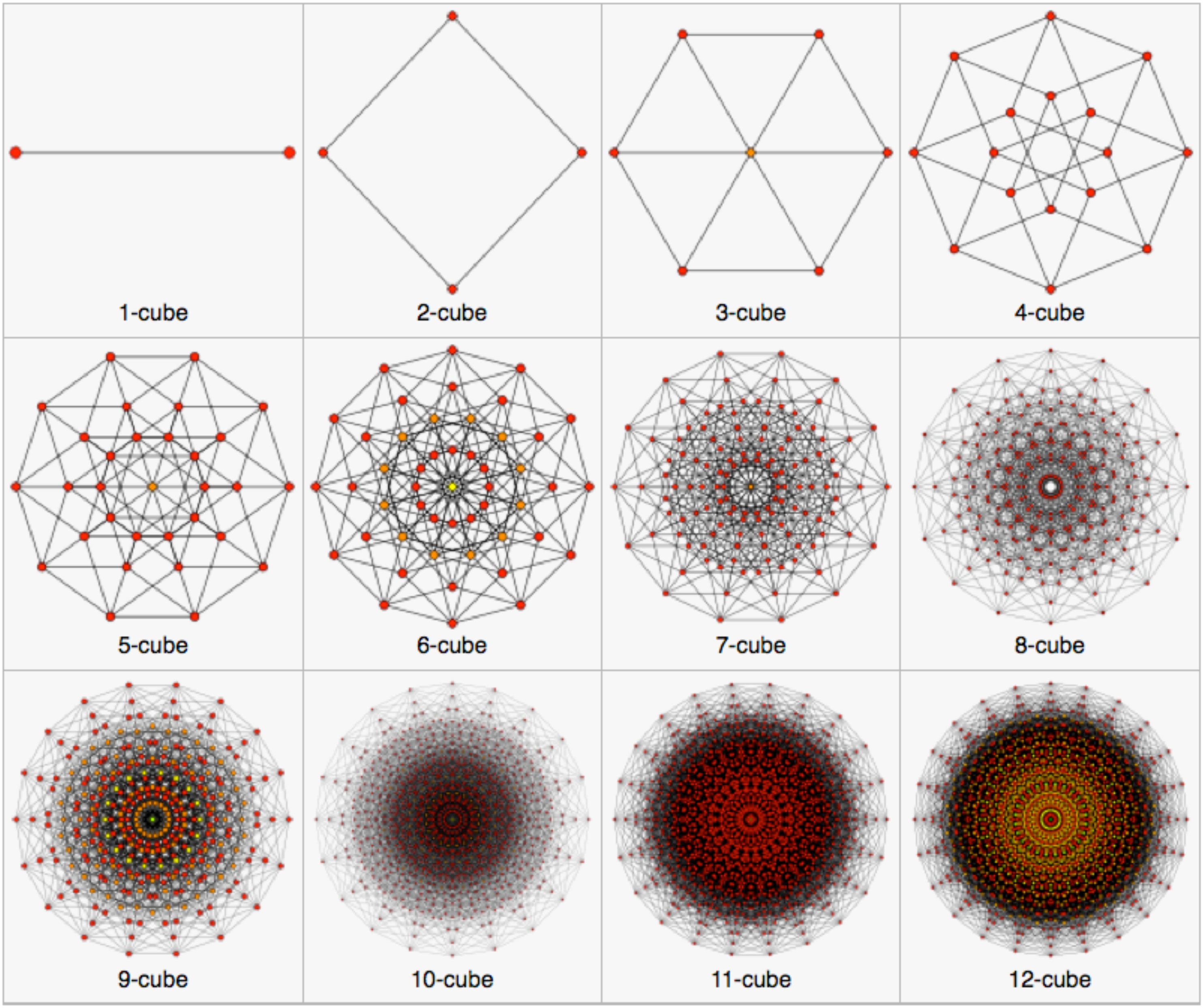
The following table gives the names for polygons with n sides. The words for polygons with $n \geq 5$ sides (e.g., pentagon, hexagon, heptagon, etc.) can refer to either regular or non-regular polygons, depending on context. It is therefore always best to specify "regular n-gon" explicitly. For some polygons, several different terms are used interchangeably, e.g., nonagon and enneagon both refer to the polygon with $n = 9$ sides.

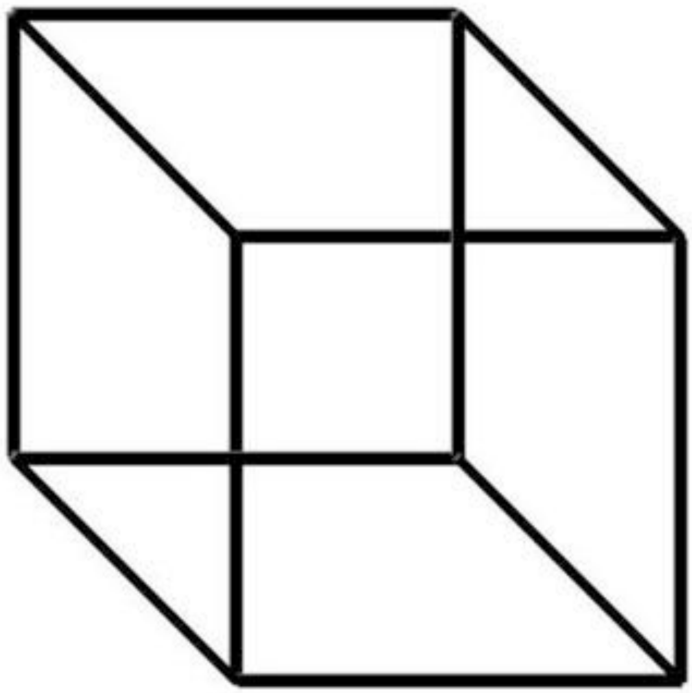
n	polygon
2	digon
3	triangle (trigon)
4	quadrilateral (tetragon)
5	pentagon
6	hexagon
7	heptagon
8	octagon
9	nonagon (enneagon)
10	decagon
11	hendecagon (undecagon)
12	dodecagon
13	tridecagon (triskaidecagon)
14	tetradecagon (tetrakaidecagon)
15	pentadecagon (pentakaidecagon)
16	hexadecagon (hexakaidecagon)
17	heptadecagon (heptakaidecagon)
18	octadecagon (octakaidecagon)
19	enneadecagon (enneakaidecagon)
20	icosagon
30	triacontagon
40	tetracontagon
50	pentacontagon
60	hexacontagon
70	heptacontagon
80	octacontagon
90	enneacontagon
100	hectogon
10000	myriagon

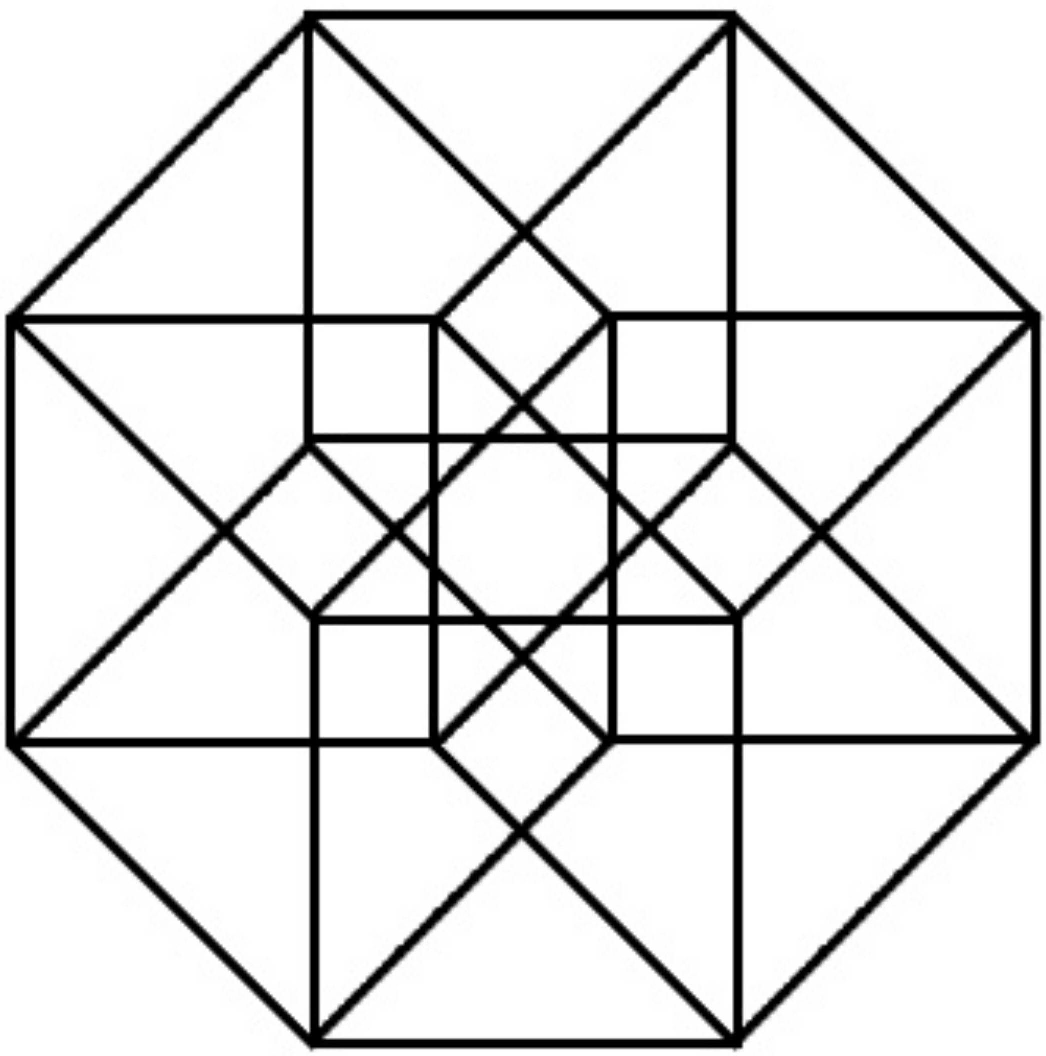


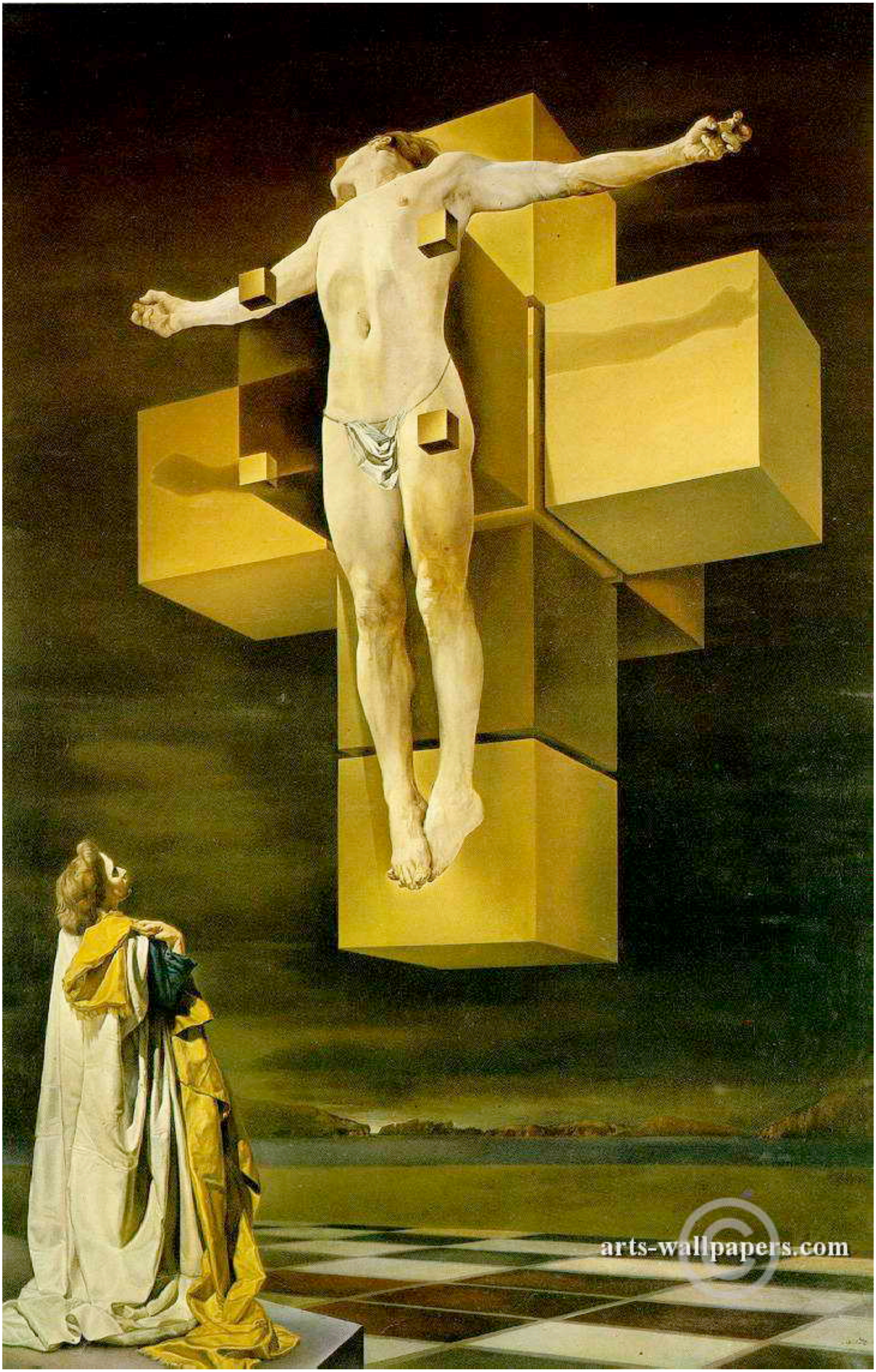


Petrie polygon Orthographic projections

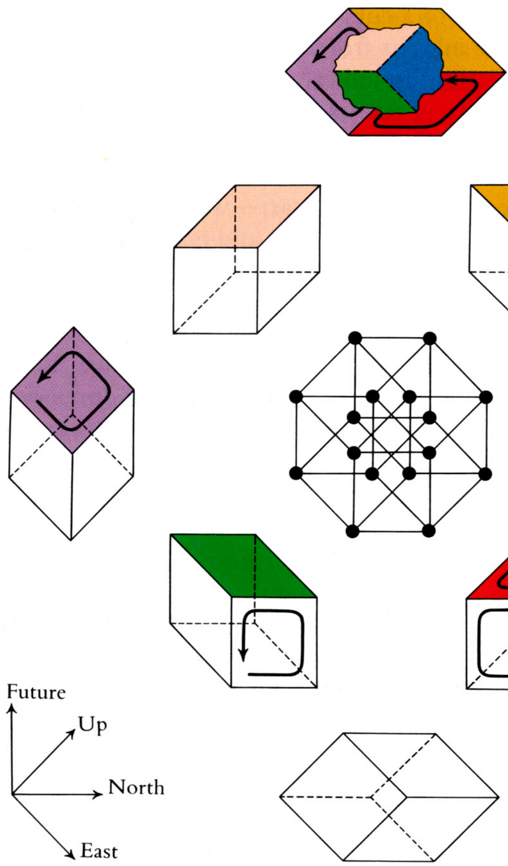







arts-wallpapers.com

The 4-D block of spacetime depicted at the center is bounded by eight 3-cubes, shown here as exploded off that block. Each 3-cube has six 2-faces. The content of momenergy—of mass-in-motion—inside each 3-cube expresses itself on its faces. Each face is endowed with an "orientation" or sense of swirl represented here—for a few sample faces—by one-dimensional arrows around its perimeter. Faces with matching colors butt up against each other. Two faces that abut have opposite orientations, opposite senses of swirl, and thus make contributions equal in magnitude but opposite in sign to the audit of momenergy creation in the 4-D block during the stretch of time from "start" (bottom of diagram) to "end" (top of diagram).



to pulse. We are getting hot on the trail. Does spacetime curvature on the *faces* of the cube grip the content of momenergy inside the cube? In turn, does the content of momenergy inside the cube grip the spacetime curvature on the *faces* of the cube? As if in answer to this reciprocal query, the cube of many colors flashes faster.

Suddenly the light of *the* great idea floods our consciousness. Before we even fully grasp what it is, we know in our bones that the glittering central mechanism of gravity at last lies here, exposed. Our eyes, blinded by the brightness, bit by bit adapt. Outlines become visible before we discern any details.

The two-dimensional boundary of the cube—that's where the action is! That is where the content of momenergy is revealed, and ultimately where the grip of gravity grabs. But how can we be sure that the action takes place at that boundary, not inside it? Suddenly a memory springs to my mind. The little fishing town of South Bristol on the Maine seacoast. Farrin's Lobster Pound. A puzzled storekeeper with two identical cardboard boxes before him. He had spent half an hour packing them with styrofoam insulation and cold wet seaweed. As I turned to go, he couldn't remember which one had a single lobster in it, destined for an unmarried friend of mine, and which contained two, as a gift for a young married couple. We had been talking too much. Muttering to himself, he started to unpack one box, thinking to check the contents. His wife stopped him, "Don't unpack, Frank; weigh them!" So he did. He never had to look inside to resolve his dilemma. Weight—a purely outside property—instantly revealed the vital information about what was inside—one lobster or two.

Likewise, never does the grip of spacetime have to reach inside the luminous Tiffany box to sense—and measure—that 3-cube's total content of momenergy. Instead the content of momenergy inside a 3-cube shows up accurately, simply, completely via "scoreboard indicators" on the two-dimensional (2-D) faces of that 3-cube. The sum of those six surface-located scoreboard readings gives the momenergy inside. Outside reveals inside. How, then, do we define and measure the "scoreboard reading" for any one 2-face? Amazingly we don't have to if we're willing to postpone knowing *how much* momenergy lies within the 3-cube and concentrate on checking *conservation* of momenergy. Why? Because any one face—say, the red face—butts up against and cancels out the oppositely swirl-oriented red face of an adjacent 3-cube (see diagram on page 112).

The swirl, the circulation, for each red face—or any 2-D face—is indicated by the arrow that runs around the one-dimensional perimeter of that face. Opposite swirl, opposite sense of circulation—or, in the word of our friends from the world of mathematics—opposite *orientation* means that these circumnavigating arrows run in opposite directions. The two abutting red faces—one belonging to one 3-D cube and the other to an adjacent 3-D cube with an oppositely directed perimeter—are mutually canceling, are mathematical negatives, the one of the other.

The momenergy scoreboard reading that goes with the one red face must be equal in magnitude but opposite in mathematical sign to the momenergy scoreboard reading of the mated red face. Because the momenergy scoreboard readings—whatever *they* may be—that belong to abutting, or mating 2-D faces are opposites, their sum—the momenergy of the interface—is zero. The momenergy scoreboard readings likewise cancel at every other 2-D interface. We feel in our bones that we begin almost to understand how nature conserves momenergy in any four-dimensional block of spacetime. But we also know that no idea is valuable until we can voice it sharp and clear: *conservation of momenergy is automatic because the forty-eight 2-D face-boundaries of the eight 3-D cube-boundaries of any 4-D spacetime region self-annihilate.*

Within a 4-D cube or block of spacetime—a region of space examined for an interval of time—momenergy is automatically conserved. How does nature audit that block of spacetime to make sure that no momenergy is created or destroyed in it? Is it enough for nature to audit the momenergy—the measure of mass in motion—contained in just one of the bounding 3-cubes? No! Instead nature must audit all eight 3-cubes that bound that 4-cube—or, rather, the content of momenergy in all eight of them. Those eight "contents of momenergy" must and do add to zero. That's nature's way to guarantee that never anywhere, in any region of space, studied for any stretch of time, is there ever any creation or destruction of momenergy.

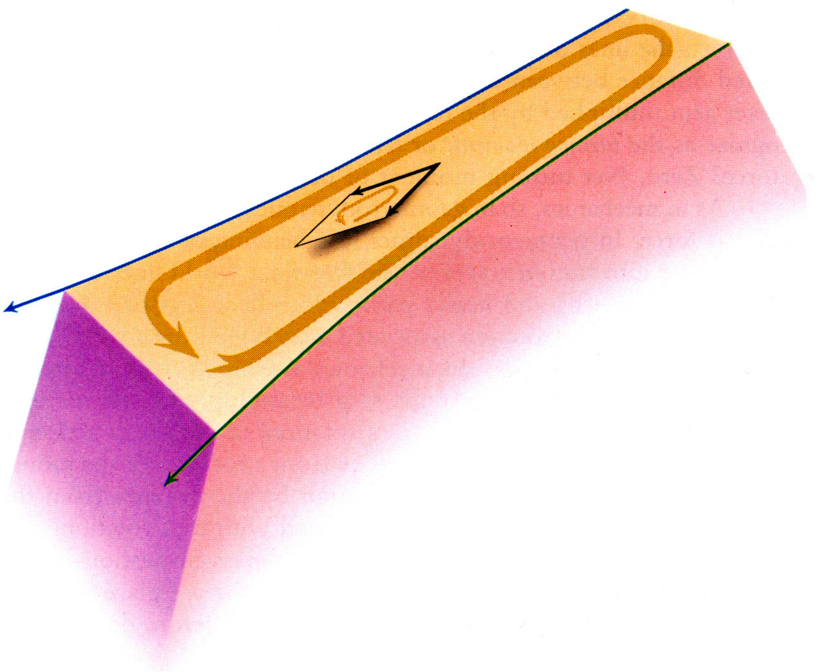
Eight cubes. Six faces for each. That's 48 faces. Each of these two-dimensional faces bears an indicator, a scoreboard, a momenergy register of some yet-unfathomed kind. Not one of those 2-faces stands exposed to any outside world. Everyone of those forty-eight 2-faces abuts another one of them. Twenty-four mated pairs!

Whatever the scoreboard of momenergy carried by any one 2-face, the scoreboard of momenergy carried by its mated face has to, and does, read the exactly opposite number. And so it is for the six faces of the next 3-cube, and the next 3-cube, and the next, all the way around all eight 3-cubes that bound that 4-D region, that block of spacetime, that audited domain, in which nature demands zero change of momenergy. This conservation condition fulfills itself automatically. How? By the principle that the 2-boundary of the 3-boundary of a 4-region is zero. That is how nature ensures and guarantees conservation of momenergy.

Action of Mass-in-Motion Inside on Spacetime Outside

Bravo! The content of momenergy inside a 3-cube, we conclude, must somehow reveal itself—without any internal probing whatsoever—through the sum of scoreboard readings on the six 2-D faces of that 3-cube. But reveal itself how? Make itself felt how? Suddenly we remember from our earlier discussions (Chapter 6) that the content, the magnitude, the measure of momenergy is mass. "Momenergy make itself felt how?" is the same as "Mass make itself felt how?" The answer to the second question and hence the first we discovered in Chapter 5: Mass—most vividly, the mass of Earth itself—makes itself felt in the curvature that it imposes on spacetime. Curvature manifested not in the free-float motion of one mass, but in the relative motion of two geodesics, two histories of free-float motion, two nearby and nearly parallel worldlines of two microscopic test masses whose directions of travel slowly bend—or twist—or bend *and* twist—relative to each other. Two such geodesics can also define for us two of the opposite edges of a 2-D face, which though nearly parallel cannot be completely so, because of the slight bending of the worldlines. This bending, this rotation of two geodesic boundaries of that face relative to each other, does it not provide at last our long-sought 2-face "scoreboard indicator" for momenergy? Do we indeed get a true count of the momentum and energy contained within a 3-D cube by adding the angles of bending associated with the six 2-D faces of that cube? What could be simpler? We have only to ask this question to feel ourselves happily launched on the wonderful way to understanding the grip of mass-in-motion on spacetime.

We can measure the bending, the angle, the rotation associated with each 2-D face, or plaquette—a two-dimensional slice of spacetime—by the method of parallel transport that we investigated in Chapter 5 (page 79). Geodesics, free-float worldlines, run along and define two nearly



The rotation, *or bending-plus-twist, associated with the 2-D face of a 3-D cube, one of the eight blocks that bound a region of 4-D spacetime. The sweeping green and blue lines mark two of the edges of the face. For simplicity we envisage both as free-float worldlines that start parallel at the remote end of the face—and think of Blue as carrying a radar mounted on a gyroscope-stabilized platform. To "start parallel" means that initially Green shows up on Blue's radar as having a zero speed of approach and a zero rate of change of direction. In consequence of the curvature of the intervening spacetime, by the end of the run Green shows up as not only approaching but also changing direction. In other words, spacetime curvature has rotated Green's spacetime velocity relative to Blue's. This rotation has been translated, in the inset at the center of the 2-D face, to two arrows of unit length and a parallelogram, which they span and thereby define. The "orientation" or sense of swirl (gold arrows) has this in common with the parallelogram and the 2-D face, that it runs concurrent with the edge or arrow that belongs to Blue.*

parallel edges of each face; the other two nearly parallel edges cut those two free-float worldlines. Thus they define the start and stop of the stretch of line in which we are interested. Moreover, each edge of any one face also is an edge for an adjacent face. Belonging to one face, that edge is traversed in parallel transport in one direction. Belonging to the other face, it is traversed a second time but in the opposite direction. The first traverse contributes to the rotation associated with the one face. The second traverse makes a contribution to the rotation of the other face equal in magnitude but opposite in sign to the rotation associated with the first face. Thus when we add all six rotations, we count the contributions of all the edges twice over—once with a plus sign and once with a minus sign. The total? Zero! Zero for our intended measure of the momenergy inside the 3-cube! Heavens! Something *must* be wrong in our thinking.

We stumble but continue. Any minute now, we're sure, we're going to discover the answer to the central question: how does mass-in-motion *inside* reveal itself in spacetime bending *outside,* at the surface? The many-colored faces of the Tiffany cube flash encouragement.

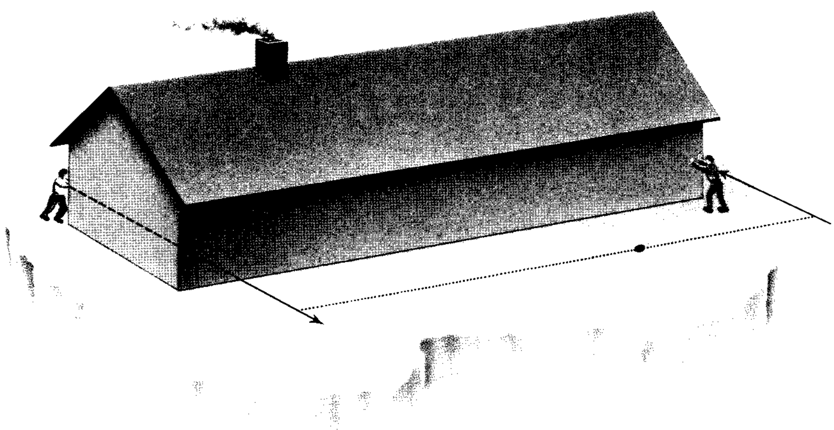
In imagination we see a shack adrift upon the ice of a frozen sea. The owner is making an urgent appeal: turn the building to face south but don't displace its center! Is this possible? Two men appear and start pushing on the shack with equal and opposite forces. No displacement of the center occurs, but the lodge slowly starts to turn. Why? The separation between the lines of action of those two forces, a perpendicularly measured distance between them, acts as a "lever arm." The length of that lever arm multiplied by the force gives what students of mechanics recognize as the *turning moment,* or *torque* responsible for that turning. Net force? Zero. Net turning moment of those forces? Non-zero!

As in mechanics, so in geometrodynamics. In mechanics, the central idea is force. In geometrodynamics it is bending or rotation—the rotation of the direction of travel of one free-float world line, one geodesic, relative to a nearby and nearly parallel free-float geodesic. Here we are concerned with the rotation associated with one of the six 2-D faces that enclose a little 3-D cube. In mechanics we move from a force to the moment of that force in a single step by multiplying the force by the distance from a fulcrum (i.e., axis of rotation) to the line of action of that force. Here, too, in one leap, we move from a rotation to the moment of that rotation by multiplying the rotation by the distance from the fulcrum to the center of the 2-face associated with that rotation.

In figuring the grip of momenergy—of mass-in-motion—on spacetime, it does not matter where, in imagination, we place the fulcrum. How come? As well ask why a new choice of fulcrum does not change the value of the total torque exerted by two equal forces upon the shack. In the case of the shack, the sum of the *forces*—due allowance being made for their counterdirectedness—is zero. Here, analogously, the sum of the *rotations* associated with the six faces of a 3-D cube is zero.

The tempo of the flashing colored faces has risen. The long-sought answer to the action of mass inside on spacetime outside, it signals, now lies in ready reach. In the turning of the shack, the key quantity was not the force—for the two countervailing forces added up to zero—but the moment of that force. Likewise, in the action of mass on spacetime, the vital quantity associated with a 2-D face is not the bending, not the angle, not the rotation, but the *moment of rotation*. The rotations, summed up for the six faces, add to zero. But the moments of rotation don't! *The sum of the moments of surface-located rotation reveals and measures the amount of mass-in-motion—the momenergy—inside the 3-cube:*

$$\begin{pmatrix} \text{sum of moments} \\ \text{of rotation for} \\ \text{the faces of a} \\ \text{little 3-cube} \end{pmatrix} = 8\pi \times \begin{pmatrix} \text{amount of} \\ \text{momenergy within} \\ \text{that 3-cube} \end{pmatrix}$$

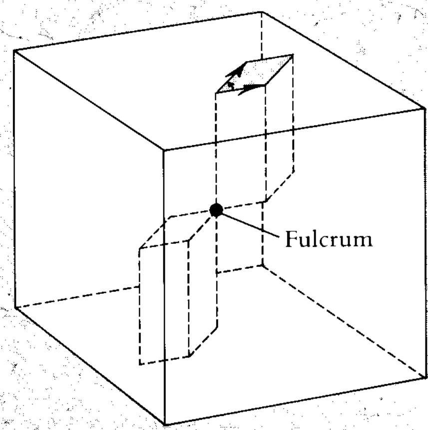


The total turning moment is independent of the location of the fulcrum. *Two men push on the opposite corners of the shack with equal but opposite forces. Each exerts a torque or turning moment about a fulcrum located in imagination wherever the beholder chooses to place it. The perpendicular distance from that fulcrum to either force governs the moment which the beholder attributes to that force. However, the sum of the turning moments, the total torque, which is what counts for swinging the shack around, does not depend on the placement of the fulcrum.*

This single simple expression—the Einstein-Cartan equation—gives us the most vivid image that mankind has ever won of the living heart of gravity. Here shines forth the influence of energy and momentum—whether of mass-in-motion, or of electromagnetic fields, or of other fields, or of all of these—on spacetime curvature at the two-dimensional boundary of any elementary 3-D region. Here glitters in a simple geometric form Einstein's 1915 battle-tested and still-standard *law of geometrodynamics*, his famous field equation. This equation holds the answers to every question about gravity that we know enough to ask.

A CLOSER LOOK AT MOMENT OF ROTATION

The moments of rotation associated with the top and bottom faces of a 3-cube are represented as parallelepipeds, three-dimensional objects. One edge of each is the line from a fulcrum, or center of rotation, inside or outside the 3-cube to the center of the face in question. The side of each parallelepiped away from the fulcrum is a little parallelogram. Spanning the parallelograms are two arrows of *unit* length. They symbolize the initial and the final orientation of a marker arrow to be carried, in imagination, in parallel transport around the perimeter of the face in question to measure the rotation associated with that face. The volume contained within the parallelepiped expresses itself, then, not in cubic meters but in meters—that is, (meters) × (dimension-free number) × (dimension-free number). And the meter, too, is the measure of mass, and of momentum and energy. As far as units of measure go, we are evidently on the right track. But momenergy has the quality of an arrow, an arrow that points in a well-defined direction in spacetime. Whoever heard of a 3-D parallelepiped pointing anywhere? Our everyday intuitions, however, unconsciously assume that the objects we examine sit in a 3-D space. In spacetime, there *is* a well-defined direction perpendicular to the 3-D parallelepiped. *That* is the direction of that parallelepiped's contribution to the arrow of momenergy! This feature of the grip of gravity is like the direction of a torque—a turning moment—such as the direction in which a screw advances through the wood when we twist it to the right! With direction now evident for each face's contribution to the momenergy, and its magnitude also evident, six directed arrows stand forth and have only to be combined, as arrows combine, to reveal the total momenergy contained in the 3-D cube.



Moments of rotation associated with the top and bottom faces of a 3-cube, represented as parallelepipeds. The sum of the volumes of the parallelepipeds associated with all six faces tells us how much momenergy there is in the 3-cube. That sum is independent of the location of the fulcrum (black dot). The perpendicular—in 4-D spacetime—to that totalized 3-volume determines the direction of the arrow of momenergy.

Geometrodynamic Field Equation

Einstein's field equation can tell us the gravity—that is, the spacetime curvature—in and around a star, a black hole, or any other spherically symmetric center of mass. And it can confirm the strength and directional properties for the tide-driving curvature outside that mass, which experience already tells us are correct (page 89).

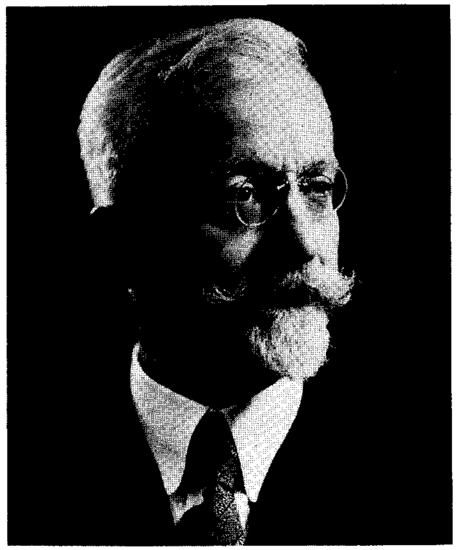
Can it tell us in a jiffy the story of the expansion of a closed model universe, and of the slowing of this expansion, when the model is so simple that in it the density of mass departs nowhere greatly—on the average—from uniformity? Yes.

Can it tell us the workings of a gravity wave—how the collapse of a distant star generates it, how it zings through space, how it makes the Earth tremble beneath our feet? Yes.

All this and more follows from the geometrodynamic field equation. This marvelous statement tells how mass—how momentum and energy—grip spacetime. It reveals how mass, there, bends spacetime geometry, there—deforms it, warps it—as a jumper warps the trampoline beneath his feet. Cupping of the fabric there, however, demands and enforces warping on the canvas in all the domain roundabout. Likewise spacetime geometry there, bent in one way where mass sits or moves, demands and enforces bending of another kind on all the surrounding empty spacetime.

Bending spreads its influence from region to region. The Einstein geometrodynamic influence equation describes and quantifies this spread of influence from region to region. In any region of emptiness, momenergy is zero. So is the moment of rotation—because zero momenergy in any locale implies and demands that there the moment of rotation also must be zero. But for the shack sitting on the ice, with two men pushing on it at different places and in opposite directions, does a zero total turning effect, a zero sum for the separate turning moments of those forces, mean that those forces themselves are individually zero? Not at all.

As in mechanics, so in geometrodynamics! In the emptiness outside a center of mass do we find a zero disturbance, a zero measure of geometric influence, a zero departure of spacetime from ideal? First, the moment of rotation. Yes, that's zero. Second, rotation itself, relative bending of nearby and nearly parallel free-float worldlines, change in their relative motion. Like forces in mechanics, the rotation is not zero. Were it zero, there'd be zero gravity outside the Earth, no planetary orbits, no gravity waves! That rotation, that bending, that change in the relative speed of nearby worldlines—that's what gravity, examined *locally, is!*



Élie Cartan

The French mathematician and geometer Élie Cartan (1869–1951) used his calculus of "exterior differential forms" to derive the simple Einstein-Cartan equation from Einstein's much more complicated geometrodynamic field equation. Cartan's boundary-of-a-boundary geometric approach to Einstein's theory provides a wonderful simplicity. In 1929, Cartan sent Einstein some of his mathematical results concerning general relativity. Einstein wrote in reply: "I didn't at all understand the explanations you gave me . . . ; still less was it clear to me how they might have been useful for physical theory." Einstein's difficulties may have stemmed from Cartan's "extremely elliptic style that . . . has baffled two generations of mathematicians," according to Jean Dieudonné.

Einstein's influence law thus displays not only the grip of mass where it is on spacetime where it is. It also tells how bent spacetime here grabs and bends adjacent spacetime.

Even beyond the grip of mass on spacetime, the moment-of-rotation equation reveals and rules the grip of spacetime on mass. Grip *on* mass? Yes, reaction to the grip *of* mass! That's the other half of the story of gravity. That grip on mass enforces the law of conservation of momenergy. Enforces it automatically. Enforces this conservation in a subtle, clever, hidden way. Enforces it via the principle that the 2-boundary of a 3-boundary of a 4-D spacetime region is zero!

We have traveled an adventurous course in these last pages. Content of momenergy *within* a 3-cube, we have found, is the sum of moments of rotation—rotation, or relative bending of worldlines, or spacetime curvature—as evidenced at the six 2-D faces of that 3-cube. Every single one of these 2-D faces, however, is paired with a totally compensating face when we consider not a single 3-D cube but all eight of the 3-D cubes that surround an elementary 4-D block of spacetime. In consequence of this pairing, this mutual annihilation of these *faces,* the sum of the momenergy contained *within* all eight 3-cubes adds up to zero. Zero total content of momentum and energy as summed up for the eight cubes that constitute the *3-D boundary of that little spacetime region.* And that's our sophisticated way to say that in that region there is total balance between what goes in and what goes out! By extension to other times and other places, this central boundary-of-a-boundary feature of geometrodynamics ensures that never at any time or any place is there ever any creation or destruction of momenergy whatsoever!

Is Physics at Bottom "Law without Law"?

This great world around us—how is it put together? Out of gears and pinions? By a corps of Swiss watchmakers? According to some multifaceted master plan embodying an all-embracing corpus of laws and regulations? Or the direct opposite? Are we destined to find that every law of physics, pushed to the extreme of experimental test, is statistical—as heat is—not mathematically perfect and precise? Is physics in the end "law without law," the very epitome of austerity?

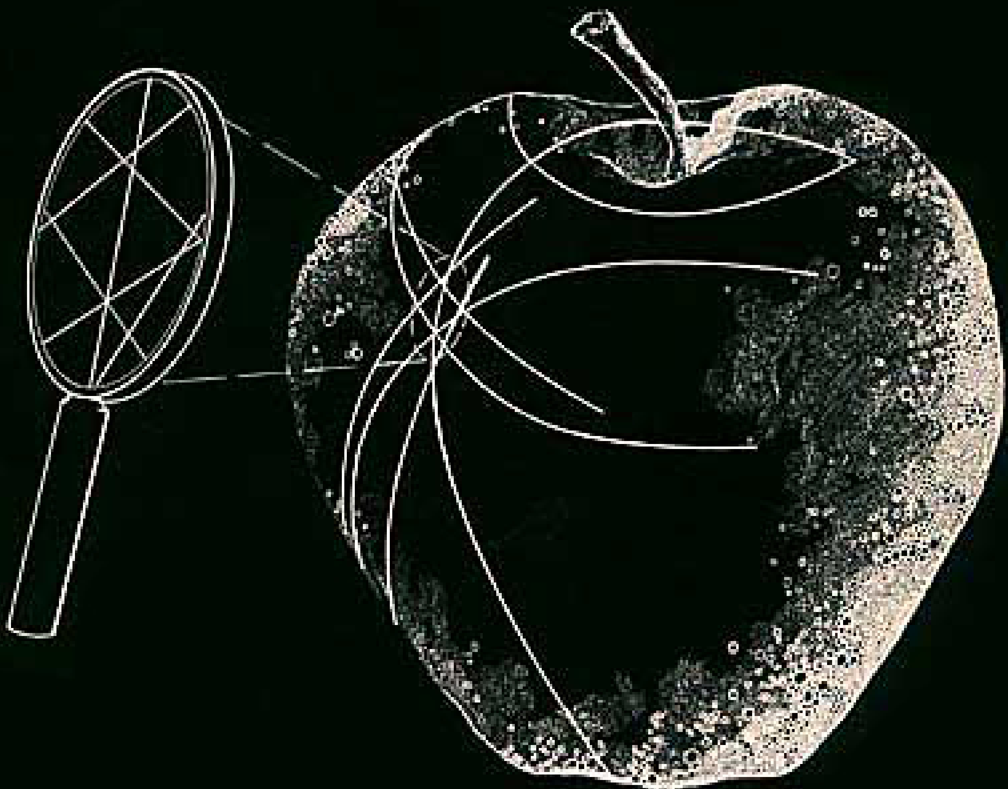
Nothing seems at first sight to conflict more violently with austerity than all the beautiful structure of the three great field theories of our age—electrodynamics, geometrodynamics, and string-theory dynamics. They are the fruit of a century of labor, hundreds of experiments, scores of gifted investigators. How can we possibly imagine all this ordered richness originating in austerity? Only a principle of organization that is no organization at all would seem to meet any demand for total austerity. In all of mathematics, nothing of this kind more obviously offers itself than the principle that the boundary of a boundary is zero. Moreover, this principle occupies a central place in all three of today's great field theories. To this extent almost all of physics founds itself on almost nothing.

Far-seeing Gottfried Wilhelm Leibniz advocated a still greater vision of existence: "For deriving everything out of nothing one principle suffices." Was he right? Underneath the workings of the world will someday a humble, thoughtful, gifted knot of searchers lay open to view the great unifying principle? Noble work for man! Rich gift to mankind!

To confront issues so cosmic is to turn back with a fresh eye to how gravity works—via the double grip of spacetime on mass and of mass on spacetime. Nowhere better than by examining that grip can we see more vividly, more instructively, and more impressively the reach and power of that austerity-flavored principle, *the boundary of a boundary is zero.* And nowhere will we find that principle operating more beautifully, more simply, and with more direct ties to everyday experience than in the warping of space—and of spacetime—around a spherically symmetric center of attraction.

GRAVITATION

Charles W. MISNER   Kip S. THORNE   John Archibald WHEELER



2

CHAPTER 15

BIANCHI IDENTITIES AND THE BOUNDARY OF A BOUNDARY

§15.1. BIANCHI IDENTITIES IN BRIEF

This chapter is entirely Track 2. As preparation, one needs to have covered (1) Chapter 4 (differential forms) and (2) Chapter 14 (computation of curvature). In reading it, one will be helped by Chapters 9–11 and 13. It is not needed as preparation for any later chapter, but it will be helpful in Chapter 17 (Einstein field equations).

Geometry gives instructions to matter, but how does matter manage to give instructions to geometry? Geometry conveys its instructions to matter by a simple handle: "pursue a world line of extremal lapse of proper time (geodesic)." What is the handle by which matter can act back on geometry? How can one identify the right handle when the metric geometry of Riemann and Einstein has scores of interesting features? Physics tells one what to look for: *a machinery of coupling between gravitation (spacetime curvature) and source* (matter; stress-energy tensor $\boldsymbol{T}$) *that will guarantee the automatic conservation of the source* ($\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$). Physics therefore asks mathematics: "What tensor-like feature of the geometry is automatically conserved?" Mathematics comes back with the answer: "The Einstein tensor." Physics queries, "How does this conservation come about?" Mathematics, in the person of Élie Cartan, replies, "Through the principle that 'the boundary of a boundary is zero'" (Box 15.1).

Actually, two features of the curvature are automatically conserved; or, otherwise stated, the curvature satisfies two Bianchi identities, the subject of this chapter. Both features of the curvature, both "geometric objects," lend themselves to representation in diagrams, moreover, diagrams that show in action the principle that "the boundary of a boundary is zero." In this respect, the geometry of spacetime shows a striking analogy to the field of Maxwell electrodynamics.

Identities and conservation of the source: electromagnetism and gravitation compared:

In electrodynamics there are four potentials that are united in the 1-form $\boldsymbol{A} \equiv A_\mu \boldsymbol{d}x^\mu$. Out of this quantity by differentiation follows the ***Faraday***, $\boldsymbol{F} = \boldsymbol{dA}$. This field satisfies the identity $\boldsymbol{dF} = 0$ (identity, yes; identity lending itself to the definition of a conserved source, no).

$\boldsymbol{dF} \equiv 0$

In gravitation there are ten potentials (metric coefficients $g_{\mu\nu}$) that are united in the metric tensor $\boldsymbol{g} = g_{\mu\nu}\, \boldsymbol{dx}^{\mu} \otimes \boldsymbol{dx}^{\nu}$. Out of this quantity by two differentiations follows the curvature operator

$$\mathcal{R} = \frac{1}{4} \boldsymbol{e}_{\mu} \wedge \boldsymbol{e}_{\nu} R^{\mu\nu}{}_{\alpha\beta}\, \boldsymbol{dx}^{\alpha} \wedge \boldsymbol{dx}^{\beta}.$$

This curvature operator satisfies the Bianchi identity $\boldsymbol{d}\mathcal{R} = 0$, where now "$\boldsymbol{d}$" is a generalization of Cartan's exterior derivative, described more fully in Chapter 14 (again an identity, but again one that does not lend itself to the definition of a conserved source).

$\boldsymbol{d}\mathcal{R} \equiv 0$

In electromagnetism, one has to go to the dual, $^*\boldsymbol{F}$, to have any feature of the field that offers a handle to the source, $\boldsymbol{d}^*\boldsymbol{F} = 4\pi\, ^*\boldsymbol{J}$. The conservation of the source, $\boldsymbol{d}^*\boldsymbol{J} = 0$, appears as a consequence of the identity $\boldsymbol{dd}^*\boldsymbol{F} = 0$; or, by a rewording of the reasoning (Box 15.1), as a consequence of the vanishing of the boundary of a boundary.

$\boldsymbol{dd}^*\boldsymbol{F} \equiv 0$ plus Maxwell equations $\Longrightarrow \boldsymbol{d}^*\boldsymbol{J} = 0$

(*continued on page 370*)

Box 15.1 THE BOUNDARY OF A BOUNDARY IS ZERO

A. The Idea in Its 1-2-3-Dimensional Form

Begin with an oriented cube or approximation to a cube (3-dimensional).

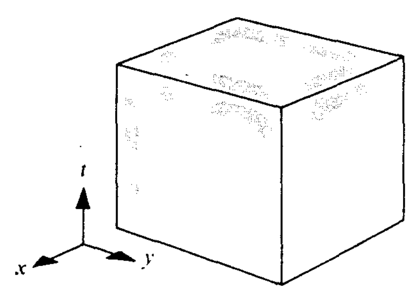
Its boundary is composed of six oriented faces, each two-dimensional. Orientation of each face is indicated by an arrow.

Boundary of any one oriented face consists of four oriented edges or arrows, each one-dimensional.

Every edge unites one face with another. No edge stands by itself in isolation.

"Sum" over all these edges, with due regard to sign. Find that any given edge is counted twice, once going one way, once going the other.

Conclude that the one-dimensional boundary of the two-dimensional boundary of the three-dimensional cube is identically zero.

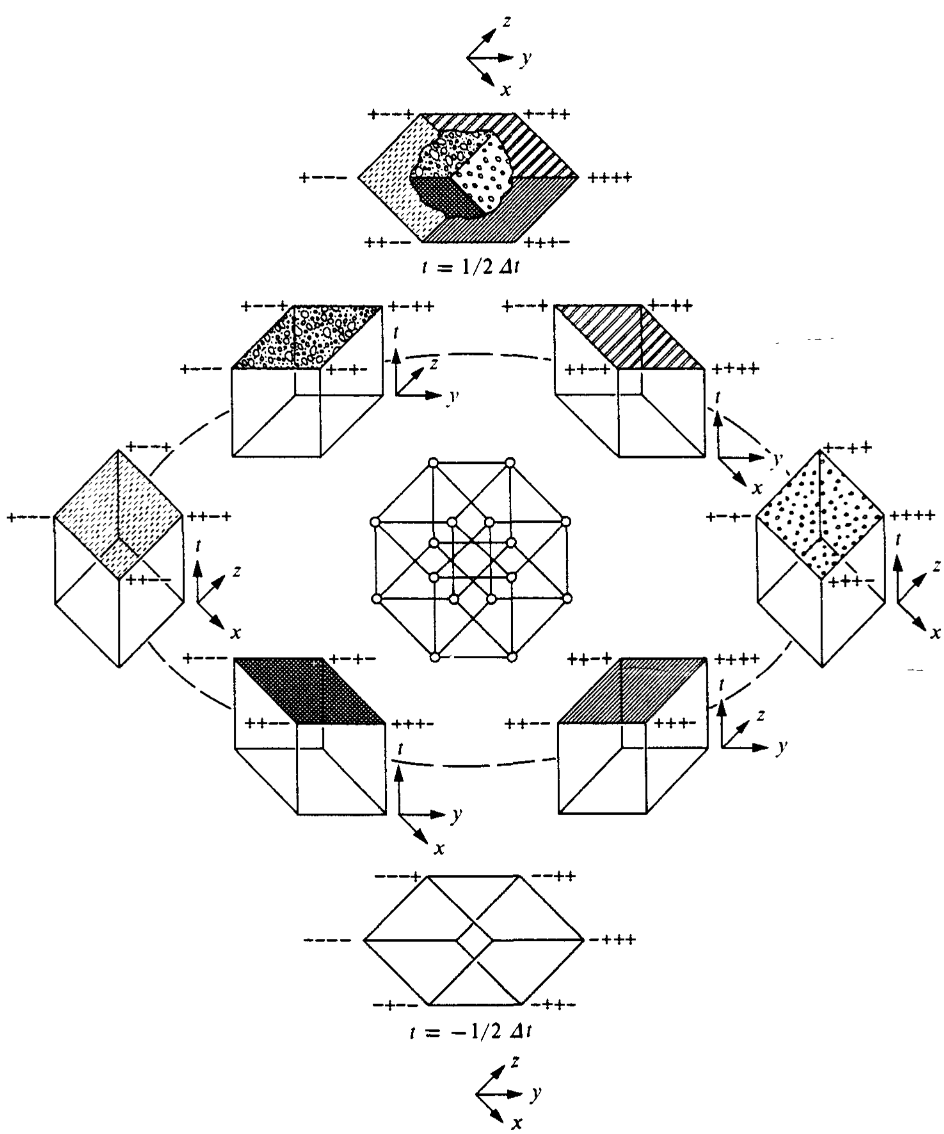


Box 15.1 ***(continued)***

B. The Idea in Its 2-3-4-Dimensional Form

Begin with an oriented four-dimensional cube or approximation thereto. The coordinates of the typical corner of the four-cube may be taken to be $(t_0 \pm \frac{1}{2}\Delta t,\ x_0 \pm \frac{1}{2}\Delta x,\ y_0 \pm \frac{1}{2}\Delta y,\ z_0 \pm \frac{1}{2}\Delta z)$; and, accordingly, a sample corner itself, in an obvious abbreviation, is conveniently abbreviated $+--+$. There are 16 of these corners. Less complicated in appearance than the 4-cube itself are its three-dimensional faces, which are "exploded off of it" into the surrounding area of the diagram, where they can be inspected in detail.

The boundary of the 4-cube is composed of eight oriented hyperfaces, each of them three-dimensional (top hyperface with extension $\Delta x\, \Delta y\, \Delta z$, for example; a "front" hyperface with extension $\Delta t\, \Delta y\, \Delta z$; etc.)



Boundary of any one hyperface ("cube") consists of six oriented faces, each two-dimensional.

Every face (for example, the hatched face $\Delta x\, \Delta y$ in the lower lefthand corner) unites one hypersurface with another (the "3-cube side face" $\Delta t\, \Delta x\, \Delta y$ in the lower lefthand corner with the "3-cube top face" $\Delta x\, \Delta y\, \Delta z$, in this example). No face stands by itself in isolation. The three-dimensional boundary of the 4-cube exposes no 2-surface to the outside world. It is *faceless.*

"Sum" over all these faces, with due regard to orientation. Find any given face is counted twice, once with one orientation, once with the opposite orientation.

Conclude that the two-dimensional boundary of the three-dimensional boundary of the four-dimensional cube is identically zero.

C. The Idea in Its General Abstract Form

$\partial\partial = 0$ (the boundary of a boundary is zero).

D. Idea Behind Application to Gravitation and Electromagnetism

The one central point is a law of conservation (conservation of charge; conservation of momentum-energy).

The other central point is "automatic fulfillment" of this conservation law.

"Automatic conservation" requires that source not be an agent free to vary arbitrarily from place to place and instant to instant.

Source needs a tie to something that, while having degrees of freedom of its own, will cut down the otherwise arbitrary degrees of freedom of the source sufficiently to guarantee that the source automatically fulfills the conservation law. Give the name "field" to this something.

Define this field and "wire it up" to the source in such a way that the conservation of the source shall be an automatic consequence of the "zero boundary of a boundary." Or, more explicitly: Conservation demands no creation or destruction of source inside the four-dimensional cube shown in the diagram. Equivalently, integral of "creation events" (integral of $\boldsymbol{d}\,{}^*\boldsymbol{J}$ for electric charge; integral of $\boldsymbol{d}\,{}^*\boldsymbol{T}$ for energy-momentum) over this four-dimensional region is required to be zero.

Integral of creation over this four-dimensional region translates into integral of source density-current (${}^*\boldsymbol{J}$ or ${}^*\boldsymbol{T}$) over three-dimensional boundary of this region. This boundary consists of eight hyperfaces, each taken with due regard to orientation. Integral over upper hyperface ("$\Delta x\, \Delta y\, \Delta z$)" gives amount of source present at later moment; over lower hyperface gives amount of souce present at earlier moment; over such hyperfaces as "$\Delta t\, \Delta x\, \Delta y$" gives outflow of source over intervening period of time. Conservation demands that sum of these eight three-dimensional integrals shall be zero (details in Chapter 5).

Box 15.1 ***(continued)***

Vanishing of this sum of three-dimensional integrals states the conservation requirement, but does not provide the machinery for "automatically" (or, in mathematical terms, "identically") meeting this requirement. For that, turn to principle that "boundary of a boundary is zero."

Demand that integral of source density-current over any oriented hyperface $\mathcal{V}$ (three-dimensional region; "cube") shall equal integral of field over faces of this "cube" (each face being taken with the appropriate orientation and the cube being infinitesimal):

$$4\pi \int_{\mathcal{V}} {}^*\boldsymbol{J} = \int_{\partial\mathcal{V}} {}^*\boldsymbol{F}; \qquad 8\pi \int_{\mathcal{V}} {}^*\boldsymbol{T} = \int_{\partial\mathcal{V}} \begin{pmatrix}\text{moment of}\\ \text{rotation}\end{pmatrix}.$$

Sum over the six faces of this cube and continue summing until the faces of all eight cubes are covered. Find that any given face (as, for example, the hatched face in the diagram) is counted twice, once with one orientation, once with the other ("boundary of a boundary is zero"). Thus is guaranteed the conservation of source: integral of source density-current over three-dimensional boundary of four-dimensional region is automatically zero, making integral of creation over interior of that four-dimensional region also identically zero.

Repeat calculation with boundary of that four-dimensional region slightly displaced in one locality [the "bubble differentiation" of Tomonaga (1946) and Schwinger (1948)], and conclude that conservation is guaranteed, not only in the four-dimensional region as a whole, but at every point within it, and, by extension, everywhere in spacetime.

E. Relation of Source to Field

One view: Source is primary. Field may have other duties, but its prime duty is to serve as "slave" of source. Conservation of source comes first; field has to adjust itself accordingly.

Alternative view: Field is primary. Field takes the responsibility of seeing to it that the source obeys the conservation law. Source would not know what to do in absence of the field, and would not even exist. Source is "built" from field. Conservation of source is consequence of this construction.

One model illustrating this view in an elementary context: Concept of "classical" electric charge as nothing but "electric lines of force trapped in the topology of a multiply connected space" [Weyl (1924b); Wheeler (1955); Misner and Wheeler (1957)].

On any view: Integral of source density-current over any three-dimensional region (a "cube" in simplified analysis above) equals integral of field over boundary of this region (the six faces of the cube above). No one has ever found any other way to understand the correlation between field law and conservation law.

F. Electromagnetism as a Model: How to "Wire Up" Source to Field to Give Automatic Conservation of Source Via "$\partial\partial = 0$" in Its 2-3-4-Dimensional Form

Conservation means zero creation of charge (zero creation in four-dimensional region Ω).

Conservation therefore demands zero value for integral of charge density-current over three-dimensional boundary of this volume; thus,

$$0 = \int_\Omega \frac{\partial J^\mu}{\partial x^\mu}\, d^4\Omega = \int_{\partial\Omega} J^\mu\, d^3\Sigma_\mu$$

in the Track-1 language of Chapters 3 and 5. Equivalently, in the coordinate-free abstract language of §§4.3-4.6, one has

$$0 = \int_\Omega \boldsymbol{d}{}^*\boldsymbol{J} = \int_{\partial\Omega} {}^*\boldsymbol{J},$$

where

$$\begin{aligned} {}^*\boldsymbol{J} = {}& {}^*J_{123}\, \boldsymbol{d}x^1 \wedge \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 + {}^*J_{023}\, \boldsymbol{d}x^0 \wedge \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 \\ & + {}^*J_{031}\, \boldsymbol{d}x^0 \wedge \boldsymbol{d}x^3 \wedge \boldsymbol{d}x^1 + {}^*J_{012}\, \boldsymbol{d}x^0 \wedge \boldsymbol{d}x^1 \wedge \boldsymbol{d}x^2 \end{aligned}$$

("eggcrate-like structure" of the 3-form of charge-density and current-density).

Fulfill this conservation requirement automatically ("identically") through the principle that "the boundary of a boundary is zero" by writing $4\pi\, {}^*\boldsymbol{J} = \boldsymbol{d}{}^*\boldsymbol{F}$; thus,

$$4\pi \int_{\partial\Omega} {}^*\boldsymbol{J} = \int_{\partial\Omega} \boldsymbol{d}{}^*\boldsymbol{F} = \int_{\partial\partial\Omega(\text{zero!})} {}^*\boldsymbol{F} \equiv 0$$

or, in Track-1 language, write $4\pi J^\mu = F^{\mu\nu}{}_{;\nu}$, and have

$$4\pi \int_{\partial\Omega} J^\mu\, d^3\Sigma_\mu = \int_{\partial\Omega} F^{\mu\nu}{}_{;\nu}\, d^3\Sigma_\mu = \int_{\partial\partial\Omega(\text{zero!})} F^{\mu\alpha}\, d^2\Sigma_{\mu\alpha} \equiv 0.$$

In other words, half of Maxwell's equations in their familiar flat-space form,

$$\operatorname{div} \boldsymbol{E} = \boldsymbol{\nabla} \cdot \boldsymbol{E} = 4\pi\rho, \qquad \operatorname{curl} \boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{B} = \dot{\boldsymbol{E}} + 4\pi \boldsymbol{J},$$

"wire up" the source to the field in such a way that the law of conservation of source follows directly from "$\partial\partial\Omega = 0$."

G. Electromagnetism Also Employs "$\partial\partial = 0$" in its 1-2-3-Dimensional Form ("No Magnetic Charge")

Magnetic charge is linked with field via $4\pi \boldsymbol{J}_{\text{mag}} = \boldsymbol{dF}$ (see point **F** above for translation of this compact Track-2 language into equivalent Track-1 terms). Absence of

Box 15.1 ***(continued)***

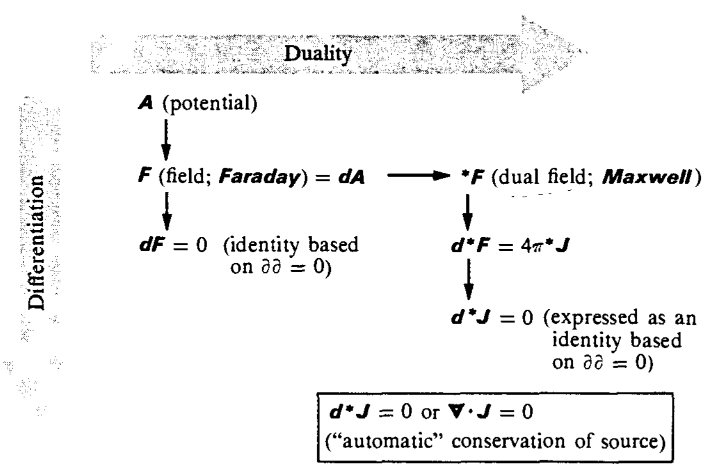
any magnetic charge says that integral of $\boldsymbol{J}_{\text{mag}}$ over any 3-volume $\mathcal{V}$ is necessarily zero; or ("integration by parts," generalized Stokes theorem)

$$0 = \int_{\mathcal{V}} \boldsymbol{dF} = \int_{\partial\mathcal{V}} \boldsymbol{F} = \begin{pmatrix}\text{total magnetic flux}\\ \text{exiting through } \partial\mathcal{V}\end{pmatrix}.$$

In order to satisfy this requirement "automatically," via principle that "the boundary of a boundary is zero," write $\boldsymbol{F} = \boldsymbol{dA}$ ("expression of field in terms of 4-potential"), and have

$$\int_{\partial\mathcal{V}} \boldsymbol{F} = \int_{\partial\mathcal{V}} \boldsymbol{dA} = \int_{\partial\partial\mathcal{V}\,(\text{zero!})} \boldsymbol{A} \equiv 0.$$

H. Structure of Electrodynamics in Outline Form

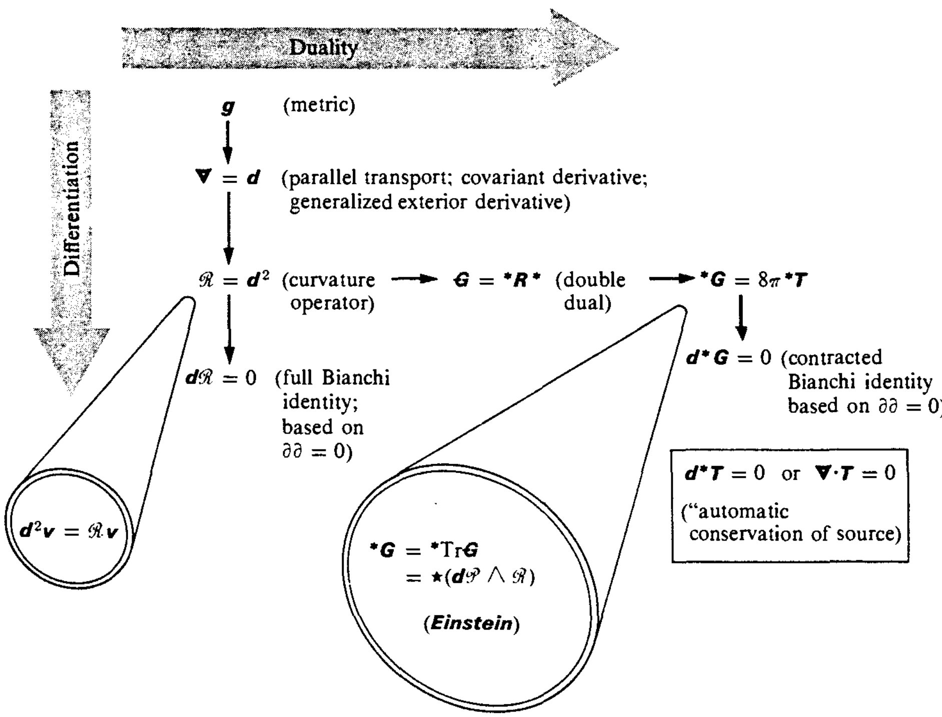


In gravitation physics, one has to go to the "double dual" (two pairs of alternating indices, two places to take the dual) $\boldsymbol{G} = {}^*\boldsymbol{R}^*$ of ***Riemann*** to have a feature of the field that offers a handle to the source:

$$\boldsymbol{G} = \text{Tr}\boldsymbol{G} = \textit{\textbf{Einstein}} = 8\pi\boldsymbol{T} = 8\pi \times (\text{density of energy-momentum}).$$

The conservation of the source $\boldsymbol{T} \equiv \boldsymbol{e}_\mu T^\mu{}_\nu \boldsymbol{\omega}^\nu$ can be stated $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$. But better suited for the present purpose is the form (see Chapter 14 and exercise 14.18)

I. Structure of Geometrodynamics in Outline Form



$$\boldsymbol{d}{}^*\boldsymbol{T} = 0,$$

$\boldsymbol{d}{}^*\boldsymbol{G} \equiv 0$ plus Einstein field equation $\Longrightarrow \boldsymbol{d}{}^*\boldsymbol{T} = 0$

where

$$^*\boldsymbol{T} \equiv \boldsymbol{e}_\mu T^\mu{}_\nu({}^*\boldsymbol{\omega}^\nu) = \boldsymbol{e}_\mu T^{\mu\nu}\, d^3\Sigma_\nu.$$

This conservation law arises as a consequence of the "contracted Bianchi identity", $\boldsymbol{d}{}^*\boldsymbol{G} = 0$, again interpretable in terms of the vanishing of the boundary of a boundary.

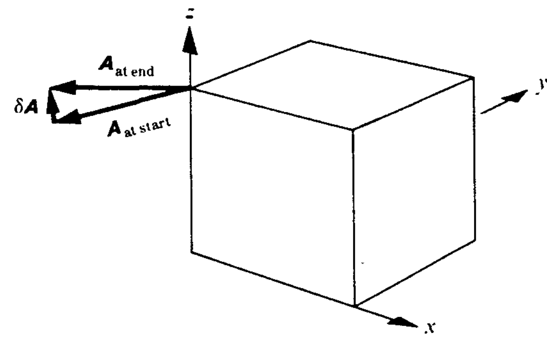


Figure 15.1.
Combine rotations associated with each of the six faces of the illustrated 3-volume and end up with zero net rotation ("full Bianchi identity"). Reason: Contribution of any face is measured by change in a test vector $\boldsymbol{A}$ carried in parallel transport around the perimeter of that face. Combine contributions of all faces and end up with each edge traversed twice, once in one direction, once in the other direction [boundary (here one-dimensional) of boundary (two-dimensional) of indicated three-dimensional figure is zero]. Detail: The vector $\boldsymbol{A}$, residing at the indicated site, is transported parallel to itself over to the indicated face, then carried around the perimeter of that face by parallel transport, experiencing in the process a rotation measured by the spacetime curvature associated with that face, then transported parallel to itself back to the original site. To the lowest relevant order of small quantities one can write

$$(\text{change in } \boldsymbol{A}) = -\Delta y\, \Delta z\, \mathcal{R}(\boldsymbol{e}_y, \boldsymbol{e}_z)\, \boldsymbol{A}$$

in operator notation; or in coordinate language,

$$-\delta A^\alpha = R^\alpha{}_{\beta yz}(\text{at } x + \Delta x)A^\beta\, \Delta y\, \Delta z.$$

§15.2. BIANCHI IDENTITY $d\mathcal{R} = 0$ AS A MANIFESTATION OF "BOUNDARY OF BOUNDARY = 0"

Bianchi identity, $\boldsymbol{d}\mathcal{R} \equiv 0$, interpreted in terms of parallel transport around the six faces of a cube.

Such is the story of the two Bianchi identities in outline form; it is now appropriate to fill in the details. Figure 15.1 illustrates the full Bianchi identity, $\boldsymbol{d}\mathcal{R} = 0$ (see exercise 14.17), saying in brief, "The sum of the curvature-induced rotations associated with the six faces of any elementary cube is zero." The change in a vector $\boldsymbol{A}$ associated with transport around the perimeter of the indicated face evaluated to the lowest relevant order of small quantities is given by

$$-\delta A^\alpha = R^\alpha{}_{\beta yz}\ (\text{at } x + \Delta x)A^\beta\, \Delta y\, \Delta z. \tag{15.1}$$

The opposite face gives a similar contribution, except that now the sign is reversed and the evaluation takes place at x rather than at $x + \Delta x$. The combination of the contributions from the two faces gives

$$\frac{\partial R^\alpha{}_{\beta yz}}{\partial x} A^\beta\, \Delta x\, \Delta y\, \Delta z, \tag{15.2}$$

when Riemann normal coordinates are in use. In such coordinates, the vanishing of the total $-\delta A^\alpha$ contributed by all six faces implies

$$R^\alpha{}_{\beta yz;x} + R^\alpha{}_{\beta zx;y} + R^\alpha{}_{\beta xy;z} = 0. \tag{15.3}$$

Here semicolons (covariant derivatives) can be and have been inserted instead of commas (ordinary derivatives), because the two are identical in the context of Riemann normal coordinates; and the covariant version (15.3) generalizes itself to arbitrary curvilinear coordinates. Turn from an *xyz* cube to a cube defined by any set of coordinate axes, and write Bianchi's identity in the form

$$R^\alpha{}_{\beta[\lambda\mu;\nu]} = 0. \tag{15.4}$$

(See exercise 14.17 for one reexpression of this identity in the abstract coordinate-independent form, $\boldsymbol{d}\mathcal{R} = 0$, and §15.3 for another.) This identity occupies much the same place in gravitation physics as that occupied by the identity $\boldsymbol{dF} = \boldsymbol{ddA} \equiv 0$ in electromagnetism:

$$F_{[\lambda\mu,\nu]} = F_{[\lambda\mu;\nu]} = 0. \tag{15.5}$$

§15.3. MOMENT OF ROTATION: KEY TO CONTRACTED BIANCHI IDENTITY

The contracted Bianchi identity, the identity that offers a "handle to couple to the source," was shown by Élie Cartan to deal with "moments of rotation" [Cartan (1928); Wheeler (1964b); Misner and Wheeler (1972)]. Moments are familiar in elementary mechanics. A rigid body will not remain at rest unless all the forces acting on it sum to zero:

$$\sum_i \boldsymbol{F}^{(i)} = 0. \tag{15.6}$$

Although necessary, this condition is not sufficient. The sum of the moments of these forces about some point $\mathcal{P}$ must also be zero:

$$\sum_i (\mathcal{P}^{(i)} - \mathcal{P}) \wedge \boldsymbol{F}^{(i)} = 0. \tag{15.7}$$

Exactly what point these moments are taken about happily does not matter, and this for a simple reason. The arbitrary point in the vector product (15.7) has for coefficient the quantity $\Sigma_i \boldsymbol{F}^{(i)}$, which already has been required to vanish. The situation is similar in the elementary cube of Figure 15.1. Here the rotation associated with a given face is the analog of the force $\boldsymbol{F}^{(i)}$ in mechanics. That the sum of these rotations vanishes when extended over all six faces of the cube is the analog of the vanishing of the sum of the forces $\boldsymbol{F}^{(i)}$.

What is the analog for curvature of the moment of the force that one encounters in mechanics? It is the *moment of the rotation associated with a given face of the*

Net moment of rotation over all six faces of a cube:

(1) described

cube. The value of any individual moment depends on the reference point $\mathscr{P}$. However, the sum of these moments taken over all six faces of the cube will have a value independent of the reference point $\mathscr{P}$, for the same reason as in mechanics. Therefore $\mathscr{P}$ can be taken where one pleases, inside the elementary cube or outside it. Moreover, the cube may be viewed as a bit of a hypersurface sliced through spacetime. Therefore $\mathscr{P}$ can as well be off the slice as on it. It is only required that all distances involved be short enough that one obtains the required precision by calculating the moments and the sum of moments in a local Riemann-normal coordinate system. One thus arrives at a $\mathscr{P}$-independent totalized moment of rotation (not necessarily zero; gravitation is not mechanics!) associated with the cube in question.

(2) equated to integral of source, $\int {}^*\boldsymbol{T}$, over interior of cube

Now comes the magic of "the boundary of the boundary is zero." Identify this net moment of rotation of the cube, evaluated by summing individual moments of rotation associated with individual faces, with the integral of the source density-current (energy-momentum tensor ${}^*\boldsymbol{T}$) over the interior of the 3-cube. Make this identification not only for the one 3-cube, but for all eight 3-cubes (hyperfaces) that bound the four-dimensional cube in Box 15.1. Sum the integrated source density-current ${}^*\boldsymbol{T}$ not only for the one hyperface of the 4-cube, but for all eight hyperfaces. Thus have

$$\int_{\text{4-cube}} \begin{pmatrix}\text{source}\\ \text{creation}\\ \boldsymbol{d}{}^*\boldsymbol{T}\end{pmatrix} = \int_{\substack{\text{3-boundary}\\ \text{of this 4-cube}}} \begin{pmatrix}\text{source current-}\\ \text{density, } {}^*\boldsymbol{T}\end{pmatrix}$$
$$= \sum_{\substack{\text{these eight}\\ \text{bounding}\\ \text{3-cubes}}} \begin{pmatrix}\text{net moment of rotation}\\ \text{associated with speci-}\\ \text{fied cube}\end{pmatrix}$$
$$= \underbrace{\sum_{\substack{\text{eight}\\ \text{bounding}\\ \text{3-cubes}}} \sum_{\substack{\text{six faces}\\ \text{bounding}\\ \text{given 3-cube}}}}_{\text{(zero!)}} \begin{pmatrix}\text{moment of rotation}\\ \text{associated with specified}\\ \text{face of specified cube}\end{pmatrix}. \qquad (15.8)$$

(3) conserved

Let the moments of rotation, not only for the six faces of one cube, but for all the faces of all the cubes, be taken with respect to one and the same point $\mathscr{P}$. Recall (Box 15.1) that any given face joins two cubes or hyperfaces. It therefore appears twice in the count of faces, once with one orientation ("sense of circumnavigation in parallel transport to evaluate rotation") and once with the opposite orientation. Therefore the double sum vanishes identically (boundary of a boundary is zero!) This identity establishes existence of a new geometric object, a feature of the curvature, that is conserved, and therefore provides a handle to which to couple a source. The desired result has been achieved. Now to translate it into standard mathematics!

§15.4. CALCULATION OF THE MOMENT OF ROTATION

It remains to find the tensorial character and value of this conserved Cartan moment of rotation that appertains to any elementary 3-volume. The rotation associated with the front face $\Delta y\, \Delta z \boldsymbol{e}_y \wedge \boldsymbol{e}_z$ of the cube in Figure 15.1 will be represented by the bivector (4) evaluated

$$\begin{pmatrix}\text{rotation associated} \\ \text{with front } \Delta y\, \Delta z \text{ face}\end{pmatrix} = \boldsymbol{e}_\lambda \wedge \boldsymbol{e}_\mu R^{|\lambda\mu|}{}_{yz}\, \Delta y\, \Delta z \qquad (15.9)$$

located at $\mathscr{P}_{\text{front}} = (t - \frac{1}{2}\Delta t,\ x + \Delta x,\ y + \frac{1}{2}\Delta y,\ z + \frac{1}{2}\Delta z)$. This equation uses Riemann normal coordinates; indices enclosed by strokes, as in $|\lambda\mu|$, are summed with the restriction $\lambda < \mu$. The moment of this rotation with respect to the point $\mathscr{P}$ will be represented by the trivector

$$\begin{pmatrix}\text{moment of rotation} \\ \text{associated with} \\ \text{front } \Delta y\, \Delta z \text{ face}\end{pmatrix} = (\mathscr{P}_{\substack{\text{center} \\ \text{of front} \\ \text{face}}} - \mathscr{P}) \wedge \boldsymbol{e}_\lambda \wedge \boldsymbol{e}_\mu R^{|\lambda\mu|}{}_{yz}\, \Delta y\, \Delta z. \qquad (15.10)$$

Here neither $\mathscr{P}_{\text{center front}}$ nor $\mathscr{P}$ has any well-defined meaning whatsoever as a vector, but their difference is a vector in the limit of infinitesimal separation, $\Delta\mathscr{P} = \mathscr{P}_{\text{center front}} - \mathscr{P}$. With the back face a similar moment of rotation is associated, with the opposite sign, and with $\mathscr{P}_{\text{center front}}$ replaced by $\mathscr{P}_{\text{center back}}$. In the difference between the two terms, the factor $\mathscr{P}$ is of no interest, because one is already assured it will cancel out [Bianchi identity (15.4); analog of $\Sigma F^{(i)} = 0$ in mechanics]. The difference $\mathscr{P}_{\text{center front}} - \mathscr{P}_{\text{center back}}$ has the value $\Delta x \boldsymbol{e}_x$. Summing over all six faces, one has

$$\begin{pmatrix}\text{net moment of} \\ \text{rotation associated} \\ \text{with cube or hyper-} \\ \text{face } \Delta x\, \Delta y\, \Delta z\end{pmatrix} =$$

$$\begin{aligned} &\boldsymbol{e}_x \wedge \boldsymbol{e}_\lambda \wedge \boldsymbol{e}_\mu R^{|\lambda\mu|}{}_{yz}\, \Delta x\, \Delta y\, \Delta z \text{ (front and back)} \\ &+ \boldsymbol{e}_y \wedge \boldsymbol{e}_\lambda \wedge \boldsymbol{e}_\mu R^{|\lambda\mu|}{}_{zx}\, \Delta y\, \Delta z\, \Delta x \text{ (sides)} \\ &+ \boldsymbol{e}_z \wedge \boldsymbol{e}_\lambda \wedge \boldsymbol{e}_\mu R^{|\lambda\mu|}{}_{xy}\, \Delta z\, \Delta x\, \Delta y \text{ (top and bottom).} \end{aligned} \qquad (15.11)$$

This sum one recognizes as the value (on the volume element $\boldsymbol{e}_x \wedge \boldsymbol{e}_y \wedge \boldsymbol{e}_z\, \Delta x\, \Delta y\, \Delta z$) of the 3-form

$$\boldsymbol{e}_\nu \wedge \boldsymbol{e}_\lambda \wedge \boldsymbol{e}_\mu R^{|\lambda\mu|}{}_{|\alpha\beta|}\, \boldsymbol{d}x^\nu \wedge \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta.$$

Moreover this 3-form is defined, and precisely defined, at a point, whereas (15.11), applying as it does to an extended region, does not lend itself to an analysis that is at the same time brief and precise. Therefore forego (15.11) in favor of the 3-form. Only remember, when it comes down to interpretation, that this 3-form is to be evaluated for the "cube" $\boldsymbol{e}_x \wedge \boldsymbol{e}_y \wedge \boldsymbol{e}_z \, \Delta x \, \Delta y \, \Delta z$. Now note that the "trivector-valued moment-of-rotation 3-form" can also be written as

(5) abstracted to give $\boldsymbol{d\mathcal{P}} \wedge \mathcal{R}$

$$\begin{pmatrix}\text{moment of}\\ \text{rotation}\end{pmatrix} = \boldsymbol{d\mathcal{P}} \wedge \mathcal{R} = \boldsymbol{e}_\nu \wedge \boldsymbol{e}_\lambda \wedge \boldsymbol{e}_\mu R^{|\lambda\mu|}{}_{|\alpha\beta|}\, \boldsymbol{d}x^\nu \wedge \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta. \quad (15.12)$$

Here

$$\boldsymbol{d\mathcal{P}} = \boldsymbol{e}_\sigma \, \boldsymbol{d}x^\sigma \quad (15.13)$$

is Cartan's $\binom{1}{1}$ unit tensor. Also $\mathcal{R}$ is the curvature operator, treated as a bivector-valued 2-form:

$$\mathcal{R} = \boldsymbol{e}_\lambda \wedge \boldsymbol{e}_\mu R^{|\lambda\mu|}{}_{|\alpha\beta|}\, \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta. \quad (15.14)$$

Using the language of components as in (15.11), or the abstract language introduced in (15.12), one finds oneself dealing with a trivector. A trivector can be left a trivector, as, in quite another context, an element of 3-volume on a hypersurface in 4-space can be left as a trivector. However, there it is more convenient to take the dual representation, and speak of the element of volume as a vector. Denote by $\star$ a duality operation that acts only on contravariant vectors, trivectors, etc. (but *not* on forms). Then in a Lorentz frame one has $\star(\boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \wedge \boldsymbol{e}_3) = \boldsymbol{e}_0$; but $\star(\boldsymbol{d}x^3) = \boldsymbol{d}x^3$. More generally,

$$\star(\boldsymbol{e}_\nu \wedge \boldsymbol{e}_\lambda \wedge \boldsymbol{e}_\mu) = \varepsilon_{\nu\lambda\mu}{}^\sigma \boldsymbol{e}_\sigma. \quad (15.15)$$

(6) abstracted to give $\star(\boldsymbol{d\mathcal{P}} \wedge \mathcal{R}) = \boldsymbol{e}_\sigma G^{\sigma\tau}\, d^3\Sigma_\tau$

In this notation, the "vector-valued moment-of-rotation 3-form" is

$$\begin{pmatrix}\text{moment}\\ \text{of rotation}\end{pmatrix} = \star(\boldsymbol{d\mathcal{P}} \wedge \mathcal{R}) = \boldsymbol{e}_\sigma \varepsilon_{\nu\lambda\mu}{}^\sigma R^{|\lambda\mu|}{}_{|\alpha\beta|}\, \boldsymbol{d}x^\nu \wedge \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta$$
$$= \boldsymbol{e}_\sigma ({}^*R)_\nu{}^\sigma{}_{|\alpha\beta|}\, \boldsymbol{d}x^\nu \wedge \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta,$$

or, in one more step,

$$\begin{pmatrix}\text{moment}\\ \text{of rotation}\end{pmatrix} = \star(\boldsymbol{d\mathcal{P}} \wedge \mathcal{R}) = \boldsymbol{e}_\sigma ({}^*R^*)_\nu{}^{\sigma\nu\tau}\, d^3\Sigma_\tau. \quad (15.16)$$

Here $d^3\Sigma_\tau$ is a notation for basis 3-forms, as in Box 5.4; thus,

$$\boldsymbol{d}x^\nu \wedge \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta = \varepsilon^{\nu\alpha\beta\tau}\, d^3\Sigma_\tau. \quad (15.17)$$

(In a local Lorentz frame, $\boldsymbol{d}x^1 \wedge \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 = d^3\Sigma_0$.)

Nothing is more central to the analysis of curvature than the formula (15.16). It starts with an element of 3-volume and ends up giving the moment of rotation in that 3-volume. The tensor that connects the starting volume with the final moment, the "contracted double-dual" of ***Riemann***, is so important that it deserves and receives a name of its own, $\boldsymbol{G} \equiv$ ***Einstein***; thus

$$(\text{Einstein})^{\sigma\tau} \equiv G^{\sigma\tau} = G_\nu{}^{\sigma\nu\tau} = ({}^*R^*)_\nu{}^{\sigma\nu\tau}. \quad (15.18)$$

This tensor received attention in §§13.5 and 14.2, and also in the examples at the end of Chapter 14. In terms of ***Einstein***, the connection between element of 3-volume and "vector-valued moment of rotation" is

$$\begin{pmatrix}\text{moment}\\ \text{of rotation}\end{pmatrix} = \star(\boldsymbol{d}\mathcal{P} \wedge \mathcal{R}) = \boldsymbol{e}_\sigma G^{\sigma\tau}\, d^3\Sigma_\tau. \tag{15.19}$$

The amount of "vector-valued moment of rotation" contained in the element of 3-volume $d^3\Sigma_\mu$ is identified by general relativity with the amount of energy-momentum contained in that 3-volume. However, defer this identification for now. Concentrate instead on the conservation properties of this moment of rotation. See them once in the formulation of integral calculus, as a consequence of the principle "$\partial\partial \equiv 0$." See them then a second time, in differential formulation, as a consequence of "$\boldsymbol{dd} \equiv 0$."

§15.5. CONSERVATION OF MOMENT OF ROTATION SEEN FROM "BOUNDARY OF A BOUNDARY IS ZERO"

Conservation of net moment of rotation:

(1) derived from "$\partial\partial = 0$"

The moment of rotation defines an automatically conserved quantity. In other words, the value of the moment of rotation for an elementary 3-volume $\Delta x\, \Delta y\, \Delta z$ after the lapse of a time Δt is equal to the value of the moment of rotation for the same 3-volume at the beginning of that time, corrected by the inflow of moment of rotation over the six faces of the 3-volume in that time interval (quantities proportional to $\Delta y\, \Delta z\, \Delta t$, etc.) Now verify this conservation of moment of rotation in the language of "the boundary of a boundary." Follow the pattern of equation (15.8), but translate the words into formulas, item by item. Evaluate the amount of moment of rotation created in the elementary 4-cube Ω, and find

$$\text{"creation"} \underset{\text{definition}}{\equiv} \int \begin{pmatrix}\text{"creation of moment of}\\ \text{rotation" in the elementary}\\ \text{4-cube of spacetime } \Omega\end{pmatrix} \underset{\text{definition}}{=} \int_\Omega \boldsymbol{d}{}^*\boldsymbol{G};$$

$$\int_\Omega \boldsymbol{d}{}^*\boldsymbol{G} \underset{\text{step 1}}{=} \int_{\partial\Omega} {}^*\boldsymbol{G} \underset{\text{step 2}}{=} \int_{\partial\Omega} \star(\boldsymbol{d}\mathcal{P} \wedge \mathcal{R}) \underset{\text{step 3}}{=} \sum_{\substack{\text{the eight}\\ \text{3-cubes}\\ \text{that bound } \Omega}} \star \begin{pmatrix}\text{moment of rotation}\\ \int_{\text{3-cube}} (\boldsymbol{d}\mathcal{P} \wedge \mathcal{R})\\ \text{associated with}\\ \text{specified 3-cube}\end{pmatrix} \underset{\text{step 4}}{=}$$

$$\underset{\text{step 4}}{=} \sum_{\substack{\text{eight bounding}\\ \text{3-cubes}}} \sum_{\substack{\text{six faces bounding}\\ \text{specified 3-cube}}} \star \begin{pmatrix}\text{moment of rotation}\\ \int_{\text{face}} (\mathcal{P} \wedge \mathcal{R})\\ \text{associated with}\\ \text{specified face of}\\ \text{specified cube}\end{pmatrix} \underset{\text{step 5}}{\equiv} 0. \tag{15.20}$$

2

Here step 1 is the theorem of Stokes. Step 2 is the identification established by (15.19) between the Einstein tensor and the moment of rotation. Step 3 breaks down the integral over the entire boundary $\partial\Omega$ into integrals over the individual 3-cubes that constitute this boundary. Moreover, in all these integrals, the star ★ is treated as a constant and taken outside the sign of integration. The reason for such treatment is simple: the duality operation ★ involves only the metric, and the metric is locally constant throughout the infinitesimal 4-cube over the boundary of which the integration extends. Step 4 uses the formula

$$\boldsymbol{d}(\mathscr{P} \wedge \mathscr{R}) = \boldsymbol{d}\mathscr{P} \wedge \mathscr{R} + \mathscr{P} \wedge \boldsymbol{d}\mathscr{R} = \boldsymbol{d}\mathscr{P} \wedge \mathscr{R} \tag{15.21}$$

and the theorem of Stokes to express each 3-cube integral as an integral of $\mathscr{P} \wedge \mathscr{R}$ over the two-dimensional boundary of that cube. The culminating step is 5. It has nothing to do with the integrand. It depends solely on the principle $\partial\partial \equiv 0$.

In brief, the conservation of moment of rotation follows from two circumstances. (1) The moment of rotation associated with any elementary 3-cube is by definition a net value, obtained by adding the six moments of rotation associated with the six faces of that cube. (2) When one sums these net values for all eight 3-cubes in (15.20), which are the boundary of the elementary 4-cube Ω, one counts the contribution of a given 2-face twice, once with one sign and once with the opposite sign. In virtue of the principle that "the boundary of a boundary is zero," the conservation of moment of rotation is thus an identity.

§15.6. CONSERVATION OF MOMENT OF ROTATION EXPRESSED IN DIFFERENTIAL FORM

(2) derived from "$\boldsymbol{dd} = 0$"

Every conservation law stated in integral form lends itself to restatement in differential form, and conservation of moment of rotation is no exception. The calculation is brief. Evaluate the generalized exterior derivative of the moment of rotation in three steps, and find that it vanishes; thus:

$$\begin{aligned} \boldsymbol{d}{}^{*}\boldsymbol{G} &= \boldsymbol{d}[\star(\boldsymbol{d}\mathscr{P} \wedge \mathscr{R})] && \} \text{ step 1} \\ &= \star[\boldsymbol{d}(\boldsymbol{d}\mathscr{P} \wedge \mathscr{R})] && \} \text{ step 2} \\ &= \star[\boldsymbol{d}^2\mathscr{P} \wedge \mathscr{R} - \boldsymbol{d}\mathscr{P} \wedge \boldsymbol{d}\mathscr{R}] && \} \text{ step 3} \\ &= 0 \end{aligned}$$

Step 1 uses the relation $\boldsymbol{d}\star = \star\boldsymbol{d}$. The star duality and the generalized exterior derivative commute because when $\boldsymbol{d}$ is applied to a contravariant vector, it acts as a covariant derivative, and when ★ is applied to a covariant vector or 1-form, it is without effect. Step 2 applies the standard rule for the action of $\boldsymbol{d}$ on a product of tensor-valued forms [see equation (14.13b)]. Step 3 deals with two terms. The first term vanishes because the first factor in it vanishes; thus, $\boldsymbol{d}^2\mathscr{P} = 0$ [Cartan's equation of structure; expresses the "vanishing torsion" of the covariant derivative; see equation (14.26)]. The second term also vanishes, in this case, because the second factor in it vanishes; thus, $\boldsymbol{d}\mathscr{R} = 0$ (the full Bianchi identity). Thus briefly is conservation of moment of rotation established.

Box 15.2 THE SOURCE OF GRAVITATION AND THE MOMENT OF ROTATION: THE TWO KEY QUANTITIES AND THE MOST USEFUL MATHEMATICAL REPRESENTATIONS FOR THEM

	Energy-momentum as source of gravitation (curvature of space-time)	Moment of rotation as automatically conserved feature of the geometry
Representation as a vector-valued 3-form, a coordinate-independent geometric object	Machine to tell how much energy-momentum is contained in an elementary 3-volume: $^*\boldsymbol{T} = \boldsymbol{e}_\sigma T^{\sigma\tau}\, d^3\Sigma_\tau$ ("dual of stress-energy tensor")	Machine to tell how much net moment of rotation—expressed as a vector—is obtained by adding the six moments of rotation associated with the six faces of the elementary 3-cube: $\star(\boldsymbol{d}\mathcal{P} \wedge \mathcal{R}) = {}^*\boldsymbol{G} = \boldsymbol{e}_\sigma G^{\sigma\tau}\, d^3\Sigma_\tau$ ("dual of Einstein")
Representation as a $\binom{2}{0}$-tensor (also a coordinate independent geometric object)	Stress-energy tensor itself: $\boldsymbol{T} = \boldsymbol{e}_\sigma T^{\sigma\tau}\boldsymbol{e}_\tau$	Einstein itself: $\boldsymbol{G} = \boldsymbol{e}_\sigma G^{\sigma\tau}\boldsymbol{e}_\tau$
Representation in language of components (values depend on choice of coordinate system)	$T^{\sigma\tau}$	$G^{\sigma\tau}$
Conservation law in language of components	$T^{\sigma\tau}{}_{;\tau} = 0$	$G^{\sigma\tau}{}_{;\tau} \equiv 0$
Conservation in abstract language. for the $\binom{2}{0}$-tensor	$\boldsymbol{\nabla}\cdot\boldsymbol{T} = 0$	$\boldsymbol{\nabla}\cdot\boldsymbol{G} \equiv 0$
Conservation in abstract language, as translated into exterior derivative of the dual tensor (vector-valued 3-form)	$\boldsymbol{d}{}^*\boldsymbol{T} = 0$	$\boldsymbol{d}{}^*\boldsymbol{G} \equiv 0$ or $\boldsymbol{d}\star(\boldsymbol{d}\mathcal{P} \wedge \mathcal{R}) \equiv 0$
Same conservation law expressed in integral form for an element of 4-volume Ω	$\int_{\partial\Omega} {}^*\boldsymbol{T} = 0$	$\int_{\partial\Omega} {}^*\boldsymbol{G} \equiv 0$ or $\star\int_{\partial\Omega} (\boldsymbol{d}\mathcal{P} \wedge \mathcal{R}) \equiv 0$ or $\star\int_{\partial\partial\Omega} (\mathcal{P} \wedge \mathcal{R}) \equiv 0$

§15.7. FROM CONSERVATION OF MOMENT OF ROTATION TO EINSTEIN'S GEOMETRODYNAMICS: A PREVIEW

Einstein field equation "derived" from demand that (conservation of net moment of rotation) $\Rightarrow$ (conservation of source)

Mass, or mass-energy, is the source of gravitation. Mass-energy is one component of the energy-momentum 4-vector. Energy and momentum are conserved. The amount of energy-momentum in the element of 3-volume $d^3\boldsymbol{\Sigma}$ is

$$^*\boldsymbol{T} = \boldsymbol{e}_\sigma T^{\sigma\tau}\, d^3\Sigma_\tau \tag{15.22}$$

(see Box 15.2). Conservation of energy-momentum for an elementary 4-cube Ω expresses itself in the form

$$\int_{\partial\Omega} {}^*\boldsymbol{T} = 0. \tag{15.23}$$

This conservation is not an accident. According to Einstein and Cartan, it is "automatic"; and automatic, moreover, as a consequence of exact equality between energy-momentum and an automatically conserved feature of the geometry. What is this feature? It is the moment of rotation, which satisfies the law of automatic conservation,

$$\int_{\partial\Omega} {}^*\boldsymbol{G} = 0. \tag{15.24}$$

In other words, the conservation of momentum-energy is to be made geometric in character and automatic in action by the following prescription: *Identify the stress-energy tensor* (up to a factor 8π, or $8\pi G/c^4$, or other factor that depends on choice of units) *with the moment of rotation;* thus,

$$\star(\boldsymbol{d}\mathcal{P} \wedge \mathcal{R}) = {}^*\boldsymbol{G} = 8\pi\, {}^*\boldsymbol{T}; \tag{15.25}$$

or equivalently (still in the language of vector-valued 3-forms)

$$\begin{pmatrix}\text{moment of}\\ \text{rotation}\end{pmatrix} = \star(\boldsymbol{d}\mathcal{P} \wedge \mathcal{R}) = \boldsymbol{e}_\sigma G^{\sigma\tau}\, d^3\Sigma_\tau = 8\pi \boldsymbol{e}_\sigma T^{\sigma\tau}\, d^3\Sigma_\tau; \tag{15.26}$$

or, in the language of tensors,

$$\boldsymbol{G} = \boldsymbol{e}_\sigma G^{\sigma\tau} \boldsymbol{e}_\tau = 8\pi \boldsymbol{e}_\sigma T^{\sigma\tau} \boldsymbol{e}_\tau = 8\pi \boldsymbol{T}; \tag{15.27}$$

or, in the language of components,

$$G^{\sigma\tau} = 8\pi T^{\sigma\tau} \tag{15.28}$$

(Einstein's field equation; more detail, and more on the question of uniqueness, will be found in Chapter 17; see also Box 15.3). Thus simply is all of general relativity tied to the principle that the boundary of a boundary is zero. No one has ever discovered a more compelling foundation for the principle of conservation of momentum and energy. No one has ever seen more deeply into that action of matter on space, and space on matter, which one calls gravitation.

In summary, *the Einstein theory realizes the conservation of energy-momentum as the identity, "the boundary of a boundary is zero."*

EXERCISES

Exercise 15.1. THE BOUNDARY OF THE BOUNDARY OF A 4-SIMPLEX

In the analysis of the development in time of a geometry lacking all symmetry, when one is compelled to resort to a computer, one can, as one option, break up the 4-geometry into simplexes [four-dimensional analog of two-dimensional triangle, three-dimensional tetrahedron; vertices of "central simplex" conveniently considered to be at $(t, x, y, z) = (0, 1, 1, 1)$, $(0, 1, -1, -1)$, $(0, -1, 1, -1)$, $(0, -1, -1, 1)$, $(5^{1/2}, 0, 0, 0)$, for example], sufficiently numerous, and each sufficiently small, that the geometry inside each can be idealized as flat (Lorentzian), with all the curvature concentrated at the join between simplices (see discussion of dynamics of geometry via Regge calculus in Chapter 42). Determine ("give a mathematical

Box 15.3 OTHER IDENTITIES SATISFIED BY THE CURVATURE

(1) The source of gravitation is energy-momentum. (2) Energy-momentum is expressed by stress-energy tensor (or by its dual) as a vector-valued 3-form ("energy-momentum per unit 3-volume"). (3) This source is conserved (no creation in an elementary spacetime 4-cube).

These principles form the background for the probe in this chapter of the Bianchi identities. That is why two otherwise most interesting identities [Allendoerfer and Weil (1943); Chern (1955, 1962)] are dropped from attention. One deals with the 4-form

$$\Pi = \frac{1}{24\pi^2} g^{\alpha\gamma} g^{\beta\delta} \mathcal{R}_{\alpha\beta} \wedge \mathcal{R}_{\gamma\delta}, \tag{1}$$

and the other with the 4-form

$$\Gamma = \frac{1}{8\pi^2 |\det g_{\mu\nu}|^{1/2}} (\mathcal{R}_{12} \wedge \mathcal{R}_{30} + \mathcal{R}_{13} \wedge \mathcal{R}_{02} + \mathcal{R}_{10} \wedge \mathcal{R}_{23}). \tag{2}$$

Both quantities are built from the tensorial "curvature 2-forms"

$$\mathcal{R}_{\alpha\gamma} = \frac{1}{2} R_{\alpha\gamma\beta\delta}\, \boldsymbol{d}x^\beta \wedge \boldsymbol{d}x^\delta. \tag{3}$$

The four-dimensional integral of either quantity over a four-dimensional region Ω has a value that (1) is a scalar, (2) is not identically equal to zero, (3) depends on the boundary of the region of spacetime over which the integral is extended, but (4) is independent of any changes made in the spacetime geometry interior to that surface (provided that these changes neither abandon the continuity nor change the connectivity of the 4-geometry in that region). Property (1) kills any possibility of identifying the integral, a scalar, with energy-momentum, a 4-vector. Property (2) kills it for the purpose of a conservation law, because it implies a non-zero creation in Ω.

Also omitted here is the Bel-Robinson tensor (see exercise 15.2), built bilinearly out of the curvature tensor, and other tensors for which see, e.g., Synge (1962).

One or all of these quantities may be found someday to have important physical content.

The integral of the 4-form Γ of equation (2) over the entire manifold gives a number, an integer, the so-called Euler-Poincaré characteristic of the manifold, whenever the integral and the integer are well-defined. This result is the four-dimensional generalization of the Gauss-Bonnet integral, widely known in the context of two-dimensional geometry:

$$\int \begin{pmatrix} \text{Riemannian scalar curvature} \\ \text{invariant (value } 2/a^2 \\ \text{for a sphere of radius } a) \end{pmatrix} g^{1/2}\, d^2x.$$

This integral has the value 8π for any closed, oriented, two-dimensional manifold with the topology of a 2-sphere, no matter how badly distorted; and the value 0 for any 2-torus, again no matter how rippled and twisted; and other equally specific values for other topologies.

description of") the boundary (three-dimensional) of such a simplex. Take one piece of this boundary and determine its boundary (two-dimensional). For one piece of this two-dimensional boundary, verify that there is at exactly one other place, and no more, in the book-keeping on the boundary of a boundary, another two-dimensional piece that cancels it ("facelessness" of the 3-boundary of the simplex).

Exercise 15.2. THE BEL-ROBINSON TENSOR [Bel (1958, 1959, 1962), Robinson (1959b), Sejnowski (1973); see also Pirani (1957) and Lichnerowicz (1962)].

Define the Bel-Robinson tensor by

$$T_{\alpha\beta\gamma\delta} = R_{\alpha\rho\gamma\sigma}R_{\beta}{}^{\rho}{}_{\delta}{}^{\sigma} + {}^{*}R_{\alpha\rho\gamma\sigma}{}^{*}R_{\beta}{}^{\rho}{}_{\delta}{}^{\sigma}. \tag{15.29}$$

Show that in empty spacetime this tensor can be rewritten as

$$T_{\alpha\beta\gamma\delta} = R_{\alpha\rho\gamma\sigma}R_{\beta}{}^{\rho}{}_{\delta}{}^{\sigma} + R_{\alpha\rho\delta\sigma}R_{\beta}{}^{\rho}{}_{\gamma}{}^{\sigma} - \frac{1}{8}g_{\alpha\beta}g_{\gamma\delta}R_{\rho\sigma\lambda\mu}R^{\rho\sigma\lambda\mu}. \tag{15.30a}$$

Show also that in empty spacetime

$$T^{\alpha}{}_{\beta\gamma\delta;\alpha} = 0, \tag{15.30b}$$

$$T_{\alpha\beta\gamma\delta} \text{ is symmetric and traceless on all pairs of indices.} \tag{15.30c}$$

Discussion: It turns out that Einstein's "canonical energy-momentum pseudotensor" (§20.3) for the gravitational field in empty spacetime has a second derivative which, in a Riemann-normal coordinate system, is

$$t_{\mathrm{E}\alpha\beta,\gamma\delta} = -\frac{4}{9}\left(T_{\alpha\beta\gamma\delta} - \frac{1}{4}S_{\alpha\beta\gamma\delta}\right). \tag{15.31a}$$

Here $T_{\alpha\beta\gamma\delta}$ is the completely symmetric Bel-Robinson tensor, and $S_{\alpha\beta\gamma\delta}$ is defined by

$$S_{\alpha\beta\gamma\delta} \equiv R_{\alpha\delta\rho\sigma}R_{\beta\gamma}{}^{\rho\sigma} + R_{\alpha\gamma\rho\sigma}R_{\beta\delta}{}^{\rho\sigma} + \frac{1}{4}g_{\alpha\beta}g_{\gamma\delta}R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}. \tag{15.31b}$$

$S_{\alpha\beta\gamma\delta}$ appears in the empty-space covariant wave equation

$$\Delta R_{\alpha\beta\gamma\delta} \equiv -R_{\alpha\beta\gamma\delta;\mu}{}^{;\mu} + R_{\alpha\beta\rho\sigma}R_{\gamma\delta}{}^{\rho\sigma} + 2(R_{\alpha\rho\gamma\sigma}R_{\beta}{}^{\rho}{}_{\delta}{}^{\sigma} - R_{\alpha\rho\delta\sigma}R_{\beta}{}^{\rho}{}_{\gamma}{}^{\sigma}) = 0, \tag{15.31c}$$

where Δ is a variant of the Lichnerowicz-de Rham wave operator [Lichnerowicz (1964)], when one rewrites this wave equation as

$$\Box R_{\alpha\beta}{}^{\gamma\delta} \equiv R_{\alpha\beta}{}^{\gamma\delta}{}_{;\mu}{}^{;\mu} = 2S_{[\alpha}{}^{[\gamma}{}_{\beta]}{}^{\delta]}. \tag{15.31d}$$

PART **IV**

EINSTEIN'S GEOMETRIC THEORY OF GRAVITY

Wherein the reader is seduced into marriage with the most elegant temptress of all—Geometrodynamics—and learns from her the magic potions and incantations that control the universe.

2

CHAPTER 42

REGGE CALCULUS

§42.1. WHY THE REGGE CALCULUS?

This chapter is entirely Track 2. As preparation for it, Chapter 21 (variational principle and initial-value formalism) is needed. It is not needed as preparation for any later chapter, though it will be helpful in Chapter 43 (dynamics of geometry).

The need for Regge calculus as a computational tool

Gravitation theory is entering an era when situations of greater and greater complexity must be analyzed. Before about 1965 the problems of central interest could mostly be handled by idealizations of special symmetry or special simplicity or both. The Schwarzschild geometry and its generalizations, the Friedmann cosmology and its generalizations, the joining together of the Schwarzschild geometry and the Friedmann geometry to describe the collapse of a bounded collection of matter, the vibrations of relativistic stars, weak gravitational waves propagating in an otherwise flat space: all these problems and others were solved by elementary means.

But today one is pressed to understand situations devoid of symmetry and not amenable to perturbation theory: How do two black holes alter in shape, and how much gravitational radiation do they emit when they collide and coalesce? What are the structures and properties of the singularities at the endpoint of gravitational collapse, predicted by the theorems of Penrose, Hawking, and Geroch? Can a Universe that begins completely chaotic smooth itself out quickly by processes such as inhomogeneous mixmaster oscillations?

To solve such problems, one needs new kinds of mathematical tools—and in response to this need, new tools are being developed. The "global methods" of Chapter 34 provide one set of tools. The Regge Calculus provides another[3] [Regge (1961); see also pp. 467–500 of Wheeler (1964a)].

§42.2. REGGE CALCULUS IN BRIEF

Approximation of smooth geometries by skeleton structures

Consider the geodesic dome that covers a great auditorium, made of a multitude of flat triangles joined edge to edge and vertex to vertex. Similarly envisage spacetime, in the Regge calculus, as made of flat-space "simplexes" (four-dimensional item in this progression: two dimensions, triangle; three dimensions, tetrahedron; four dimensions, simplex) joined face to face, edge to edge, and vertex to vertex. To specify the lengths of the edges is to give the engineer all he needs in order to know the shape of the roof, and the scientist all he needs in order to know the geometry of the spacetime under consideration. A smooth auditorium roof can be approximated arbitrarily closely by a geodesic dome constructed of sufficiently small triangles. A smooth spacetime manifold can be approximated arbitrarily closely by a locked-together assembly of sufficiently small simplexes. Thus the Regge calculus, reaching beyond ordinary algebraic expressions for the metric, provides a way to analyze physical situations deprived, as so many situations are, of spherical symmetry, and systems even altogether lacking in symmetry.

Role of Einstein field equation in fixing the skeleton structure

If the designer can give the roof any shape he pleases, he has more freedom than the analyst who is charting out the geometry of spacetime. Given the geometry of spacetime up to some spacelike slice that, for want of a better name, one may call "now," one has no freedom at all in the geometry from that instant on. Einstein's geometrodynamic law is fully deterministic. Translated into the language of the Regge calculus, it provides a means to calculate the edge lengths of new simplexes from the dimensions of the simplexes that have gone before. Though the geometry is deterministically specified, how it will be approximated is not. The original spacelike hypersurface ("now") is approximated as a collection of tetrahedrons joined together face to face; but how many tetrahedrons there will be and where their vertices will be placed is the option of the analyst. He can endow the skeleton more densely with bones in a region of high curvature than in a region of low curvature to get the most "accuracy profit" from a specified number of points. Some of this freedom of choice for the lengths of the bones remains as one applies the geometrodynamic law in the form given by Regge (1961) to calculate the future from the past. This freedom would be disastrous to any computer program that one tried to write, unless the programmer removed all indefiniteness by adding supplementary conditions of his own choice, either tailored to give good "accuracy profit," or otherwise fixed.

Having determined the lengths of all the bones in the portion of skeletonized spacetime of interest, one can examine any chosen local cluster of bones in and by themselves. In this way one can find out all there is to be learned about the geometry in that region. Of course, the accuracy of one's findings will depend on the fineness with which the skeletonization has been carried out. But in principle that is no limit to the fineness, or therefore to the accuracy, so long as one is working in the context of classical physics. Thus one ends up with a catalog of all the bones, showing the lengths of each. Then one can examine the geometry of whatever spacelike surface one pleases, and look into many other questions besides. For this purpose one has only to pick out the relevant bones and see how they fit together.

§42.3. SIMPLEXES AND DEFICIT ANGLES

Figure 42.1 recalls how a smoothly curved surface can be approximated by flat triangles. All the curvature is concentrated at the vertices. No curvature resides at

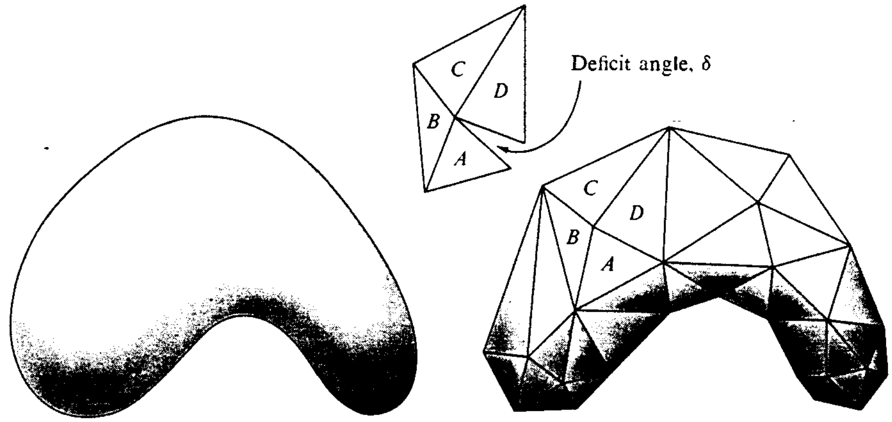


Figure 42.1.
A 2-geometry with continuously varying curvature can be approximated arbitrarily closely by a polyhedron built of triangles, provided only that the number of triangles is made sufficiently great and the size of each sufficiently small. The geometry in each triangle is Euclidean. The curvature of the surface shows up in the amount of deficit angle at each vertex (portion $ABCD$ of polyhedron laid out above on a flat surface).

Deficit angle as a skeletonized measure of curvature:

(1) in two dimensions

the edge between one triangle and the next, despite one's first impression. A vector carried by parallel transport from A through B and C to D, and then carried back by another route through C and B to A returns to its starting point unchanged in direction, as one sees most easily by laying out this complex of triangles on a flat surface. Only if the route is allowed to encircle the vertex common to A, B, C, and D does the vector experience a net rotation. The magnitude of the rotation is equal to the indicated deficit angle, δ, at the vertex. The sum of the deficit angles over all the vertices has the same value, 4π, as does the half-integral of the continuously distributed scalar curvature (${}^{(2)}R = 2/a^2$ for a sphere of radius a) taken over the entirety of the original smooth figure,

$$\sum_{\substack{\text{skeleton}\\ \text{geometry}}} \delta_i = \frac{1}{2}\int_{\substack{\text{actual smooth}\\ \text{geometry}}} {}^{(2)}R \, d(\text{surface}) = 4\pi. \tag{42.1}$$

(2) in n (or four) dimensions

Generalizing from the example of a 2-geometry, Regge calculus approximates a smoothly curved n-dimensional Riemannian manifold as a collection of n-dimensional blocks, each free of any curvature at all, joined by $(n-2)$-dimensional regions in which all the curvature is concentrated (Box 42.1). For the four-dimensional spacetime of general relativity, the "hinge" at which the curvature is concentrated has the shape of a triangle, as indicated schematically in the bottom row of Figure 42.2. In the example illustrated there, ten tetrahedrons have that triangle in common. Between one of these tetrahedrons and the next fits a four-dimensional simplex. Every feature of this simplex is determined by the lengths of its ten edges. One of the features is the angle α between one of the indicated tetrahedrons or "faces" of the simplex and the next. Thus α represents the angle subtended by this simplex

Box 42.1 THE HINGES WHERE THE CURVATURE IS CONCENTRATED IN THE "ANGLE OF RATTLE" BETWEEN BUILDING BLOCKS IN A SKELETON MANIFOLD

Dimensionality of manifold	*2*	*3*	*4*
Elementary flat-space building block:	triangle	tetrahedron	simplex
Edge lengths to define it:	3	4	5
Hinge where cycle of such blocks meet with a deficit angle or "angle of rattle" δ:	vertex	edge	triangle
Dimensionality of hinge:	0	1	2
"Content" of such a hinge:	1	length l	area A
Contribution from all hinges within a given small region to curvature of manifold:	$\sum_{\substack{\text{that}\\\text{region}}} \delta_i$	$\sum_{\substack{\text{that}\\\text{region}}} l_i\,\delta_i$	$\sum_{\substack{\text{that}\\\text{region}}} A_i\,\delta_i$
Continuum limit of this quantity expressed as an integral over the same small region:	$\frac{1}{2}\int {}^{(2)}R({}^{(2)}g)^{1/2}\,d^2x$	$\frac{1}{2}\int {}^{(3)}R({}^{(3)}g)^{1/2}\,d^3x$	$\frac{1}{2}\int {}^{(4)}R(-{}^{(4)}g)^{1/2}\,d^4x$

at the hinge. Summing the angles α for all the simplexes that meet on the given hinge $\mathscr{P}\mathscr{Q}\mathscr{R}$, and subtracting from 2π, one gets the deficit angle associated with that hinge. And by then summing the deficit angles in a given small n-volume with appropriate weighting (Box 42.1), one obtains a number equal to the volume integral of the scalar curvature of the original smooth n-geometry. See Box 42.2.

§42.4. SKELETON FORM OF FIELD EQUATIONS

Einstein-Hilbert variational principle reduced to skeleton form

Rather than translate Einstein's field equations directly into the language of the skeleton calculus, Regge turns to a standard variational principle from which Einstein's law lets itself be derived. It says (see §§21.2 and 43.3) adjust the 4-geometry throughout an extended region of spacetime, subject to certain specified conditions on the boundary, so that the dimensionless integral (action in units of $\hbar$!),

$$I = (c^3/16\pi\hbar G)\int R(-g)^{1/2}\,d^4x, \qquad (42.2)$$

is an extremum. This statement applies when space is free of matter and electromag-

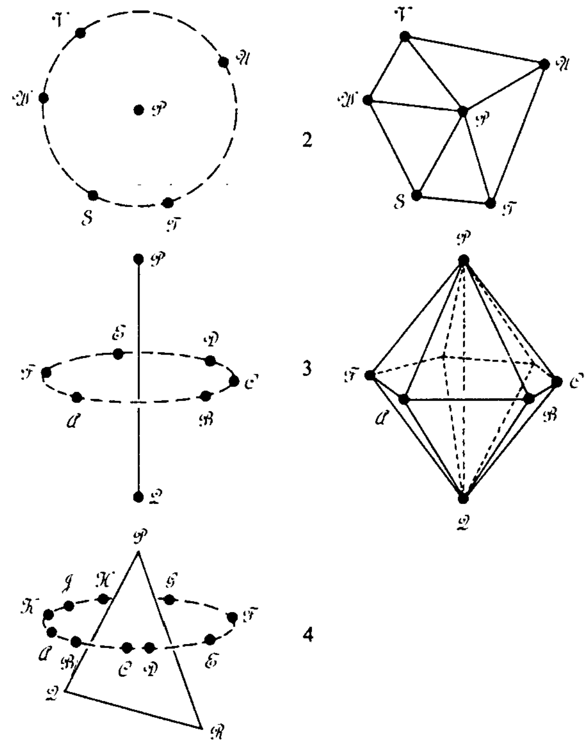


Figure 42.2.
Cycle of building blocks associated with a single hinge. Top row, two dimensions: left, schematic association of vertices $\mathscr{S}$, $\mathscr{T}$, $\mathscr{U}$, $\mathscr{V}$, $\mathscr{W}$ with "hinge" at the vertex $\mathscr{P}$; right, same, but with elementary triangles indicated in full. Middle row, three dimensions: left, schematic; right, perspective representation of the six tetrahedrons that meet on the "hinge" $\mathscr{PQ}$. Bottom row, four dimensions; shown only schematically. The five vertices $\mathscr{PQRCD}$ belong to one simplex, a four-dimensional region throughout the interior of which space is flat. The five vertices $\mathscr{PQRDE}$ belong to the next simplex; and so on around the cycle of simplexes. The two simplexes just named interface at the tetrahedron $\mathscr{PQRD}$, inside which the geometry is also flat. Between that tetrahedron and the next, $\mathscr{PQRE}$, there is a certain hyperdihedral angle α subtended at the "hinge" $\mathscr{PQR}$. The value of this angle is completely fixed by the ten edge lengths of the intervening simplex $\mathscr{PQRDE}$. This dihedral angle, plus the corresponding dihedral angles subtended at the hinge $\mathscr{PQR}$ by the other simplexes of the cycle, do not in general add up to 2π. The deficit, the "angle of rattle" or deficit angle δ, gives the amount of curvature concentrated at the hinge $\mathscr{PQR}$. There is no actual rattle or looseness of fit, unless one tries to imbed the cycle into an over-all flat four-dimensional space (analog of "stamping on" the collection of triangles, and seeing them open out by the amount of the deficit angle, as indicated in inset in Figure 42.1).

netic fields, a simplification that will be made in the subsequent discussion to keep it from becoming too extended. When in addition all lengths are expressed in units of the Planck length

$$L^* = (\hbar G/c^3)^{1/2} = 1.6 \times 10^{-33}\,\text{cm}, \tag{42.3}$$

and the curvature integral is approximated by its expression in terms of deficit angles, Regge shows that the statement $\delta I = 0$ (condition for an extremum!) becomes

$$(1/8\pi)\,\delta \sum_{\substack{\text{hinges}\\ h=1}}^{H} A_h\,\delta_h = 0. \tag{42.4}$$

Box 42.2 FLOW DIAGRAMS FOR REGGE CALCULUS

A skeleton 4-geometry is completely determined by all its edge lengths. From the edge lengths one gets the integrated curvature by pursuing, for each hinge in the 4-geometry, the following flow diagram:

hinge

↓

cycle of blocks swinging on this hinge

↓

one of these blocks

↓

the two tetrahedral "faces" that set this block off from the blocks before and after it in the cycle of blocks

↓

angle α between these two faces fixed by the block's $n(n-1)/2$ edge lengths

↓

deficit angle at the given hinge is

$$\delta = 2\pi - \sum_{\substack{\text{blocks swinging}\\ \text{on that hinge}}} \alpha$$

↓

contribution to integrated curvature (Box 42.1) is δ times area of hinge

One finds it natural to apply this analysis in either of two ways. First, one can probe a given 4-geometry (given set of edge lengths!) in the sense

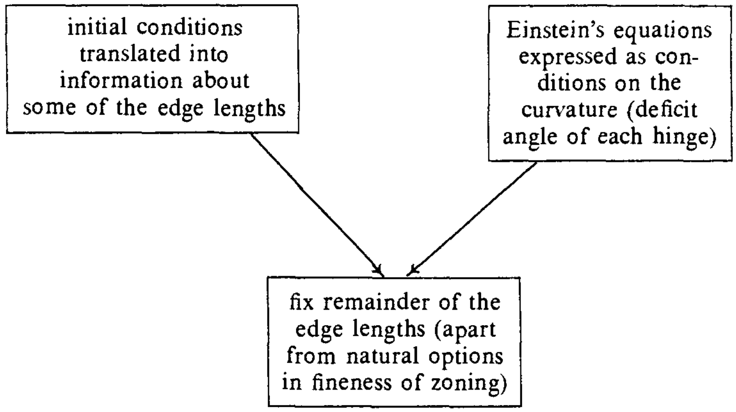
edge lengths

↓

curvature

Box 42.2 (*continued*)

Second—and this is the rationale of Regge calculus—one can use the skeleton calculus to deduce a previously unknown 4-geometry from Einstein's geometrodynamic law, proceeding in the direction



In the changes contemplated in this variational principle, certain edge lengths are thought of as being fixed. They have to do with the conditions specified at the boundaries of the region of spacetime under study. It is not necessary here to enter into the precise formulation of these boundary conditions, fortunately, since some questions of principle still remain to be clarified about the precise formulation of boundary conditions in general relativity (see §21.12). Rather, what is important is the effect of changes in the lengths of the edges of the blocks in the interior of the region being analyzed, as they augment or decrease the deficit angles at the various hinges. In his basic paper on the subject, Regge (1961) notes that the typical deficit angle δ_h depends in a complicated trigonometric way on the values of numerous edge lengths ℓ_p. However, he proves (Appendix of his paper) that "quite remarkably, we can carry out the variation as if the δ_h were constants," thus reducing the variational principle to the form

$$(1/8\pi) \sum_{\substack{\text{hinges}\\ h=1}}^{H} \delta_h \, \delta A_h = 0. \tag{42.5}$$

Here the change in area of the h-th triangle-shaped hinge, according to elementary trigonometry, is

$$\delta A_h = \frac{1}{2} \sum_p \ell_p \, \delta \ell_p \cot\!\text{an}\, \theta_{ph}. \tag{42.6}$$

In this equation θ_{ph} is the angle opposite to the p-th edge in the triangle. Consequently, Einstein's equations in empty space reduce in skeleton geometry to the form

$$\sum_{\substack{\text{hinges that}\\ \text{have the}\\ \text{given edge}\\ p \text{ in common}}} \delta_h \cot \theta_{ph} = 0, \qquad (p = 1, 2, \ldots), \tag{42.7}$$

Einstein field equation reduced to skeleton form

one equation for each edge length in the interior of the region of spacetime being analyzed.

§42.5. THE CHOICE OF LATTICE STRUCTURE

Two questions arise in the actual application of Regge calculus, and it is not clear that either has yet received the resolution which is most convenient for practical applications of this skeleton analysis: What kind of lattice to use? How best to capitalize on the freedom that exists in the choice of edge lengths? The first question is discussed in this section, the second in the next section.

The choice of lattice structure: (1) avoiding floppiness

It might seem most natural to use a lattice made of small, nearly rectangular blocks, the departure of each from rectangularity being conditioned by the amount and directionality of the local curvature. However, such building blocks are "floppy." One could give them rigidity by specifying certain angles as well as the edge lengths. But then one would lose the cleanness of Regge's prescription: give edge lengths, and give only edge lengths, and give each edge length freely and independently, in order to define a geometry. In addition one would have to rederive the Regge equations, including new equations for the determination of the new angles. Therefore one discards the quasirectangle in favor of the simplex with its $5 \cdot 4/2 = 10$ edge lengths. This decided, one also concludes that even in flat spacetime the simplexes cannot all have identical edge lengths. Two-dimensional flat space can be filled with identical equilateral triangles, but already at three dimensions it ceases to be possible to fill out the manifold with identical equilateral tetrahedrons. One knows that a given carbon atom in diamond is joined to its nearest neighbors with tetrahedral bonds, but a little reflection shows that the cell assignable to the given atom is far from having the shape of an equilateral tetrahedron.

(2) necessity for unequal edge lengths

Synthesis would appear to be a natural way to put together the building blocks: first make one-dimensional structures; assemble these into two-dimensional structures; these, into three-dimensional ones; and these, into the final four-dimensional structure. The one-dimensional structure is made of points, 1, 2, 3, . . . , alternating with line segments, 12, 23, 34, ... To start building a two-dimensional structure, pick up a second one-dimensional structure. It might seem natural to label its points 1′, 2′, 3′, . . . , etc. However, that labeling would imply a cross-connection between 1 and 1′, between 2 and 2′, etc., after the fashion of a ladder. Then the elementary cells would be quasirectangles. They would have the "floppiness" that is to be excluded. Therefore relabel the points of the second one-dimensional structure as $1\frac{1}{2}'$, $2\frac{1}{2}'$, $3\frac{1}{2}'$, etc. The implication is that one cross-connects $2\frac{1}{2}'$ with points 2 and 3 of the original one-dimensional structure, etc. One ends up with something like the

(3) construction of two-dimensional structures

girder structure of a bridge, fully rigid in the context of two dimensions, as desired. The same construction, extended, fills out the plane with triangles. One now has a simple, standard two-dimensional structure. One might mistakenly conclude that one is ready to go ahead to build up a three-dimensional structure: the mistake lies in the tacit assumption that the flat-space topology is necessarily correct.

Let it be the problem, for example, to determine the development in time of a 3-geometry that has the topology of a 3-sphere. This 3-sphere is perhaps strongly deformed from ideality by long-wavelength gravitational waves. A right arrangement of the points is the immediate desideratum. Therefore put aside for the present any consideration of the deformation of the geometry by the waves (alteration of edge lengths from ideality). Ask how to divide a perfect 3-sphere into two-dimensional sheets. Here each sheet is understood to be separated from the next by a certain distance. At this point two alternative approaches suggest themselves that one can call for brevity "blocks" and "spheres."

(4) 3-D structures built from 2-D structures by "method of blocks"

(1) *Blocks*. Note that a 3-sphere lets itself be decomposed into 5 identical, tetrahedron-like solid blocks (5 vertices; 5 ways to leave out any one of these vertices!) Fix on one of these "tetrahedrons." Select one vertex as summit and the face through the other three vertices as base. Give that base the two-dimensional lattice structure already described. Introduce a multitude of additional sheets piled above it as evenly spaced layers reaching to the summit. Each layer has fewer points than the layer before. The decomposition of the 3-geometry inside one "tetrahedron" is thereby accomplished. However, an unresolved question remains; not merely how to join on this layered structure in a regular way to the corresponding structure in the adjacent "tetrahedrons"; but even whether such a regular joinup is at all possible. The same question can be asked about the other two ways to break up the 3-sphere into identical "tetrahedrons" [Coxeter (1948), esp. pp. 292–293: 16 tetrahedrons defined by a total of 8 vertices or 600 tetrahedrons defined by a total of 120 vertices]. One can eliminate the question of joinup of structure in a simple way, but at the price of putting a ceiling on the accuracy attainable: take the stated number of vertices (5 or 8 or 120) as the total number of points that will be employed in the skeletonization of the 3-geometry (no further subdivision required or admitted). Considering the boundedness of the memory capacity of any computer, it is hardly ridiculous to contemplate a limitation to 120 tracer points in exploratory calculations!

(5) 3-D structures from 2-D by "method of spheres"

(2) *Spheres*. An alternative approach to the "atomization" of the 3-sphere begins by introducing on the 3-sphere a North Pole and a South Pole and the hyperspherical angle χ ($\chi = 0$ at the first pole, $\chi = \pi$ at the second, $\chi = \pi/2$ at the equator; see Box 27.2). Let each two-dimensional layer lie on a surface of constant χ (χ equal to some integer times some interval $\Delta\chi$). The structure of this 2-sphere is already to be regarded as skeletonized into elementary triangles ("fully complete Buckminster Fuller geodesic dome"). Therefore the number of "faces" or triangles F, the number of edge lengths E, and the number of vertices V must be connected by the relation of Euler:

$$F - E + V = \begin{pmatrix}\text{a topology-dependent} \\ \text{number or "Euler character"}\end{pmatrix} = \begin{cases} 2 \text{ for 2-sphere,} \\ 0 \text{ for 2-torus.}\end{cases} \tag{42.8}$$

It follows from this relation that it is impossible for each vertex to sit at the center of a hexagon (each vertex the point of convergence of 6 triangles). This being the case, one is not astonished that a close inspection of the pattern of a geodesic dome shows several vertices where only 5 triangles meet. It is enough to have 12 such 5-triangle vertices among what are otherwise all 6-triangle vertices in order to meet the requirements of the Euler relation:

$$\begin{aligned} n &\quad \text{5-triangle vertices} \\ V - n &\quad \text{6-triangle vertices} \\ F &= (V - n)(6/3) + n(5/3) \text{ triangles} \\ E &= (V - n)(6/2) + n(5/2) \text{ edges} \\ V &= (V - n)(6/6) + n \text{ vertices} \\ 2 = F - E + V &= n/6 \quad \text{Euler characteristic} \\ n &= 12 \end{aligned} \tag{42.9}$$

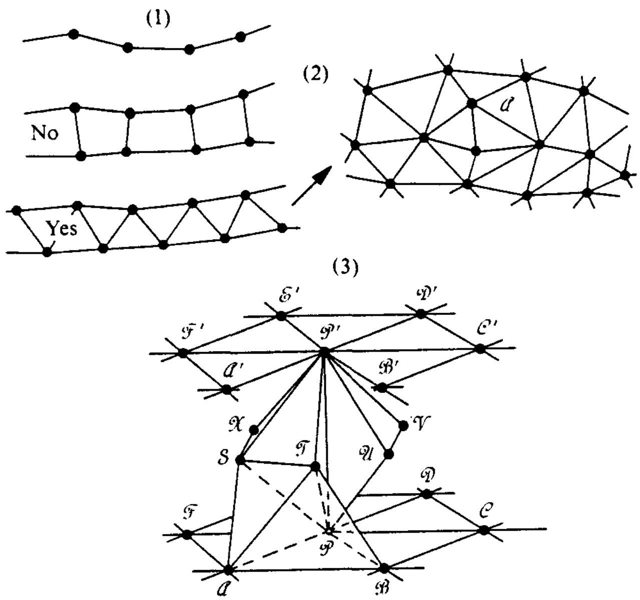
Among all figures with triangular faces, the icosahedron is the one with the smallest number of faces that meets this condition (5-triangle vertices exclusively!)

If each 2-surface has the pattern of vertices of a geodesic dome, how is one dome to be joined to the next to make a rigid skeleton 3-geometry? Were the domes imbedded in a flat 3-geometry, rigidity would be no issue. Each dome would already be rigid in and by itself. However, the 3-geometry is not given to be flat. Only by a completely deterministic skeletonization of the space between the two 2-spheres will they be given rigidity in the context of curved space geometry. (1) Not by running a single connector from each vertex in one surface to the corresponding vertex in the next ("floppy structure"!) (2) Not by displacing one surface so each of its vertices comes above, or nearly above, the center of a triangle in the surface "below." First, the numbers of vertices and triangles ordinarily will not agree. Second, even when they do, it will not give the structure the necessary rigidity to connect the vertex of the surface above to the three vertices of the triangle below. The space between will contain some tetrahedrons, but it will not be throughout decomposed into tetrahedrons. (3) A natural and workable approach to the skeletonization of the 3-geometry is to run a connector from each vertex in the one surface to the corresponding vertex in the next, but to flesh out this connection with additional structure that will give rigidity to the 3-geometry: intervening vertices and connectors as illustrated in Box 42.3.

(6) 4-D structures built from 3-D structures

In working up from the skeletonization of a 3-geometry to the skeletonization of a 4-geometry, it is natural to proceed similarly. (1) Use identical patterns of points in the two 3-geometries. (2) Tie corresponding points together by single connectors. (3) Halfway, or approximately half way between the two 3-geometries insert a whole additional pattern of vertices. Each of these supplementary vertices is "dual" to and lies nearly "below" the center of a tetrahedron in the 3-geometry immediately above. (4) Connect each supplementary vertex to the vertices of the tetrahedron immediately above, to the vertices of the tetrahedron immediately below, and to those other supplementary vertices that are its immediate neighbors. (5) In this way get the edge lengths needed to divide the 4-geometry into simplexes, each of rigidly defined dimensions.

Box 42.3 SYNTHESIS OF HIGHER-DIMENSIONAL SKELETON GEOMETRIES OUT OF LOWER-DIMENSIONAL SKELETON GEOMETRIES



(1) One-dimensional structure as alternation of points and line segments. (2) Two-dimensional structure (a) "floppy" (unacceptable) and (b) rigidified (angles of triangles fully determined by edge lengths). When this structure is extended, as at right, the "normal" vertex has six triangles hinging on it. However, at least twelve 5-triangle vertices of the type indicated at $\mathcal{A}$ are to be interpolated if the 2-geometry is to be able to close up into a 2-sphere. (3) Skeleton 3-geometry obtained by filling in between the skeleton 2-geometry . . . $\mathcal{A}\mathcal{B}$ . . . $\mathcal{F}\mathcal{P}\mathcal{C}$ . . . $\mathcal{E}\mathcal{D}$ . . . and the similar structure . . . $\mathcal{A}'\mathcal{B}'$ . . . $\mathcal{F}'\mathcal{P}'\mathcal{C}'$ . . . $\mathcal{E}'\mathcal{D}'$ . . . as follows. (a) Insert direct connectors such as $\mathcal{P}\mathcal{P}'$ between corresponding points in the two 2-geometries. (b) Insert an intermediate layer of "supplementary vertices" such as $\mathcal{S}\mathcal{T}\mathcal{U}\mathcal{V}\mathcal{W}\mathcal{X}$. . . . Each of these supplementary vertices lies roughly halfway between the center of the triangle "above" it and the center of the corresponding triangle "below" it. (c) Connect each such "supplementary vertex" with its immediate neighbors above, below, and in the same plane. (d) Give all edge lengths. (e) Then the skeleton 3-geometry between the two 2-geometries is rigidly specified. It is made up of five types of tetrahedrons, as follows. (1) "Right-through blocks," such as $\mathcal{P}\mathcal{P}'\mathcal{S}\mathcal{T}$ (six of these hinge on $\mathcal{P}\mathcal{P}'$ when $\mathcal{P}$ is a normal vertex; five, when it is a 5-fold vertex, such as indicated by $\mathcal{A}$ at the upper right). (2) "Lower-facing blocks," such as $\mathcal{A}\mathcal{B}\mathcal{P}\mathcal{T}$. (3) "Lower-packing blocks," such as $\mathcal{A}\mathcal{P}\mathcal{S}\mathcal{T}$. (4, 5) Corresponding "upper-facing blocks" and "upper-packing blocks" (not shown). The number of blocks of each kind is appropriately listed here for the two extreme cases of a 2-geometry that consists (a) of a normal hexagonal lattice extending indefinitely in a plane and (b) of a lattice consisting of the minimum number of 5-fold vertices ("type $\mathcal{A}$ vertices") that will permit close-up into a 2-sphere.

2-geometry of upper (or lower) face	*Hexagonal pattern of triangles*	*Icosahedron made of triangles*
Its topology	Infinite 2-plane	2-sphere
Vertices on upper face	V	12
Nature of these vertices	6-fold	5-fold
Edge lengths on upper face	$3V$	$\frac{5}{2}V = 30$
Triangles on upper face	$2V$	20
Number of "supplementary vertices"	$2V$	20
Outer facing blocks	$2V$	20
Outer packing blocks	$3V$	30
Right through blocks	$6V$	60
Inner packing blocks	$3V$	30
Inner facing blocks	$2V$	20

§42.6. THE CHOICE OF EDGE LENGTHS

So much for the lattice structure of the 4-geometry; now for the other issue, the freedom that exists in the choice of edge lengths. Why not make the simplest choice and let all edges be light rays? Because the 4-geometry would not then be fully determined. The geometry $g_{\alpha\beta}(x^\mu)$ differs from the geometry $\lambda(x^\mu)\, g_{\alpha\beta}(x^\mu)$, even though the same points that are connected by light rays in the one geometry are also connected by light rays in the other geometry.

The choice of edge lengths:

If none of the edges is null, it is nevertheless natural to take some of the edge lengths to be spacelike and some to be timelike. In consequence the area A of the triangle in some cases will be real, in other cases imaginary. In 3-space the parallelogram (double triangle) spanned by two vectors $\boldsymbol{B}$ and $\boldsymbol{C}$ is described by a vector

(1) choose some timelike, others spacelike

$$2\boldsymbol{A} = \boldsymbol{B} \times \boldsymbol{C}$$

perpendicular to the two vectors. One obtains the magnitude of A from the formula

$$4A^2 = \boldsymbol{B}^2\boldsymbol{C}^2 - (\boldsymbol{B} \cdot \boldsymbol{C})^2.$$

In 4-space, let $\boldsymbol{B}$ and $\boldsymbol{C}$ be two edges of the triangle. Then, as in three dimensions, $2\boldsymbol{A}$ is *dual* to the bivector built from $\boldsymbol{B}$ and $\boldsymbol{C}$. In other words, if $\boldsymbol{B}$ goes in the t direction and $\boldsymbol{C}$ in the z direction, then $\boldsymbol{A}$ is a bivector lying in the (x, y) plane. Consequently its magnitude A is to be thought of as a *real* quantity. Therefore the appropriate formula for the area A is (Tullio Regge)

$$4A^2 = (\boldsymbol{B} \cdot \boldsymbol{C})^2 - \boldsymbol{B}^2\boldsymbol{C}^2. \tag{42.10}$$

The quantity A is real when the deficit angle δ is real. Thus the geometrically important product $A\delta$ is also real.

When the hinge lies in the (x, y) plane, on the other hand, the quantity A is purely imaginary. In that instance a test vector taken around the cycle of simplexes that swing on this hinge has undergone change only in its z and t components; that is, it has experienced a Lorentz boost; that is, the deficit angle δ is also purely imaginary. So again the product $A\delta$ is a purely real quantity.

(2) choose timelike lengths comparable to spacelike lengths

Turn now from character of edge lengths to magnitude of edge lengths. It is desirable that the elementary building blocks sample the curvatures of space in different directions on a roughly equal basis. In other words, it is desirable not to have long needle-shaped building blocks nor pancake-shaped tetrahedrons and simplexes. This natural requirement means that the step forward in time should be comparable to the steps "sidewise" in space. The very fact that one should have to state such a requirement brings out one circumstance that should have been obvious before: the "hinge equations"

$$\sum_{\substack{\text{hinges } h \text{ that} \\ \text{have edge } p \\ \text{in common}}} \delta_h \cot\!\operatorname{an} \theta_{ph} = 0 \qquad (p = 1, 2, \ldots), \tag{42.7}$$

(3) why some lengths must be chosen arbitrarily

though they are as numerous as the edges, cannot be regarded as adequate to determine all edge lengths. There are necessarily relations between these equations that keep them from being independent. The equations cannot determine all the details of the necessarily largely arbitrary skeletonization process. They cannot do so any more than the field equations of general relativity can determine the coordinate system. With a given pattern of vertices (four-dimensional generalization of drawings in Box 42.3), one still has (a) the option how close together one will take successive layers of the structure and (b) how one will distribute a given number of points in space on a given layer to achieve the maximum payoff in accuracy (greater density of points in regions of greater curvature). To prepare a practical computer program founded on Regge calculus, one has to supply the machine not only with the hinge equations and initial conditions, but also with definite algorithms to remove all the arbitrariness that resides in options (a) and (b).

Deficit angles in terms of edge lengths

Formulas from solid geometry and four-dimensional geometry, out of which to determine the necessary hyperdihedral angles α and the deficit angles δ in terms of edge lengths and nothing but edge lengths, are summarized by Wheeler (1964a, pp. 469, 470, and 490) and by C. Y. Wong (1971). Regge (1961) also gives a formula for the Riemann curvature tensor itself in terms of deficit angles and number of edges running in a given direction [see also Wheeler (1964a, p. 471)].

§42.7. PAST APPLICATIONS OF REGGE CALCULUS

Past applications of Regge calculus

Wong (1971) has applied Regge calculus to a problem where no time development shows itself, where the geometry can therefore be treated as static, and where in addition it is spherically symmetric. He determined the Schwarzschild and Reissner-Nordstrøm geometries by the method of skeletonization. Consider successive spheres surrounding the center of attraction. Wong approximates each as an icosahedron. The condition

$$^{(3)}R = 16\pi \begin{pmatrix}\text{energy density}\\ \text{on the 3-space}\end{pmatrix}$$

(§21.5) gives a recursion relation that determines the dimension of each icosahedron in terms of the two preceding icosahedra. Errors in the skeleton representation of the exact geometry range from roughly 10 percent to less than 1 percent, depending on the method of analysis, the quantity under analysis, and the fineness of the subdivision.

Partial skeletonization

Skeletonization of geometry is to be distinguished from mere rewriting of partial differential equations as difference equations. One has by now three illustrations that one can capitalize on skeletonization without fragmenting spacetime all the way to the level of individual simplexes. The first illustration is the first part of Wong's work, where the time dimension never explicitly makes an appearance, so that the building blocks are three-dimensional only. The second is an alternative treatment, also given by Wong, that goes beyond the symmetry in t to take account of the symmetry in θ and ϕ. It divides space into spherical shells, in each of which the geometry is "pseudo-flat" in much the same sense that the geometry of a paper cone is flat. The third is the numerical solution for the gravitational collapse of a spherical star by May and White (1966), in which there is symmetry in θ and ϕ, but not in r or t. This zoning takes place exclusively in the r, t-plane. Each zone is a spherical shell. The difference as compared to Regge calculus (flat geometry within each building block) is the adjustable "conicity" given to each shell. The examples show that the decision about skeletonizing the geometry in a calculation is ordinarily not "whether" but "how much."

§42.8. THE FUTURE OF REGGE CALCULUS

Hopes for the future

In summary, Regge's skeleton calculus puts within the reach of computation problems that in practical terms are beyond the power of normal analytical methods. It affords any desired level of accuracy by sufficiently fine subdivision of the spacetime region under consideration. By way of its numbered building blocks, it also offers a practical way to display the results of such calculations. Finally, one can hope that Regge's truly geometric way of formulating general relativity will someday make the content of the Einstein field equations (Cartan's "moment of rotation"; see Chapter 15) stand out sharp and clear, and unveil the geometric significance of the so-called "geometrodynamic field momentum" (analysis of the boundary-value problem associated with the variational problem of general relativity in Regge calculus; see §21.12).

CHAPTER 4

ELECTROMAGNETISM AND DIFFERENTIAL FORMS

The ether trembled at his agitations
In a manner so familiar that I only need to say,
In accordance with Clerk Maxwell's six equations
It tickled peoples' optics far away.
You can feel the way it's done,
You may trace them as they run—
dγ *by* dy *less* dβ *by* dz *is equal* KdX/dt . . .

While the curl of (X, Y, Z) *is the*
minus d/dt *of the vector* (a, b, c).

From *The Revolution of the Corpuscle,*
written by A. A. Robb
(to the tune of *The Interfering Parrott*)
for a dinner of the research students
of the Cavendish Laboratory
in the days of the old mathematics.

This chapter is all Track 2. It is needed as preparation for §§14.5 and 14.6 (computation of curvature using differential forms) and for Chapter 15 (Bianchi identities and boundary of a boundary), but is not needed for the rest of the book.

§4.1. EXTERIOR CALCULUS

Stacks of surfaces, individually or intersecting to make "honeycombs," "egg crates," and other such structures ("differential forms"), give unique insight into the geometry of electromagnetism and gravitation. However, such insight comes at some cost in time. Therefore, most readers should skip this chapter and later material that depends on it during a first reading of this book.

Analytically speaking, differential forms are completely antisymmetric tensors; pictorially speaking, they are intersecting stacks of surfaces. The mathematical formalism for manipulating differential forms with ease, called "exterior calculus," is summarized concisely in Box 4.1; its basic features are illustrated in the rest of this chapter by rewriting electromagnetic theory in its language. An effective way to tackle this chapter might be to (1) scan Box 4.1 to get the flavor of the formalism; (2) read the rest of the chapter in detail; (3) restudy Box 4.1 carefully; (4) get practice in manipulating the formalism by working the exercises.*

(continued on page 99)

*Exterior calculus is treated in greater detail than here by: É. Cartan (1945); de Rham (1955); Nickerson, Spencer, and Steenrod (1959); Hauser (1970); Israel (1970); especially Flanders (1963, relatively easy, with many applications); Spivak (1965, sophomore or junior level, but fully in tune with modern mathematics); H. Cartan (1970); and Choquet-Bruhat (1968a).

Box 4.1 DIFFERENTIAL FORMS AND EXTERIOR CALCULUS IN BRIEF

The fundamental definitions and formulas of exterior calculus are summarized here for ready reference. Each item consists of a general statement (at left of page) plus a leading application (at right of page). This formalism is applicable not only to spacetime, but also to more general geometrical systems (see heading of each section). No attempt is made here to demonstrate the internal consistency of the formalism, nor to derive it from any set of definitions and axioms. For a systematic treatment that does so, see, e.g., Spivak (1965), or Misner and Wheeler (1957).

A. Algebra I (applicable to any vector space)

1. *Basis 1-forms.*
   a. Coordinate basis $\boldsymbol{\omega}^j = \boldsymbol{d}x^j$
   (j tells which 1-form, not which component).
   b. General basis $\boldsymbol{\omega}^j = L^j{}_{k'}\, \boldsymbol{d}x^{k'}$.

An application
Simple basis 1-forms for analyzing Schwarzschild geometry around static spherically symmetric center of attraction:

$$\boldsymbol{\omega}^0 = (1 - 2m/r)^{1/2}\, \boldsymbol{d}t;$$
$$\boldsymbol{\omega}^1 = (1 - 2m/r)^{-1/2}\, \boldsymbol{d}r;$$
$$\boldsymbol{\omega}^2 = r\, \boldsymbol{d}\theta;$$
$$\boldsymbol{\omega}^3 = r \sin\theta\, \boldsymbol{d}\phi.$$

2. *General* p*-form* (*or* p*-vector*) is a completely antisymmetric tensor of rank $\binom{0}{p}$ [or $\binom{p}{0}$]. It can be expanded in terms of wedge products (see §3.5 and exercise 4.12):

$$\boldsymbol{\alpha} = \frac{1}{p!}\alpha_{i_1 i_2 \ldots i_p}\boldsymbol{\omega}^{i_1} \wedge \boldsymbol{\omega}^{i_2} \wedge \cdots \wedge \boldsymbol{\omega}^{i_p}$$
$$\equiv \alpha_{|i_1 i_2 \ldots i_p|}\boldsymbol{\omega}^{i_1} \wedge \boldsymbol{\omega}^{i_2} \wedge \cdots \wedge \boldsymbol{\omega}^{i_p}.$$

(Note: Vertical bars around the indices mean summation extends only over $i_1 < i_2 < \cdots < i_p$.)

Two applications
Energy-momentum 1-form is of type $\boldsymbol{\alpha} = \alpha_i \boldsymbol{\omega}^i$ or

$$\boldsymbol{p} = -E\, \boldsymbol{d}t + p_x\, \boldsymbol{d}x + p_y\, \boldsymbol{d}y + p_z\, \boldsymbol{d}z.$$

Faraday is a 2-form of type $\boldsymbol{\beta} = \beta_{|\mu\nu|}\boldsymbol{\omega}^\mu \wedge \boldsymbol{\omega}^\nu$ or in flat spacetime

$$\boldsymbol{F} = -E_x\, \boldsymbol{d}t \wedge \boldsymbol{d}x - E_y\, \boldsymbol{d}t \wedge \boldsymbol{d}y - E_z\, \boldsymbol{d}t \wedge \boldsymbol{d}z$$
$$+ B_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z + B_y\, \boldsymbol{d}z \wedge \boldsymbol{d}x + B_z\, \boldsymbol{d}x \wedge \boldsymbol{d}y$$

Box 4.1 (*continued*)

3. *Wedge product.*
All familiar rules of addition and multiplication hold, such as

$$(a\boldsymbol{\alpha} + b\boldsymbol{\beta}) \wedge \boldsymbol{\gamma} = a\boldsymbol{\alpha} \wedge \boldsymbol{\gamma} + b\boldsymbol{\beta} \wedge \boldsymbol{\gamma},$$
$$(\boldsymbol{\alpha} \wedge \boldsymbol{\beta}) \wedge \boldsymbol{\gamma} = \boldsymbol{\alpha} \wedge (\boldsymbol{\beta} \wedge \boldsymbol{\gamma}) \equiv \boldsymbol{\alpha} \wedge \boldsymbol{\beta} \wedge \boldsymbol{\gamma},$$

except for a modified commutation law between a p-form $\boldsymbol{\alpha}$ and a q-form $\boldsymbol{\beta}$:

$$\underset{p}{\boldsymbol{\alpha}} \wedge \underset{q}{\boldsymbol{\beta}} = (-1)^{pq} \underset{q}{\boldsymbol{\beta}} \wedge \underset{p}{\boldsymbol{\alpha}}.$$

Applications to 1-forms $\boldsymbol{\alpha}, \boldsymbol{\beta}$:

$$\boldsymbol{\alpha} \wedge \boldsymbol{\beta} = -\boldsymbol{\beta} \wedge \boldsymbol{\alpha}, \qquad \boldsymbol{\alpha} \wedge \boldsymbol{\alpha} = 0;$$
$$\boldsymbol{\alpha} \wedge \boldsymbol{\beta} = (\alpha_j \boldsymbol{\omega}^j) \wedge (\beta_k \boldsymbol{\omega}^k) = \alpha_j \beta_k \boldsymbol{\omega}^j \wedge \boldsymbol{\omega}^k$$
$$= \frac{1}{2}(\alpha_j \beta_k - \beta_j \alpha_k) \boldsymbol{\omega}^j \wedge \boldsymbol{\omega}^k.$$

4. *Contraction of* p-*form on* p-*vector.*

$$\langle \underset{p}{\boldsymbol{\alpha}}, \underset{p}{\boldsymbol{A}} \rangle$$
$$= \alpha_{|i_1 \ldots i_p|} A^{|j_1 \ldots j_p|} \underbrace{\langle \boldsymbol{\omega}^{i_1} \wedge \cdots \wedge \boldsymbol{\omega}^{i_p}, \boldsymbol{e}_{j_1} \wedge \cdots \wedge \boldsymbol{e}_{j_p} \rangle}_{[\equiv \delta^{i_1 \ldots i_p}_{j_1 \ldots j_p} \text{ (see exercises 3.13 and 4.12)}]}$$
$$= \alpha_{|i_1 \ldots i_p|} A^{i_1 \ldots i_p}.$$

Four applications

a. Contraction of a particle's energy-momentum 1-form $\boldsymbol{p} = p_\alpha \boldsymbol{\omega}^\alpha$ with 4-velocity $\boldsymbol{u} = u^\alpha \boldsymbol{e}_\alpha$ of observer (a 1-vector):

$$-\langle \boldsymbol{p}, \boldsymbol{u} \rangle = -p_\alpha u^\alpha = \text{energy of particle.}$$

b. Contraction of ***Faraday*** 2-form $\boldsymbol{F}$ with bivector $\delta\mathscr{P} \wedge \Delta\mathscr{P}$ [where $\delta\mathscr{P} = (d\mathscr{P}/d\lambda_1)\Delta\lambda_1$ and $\Delta\mathscr{P} = (d\mathscr{P}/d\lambda_2)\Delta\lambda_2$ are two infinitesimal vectors in a 2-surface $\mathscr{P}(\lambda_1, \lambda_2)$, and the bivector represents the surface element they span] is the magnetic flux $\Phi = \langle \boldsymbol{F}, \delta\mathscr{P} \wedge \Delta\mathscr{P} \rangle$ through that surface element.

c. More generally, a p-dimensional parallelepiped with vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p$ for legs has an oriented volume described by the "*simple*" p-vector $\boldsymbol{a}_1 \wedge \boldsymbol{a}_2 \wedge \cdots \boldsymbol{a}_p$ (oriented because interchange of two legs changes its sign). An egg-crate type of structure with walls made from the hyperplanes of p different 1-forms $\boldsymbol{\sigma}^1$,

$\boldsymbol{\sigma}^2, \ldots, \boldsymbol{\sigma}^p$ is described by the *"simple"* p-form $\boldsymbol{\sigma}^1 \wedge \boldsymbol{\sigma}^2 \wedge \cdots \wedge \boldsymbol{\sigma}^p$. The number of cells of $\boldsymbol{\sigma}^1 \wedge \boldsymbol{\sigma}^2 \wedge \cdots \wedge \boldsymbol{\sigma}^p$ sliced through by the infinitesimal p-volume $\boldsymbol{a}_1 \wedge \boldsymbol{a}_2 \wedge \cdots \wedge \boldsymbol{a}_p$ is

$$\langle \boldsymbol{\sigma}^1 \wedge \boldsymbol{\sigma}^2 \wedge \cdots \wedge \boldsymbol{\sigma}^p, \boldsymbol{a}_1 \wedge \boldsymbol{a}_2 \wedge \cdots \wedge \boldsymbol{a}_p \rangle.$$

d. The Jacobian determinant of a set of p functions $f^k(x^1, \ldots, x^n)$ with respect to p of their arguments is

$$\left\langle \boldsymbol{d}f^1 \wedge \boldsymbol{d}f^2 \wedge \cdots \wedge \boldsymbol{d}f^p, \frac{\partial \mathscr{P}}{\partial x^1} \wedge \frac{\partial \mathscr{P}}{\partial x^2} \wedge \cdots \wedge \frac{\partial \mathscr{P}}{\partial x^p} \right\rangle = \det \left\| \left(\frac{\partial f^k}{\partial x^j} \right) \right\| \equiv \frac{\partial(f^1, f^2, \ldots, f^p)}{\partial(x^1, x^2, \ldots, x^p)}.$$

5. *Simple forms.*
   a. A simple p-form is one that can be written as a wedge product of p 1-forms:

$$\underset{p}{\boldsymbol{\sigma}} = \underbrace{\boldsymbol{\alpha} \wedge \boldsymbol{\beta} \wedge \cdots \wedge \boldsymbol{\gamma}}_{p \text{ factors}}.$$

   b. A simple p-form $\boldsymbol{\alpha} \wedge \boldsymbol{\beta} \wedge \cdots \wedge \boldsymbol{\gamma}$ is represented by the intersecting families of surfaces of $\boldsymbol{\alpha}, \boldsymbol{\beta}, \ldots, \boldsymbol{\gamma}$ (egg-crate structure) plus a sense of circulation (orientation).

Applications:

a. In four dimensions (e.g., spacetime) all 0-forms, 1-forms, 3-forms, and 4-forms are simple. A 2-form $\boldsymbol{F}$ is generally a sum of two simple forms, e.g., $\boldsymbol{F} = -e\,\boldsymbol{d}t \wedge \boldsymbol{d}x + h\,\boldsymbol{d}y \wedge \boldsymbol{d}z$; it is simple if and only if $\boldsymbol{F} \wedge \boldsymbol{F} = 0$.

b. A set of 1-forms $\boldsymbol{\alpha}, \boldsymbol{\beta}, \ldots, \boldsymbol{\gamma}$ is linearly dependent (one a linear combination of the others) if and only if

$$\boldsymbol{\alpha} \wedge \boldsymbol{\beta} \wedge \cdots \wedge \boldsymbol{\gamma} = 0 \qquad \text{(egg crate collapsed)}.$$

B. Exterior Derivative (applicable to any "differentiable manifold," with or without metric)

1. $\boldsymbol{d}$ produces a $(p+1)$-form $\boldsymbol{d\sigma}$ from a p-form $\boldsymbol{\sigma}$.
2. Effect of $\boldsymbol{d}$ is defined by induction using the

Box 4.1 ***(continued)***

(Chapter 2) definition of $\boldsymbol{d}f$, and f a function (0-form), plus

$$\boldsymbol{d}(\underset{p}{\boldsymbol{\alpha}} \wedge \underset{q}{\boldsymbol{\beta}}) = \boldsymbol{d\alpha} \wedge \boldsymbol{\beta} + (-1)^p \boldsymbol{\alpha} \wedge \boldsymbol{d\beta},$$

$$\boldsymbol{d}^2 = \boldsymbol{dd} = 0.$$

Two applications

$$\boldsymbol{d}(\boldsymbol{\alpha} \wedge \boldsymbol{d\beta}) = \boldsymbol{d\alpha} \wedge \boldsymbol{d\beta}.$$

For the p-form $\boldsymbol{\phi}$, with

$$\boldsymbol{\phi} = \phi_{|i_1 \ldots i_p|}\, \boldsymbol{d}x^{i_1} \wedge \cdots \wedge \boldsymbol{d}x^{i_p},$$

one has (alternative and equivalent definition of $\boldsymbol{d\phi}$)

$$\boldsymbol{d\phi} = \boldsymbol{d}\phi_{|i_1 \ldots i_p|} \wedge \boldsymbol{d}x^{i_1} \wedge \cdots \wedge \boldsymbol{d}x^{i_p}.$$

C. Integration (applicable to any "differentiable manifold," with or without metric)

1. *Pictorial interpretation.*
   Text and pictures of Chapter 4 interpret $\int \boldsymbol{\alpha}$ (integral of specified 1-form $\boldsymbol{\alpha}$ along specified curve from specified starting point to specified end point) as "number of $\boldsymbol{\alpha}$-surfaces pierced on that route"; similarly, they interpret $\int \boldsymbol{\phi}$ (integral of specified 2-form $\boldsymbol{\phi}$ over specified bit of surface on which there is an assigned sense of circulation or "orientation") as "number of cells of the honeycomb-like structure $\boldsymbol{\phi}$ cut through by that surface"; similarly for the egg-crate-like structures that represent 3-forms; etc.
2. *Computational rules for integration.*
   To evaluate $\int \boldsymbol{\alpha}$, the integral of a p-form

   $$\boldsymbol{\alpha} = \alpha_{|i_1 \ldots i_p|}(x^1, \ldots, x^n)\, \boldsymbol{d}x^{i_1} \wedge \cdots \wedge \boldsymbol{d}x^{i_p},$$

   over a p-dimensional surface, proceed in two steps.
   a. Substitute a parameterization of the surface,

   $$x^k(\lambda^1, \ldots, \lambda^p)$$

   into $\boldsymbol{\alpha}$, and collect terms in the form

   $$\boldsymbol{\alpha} = a(\lambda^j)\, \boldsymbol{d}\lambda^1 \wedge \cdots \wedge \boldsymbol{d}\lambda^p$$

   (this is $\boldsymbol{\alpha}$ viewed as a p-form in the p-dimensional surface);

b. Integrate

$$\int \boldsymbol{\alpha} = \int a(\lambda^j)\, d\lambda^1\, d\lambda^2 \ldots d\lambda^p$$

using elementary definition of integration.

Example: See equations (4.12) to (4.14).

3. *The differential geometry of integration.*
Calculate $\int \boldsymbol{\alpha}$ for a p-form $\boldsymbol{\alpha}$ as follows.
   a. Choose the p-dimensional surface $\mathcal{S}$ over which to integrate.
   b. Represent $\mathcal{S}$ by a parametrization giving the generic point of the surface as a function of the parameters, $\mathscr{P}(\lambda^1, \lambda^2, \ldots \lambda^p)$. This fixes the orientation. The same function with $\lambda^1 \leftrightarrow \lambda^2$, $\mathscr{P}(\lambda^2, \lambda^1, \ldots, \lambda^p)$, describes a different (i.e., oppositely oriented) surface, $-\mathcal{S}$.
   c. The infinitesimal parallelepiped

$$\left(\frac{\partial \mathscr{P}}{\partial \lambda^1} \Delta\lambda^1\right) \wedge \left(\frac{\partial \mathscr{P}}{\partial \lambda^2} \Delta\lambda^2\right) \wedge \cdots \wedge \left(\frac{\partial \mathscr{P}}{\partial \lambda^p} \Delta\lambda^p\right)$$

is tangent to the surface. The number of cells of $\boldsymbol{\alpha}$ it slices is

$$\left\langle \boldsymbol{\alpha}, \frac{\partial \mathscr{P}}{\partial \lambda^1} \wedge \cdots \wedge \frac{\partial \mathscr{P}}{\partial \lambda^p} \right\rangle \Delta\lambda^1 \ldots \Delta\lambda^p.$$

This number changes sign if two of the vectors $\partial\mathscr{P}/\partial\lambda^k$ are interchanged, as for an oppositely oriented surface.
   d. The above provides an interpretation motivating the definition

$$\int \boldsymbol{\alpha} \equiv \int\int \cdots \int \left\langle \boldsymbol{\alpha}, \frac{\partial \mathscr{P}}{\partial \lambda^1} \wedge \frac{\partial \mathscr{P}}{\partial \lambda^2} \wedge \cdots \wedge \frac{\partial \mathscr{P}}{\partial \lambda^p} \right\rangle d\lambda^1\, d\lambda^2 \ldots d\lambda^p.$$

This definition is identified with the computational rule of the preceding section (C.2) in exercise 4.9.

An application
Integrate a gradient $\boldsymbol{d}f$ along a curve, $\mathscr{P}(\lambda)$ from $\mathscr{P}(0)$ to $\mathscr{P}(1)$:

$$\int \boldsymbol{d}f = \int_0^1 \langle \boldsymbol{d}f, d\mathscr{P}/d\lambda \rangle\, d\lambda = \int_0^1 (df/d\lambda)\, d\lambda = f[\mathscr{P}(1)] - f[\mathscr{P}(0)].$$

   e. Three different uses for symbol "d": *First,* lightface d in explicit derivative expressions such as

Box 4.1 ***(continued)***

d/da, or df/da, or $d\mathscr{P}/da$; neither numerator nor denominator alone has any meaning, but only the full string of symbols. *Second,* lightface d inside an integral sign; e.g., $\int f\,da$. This is an instruction to perform integration, and has no meaning whatsoever without an integral sign; "$\int \ldots d \ldots$" lives as an indivisible unit. *Third,* sans-serif $\mathbf{d}$; e.g., $\mathbf{d}$ alone, or $\mathbf{d}f$, or $\mathbf{d}a$. This is an exterior derivative, which converts a p-form into a $(p+1)$-form. Sometimes lightface d is used for the same purpose. Hence, d alone, or df, or dx, is always an exterior derivative unless coupled to an $\int$ sign (*second* use), or coupled to a / sign (*first* use).

4. *The generalized Stokes theorem* (see Box 4.6).
   a. Let $\partial\mathscr{V}$ be the closed p-dimensional boundary of a $(p+1)$-dimensional surface $\mathscr{V}$. Let $\boldsymbol{\sigma}$ be a p-form defined throughout $\mathscr{V}$.

   Then

$$\boxed{\int_{\mathscr{V}} \mathbf{d}\boldsymbol{\sigma} = \int_{\partial\mathscr{V}} \boldsymbol{\sigma}}$$

   [integral of p-form $\boldsymbol{\sigma}$ over boundary $\partial\mathscr{V}$ equals integral of $(p+1)$-form $\mathbf{d}\boldsymbol{\sigma}$ over interior $\mathscr{V}$].

   b. For the sign to come out right, orientations of $\mathscr{V}$ and $\partial\mathscr{V}$ must agree in this sense: choose coordinates $y^0, y^1, \ldots, y^p$ on a portion of $\mathscr{V}$, with y^0 specialized so $y^0 \leq 0$ in $\mathscr{V}$, and $y^0 = 0$ at the boundary $\partial\mathscr{V}$; then the orientation

$$\frac{\partial\mathscr{P}}{\partial y^0} \wedge \frac{\partial\mathscr{P}}{\partial y^1} \wedge \cdots \wedge \frac{\partial\mathscr{P}}{\partial y^p}$$

   for $\mathscr{V}$ demands the orientation

$$\frac{\partial\mathscr{P}}{\partial y^1} \wedge \cdots \wedge \frac{\partial\mathscr{P}}{\partial y^p}$$

   for $\partial\mathscr{V}$.

   c. Note: For a nonorientable surface, such as a Möbius strip, where a consistent and continuous choice of orientation is impossible, more intricate mathematics is required to give a definition of "∂" for which the Stokes theorem holds.

Applications: Includes as special cases all integral theorems for surfaces of arbitrary dimension in spaces of arbitrary dimension, with or without metric, generalizing all versions of theorems of Stokes and Gauss. Examples:

a. $\mathcal{V}$ a curve, $\partial\mathcal{V}$ its endpoints, $\boldsymbol{\sigma} = f$ a 0-form (function):

$$\int_{\mathcal{V}} \boldsymbol{d}f = \int_0^1 (df/d\lambda)\, d\lambda = \int_{\partial\mathcal{V}} f = f(1) - f(0).$$

b. $\mathcal{V}$ a 2-surface in 3-space, $\partial\mathcal{V}$ its closed-curve boundary, $\boldsymbol{v}$ a 1-form; translated into Euclidean vector notation, the two integrals are

$$\int_{\mathcal{V}} \boldsymbol{dv} = \int_{\mathcal{V}} (\nabla \times v)\cdot \boldsymbol{dS}; \int_{\partial\mathcal{V}} \boldsymbol{v} = \int_{\partial\mathcal{V}} v \cdot \boldsymbol{dl}.$$

c. Other applications in §§5.8, 20.2, 20.3, 20.5, and exercises 4.10, 4.11, 5.2, and below.

D. Algebra II (applicable to any vector space with metric)

1. *Norm of a p-form.*

$||\boldsymbol{\alpha}||^2 \equiv \alpha_{|i_1\ldots i_p|}\alpha^{i_1\ldots i_p}$.

Two applications: Norm of a 1-form equals its squared length, $||\boldsymbol{\alpha}||^2 = \boldsymbol{\alpha}\cdot\boldsymbol{\alpha}$. Norm of electromagnetic 2-form or ***Faraday***: $||\boldsymbol{F}||^2 = B^2 - E^2$.

2. *Dual of a p-form.*

a. In an n-dimensional space, the dual of a p-form $\boldsymbol{\alpha}$ is the $(n-p)$-form $^*\boldsymbol{\alpha}$, with components

$$(^*\alpha)_{k_1\ldots k_{n-p}} = \alpha^{|i_1\ldots i_p|}\epsilon_{i_1\ldots i_p k_1\ldots k_{n-p}}.$$

b. Properties of duals:

$$^{**}\boldsymbol{\alpha} = (-1)^{p-1}\boldsymbol{\alpha} \text{ in spacetime;}$$
$$\boldsymbol{\alpha} \wedge {}^*\boldsymbol{\alpha} = ||\boldsymbol{\alpha}||^2\boldsymbol{\varepsilon} \text{ in general.}$$

c. Note: the definition of $\boldsymbol{\varepsilon}$ (exercise 3.13) entails choosing an orientation of the space, i.e., deciding which orthonormal bases (1) are "right-handed" and thus (2) have $\boldsymbol{\varepsilon}(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n) = +1$.

Applications

a. For f a 0-form, $^*f = f\boldsymbol{\varepsilon}$, and $\int f d(\text{volume}) = \int {}^*f$.

b. Dual of charge-current 1-form $\boldsymbol{J}$ is charge-current 3-form $^*\boldsymbol{J}$. The total charge Q in a 3-dimensional hypersurface region $\mathcal{S}$ is

$$Q(\mathcal{S}) = \int_{\mathcal{S}} {}^*\boldsymbol{J}.$$

Box 4.1 *(continued)*

Conservation of charge is stated locally by $\boldsymbol{d}{}^*\boldsymbol{J} = 0$. Stokes' Theorem goes from this differential conservation law to the integral conservation law,

$$0 = \int_{\mathcal{V}} \boldsymbol{d}{}^*\boldsymbol{J} = \int_{\partial\mathcal{V}} {}^*\boldsymbol{J}.$$

This law is of most interest when $\partial\mathcal{V} = \mathcal{S}_2 - \mathcal{S}_1$ consists of the future $\mathcal{S}_2$ and past $\mathcal{S}_1$ boundaries of a spacetime region, in which case it states $Q(\mathcal{S}_2) = Q(\mathcal{S}_1)$; see exercise 5.2.

c. Dual of electromagnetic field tensor $\boldsymbol{F} =$ ***Faraday*** is ${}^*\boldsymbol{F} =$ ***Maxwell***. From the $\boldsymbol{d}{}^*\boldsymbol{F} = 4\pi\, {}^*\boldsymbol{J}$ Maxwell equation, find $4\pi Q = 4\pi\int_{\mathcal{S}} {}^*\boldsymbol{J} = \int_{\mathcal{S}} \boldsymbol{d}{}^*\boldsymbol{F} = \int_{\partial\mathcal{S}} {}^*\boldsymbol{F}$.

3. *Simple forms revisited.*
   a. The dual of a simple form is simple.
   b. Egg crate of ${}^*\boldsymbol{\sigma}$ is perpendicular to egg crate of $\boldsymbol{\sigma} = \boldsymbol{\alpha} \wedge \boldsymbol{\beta} \wedge \cdots \wedge \boldsymbol{\mu}$ in this sense:
      (1) pick any vector $\boldsymbol{V}$ lying in intersection of surfaces of $\boldsymbol{\sigma}$

      $$(\langle \boldsymbol{\alpha}, \boldsymbol{V}\rangle = \langle \boldsymbol{\beta}, \boldsymbol{V}\rangle = \cdots = \langle \boldsymbol{\mu}, \boldsymbol{V}\rangle = 0);$$

      (2) pick any vector $\boldsymbol{W}$ lying in intersection of surfaces of ${}^*\boldsymbol{\sigma}$;
      (3) then $\boldsymbol{V}$ and $\boldsymbol{W}$ are necessarily perpendicular: $\boldsymbol{V}\cdot\boldsymbol{W} = 0$.

Example: $\boldsymbol{\sigma} = 3\,\boldsymbol{d}t$ is a simple 1-form in spacetime.
a. ${}^*\boldsymbol{\sigma} = -3\,\boldsymbol{d}x \wedge \boldsymbol{d}y \wedge \boldsymbol{d}z$ is a simple 3-form.
b. General vector in surfaces of $\boldsymbol{\sigma}$ is

$$\boldsymbol{V} = V^x\boldsymbol{e}_x + V^y\boldsymbol{e}_y + V^z\boldsymbol{e}_z.$$

c. General vector in intersection of surfaces of ${}^*\boldsymbol{\sigma}$ is

$$\boldsymbol{W} = W^t\boldsymbol{e}_t.$$

d. $\boldsymbol{W}\cdot\boldsymbol{V} = 0$.

§4.2. ELECTROMAGNETIC 2-FORM AND LORENTZ FORCE

The electromagnetic field tensor, ***Faraday*** = $\boldsymbol{F}$, is an antisymmetric second-rank tensor (i.e., 2-form). Instead of expanding it in terms of the tensor products of basis 1-forms,

$$\boldsymbol{F} = F_{\alpha\beta}\, \boldsymbol{d}x^{\alpha} \otimes \boldsymbol{d}x^{\beta},$$

the exterior calculus prefers to expand in terms of antisymmetrized tensor products (*"exterior products,"* exercise 4.1):

Electromagnetic 2-form expressed in terms of exterior products

$$\boldsymbol{F} = \frac{1}{2} F_{\alpha\beta}\, \boldsymbol{d}x^{\alpha} \wedge \boldsymbol{d}x^{\beta}, \tag{4.1}$$

$$\boldsymbol{d}x^{\alpha} \wedge \boldsymbol{d}x^{\beta} \equiv \boldsymbol{d}x^{\alpha} \otimes \boldsymbol{d}x^{\beta} - \boldsymbol{d}x^{\beta} \otimes \boldsymbol{d}x^{\alpha}. \tag{4.2}$$

Any 2-form (antisymmetric, second-rank tensor) can be so expanded. The symbol "$\wedge$" is variously called a "wedge," a "hat," or an "exterior product sign"; and $\boldsymbol{d}x^{\alpha} \wedge \boldsymbol{d}x^{\beta}$ are the "basis 2-forms" of a given Lorentz frame (see §3.5, exercise 3.12, and Box 4.1).

There is no simpler way to illustrate this 2-form representation of the electromagnetic field than to consider a magnetic field in the x-direction:

$$\begin{aligned} F_{yz} &= -F_{zy} = B_x, \\ \boldsymbol{F} &= B_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z. \end{aligned} \tag{4.3}$$

The 1-form $\boldsymbol{d}y$ = grad y is the set of surfaces (actually hypersurfaces) $y = 18$ (all t, x, z), $y = 19$ (all t, x, z), $y = 20$ (all t, x, z), etc.; and surfaces uniformly interpolated between them. Similarly for the 1-form $\boldsymbol{d}z$. The intersection between these two sets of surfaces produces a honeycomb-like structure. That structure becomes a "2-form" when it is supplemented by instructions (see arrows in Figure 4.1) that give a "sense of circulation" to each tube of the honeycomb (order of factors in the "wedge product" of equation 4.2; $\boldsymbol{d}y \wedge \boldsymbol{d}z = -\boldsymbol{d}z \wedge \boldsymbol{d}y$). The 2-form $\boldsymbol{F}$ in the example differs from this "basis 2-form" $\boldsymbol{d}y \wedge \boldsymbol{d}z$ only in this respect, that where $\boldsymbol{d}y \wedge \boldsymbol{d}z$ had one tube, the field 2-form has B_x tubes.

A 2-form as a honeycomb of tubes with a sense of circulation

When one considers a tubular structure that twists and turns on its way through spacetime, one must have more components to describe it. The 2-form for the general electromagnetic field can be written as

$$\begin{aligned} \boldsymbol{F} = {} & E_x\, \boldsymbol{d}x \wedge \boldsymbol{d}t + E_y\, \boldsymbol{d}y \wedge \boldsymbol{d}t + E_z\, \boldsymbol{d}z \wedge \boldsymbol{d}t + B_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z \\ & + B_y\, \boldsymbol{d}z \wedge \boldsymbol{d}x + B_z\, \boldsymbol{d}x \wedge \boldsymbol{d}y \end{aligned} \tag{4.4}$$

(6 components, 6 basis 2-forms).

A 1-form is a machine to produce a number out of a vector (bongs of a bell as the vector pierces successive surfaces). A 2-form is a machine to produce a number out of an oriented surface (surface with a sense of circulation indicated on it: Figure 4.1, lower right). The meaning is as clear here as it is in elementary magnetism:

A 2-form as a machine to produce a number out of an oriented surface

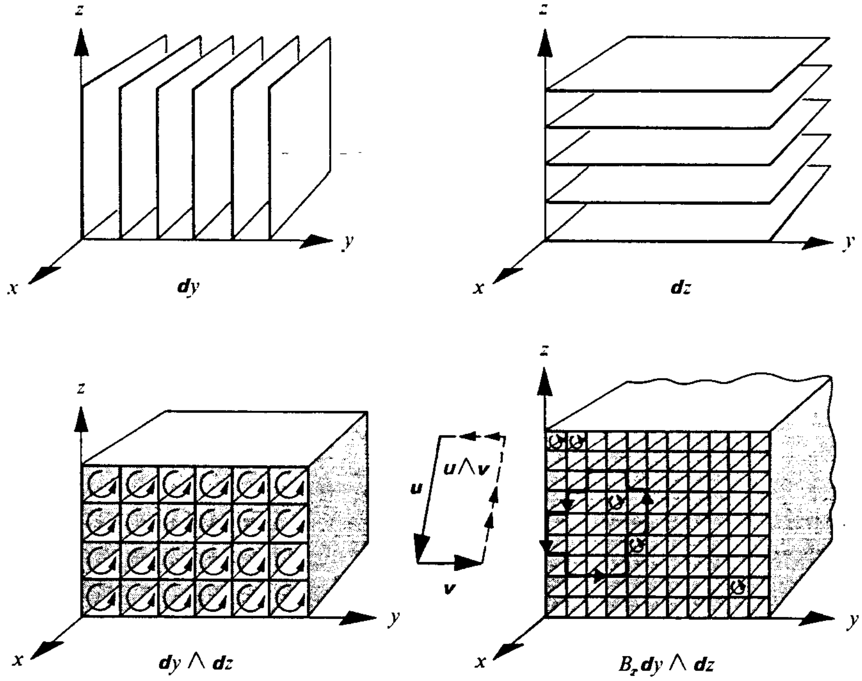


Figure 4.1.
Construction of the 2-form for the electromagnetic field $\boldsymbol{F} = B_z\, \boldsymbol{d}y \wedge \boldsymbol{d}z$ out of the 1-forms $\boldsymbol{d}y$ and $\boldsymbol{d}z$ by "wedge multiplication" (formation of honeycomb-like structure with sense of circulation indicated by arrows). A 2-form is a "machine to construct a number out of an oriented surface" (illustrated by sample surface enclosed by arrows at lower right; number of tubes intersected by this surface is

$$\int_{\text{(this surface)}} \boldsymbol{F} = 18;$$

Faraday's concept of "magnetic flux"). This idea of 2-form machinery can be connected to the "tensor-as-machine" idea of Chapter 3 as follows. The shape of the oriented surface over which one integrates $\boldsymbol{F}$ does not matter, for small surfaces. All that affects $\int \boldsymbol{F}$ is the area of the surface, and its orientation. Choose two vectors, $\boldsymbol{u}$ and $\boldsymbol{v}$, that lie in the surface. They form two legs of a parallelogram, whose orientation ($\boldsymbol{u}$ followed by $\boldsymbol{v}$) and area are embodied in the exterior product $\boldsymbol{u} \wedge \boldsymbol{v}$. Adjust the lengths of $\boldsymbol{u}$ and $\boldsymbol{v}$ so their parallelogram, $\boldsymbol{u} \wedge \boldsymbol{v}$, has the same area as the surface of integration. Then

$$\underbrace{\int_{\text{surface}} \boldsymbol{F} = \int_{\boldsymbol{u}\wedge\boldsymbol{v}} \boldsymbol{F}}_{\text{machinery idea of this chapter}} = \underbrace{\boldsymbol{F}(\boldsymbol{u}, \boldsymbol{v})}_{\text{machinery idea of Chapter 3}}.$$

Exercise: derive this result, for an infinitesimal surface $\boldsymbol{u} \wedge \boldsymbol{v}$ and for general $\boldsymbol{F}$, using the formalism of Box 4.1.

the number of Faraday tubes cut by that surface. The electromagnetic 2-form $\boldsymbol{F}$ or ***Faraday*** described by such a "tubular structure" (suitably abstracted; Box 4.2) has a reality and a location in space that is independent of all coordinate systems and all artificial distinctions between "electric" and "magnetic" fields. Moreover, those tubes provide the most direct geometric representation that anyone has ever been able to give for the machinery by which the electromagnetic field acts on a charged particle. Take a particle of charge e and 4-velocity

$$\boldsymbol{u} = \frac{dx^\alpha}{d\tau}\boldsymbol{e}_\alpha. \tag{4.5}$$

Let this particle go through a region where the electromagnetic field is described by the 2-form

$$\boldsymbol{F} = B_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z \tag{4.6}$$

of Figure 4.1. Then the force exerted on the particle (regarded as a 1-form) is the contraction of this 2-form with the 4-velocity (and the charge);

Lorentz force as contraction of electromagnetic 2-form with particle's 4-velocity

$$\dot{\boldsymbol{p}} = \boldsymbol{d}\boldsymbol{p}/d\tau = e\boldsymbol{F}(\boldsymbol{u}) \equiv e\langle \boldsymbol{F}, \boldsymbol{u}\rangle, \tag{4.7}$$

as one sees by direct evaluation, letting the two factors in the 2-form act in turn on the tangent vector $\boldsymbol{u}$:

$$\begin{aligned}\dot{\boldsymbol{p}} &= eB_x\langle \boldsymbol{d}y \wedge \boldsymbol{d}z, \boldsymbol{u}\rangle \\ &= eB_x\{\boldsymbol{d}y\langle \boldsymbol{d}z, \boldsymbol{u}\rangle - \boldsymbol{d}z\langle \boldsymbol{d}y, \boldsymbol{u}\rangle\} \\ &= eB_x\{\boldsymbol{d}y\langle \boldsymbol{d}z, u^z\boldsymbol{e}_z\rangle - \boldsymbol{d}z\langle \boldsymbol{d}y, u^y\boldsymbol{e}_y\rangle\}\end{aligned}$$

or

$$\dot{p}_\alpha\, \boldsymbol{d}x^\alpha = eB_x u^z\, \boldsymbol{d}y - eB_x u^y\, \boldsymbol{d}z. \tag{4.8}$$

Comparing coefficients of the separate basis 1-forms on the two sides of this equation, one sees reproduced all the detail of the Lorentz force exerted by the magnetic field B_x:

$$\begin{aligned}\dot{p}_y &= \frac{dp_y}{d\tau} = eB_x\frac{dz}{d\tau}, \\ \dot{p}_z &= \frac{dp_z}{d\tau} = -eB_x\frac{dy}{d\tau}.\end{aligned} \tag{4.9}$$

By simple extension of this line of reasoning to the general electromagnetic field, one concludes that *the time-rate of change of momentum (1-form) is equal to the charge multiplied by the contraction of the **Faraday** with the 4-velocity.* Figure 4.2 illustrates pictorially how the 2-form, $\boldsymbol{F}$, serves as a machine to produce the 1-form, $\dot{\boldsymbol{p}}$, out of the tangent vector, $e\boldsymbol{u}$.

(continued on page 105)

Box 4.2 ABSTRACTING A 2-FORM FROM THE CONCEPT OF "HONEYCOMB-LIKE STRUCTURE," IN 3-SPACE AND IN SPACETIME

Open up a cardboard carton containing a dozen bottles, and observe the honeycomb structure of intersecting north-south and east-west cardboard separators between the bottles. That honeycomb structure of "tubes" ("channels for bottles") is a fairly apt illustration of a 2-form in the context of everyday 3-space. It yields a number (number of tubes cut) for each choice of smooth element of 2-surface slicing through the three-dimensional structure. However, the intersecting cardboard separators are rather too specific. All that a true 2-form can ever give is the number of tubes sliced through, not the "shape" of the tubes. Slew the carton around on the floor by 45°. Then half the separators run NW-SE and the other half run NE-SW, but through a given bit of 2-surface fixed in 3-space the count of tubes is unchanged. Therefore, one should be careful to make the concept of tubes in the mind's eye abstract enough that one envisages direction of tubes (vertical in the example) and density of tubes, but not any specific location or orientation for the tube walls. Thus all the following representations give one and the same 2-form, $\boldsymbol{\sigma}$:

$$\boldsymbol{\sigma} = B\, \mathbf{d}x \wedge \mathbf{d}y;$$

$$\boldsymbol{\sigma} = B(2\, \mathbf{d}x) \wedge \left(\frac{1}{2}\, \mathbf{d}y\right)$$

(NS cardboards spaced twice as close as before; EW cardboards spaced twice as wide as before);

$$\boldsymbol{\sigma} = B\, \mathbf{d}\left(\frac{x-y}{\sqrt{2}}\right) \wedge \mathbf{d}\left(\frac{x+y}{\sqrt{2}}\right)$$

(cardboards rotated through 45°);

$$\boldsymbol{\sigma} = B \frac{\alpha\, \mathbf{d}x + \beta\, \mathbf{d}y}{(\alpha\delta - \beta\gamma)^{1/2}} \wedge \frac{\gamma\, \mathbf{d}x + \delta\, \mathbf{d}y}{(\alpha\delta - \beta\gamma)^{1/2}}$$

(both orientation and spacing of "cardboards" changing from point to point, with all four functions, α, β, γ, and δ, depending on position).

What has physical reality, and constitutes the real geometric object, is not any one of the 1-forms just encountered individually, but only the 2-form $\boldsymbol{\sigma}$ itself. This circumstance helps to explain why in the physical literature one sometimes refers to "tubes of force" and sometimes to "lines of force." The two terms for the same structure have this in common, that each yields a number when sliced by a bit of surface. The line-of-force picture has the advantage of not imposing on the mind any specific structure of "sheets of cardboard"; that is, any specific decomposition of the 2-form into the product of 1-forms. However, that very feature is also a disadvantage, for in a calculation one often finds it useful to have a well-defined representation of the 2-form as the wedge product of 1-forms. Moreover, the tube picture, abstract though it must be if it is to be truthful, also has this advantage, that the orientation of the elementary tubes (sense of circulation as indicated by arrows in Figures 4.1 and 4.5, for example) lends itself to ready visualization. Let the "walls" of the tubes therefore remain in all pictures drawn in this book as a reminder that 2-forms can be built out of 1-forms; but let it be understood here and hereafter how manyfold are the options for the individual 1-forms!

Turn now from three dimensions to four, and find that the concept of "honeycomb-like structure" must be made still more abstract. In three dimensions the arbitrariness of the decomposition of the 2-form into 1-forms showed in the slant and packing of the "cardboards," but had no effect on the verticality of the "channels for the bottles" ("direction of Faraday lines of force or tubes of force"); not so in four dimensions, or at least not in the generic case in four dimensions.

In special cases, the story is almost as simple in four dimensions as in three. An example of a special case is once again the 2-form $\boldsymbol{\sigma} = B\, \boldsymbol{d}x \wedge \boldsymbol{d}y$, with all the options for decomposition into 1-forms that have already been mentioned, but with every option giving the same "direction" for the tubes. If the word "direction" now rises in status from "tube walls unpierced by motion in the direction of increasing z" to "tube walls unpierced either by motion in the direction of increasing z, or by motion in the direction of increasing t, or by any linear combination of such motions," that is a natural enough consequence of adding the new dimension. Moreover, the same simplicity prevails for an electromagnetic plane wave. For example, let the wave be advancing in the z-direction, and let the electric polarization point in the x-direction; then for a monochromatic wave, one has

$$E_x = B_y = E_0 \cos \omega(z - t) = -F_{01} = F_{31},$$

and all components distinct from these equal zero. ***Faraday*** is

$$\begin{aligned} \boldsymbol{F} &= F_{01}\, \boldsymbol{d}t \wedge \boldsymbol{d}x + F_{31}\, \boldsymbol{d}z \wedge \boldsymbol{d}x \\ &= E_0 \cos \omega(z - t)\, \boldsymbol{d}(z - t) \wedge \boldsymbol{d}x, \end{aligned}$$

which is again representable as a single wedge product of two 1-forms.

Not so in general! The general 2-form in four dimensions consists of six distinct wedge products,

$$\boldsymbol{F} = F_{01}\, \boldsymbol{d}t \wedge \boldsymbol{d}x + F_{02}\, \boldsymbol{d}t \wedge \boldsymbol{d}y + \cdots + F_{23}\, \boldsymbol{d}y \wedge \boldsymbol{d}z.$$

It is too much to hope that this expression will reduce in the generic case to a single wedge product of two 1-forms (*"simple"* 2-form). It is not even true that it will. It is only remarkable that it can be reduced from six exterior products to two (details in exercise 4.1); thus,

$$\boldsymbol{F} = \boldsymbol{n}^1 \wedge \boldsymbol{\xi}^1 + \boldsymbol{n}^2 \wedge \boldsymbol{\xi}^2.$$

Each product $\boldsymbol{n}^i \wedge \boldsymbol{\xi}^i$ individually can be visualized as a honeycomb-like structure like those depicted in Figures 4.1, 4.2, 4.4, and 4.5. Each such structure individually can be pictured as built out of intersecting sheets (1-forms), but with such details as the tilt and packing of these 1-forms abstracted away. Each such structure individually gives a number when sliced by an element of surface. What counts for the 2-form $\boldsymbol{F}$, however, is neither the number of tubes of $\boldsymbol{n}^1 \wedge \boldsymbol{\xi}^1$ cut by the surface, nor the number of tubes of $\boldsymbol{n}^2 \wedge \boldsymbol{\xi}^2$ cut by the surface, but only the sum of the two. This sum is what is referred to in the text as the "number of tubes of $\boldsymbol{F}$" cut by the surface. The contribution of either wedge product individually is not well-defined, for a simple reason: the decomposition of a six-wedge-product object into two wedge products, miraculous though it seems, is actually far from unique (details in exercise 4.2).

In keeping with the need to have two products of 1-forms to represent the general 2-form note that the vanishing of $\boldsymbol{dF}$ ("no magnetic charges") does not automatically imply that $\boldsymbol{d}(\boldsymbol{n}^1 \wedge \boldsymbol{\xi}^1)$ or $\boldsymbol{d}(\boldsymbol{n}^2 \wedge \boldsymbol{\xi}^2)$ separately vanish. Note also that any spacelike slice through the general 2-form $\boldsymbol{F}$ (reduction from four dimensions to three) can always be represented in terms of a honeycomb-like structure ("simple" 2-form in three dimensions; Faraday's picture of magnetic tubes of force).

Despite the abstraction that has gone on in seeing in all generality what a 2-form is, there is no bar to continuing to use the term "honeycomb-like structure" in a broadened sense to describe this object; and that is the practice here and hereafter.

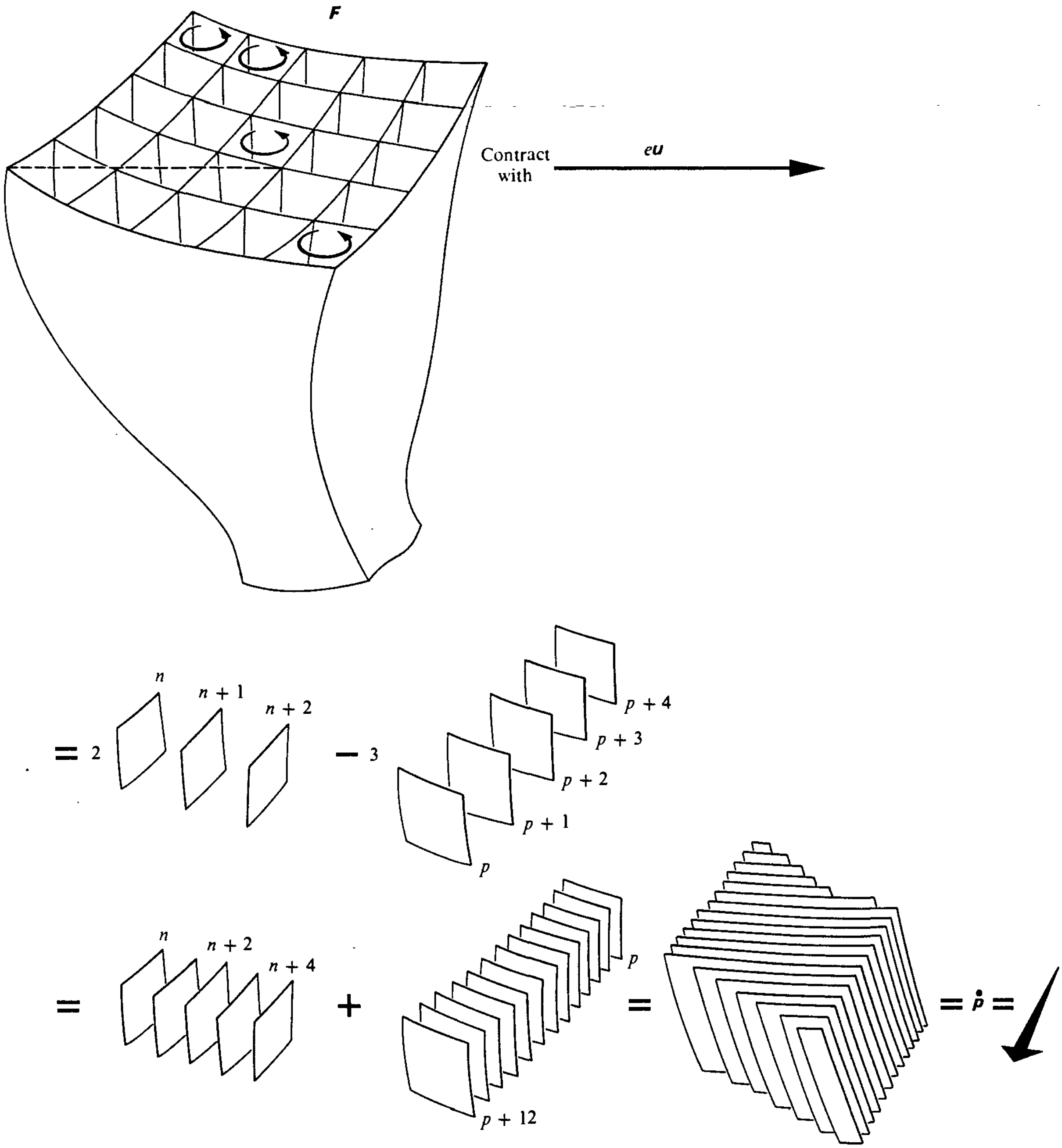


Figure 4.2.
The ***Faraday*** or 2-form ***F*** of the electromagnetic field is a machine to produce a 1-form (the time-rate of change of momentum $\dot{\boldsymbol{p}}$ of a charged particle) out of a tangent vector (product of charge e of the particle and its 4-velocity $\boldsymbol{u}$). In spacetime the general 2-form is the "superposition" (see Box 4.2) of two structures like that illustrated at the top of this diagram, the tubes of the first being tilted and packed as indicated, the tubes of the second being tilted in another direction and having a different packing density.

§4.3. FORMS ILLUMINATE ELECTROMAGNETISM, AND ELECTROMAGNETISM ILLUMINATES FORMS

All electromagnetism allows itself to be summarized in the language of 2-forms, honeycomb-like "structures" (again in the abstract sense of "structure" of Box 4.2) of tubes filling all spacetime, as well when spacetime is curved as when it is flat. In brief, there are two such structures, one ***Faraday*** = ***F***, the other ***Maxwell*** = ****F***, each dual ("perpendicular," the only place where metric need enter the discussion) to the other, each satisfying an elementary equation:

$$\boldsymbol{dF} = 0 \tag{4.10}$$

("no tubes of ***Faraday*** ever end") and

$$\boldsymbol{d}{}^*\boldsymbol{F} = 4\pi\, {}^*\boldsymbol{J} \tag{4.11}$$

("the number of tubes of ***Maxwell*** that end in an elementary volume is equal to the amount of electric charge in that volume"). To see in more detail how this machinery shows up in action, look in turn at: (1) the definition of a 2-form; (2) the appearance of a given electromagnetic field as ***Faraday*** and as ***Maxwell***; (3) the ***Maxwell*** structure for a point-charge at rest; (4) the same for a point-charge in motion; (5) the nature of the field of a charge that moves uniformly except during a brief instant of acceleration; (6) the ***Faraday*** structure for the field of an oscillating dipole; (7) the concept of exterior derivative; (8) Maxwell's equations in the language of forms; and (9) the solution of Maxwell's equations in flat spacetime, using a 1-form ***A*** from which the Liénard-Wiechert 2-form ***F*** can be calculated via ***F*** = ***dA***.

Preview of key points in electromagnetism

A 2-form, as illustrated in Figure 4.1, is a machine to construct a number ("net number of tubes cut") out of any "oriented 2-surface" (2-surface with "sense of circulation" marked on it):

A 2-form as machine for number of tubes cut

$$\begin{pmatrix}\text{number}\\ \text{of tubes}\\ \text{cut}\end{pmatrix} = \int_{\text{surface}} \boldsymbol{F}. \tag{4.12}$$

For example, let the 2-form be the one illustrated in Figure 4.1

Number of tubes cut calculated in one example

$$\boldsymbol{F} = B_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z,$$

and let the surface of integration be the portion of the surface of the 2-sphere $x^2 + y^2 + z^2 = a^2$, t = constant, bounded between $\theta = 70°$ and $\theta = 110°$ and between $\varphi = 0°$ and $\varphi = 90°$ ("Atlantic region of the tropics"). Write

$$\begin{aligned} y &= a \sin\theta \sin\varphi, \\ z &= a\cos\theta, \\ \boldsymbol{d}y &= a(\cos\theta \sin\varphi\, \boldsymbol{d}\theta + \sin\theta\cos\varphi\, \boldsymbol{d}\varphi), \\ \boldsymbol{d}z &= -a\sin\theta\, \boldsymbol{d}\theta, \\ \boldsymbol{d}y\wedge\boldsymbol{d}z &= a^2\sin^2\theta\cos\varphi\, \boldsymbol{d}\theta \wedge \boldsymbol{d}\varphi. \end{aligned} \tag{4.13}$$

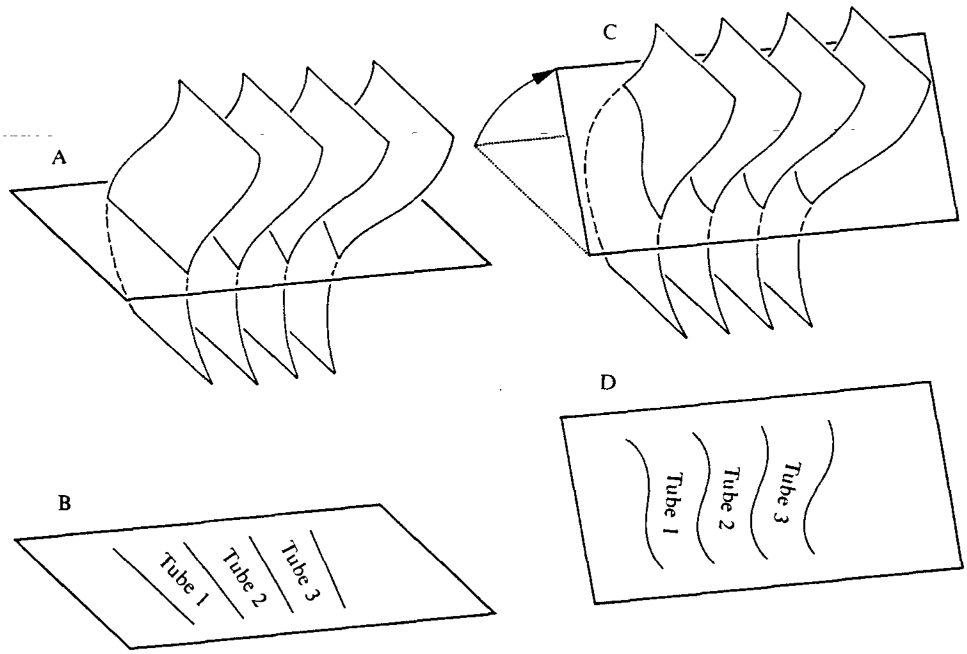


Figure 4.3.
Spacelike slices through ***Faraday***, the electromagnetic 2-form, a geometric object, a honeycomb of tubes that pervades all spacetime ("honeycomb" in the abstract sense spelled out more precisely in Box 4.2). The surfaces in the drawing do not look like a 2-form (honeycomb), because the second family of surfaces making up the honeycomb extends in the spatial direction that is suppressed from the drawing. Diagram A shows one spacelike slice through the 2-form (time increases upwards in the diagram). In diagram B, a projection of the 2-form on this spacelike hypersurface gives the Faraday tubes of magnetic force in this three-dimensional geometry (if the suppressed dimension were restored, the tubes would be tubes, not channels between lines). Diagram C shows another spacelike slice (hypersurface of simultaneity for an observer in a different Lorentz frame). Diagram D shows the very different pattern of magnetic tubes in this reference system. The demand that magnetic tubes of force shall not end ($\nabla \cdot \boldsymbol{B} = 0$), repeated over and over for every spacelike slice through ***Faraday***, gives everywhere the result $\partial \boldsymbol{B}/\partial t = -\nabla \times \boldsymbol{E}$. Thus (magnetostatics) + (covariance) $\longrightarrow$ (magnetodynamics). Similarly—see Chapters 17 and 21—(geometrostatics) + (covariance) $\longrightarrow$ (geometrodynamics).

The structure $\boldsymbol{d}\theta \wedge \boldsymbol{d}\theta$ looks like a "collapsed egg-crate" (Figure 1.4, upper right) and has zero content, a fact formally evident from the vanishing of $\boldsymbol{\alpha} \wedge \boldsymbol{\beta} = -\boldsymbol{\beta} \wedge \boldsymbol{\alpha}$ when $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are identical. The result of the integration, assuming constant B_x, is

$$\int_{\text{surface}} \boldsymbol{F} = a^2 B_x \int_{70^\circ}^{110^\circ} \sin^2\theta \, d\theta \int_{0^\circ}^{90^\circ} \cos\varphi \, d\varphi \tag{4.14}$$

It is not so easy to visualize a pure electric field by means of its 2-form $\boldsymbol{F}$ (Figure 4.4, left) as it is to visualize a pure magnetic field by means of its 2-form $\boldsymbol{F}$ (Figures 4.1, 4.2, 4.3). Is there not some way to treat the two fields on more nearly the same footing? Yes, construct the 2-form $^*\boldsymbol{F}$ (Figure 4.4, right) that is *dual* ("perpendicular"; Box 4.3; exercise 3.14) to $\boldsymbol{F}$.

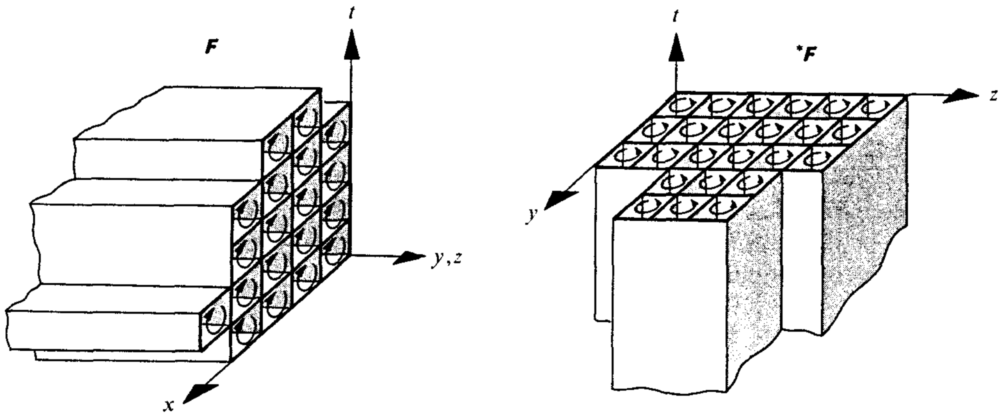


Figure 4.4.
The ***Faraday*** structure

$$\boldsymbol{F} = \frac{1}{2} F_{\mu\nu}\, \boldsymbol{d}x^\mu \wedge \boldsymbol{d}x^\nu = \frac{1}{2} F_{01}\, \boldsymbol{d}t \wedge \boldsymbol{d}x + \frac{1}{2} F_{10}\, \boldsymbol{d}x \wedge \boldsymbol{d}t = E_x\, \boldsymbol{d}x \wedge \boldsymbol{d}t$$

associated with an electric field in the x-direction, and the dual ("perpendicular") ***Maxwell*** honeycomb-like 2-form

$$^*\boldsymbol{F} = \frac{1}{2}\, {}^*F_{\mu\nu}\, \boldsymbol{d}x^\mu \wedge \boldsymbol{d}x^\nu = {}^*F_{23}\, \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 = F^{01}\, \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 = F_{10}\, \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 = E_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z.$$

Represent in geometric form the field of a point-charge of strength e at rest at the origin. Operate in flat space with spherical polar coordinates:

$$\begin{aligned} ds^2 &= -d\tau^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu \\ &= -dt^2 + dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\varphi^2. \end{aligned} \tag{4.15}$$

The electric field in the r-direction being $E_r = e/r^2$, it follows that the 2-form ***F*** or ***Faraday*** is

$$\boldsymbol{F} = \frac{1}{2} F_{\mu\nu}\, \boldsymbol{d}x^\mu \wedge \boldsymbol{d}x^\nu = -E_r\, \boldsymbol{d}t \wedge \boldsymbol{d}r = -\frac{e}{r^2}\, \boldsymbol{d}t \wedge \boldsymbol{d}r. \tag{4.16}$$

Its dual, according to the prescription in exercise 3.14, is ***Maxwell***:

Pattern of tubes in dual structure ***Mexwell*** for point-charge at rest

$$\boldsymbol{Maxwell} = {}^*\boldsymbol{F} = e \sin\theta\, \boldsymbol{d}\theta \wedge \boldsymbol{d}\varphi, \tag{4.17}$$

as illustrated in Figure 4.5.

Take a tour in the positive sense around a region of the surface of the sphere illustrated in Figure 4.5. The number of tubes of ****F*** encompassed in the route will be precisely

$$\begin{pmatrix}\text{number}\\ \text{of tubes}\end{pmatrix} = e \begin{pmatrix}\text{solid}\\ \text{angle}\end{pmatrix}.$$

The whole number of tubes of ****F*** emergent over the entire sphere will be $4\pi e$, in conformity with Faraday's picture of tubes of force.

Box 4.3 DUALITY OF 2-FORMS IN SPACETIME

Given a general 2-form (containing six exterior or wedge products)

$$\boldsymbol{F} = E_x\, \boldsymbol{d}x \wedge \boldsymbol{d}t + E_y\, \boldsymbol{d}y \wedge \boldsymbol{d}t + \cdots + B_z\, \boldsymbol{d}x \wedge \boldsymbol{d}y,$$

one gets to its dual ("perpendicular") by the prescription

$${}^*\boldsymbol{F} = -B_x\, \boldsymbol{d}x \wedge \boldsymbol{d}t - \cdots + E_y\, \boldsymbol{d}z \wedge \boldsymbol{d}x + E_z\, \boldsymbol{d}x \wedge \boldsymbol{d}y.$$

Duality Rotations

Note that the dual of the dual is the negative of the original 2-form; thus

$${}^{**}\boldsymbol{F} = -E_x\, \boldsymbol{d}x \wedge \boldsymbol{d}t - \cdots - B_z\, \boldsymbol{d}x \wedge \boldsymbol{d}y = -\boldsymbol{F}.$$

In this sense * has the same property as the imaginary number i: $** = ii = -1$. Thus one can write

$$e^{*\alpha} = \cos\alpha + {}^*\sin\alpha.$$

This operation, applied to $\boldsymbol{F}$, carries attention from the generic 2-form in its simplest representation (see exercise 4.1)

$$\boldsymbol{F} = E_x\, \boldsymbol{d}x \wedge \boldsymbol{d}t + B_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z$$

to another "duality rotated electromagnetic field"

$$e^{*\alpha}\boldsymbol{F} = (E_x \cos\alpha - B_x \sin\alpha)\, \boldsymbol{d}x \wedge \boldsymbol{d}t + (B_x \cos\alpha + E_x \sin\alpha)\, \boldsymbol{d}y \wedge \boldsymbol{d}z.$$

If the original field satisfied Maxwell's empty-space field equations, so does the new field. With suitable choice of the "complexion" α, one can annul one of the two wedge products at any chosen point in spacetime and have for the other

$$(B_x{}^2 + E_x{}^2)^{1/2}\, \boldsymbol{d}y \wedge \boldsymbol{d}z.$$

Field of a point-charge in motion

How can one determine the structure of tubes associated with a charged particle moving at a uniform velocity? First express ${}^*\boldsymbol{F}$ in rectangular coordinates moving with the particle (barred coordinates in this comoving "rocket" frame of reference; unbarred coordinates will be used later for a laboratory frame of reference). The relevant steps can be listed:

(a)

$${}^*\boldsymbol{F} = e \sin\bar\theta\, \boldsymbol{d}\bar\theta \wedge \boldsymbol{d}\bar\varphi = -e(\boldsymbol{d}\cos\bar\theta) \wedge \boldsymbol{d}\bar\varphi;$$

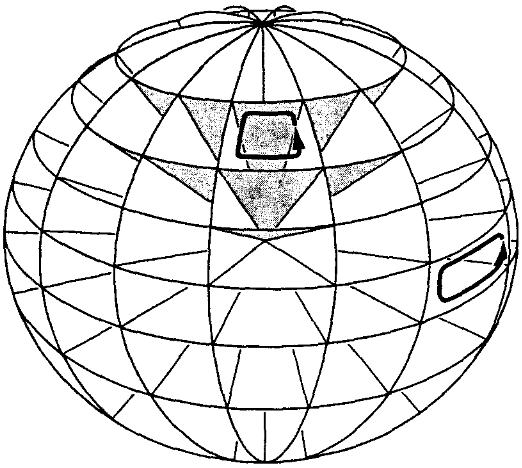


Figure 4.5.
The field of 2-forms ***Maxwell*** $= {}^*\boldsymbol{F} = e \sin\theta\, \boldsymbol{d}\theta \wedge \boldsymbol{d}\phi$ that describes the electromagnetic field of a charge e at rest at the origin. This picture is actually the intersection of ${}^*\boldsymbol{F}$ with a 3-surface of constant time t; i.e., the time direction is suppressed from the picture.

(b)

$$\bar{\varphi} = \arctan \frac{\bar{y}}{\bar{x}}; \qquad \boldsymbol{d}\bar{\varphi} = \frac{\bar{x}\, \boldsymbol{d}\bar{y} - \bar{y}\, \boldsymbol{d}\bar{x}}{\bar{x}^2 + \bar{y}^2};$$

(c)

$$\cos\bar{\theta} = \frac{\bar{z}}{\bar{r}}; \qquad -\boldsymbol{d}(\cos\bar{\theta}) = \frac{-\boldsymbol{d}\bar{z}}{\bar{r}} + \frac{\bar{z}}{\bar{r}^3}(\bar{x}\, \boldsymbol{d}\bar{x} + \bar{y}\, \boldsymbol{d}\bar{y} + \bar{z}\, \boldsymbol{d}\bar{z});$$

(d) combine to find

$${}^*\boldsymbol{F} = (e/\bar{r}^3)(\bar{x}\, \boldsymbol{d}\bar{y} \wedge \boldsymbol{d}\bar{z} + \bar{y}\, \boldsymbol{d}\bar{z} \wedge \boldsymbol{d}\bar{x} + \bar{z}\, \boldsymbol{d}\bar{x} \wedge \boldsymbol{d}\bar{y}) \qquad (4.18)$$

(electromagnetic field of point charge in a comoving Cartesian system; spherically symmetric). Now transform to laboratory coordinates:

velocity parameter α

velocity $\beta = \tanh\alpha$

$$\frac{1}{\sqrt{1-\beta^2}} = \cosh\alpha, \qquad \frac{\beta}{\sqrt{1-\beta^2}} = \sinh\alpha$$

(a)

$$\begin{cases} \bar{t} = t\cosh\alpha - x\sinh\alpha, \\ \bar{x} = -t\sinh\alpha + x\cosh\alpha, \\ \bar{y} = y \qquad \bar{z} = z; \end{cases}$$

(b)

$$\bar{r} = [(x\cosh\alpha - t\sinh\alpha)^2 + y^2 + z^2]^{1/2};$$

(c)

$${}^*\boldsymbol{F} = (e/\bar{r}^3)[(x\cosh\alpha - t\sinh\alpha)\, \boldsymbol{d}y \wedge \boldsymbol{d}z + y\, \boldsymbol{d}z \wedge (\cosh\alpha\, \boldsymbol{d}x - \sinh\alpha\, \boldsymbol{d}t) + z(\cosh\alpha\, \boldsymbol{d}x - \sinh\alpha\, \boldsymbol{d}t) \wedge \boldsymbol{d}y]; \qquad (4.19)$$

(d) compare with the general dual 2-form,

$$\begin{aligned} {}^*\boldsymbol{F} = {} & E_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z + E_y\, \boldsymbol{d}z \wedge \boldsymbol{d}x + E_z\, \boldsymbol{d}x \wedge \boldsymbol{d}y \\ & + B_x\, \boldsymbol{d}t \wedge \boldsymbol{d}x + B_y\, \boldsymbol{d}t \wedge \boldsymbol{d}y + B_z\, \boldsymbol{d}t \wedge \boldsymbol{d}z; \end{aligned}$$

and get the desired individual field components

(e)
$$\begin{cases} E_x = (e/\bar{r}^3)(x \cosh\alpha - t \sinh\alpha), & B_x = 0, \\ E_y = (e/\bar{r}^3)y \cosh\alpha, & B_y = -(e/\bar{r}^3)z \sinh\alpha, \\ E_z = (e/\bar{r}^3)z \cosh\alpha, & B_z = (e/\bar{r}^3)y \sinh\alpha. \end{cases} \tag{4.20}$$

One can verify that the invariants

$$\boldsymbol{B}^2 - \boldsymbol{E}^2 = \frac{1}{2} F_{\alpha\beta} F^{\alpha\beta}, \tag{4.21}$$

$$\boldsymbol{E} \cdot \boldsymbol{B} = \frac{1}{4} F_{\alpha\beta} {}^*F^{\alpha\beta} \tag{4.22}$$

have the same value in the laboratory frame as in the rocket frame, as required. Note that the honeycomb structure of the differential form is not changed when one goes from the rocket frame to the laboratory frame. What changes is only the mathematical formula that describes it.

§4.4. RADIATION FIELDS

The ***Maxwell*** structure of tubes associated with a charge in uniform motion is more remarkable than it may seem at first sight, and not only because of the Lorentz contraction of the tubes in the direction of motion. The tubes arbitrarily far away move on in military step with the charge on which they center, despite the fact that there is no time for information "emitted" from the charge "right now" to get to the faraway tube "right now." The structure of the faraway tubes "right now" must therefore derive from the charge at an earlier moment on its uniform-motion, straight-line trajectory. This circumstance shows up nowhere more clearly than in what happens to the field in consequence of a sudden change, in a short time $\Delta\tau$, from one uniform velocity to another uniform velocity (Figure 4.6). The tubes have the standard patterns for the two states of motion, one pattern within a sphere of radius r, the other outside that sphere, where r is equal to the lapse of time ("cm of light-travel time") since the acceleration took place. The necessity for the two patterns to fit together in the intervening zone, of thickness $\Delta r = \Delta\tau$, forces the field there to be multiplied up by a "stretching factor," proportional to r. This factor is responsible for the well-known fact that radiative forces fall off inversely only as the first power of the distance (Figure 4.6).

How an acceleration causes radiation

When the charge continuously changes its state of motion, the structure of the electromagnetic field, though based on the same simple principles as those illustrated in Figure 4.6, nevertheless looks more complex. The following is the ***Faraday*** 2-form

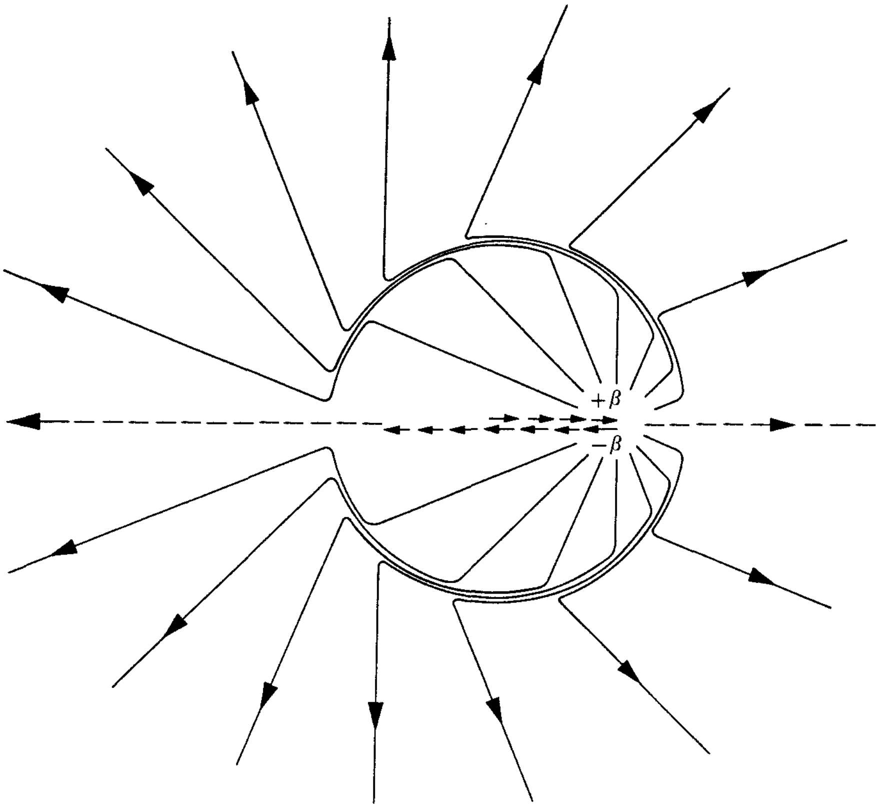


Figure 4.6.
Mechanism of radiation. J. J. Thomson's way to understand why the strength of an electromagnetic wave falls only as the inverse first power of distance r and why the amplitude of the wave varies (for low velocities) as $\sin\theta$ (maximum in the plane perpendicular to the line of acceleration). The charge was moving to the left at uniform velocity. Far away from it, the lines of force continue to move as if this uniform velocity were going to continue forever (Coulomb field of point-charge in slow motion). However, closer up the field is that of a point-change moving to the right with uniform velocity ($1/r^2$ dependence of strength upon distance). The change from the one field pattern to another is confined to a shell of thickness $\Delta\tau$ located at a distance r from the point of acceleration (amplification of field by "stretching factor" $r\sin\theta\,\Delta\beta/\Delta\tau$; see text). We thank C. Teitelboim for the construction of this diagram.

for the field of an electric dipole of magnitude p_1 oscillating up and down parallel to the z-axis:

Field of an oscillating dipole

$$\begin{aligned}\boldsymbol{F} = E_x\,\boldsymbol{d}x \wedge \boldsymbol{d}t + \cdots + B_x\,\boldsymbol{d}y \wedge \boldsymbol{d}z + \cdots = \text{real part of } \{p_1 e^{i\omega r - i\omega t} \\ \underbrace{[2\cos\theta\left(\frac{1}{r^3} - \frac{i\omega}{r^2}\right)}_{\text{gives } E_r}\boldsymbol{d}r \wedge \boldsymbol{d}t + \underbrace{\sin\theta\left(\frac{1}{r^3} - \frac{i\omega}{r^2} - \frac{\omega^2}{r}\right)}_{\text{gives } E_\theta} r\,\boldsymbol{d}\theta \wedge \boldsymbol{d}t \\ \underbrace{+ \sin\theta\left(\frac{-i\omega}{r^2} - \frac{\omega^2}{r}\right)}_{\text{gives } B_\phi}\boldsymbol{d}r \wedge r\,\boldsymbol{d}\theta]\}\end{aligned} \tag{4.23}$$

and the dual 2-form ***Maxwell*** = ****F*** is

$$\begin{aligned}
{}^*\boldsymbol{F} = -B_x\, \boldsymbol{dx} \wedge \boldsymbol{dt} - \cdots + E_x\, \boldsymbol{dy} \wedge \boldsymbol{dz} + \cdots = \text{real part of } \{p_1 e^{i\omega r - i\omega t} \\
\underbrace{[\sin\theta\left(\frac{-i\omega}{r^2} - \frac{\omega^2}{r}\right)}_{\text{gives } B_\phi} \boldsymbol{dt} \wedge r\sin\theta\, \boldsymbol{d\phi} \\
\underbrace{+\, 2\cos\theta\left(\frac{1}{r^3} - \frac{i\omega}{r^2}\right)}_{\text{gives } E_r} r\, \boldsymbol{d\theta} \wedge r\sin\theta\, \boldsymbol{d\phi} \\
\underbrace{+\, \sin\theta\left(\frac{1}{r^3} - \frac{i\omega}{r^2} - \frac{\omega^2}{r}\right)}_{\text{gives } E_\theta} r\sin\theta\, \boldsymbol{d\phi} \wedge \boldsymbol{dr}]\}.
\end{aligned} \tag{4.24}$$

§4.5. MAXWELL'S EQUATIONS

The general 2-form $\boldsymbol{F}$ is written as a superposition of wedge products with a factor $\frac{1}{2}$,

$$\boldsymbol{F} = \frac{1}{2} F_{\mu\nu}\, \boldsymbol{dx}^\mu \wedge \boldsymbol{dx}^\nu, \tag{4.25}$$

because the typical term appears twice, once as $F_{xy}\, \boldsymbol{dx} \wedge \boldsymbol{dy}$ and the second time as $F_{yx}\, \boldsymbol{dy} \wedge \boldsymbol{dx}$, with $F_{yx} = -F_{xy}$ and $\boldsymbol{dy} \wedge \boldsymbol{dx} = -\boldsymbol{dx} \wedge \boldsymbol{dy}$.

If differentiation ("taking the gradient"; the operator $\boldsymbol{d}$) produced out of a scalar a 1-form, it is also true that differentiation (again the operator $\boldsymbol{d}$, but now generally known under Cartan's name of "exterior differentiation") produces a 2-form out of the general 1-form; and applied to a 2-form produces a 3-form; and applied to a 3-form produces a 4-form, the form of the highest order that spacetime will accommodate. Write the general f-form as

$$\boldsymbol{\phi} = \frac{1}{f!} \phi_{\alpha_1\alpha_2\ldots\alpha_f}\, \boldsymbol{dx}^{\alpha_1} \wedge \boldsymbol{dx}^{\alpha_2} \wedge \cdots \wedge \boldsymbol{dx}^{\alpha_f} \tag{4.26}$$

Taking exterior derivative

where the coefficient $\phi_{\alpha_1\alpha_2\ldots\alpha_f}$, like the wedge product that follows it, is antisymmetric under interchange of any two indices. Then the exterior derivative of $\boldsymbol{\phi}$ is

$$\boldsymbol{d\phi} \equiv \frac{1}{f!} \frac{\partial \phi_{\alpha_1\alpha_2\ldots\alpha_f}}{\partial x^{\alpha_0}}\, \boldsymbol{dx}^{\alpha_0} \wedge \boldsymbol{dx}^{\alpha_1} \wedge \boldsymbol{dx}^{\alpha_2} \wedge \cdots \wedge \boldsymbol{dx}^{\alpha_f}. \tag{4.27}$$

Take the exterior derivative of ***Faraday*** according to this rule and find that it vanishes, not only for the special case of the dipole oscillator, but also for a general electromagnetic field. Thus, in the coordinates appropriate for a local Lorentz frame, one has

$$\mathbf{dF} = \mathbf{d}(E_x\, dx \wedge \mathbf{dt} + \cdots + B_x\, \mathbf{dy} \wedge \mathbf{dz} + \cdots)$$
$$= \left(\frac{\partial E_x}{\partial t}\mathbf{dt} + \frac{\partial E_x}{\partial x}\mathbf{dx} + \frac{\partial E_x}{\partial y}\mathbf{dy} + \frac{\partial E_x}{\partial z}\mathbf{dz}\right) \wedge \mathbf{dx} \wedge \mathbf{dt}$$
$$+ \cdots \text{(5 more such sets of 4 terms each)}\ldots. \tag{4.28}$$

Note that such a term as $\mathbf{dy} \wedge \mathbf{dy} \wedge \mathbf{dz}$ is automatically zero ("collapse of egg-crate cell when stamped on"). Collect the terms that do not vanish and find

$$\mathbf{dF} = \left(\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z}\right)\mathbf{dx} \wedge \mathbf{dy} \wedge \mathbf{dz}$$
$$+ \left(\frac{\partial B_x}{\partial t} + \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}\right)\mathbf{dt} \wedge \mathbf{dy} \wedge \mathbf{dz}$$
$$+ \left(\frac{\partial B_y}{\partial t} + \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}\right)\mathbf{dt} \wedge \mathbf{dz} \wedge \mathbf{dx}$$
$$+ \left(\frac{\partial B_z}{\partial t} + \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}\right)\mathbf{dt} \wedge \mathbf{dx} \wedge \mathbf{dy}. \tag{4.29}$$

Each term in this expression is familiar from Maxwell's equations

$$\operatorname{div} \boldsymbol{B} = \boldsymbol{\nabla} \cdot \boldsymbol{B} = 0$$

and

$$\operatorname{curl} E = \boldsymbol{\nabla} \times E = -\dot{\boldsymbol{B}}.$$

Each vanishes, and with their vanishing ***Faraday*** itself is seen to have zero exterior derivative:

$$\mathbf{dF} = 0. \tag{4.30}$$

In other words, "***Faraday*** is a closed 2-form"; "the tubes of $\boldsymbol{F}$ nowhere come to an end."

Faraday structure: tubes nowhere end

A similar calculation gives for the exterior derivative of the dual 2-form ***Maxwell*** the result

$$\mathbf{d}{}^*\mathbf{F} = \mathbf{d}(-B_x\, \mathbf{dx} \wedge \mathbf{dt} - \cdots + E_x\, \mathbf{dy} \wedge \mathbf{dz} + \cdots)$$
$$= \left(\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}\right)\mathbf{dx} \wedge \mathbf{dy} \wedge \mathbf{dz}$$
$$+ \left(\frac{\partial E_x}{\partial t} - \frac{\partial B_z}{\partial y} + \frac{\partial B_y}{\partial z}\right)\mathbf{dt} \wedge \mathbf{dy} \wedge \mathbf{dz}$$
$$+ \cdots$$
$$= 4\pi(\rho\, \mathbf{dx} \wedge \mathbf{dy} \wedge \mathbf{dz}$$
$$- J_x\, \mathbf{dt} \wedge \mathbf{dy} \wedge \mathbf{dz}$$
$$- J_y\, \mathbf{dt} \wedge \mathbf{dz} \wedge \mathbf{dx}$$
$$- J_z\, \mathbf{dt} \wedge \mathbf{dx} \wedge \mathbf{dy}) = 4\pi\, {}^*\mathbf{J};$$

$$\mathbf{d}{}^*\mathbf{F} = 4\pi\, {}^*\mathbf{J}. \tag{4.31}$$

Maxwell structure: density of tube endings given by charge-current 3-form

In empty space this exterior derivative, too, vanishes; there ***Maxwell*** is a closed 2-form; the tubes of $^*\boldsymbol{F}$, like the tubes of $\boldsymbol{F}$, nowhere come to an end.

In a region where charge is present, the situation changes. Tubes of ***Maxwell*** take their origin in such a region. The density of endings is described by the 3-form $^*\boldsymbol{J}$ = ***charge***, a "collection of eggcrate cells" collected along bundles of world lines.

The two equations

$$\boldsymbol{dF} = 0$$

and

$$\boldsymbol{d}^*\boldsymbol{F} = 4\pi\, ^*\boldsymbol{J}$$

summarize the entire content of Maxwell's equations in geometric language. The forms $\boldsymbol{F}$ = ***Faraday***, and $^*\boldsymbol{F}$ = ***Maxwell***, can be described in any coordinates one pleases—or in a language (honeycomb and egg-crate structures) free of any reference whatsoever to coordinates. Remarkably, neither equation makes any reference whatsoever to *metric*. As Hermann Weyl was one of the most emphatic in stressing (see also Chapters 8 and 9), the concepts of form and exterior derivative are metric-free. Metric made an appearance only in one place, in the concept of duality ("perpendicularity") that carried attention from $\boldsymbol{F}$ to the dual structure $^*\boldsymbol{F}$.

Duality: the only place in electromagnetism where metric must enter

§4.6. EXTERIOR DERIVATIVE AND CLOSED FORMS

The words "honeycomb" and "egg crate" may have given some feeling for the geometry that goes with electrodynamics. Now to spell out these concepts more clearly and illustrate in geometric terms, with electrodynamics as subject matter, what it means to speak of "exterior differentiation." Marching around a boundary, yes; but how and why and with what consequences? It is helpful to return to functions and 1-forms, and see them and the 2-forms ***Faraday*** and ***Maxwell*** and the 3-form ***charge*** as part of an ordered progression (see Box 4.4). Two-forms are seen in this box to be of two kinds: (1) a special 2-form, known as a "closed" 2-form, which has the property that as many tubes enter a closed 2-surface as emerge from it (exterior derivative of 2-form zero; no 3-form derivable from it other than the trivial *zero* 3-form!); and (2) a general 2-form, which sends across a closed 2-surface a non-zero net number of tubes, and therefore permits one to define a nontrivial 3-form ("exterior derivative of the 2-form"), which has precisely as many egg-crate cells in any closed 2-surface as the net number of tubes of the 2-form emerging from that same closed 2-surface (generalization of Faraday's concept of tubes of force to the world of spacetime, curved as well as flat).

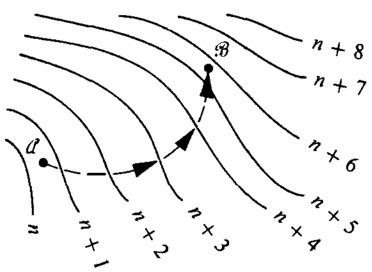
Closed 2-form contrasted with general 2-form

(continued on page 120)

Box 4.4 THE PROGRESSION OF FORMS AND EXTERIOR DERIVATIVES

0-Form or Scalar, f

An example in the context of 3-space and Newtonian physics is temperature, $T(x, y, z)$, and in the context of spacetime, a scalar potential, $\phi(t, x, y, z)$.

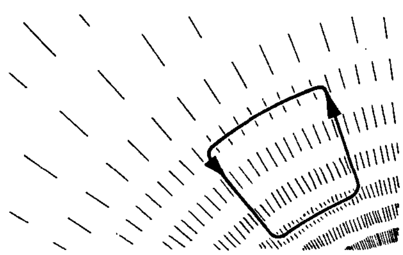


From Scalar to *1*-Form

Take the gradient or "exterior derivative" of a scalar f to obtain a special 1-form, $\boldsymbol{\gamma} = \boldsymbol{d}f$. Comments: (a) Any additive constant included in f is erased in the process of differentiation; the quantity n in the diagram at the left is unknown and irrelevant. (b) The 1-form $\boldsymbol{\gamma}$ is special in the sense that surfaces in one region "mesh" with surfaces in a neighboring region ("closed 1-form"). (c) Line integral $\int_{\mathcal{A}}^{\mathcal{B}} \boldsymbol{d}f$ is independent of path for any class of paths equivalent to one another under continuous deformation. (d) The 1-form is a machine to produce a number ("bongs of bell" as each successive integral surface is crossed) out of a displacement (approximation to concept of a tangent vector).

General *1*-Form $\boldsymbol{\beta} = \beta_\alpha \, \boldsymbol{d}x^\alpha$

This is a pattern of surfaces, as illustrated in the diagram at the right; i.e., a machine to produce a number ("bongs of bell"; $\langle \boldsymbol{\beta}, \boldsymbol{u} \rangle$) out of a vector. A 1-form has a reality and position in space independent of all choice of coordinates. Surfaces do not ordinarily mesh. Integral $\int \boldsymbol{\beta}$ around indicated closed loop does not give zero ("more bongs than antibongs").

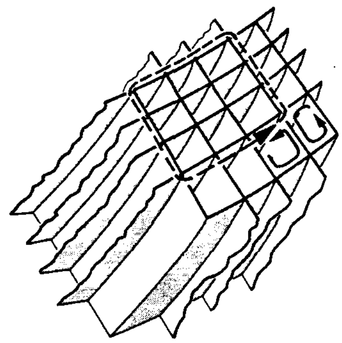


From *1*-Form to *2*-Form $\boldsymbol{\xi} = \boldsymbol{d\beta} = \dfrac{\partial \beta_\alpha}{\partial x^\mu} \boldsymbol{d}x^\mu \wedge \boldsymbol{d}x^\alpha$

$\boldsymbol{\xi}$ is a pattern of honeycomb-like cells, with a direction of circulation marked on each, so stationed

Box 4.4 (*continued*)

that the number of cells encompassed in the dotted closed path is identical to the net contribution (excess of bongs over antibongs) for the same path in the diagram of $\boldsymbol{\beta}$ above. The "exterior derivative" is *defined* so this shall be so; the generalized Stokes theorem codifies it. The word "exterior" comes from the fact that the path goes around the periphery of the region under analysis. Thus the 2-form is a machine to get a number (number of tubes, $\langle \boldsymbol{\xi}, \boldsymbol{u} \wedge \boldsymbol{v} \rangle$) out of a bit of surface ($\boldsymbol{u} \wedge \boldsymbol{v}$) that has a sense of circulation indicated upon it. The 2-form thus defined is special in this sense: a rubber sheet "supported around its edges" by the dotted curve or any other closed curve is crossed by the same number of tubes when: (a) it bulges up in the middle; (b) it is pushed down in the middle; (c) it experiences any other continuous deformation. The ***Faraday*** or 2-form $\boldsymbol{F}$ of electromagnetism, always expressible as $\boldsymbol{F} = \boldsymbol{dA}$ ($\boldsymbol{A}$ = 4-potential, a 1-form), also has always this special property ("conservation of tubes").

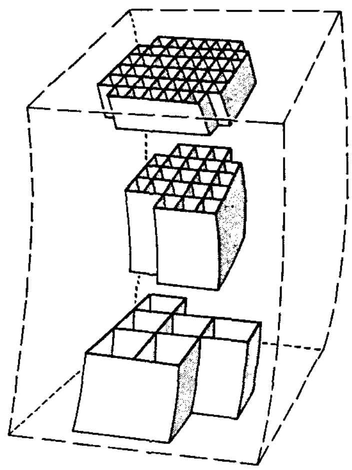


0-Form to *1*-Form to *2*-Form? No!

Go from scalar f to 1-form $\boldsymbol{\gamma} = \boldsymbol{d}f$. The next step to a 2-form $\boldsymbol{\alpha}$ is vacuous. The net contribution of the line integral $\int \boldsymbol{\gamma}$ around the dotted closed path is automatically zero. To reproduce that zero result requires a zero 2-form. Thus $\boldsymbol{\alpha} = \boldsymbol{d\gamma} = \boldsymbol{dd}f$ has to be the zero 2-form. This result is a special instance of the general result $\boldsymbol{dd} = 0$.

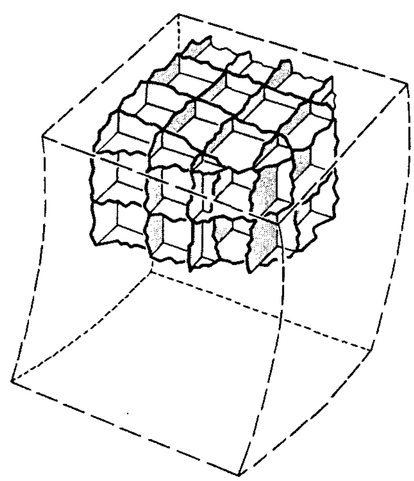
General *2*-Form $\boldsymbol{\sigma} = \frac{1}{2}\sigma_{\alpha\beta}\, \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta$, with $\sigma_{\alpha\beta} = -\sigma_{\beta\alpha}$

Again, this is a honeycomb-like structure, and again a machine to get a number (number of tubes, $\langle \boldsymbol{\sigma}, \boldsymbol{u} \wedge \boldsymbol{v} \rangle$) out of a surface ($\boldsymbol{u} \wedge \boldsymbol{v}$) that has a sense of circulation indicated on it. It is general in the sense that the honeycomb structures in one region do not ordinarily mesh with those in a neighboring region. In consequence, a closed 2-surface, such as the box-like surface indicated by dotted lines at the right, is ordinarily crossed by a non-zero net number of tubes. The net number of tubes emerging from such a closed surface is, however, exactly zero when the 2-form is the exterior derivative of a 1-form.



From 2-Form to 3-Form $\boldsymbol{\mu} = \boldsymbol{d\sigma} = \dfrac{\partial \sigma_{|\alpha\beta|}}{\partial x^\gamma} \boldsymbol{d}x^\gamma \wedge \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta$,
where $\boldsymbol{d}x^\gamma \wedge \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta \equiv 3!\, \boldsymbol{d}x^{[\gamma} \otimes \boldsymbol{d}x^\alpha \otimes \boldsymbol{d}x^{\beta]}$

This egg-crate type of structure is a machine to get a number (number of cells $\langle \boldsymbol{\mu}, \boldsymbol{u} \wedge \boldsymbol{v} \wedge \boldsymbol{w} \rangle$) from a volume (volume $\boldsymbol{u} \wedge \boldsymbol{v} \wedge \boldsymbol{w}$ within which one counts the cells). A more complete diagram would provide each cell and the volume of integration itself with an indicator of orientation (analogous to the arrow of circulation shown for cells of the 2-form). The contribution of a given cell to the count of cells is $+1$ or -1, according as the orientation indicators have same sense or opposite sense. The number of egg-crate cells of $\boldsymbol{\mu} = \boldsymbol{d\sigma}$ in any given volume (such as the volume indicated by the dotted lines) is tailored to give precisely the same number as the net number of tubes of the 2-form $\boldsymbol{\sigma}$ (diagram above) that emerge from that volume (generalized Stokes theorem). For electromagnetism, the exterior derivative of ***Faraday*** or 2-form $\boldsymbol{F}$ gives a null 3-form, but the exterior derivative of ***Maxwell*** or 2-form $^*\boldsymbol{F}$ gives 4π times the 3-form $^*\boldsymbol{J}$ of charge:



$$^*\boldsymbol{J} = \rho\, \boldsymbol{d}x \wedge \boldsymbol{d}y \wedge \boldsymbol{d}z - J_x\, \boldsymbol{d}t \wedge \boldsymbol{d}y \wedge \boldsymbol{d}z - J_y\, \boldsymbol{d}t \wedge \boldsymbol{d}z \wedge \boldsymbol{d}x - J_z\, \boldsymbol{d}t \wedge \boldsymbol{d}x \wedge \boldsymbol{d}y.$$

Box 4.4 (*continued*)

From *1*-Form to *2*-Form to *3*-Form? No!

Starting with a 1-form (electromagnetic 4-potential), construct its exterior derivative, the 2-form $\boldsymbol{F} = \boldsymbol{dA}$ (***Faraday***). The tubes in this honeycomb-like structure never end. So the number of tube endings in any elementary volume, and with it the 3-form $\boldsymbol{dF} = \boldsymbol{ddA}$, is automatically zero. This is another example of the general result that $\boldsymbol{dd} = 0$.

From *2*-Form to *3*-Form to *4*-Form? No!

Starting with 2-form $^*\boldsymbol{F}$ (***Maxwell***), construct its exterior derivative, the 3-form $4\pi\, ^*\boldsymbol{J}$. The cells in this egg-crate type of structure extend in a fourth dimension ("hypertube"). The number of these hypertubes that end in any elementary 4-volume, and with it the 4-form

$$\boldsymbol{d}(4\pi\, ^*\boldsymbol{J}) = \boldsymbol{dd}\,^*\boldsymbol{F},$$

is automatically zero, still another example of the general result that $\boldsymbol{dd} = 0$. This result says that

$$\boldsymbol{d}\,^*\boldsymbol{J} = \left(\frac{\partial \rho}{\partial t} + \frac{\partial J_x}{\partial x} + \frac{\partial J_y}{\partial y} + \frac{\partial J_z}{\partial z}\right) \boldsymbol{d}t \wedge \boldsymbol{d}x \wedge \boldsymbol{d}y \wedge \boldsymbol{d}z = 0$$

("law of conservation of charge"). Note:

$$\boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta \wedge \boldsymbol{d}x^\gamma \wedge \boldsymbol{d}x^\delta \equiv 4!\, \boldsymbol{d}x^{[\alpha} \otimes \boldsymbol{d}x^\beta \otimes \boldsymbol{d}x^\gamma \otimes \boldsymbol{d}x^{\delta]}.$$

This implies $\boldsymbol{d}t \wedge \boldsymbol{d}x \wedge \boldsymbol{d}y \wedge \boldsymbol{d}z = \boldsymbol{\varepsilon}$.

From *3*-Form to *4*-Form $\boldsymbol{\tau} = \boldsymbol{d\nu} = \dfrac{\partial \nu_{|\alpha\beta\gamma|}}{\partial x^\delta}\, \boldsymbol{d}x^\delta \wedge \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta \wedge \boldsymbol{d}x^\gamma$

This four-dimensional "super-egg-crate" type structure is a machine to get a number (number of cells, $\langle \boldsymbol{\tau}, \boldsymbol{n} \wedge \boldsymbol{u} \wedge \boldsymbol{v} \wedge \boldsymbol{w}\rangle$) from a 4-volume $\boldsymbol{n} \wedge \boldsymbol{u} \wedge \boldsymbol{v} \wedge \boldsymbol{w}$.

From *4*-Form to *5*-Form? No!

Spacetime, being four-dimensional, cannot accommodate five-dimensional egg-crate structures. At least two of the $\boldsymbol{d}x^\mu$'s in

$$\boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta \wedge \boldsymbol{d}x^\gamma \wedge \boldsymbol{d}x^\delta \wedge \boldsymbol{d}x^\epsilon$$

must be the same; so, by antisymmetry of "$\wedge$," this "basis 5-form" must vanish.

Results of Exterior Differentiation, Summarized

0-form	f				
1-form	$\boldsymbol{d}f$	$\boldsymbol{A}$			
2-form	$\boldsymbol{dd}f \equiv 0$	$\boldsymbol{F} = \boldsymbol{dA}$	$^*\boldsymbol{F}$		
3-form		$\boldsymbol{dF} = \boldsymbol{ddA} \equiv 0$	$4\pi\, ^*\boldsymbol{J} = \boldsymbol{d}\,^*\boldsymbol{F}$	$\boldsymbol{\nu}$	
4-form			$\boldsymbol{d}(4\pi\, ^*\boldsymbol{J}) = \boldsymbol{dd}\,^*\boldsymbol{F} \equiv 0$	$\boldsymbol{\tau} = \boldsymbol{d\nu}$	$\boldsymbol{\mu}$
5-form?	No!			$\boldsymbol{d\tau} \equiv 0$	$\boldsymbol{d\mu} \equiv 0$

New Forms from Old by Taking Dual (see exercise 3.14)

Dual of scalar f is 4-form: $^*f = f\,\boldsymbol{d}x^0 \wedge \boldsymbol{d}x^1 \wedge \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 = f\boldsymbol{\varepsilon}$.

Dual of 1-form $\boldsymbol{J}$ is 3-form: $^*\boldsymbol{J} = J^0\,\boldsymbol{d}x^1 \wedge \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 - J^1\,\boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 \wedge \boldsymbol{d}x^0 + J^2\,\boldsymbol{d}x^3 \wedge \boldsymbol{d}x^0 \wedge \boldsymbol{d}x^1 - J^3\,\boldsymbol{d}x^0 \wedge \boldsymbol{d}x^1 \wedge \boldsymbol{d}x^2$.

Dual of 2-form $\boldsymbol{F}$ is 2-form: $^*\boldsymbol{F} = F^{|\alpha\beta|}\varepsilon_{\alpha\beta|\mu\nu|}\,\boldsymbol{d}x^\mu \wedge \boldsymbol{d}x^\nu$, where $F^{\alpha\beta} = \eta^{\alpha\lambda}\eta^{\beta\delta}F_{\lambda\delta}$.

Dual of 3-form $\boldsymbol{K}$ is 1-form: $^*\boldsymbol{K} = K^{012}\,\boldsymbol{d}x^3 - K^{123}\,\boldsymbol{d}x^0 + K^{230}\,\boldsymbol{d}x^1 - K^{301}\,\boldsymbol{d}x^2$, where $K^{\alpha\beta\gamma} = \eta^{\alpha\mu}\eta^{\beta\nu}\eta^{\gamma\lambda}K_{\mu\nu\lambda}$.

Dual of 4-form $\boldsymbol{L}$ is a scalar: $\boldsymbol{L} = L_{0123}\,\boldsymbol{d}x^0 \wedge \boldsymbol{d}x^1 \wedge \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3$; $^*\boldsymbol{L} = L^{0123} = -L_{0123}$.

Note 1: This concept of duality between *one form and another* is to be distinguished from the concept of duality between the *vector basis* $\boldsymbol{e}_\alpha$ and the *1-form basis* $\boldsymbol{\omega}^\alpha$ of a given frame. The two types of duality have nothing whatsoever to do with each other!

Box 4.4 (*continued*)

Note 2: In spacetime, the operation of taking the dual, applied twice, leads back to the original form for forms of odd order, and to the negative thereof for forms of even order. In Euclidean 3-space the operation reproduces the original form, regardless of its order.

Duality Plus Exterior Differentiation

Start with scalar ϕ. Its gradient $\boldsymbol{d}\phi$ is a 1-form. Take its dual, to get the 3-form $^*\boldsymbol{d}\phi$. Take its exterior derivative, to get the 4-form $\boldsymbol{d}^*\boldsymbol{d}\phi$. Take its dual, to get the scalar $\Box\phi \equiv -^*\boldsymbol{d}^*\boldsymbol{d}\phi$. Verify by index manipulations that $\Box$ as defined here is the wave operator; i.e., in any Lorentz frame, $\Box\phi = \phi_{,\alpha}{}^{,\alpha} = -(\partial^2\phi/\partial t^2) + \nabla^2\phi$.

Start with 1-form $\boldsymbol{A}$. Get 2-form $\boldsymbol{F} = \boldsymbol{dA}$. Take its dual $^*\boldsymbol{F} = {}^*\boldsymbol{dA}$, also a 2-form. Take its exterior derivative, obtaining the 3-form $\boldsymbol{d}^*\boldsymbol{F}$ (has value $4\pi\, ^*\boldsymbol{J}$ in electromagnetism). Take its dual, obtaining the 1-form $^*\boldsymbol{d}^*\boldsymbol{F} = {}^*\boldsymbol{d}^*\boldsymbol{dA} = 4\pi\boldsymbol{J}$ ("Wave equation for electromagnetic 4-potential"). Reduce in index notation to

$$F_{\mu\nu}{}^{,\nu} = A_{\nu,\mu}{}^{,\nu} - A_{\mu,\nu}{}^{,\nu} = 4\pi J_\mu.$$

[More in Flanders (1963) or Misner and Wheeler (1957); see also exercise 3.17.]

§4.7. DISTANT ACTION FROM LOCAL LAW

Differential forms are a powerful tool in electromagnetic theory, but full power requires mastery of other tools as well. Action-at-a-distance techniques ("Green's functions," "propagators") are of special importance. Moreover, the passage from Maxwell field equations to electromagnetic action at a distance provides a preview of how Einstein's local equations will reproduce (approximately) Newton's $1/r^2$ law.

In flat spacetime and in a Lorentz coordinate system, express the coordinates of particle A as a function of its proper time α, thus:

$$a^\mu = a^\mu(\alpha), \qquad \frac{da^\mu}{d\alpha} = \dot{a}^\mu(\alpha), \qquad \frac{d^2a^\mu}{d\alpha^2} = \ddot{a}^\mu(\alpha). \tag{4.32}$$

Dirac found it helpful to express the distribution of charge and current for a particle of charge e following such a motion as a superposition of charges that momentarily flash into existence and then flash out of existence. Any such flash has a localization in space and time that can be written as the product of four Dirac delta functions [see, for example, Schwartz (1950–1951), Lighthill (1958)]:

World line of charge regarded as succession of flash-on, flash-off charges

$$\delta^4(x^\mu - a^\mu) = \delta[x^0 - a^0(\alpha)]\,\delta[x^1 - a^1(\alpha)]\,\delta[x^2 - a^2(\alpha)]\,\delta[x^3 - a^3(\alpha)]. \quad (4.33)$$

Here any single Dirac function $\delta(x)$ ("symbolic function"; "distribution"; "limit of a Gauss error function" as width is made indefinitely narrow and peak indefinitely high, with integrated value always unity) both (1) vanishes for $x \neq 0$, and (2) has the integral $\int_{-\infty}^{+\infty} \delta(x)\,dx = 1$. Described in these terms, the density-current vector for the particle has the value ("superposition of flashes")

$$J^\mu = e\int \delta^4[x^\nu - a^\nu(\alpha)]\dot{a}^\mu(\alpha)\,d\alpha. \quad (4.34)$$

The density-current (4.34) drives the electromagnetic field, $\boldsymbol{F}$. Write $\boldsymbol{F} = \boldsymbol{dA}$ to satisfy automatically half of Maxwell's equations ($\boldsymbol{dF} = \boldsymbol{ddA} \equiv 0$):

$$F_{\mu\alpha} = \frac{\partial A_\alpha}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\alpha}. \quad (4.35)$$

In flat space, the remainder of Maxwell's equations ($\boldsymbol{d}{}^*\boldsymbol{F} = 4\pi\,{}^*\boldsymbol{J}$) become

$$\frac{\partial F_\mu{}^\nu}{\partial x^\nu} = 4\pi J_\mu$$

or

$$\frac{\partial}{\partial x^\mu}\frac{\partial A^\nu}{\partial x^\nu} - \eta^{\nu\alpha}\frac{\partial^2 A_\mu}{\partial x^\nu\,\partial x^\alpha} = 4\pi J_\mu. \quad (4.36)$$

Make use of the freedom that exists in the choice of 4-potentials A^ν to demand

$$\frac{\partial A^\nu}{\partial x^\nu} = 0 \quad (4.37)$$

(Lorentz gauge condition; see exercise 3.17). Thus get

$$\Box A_\mu = -4\pi J_\mu. \quad (4.38)$$

The electromagnetic wave equation

The density-current being the superposition of "flashes," the effect ($\boldsymbol{A}$) of this density-current can be expressed as the superposition of the effects E of elementary flashes; thus

$$A^\mu(x) = \int E[x - a(\alpha)]\dot{a}^\mu(\alpha)\,d\alpha, \quad (4.39)$$

The solution of the wave equation

where the "elementary effect" E ("kernel"; "Green's function") satisfies the equation

$$\Box E(x) = -4\pi\,\delta^4(x). \quad (4.40)$$

One solution is the "half-advanced-plus-half-retarded potential,"

$$E(x) = \delta(\eta_{\alpha\beta}x^\alpha x^\beta). \quad (4.41)$$

It vanishes everywhere except on the backward and forward light cones, where it has equal strength. Normally more useful is the retarded solution,

$$R(x) = \begin{cases} 2E(x) & \text{if } x^0 > 0, \\ 0 & \text{if } x^0 < 0, \end{cases} \tag{4.42}$$

which is obtained by doubling (4.41) in the region of the forward light cone and nullifying it in the region of the backward light cone. All electrodynamics (Coulomb forces, Ampère's law, electromagnetic induction, radiation) follows from the simple expression (4.39) for the vector potential [see, e.g., Wheeler and Feynman (1945) and (1949), also Rohrlich (1965)].

EXERCISES

Exercise 4.1. GENERIC LOCAL ELECTROMAGNETIC FIELD EXPRESSED IN SIMPLEST FORM

In the laboratory Lorentz frame, the electric field is $\boldsymbol{E}$, the magnetic field $\boldsymbol{B}$. Special cases are: (1) pure electric field ($\boldsymbol{B} = 0$); (2) pure magnetic field ($\boldsymbol{E} = 0$); and (3) "radiation field" or "null field" ($\boldsymbol{E}$ and $\boldsymbol{B}$ equal in magnitude and perpendicular in direction). All cases other than (1), (2), and (3) are "generic." In the generic case, calculate the Poynting density of flow of energy $\boldsymbol{E} \times \boldsymbol{B}/4\pi$ and the density of energy $(E^2 + B^2)/8\pi$. Define the direction of a unit vector $\boldsymbol{n}$ and the magnitude of a velocity parameter α by the ratio of energy flow to energy density:

$$\boldsymbol{n} \tanh 2\alpha = \frac{2\boldsymbol{E} \times \boldsymbol{B}}{E^2 + B^2}.$$

View the same electromagnetic field in a rocket frame moving in the direction of $\boldsymbol{n}$ with the velocity parameter α (not 2α; factor 2 comes in because energy flow and energy density are components, not of a vector, but of a tensor). By employing the formulas for a Lorentz transformation (equation 3.23), or otherwise, show that the energy flux vanishes in the rocket frame, with the consequence that $\bar{\boldsymbol{E}}$ and $\bar{\boldsymbol{B}}$ are parallel. No one can prevent the $\bar{z}$-axis from being put in the direction common to $\bar{\boldsymbol{E}}$ and $\bar{\boldsymbol{B}}$. Show that with this choice of direction, ***Faraday*** becomes

$$\boldsymbol{F} = \bar{E}_z\, \boldsymbol{d}\bar{z} \wedge \boldsymbol{d}\bar{t} + \bar{B}_z\, \boldsymbol{d}\bar{x} \wedge \boldsymbol{d}\bar{y}$$

(only two wedge products needed to represent the generic local field; "canonical representation"; valid in one frame, valid in any frame).

Exercise 4.2. FREEDOM OF CHOICE OF 1-FORMS IN CANONICAL REPRESENTATION OF GENERIC LOCAL FIELD

Deal with a region so small that the variation of the field from place to place can be neglected. Write ***Faraday*** in canonical representation in the form

$$\boldsymbol{F} = \boldsymbol{d}p_I \wedge \boldsymbol{d}q^I + \boldsymbol{d}p_{II} \wedge \boldsymbol{d}q^{II},$$

where p_A ($A = I$ or II) and q^A are scalar functions of position in spacetime. Define a "canonical transformation" to new scalar functions of position $p_{\bar{A}}$ and $q^{\bar{A}}$ by way of the "equation of transformation"

$$p_A\, \boldsymbol{d}q^A = \boldsymbol{d}S + p_{\bar{A}}\, \boldsymbol{d}q^{\bar{A}},$$

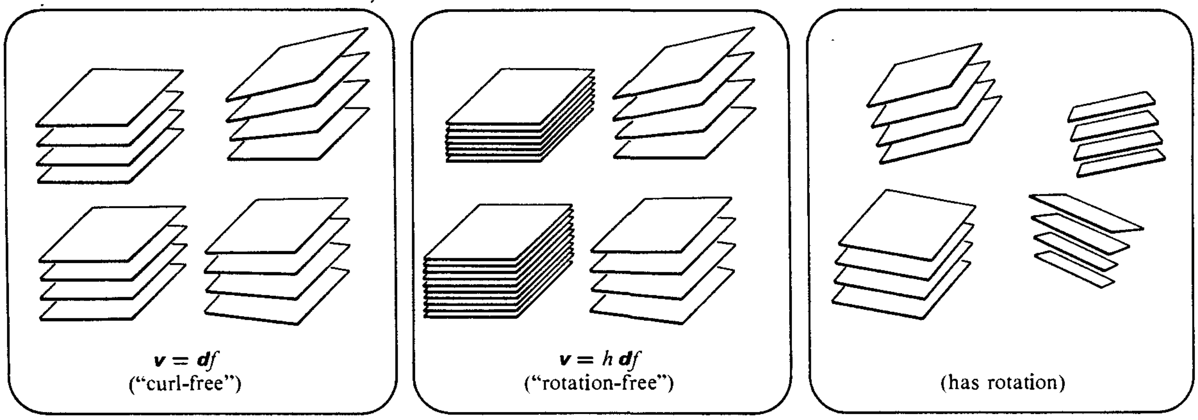


Figure 4.7.
Some simple types of 1-forms compared and contrasted.

where the "generating function" S of the transformation is an arbitrary function of the q^A and the $q^{\bar{A}}$:

$$dS = (\partial S/\partial q^A)\, dq^A + (\partial S/\partial q^{\bar{A}})\, dq^{\bar{A}}.$$

(a) Derive expressions for the two p_A's and the two $p_{\bar{A}}$'s in terms of S by equating coefficients of dq^I, dq^{II}, $dq^{\bar{I}}$. $dq^{\bar{II}}$ individually on the two sides of the equation of transformation.

(b) Use these expressions for the p_A's and $p_{\bar{A}}$'s to show that $\boldsymbol{F} = \boldsymbol{d}p_A \wedge \boldsymbol{d}q^A$ and $\bar{\boldsymbol{F}} = \boldsymbol{d}p_{\bar{A}} \wedge \boldsymbol{d}q^{\bar{A}}$, ostensibly different, are actually expressions for one and the same 2-form in terms of alternative sets of 1-forms.

Exercise 4.3. A CLOSED OR CURL-FREE 1-FORM IS A GRADIENT

Given a 1-form $\boldsymbol{\sigma}$ such that $\boldsymbol{d\sigma} = 0$, show that $\boldsymbol{\sigma}$ can be expressed in the form $\boldsymbol{\sigma} = \boldsymbol{d}f$, where f is some scalar. The 1-form $\boldsymbol{\sigma}$ is said to be "curl-free," a narrower category of 1-form than the "rotation-free" 1-form of the next exercise (expressible as $\boldsymbol{\sigma} = h\,\boldsymbol{d}f$), and it in turn is narrower (see Figure 4.7) than the category of "1-forms with rotation" (not expressible in the form $\boldsymbol{\sigma} = h\,\boldsymbol{d}f$). When the 1-form $\boldsymbol{\sigma}$ is expressed in terms of basis 1-forms $\boldsymbol{d}x^\alpha$, multiplied by corresponding components σ_α, show that "curl-free" implies $\sigma_{[\alpha,\beta]} = 0$.

Exercise 4.4. CANONICAL EXPRESSION FOR A ROTATION-FREE 1-FORM

In three dimensions a rigid body turning with angular velocity ω about the z-axis has components of velocity $v_y = \omega x$, and $v_x = -\omega y$. The quantity curl $\boldsymbol{v} = \boldsymbol{\nabla} \times \boldsymbol{v}$ has z-component equal to 2ω, and all other components equal to zero. Thus the scalar product of $\boldsymbol{v}$ and curl $\boldsymbol{v}$ vanishes:

$$v_{[i,j}v_{k]} = 0.$$

The same concept generalizes to four dimensions,

$$v_{[\alpha,\beta}v_{\gamma]} = 0.$$

and lends itself to expression in coordinate-free language, as the requirement that a certain 3-form must vanish:

$$\boldsymbol{dv} \wedge \boldsymbol{v} = 0.$$

Any 1-form $\boldsymbol{v}$ satisfying this condition is said to be "rotation-free." Show that a 1-form is rotation-free if and only if it can be written in the form

$$\boldsymbol{v} = h\,\boldsymbol{d}f,$$

where h and f are scalar functions of position (the "Frobenius theorem").

Exercise 4.5. FORMS ENDOWED WITH POLAR SINGULARITIES

List the principal results on how such forms are representable, such as

$$\boldsymbol{\Phi}_1 = \frac{\boldsymbol{d}S}{S} \wedge \boldsymbol{\psi}_1 + \boldsymbol{\theta}_1,$$

and the conditions under which each applies [for the meaning and answer to this exercise, see Lascoux (1968)].

Exercise 4.6. THE FIELD OF THE OSCILLATING DIPOLE

Verify that the expressions given for the electromagnetic field of an oscillating dipole in equations (4.23) and (4.24) satisfy $\boldsymbol{d}\boldsymbol{F} = 0$ everywhere and $\boldsymbol{d}{}^*\boldsymbol{F} = 0$ everywhere except at the origin.

Exercise 4.7. THE 2-FORM MACHINERY TRANSLATED INTO TENSOR MACHINERY

This exercise is stated at the end of the legend caption of Figure 4.1.

Exercise 4.8. PANCAKING THE COULOMB FIELD

Figure 4.5 shows a spacelike slice, $t =$ const, through the ***Maxwell*** of a point-charge at rest. By the following pictorial steps, verify that the electric-field lines get compressed into the transverse direction when viewed from a moving Lorentz frame: (1) Draw a picture of an equatorial slice ($\theta = \pi/2$; t, r, ϕ variable) through ***Maxwell*** $= {}^*\boldsymbol{F}$. (2) Draw various spacelike slices, corresponding to constant time in various Lorentz frames, through the resultant geometric structure. (3) Interpret the intersection of ***Maxwell*** $= {}^*\boldsymbol{F}$ with each Lorentz slice in the manner of Figure 4.3.

Exercise 4.9. COMPUTATION OF SURFACE INTEGRALS

In Box 4.1 the definition

$$\int \boldsymbol{\alpha} = \int \ldots \int \left\langle \boldsymbol{\alpha}, \frac{\partial \mathscr{P}}{\partial \lambda^1} \wedge \cdots \wedge \frac{\partial \mathscr{P}}{\partial \lambda^p} \right\rangle d\lambda^1 \ldots d\lambda^p$$

is given for the integral of a p-form $\boldsymbol{\alpha}$ over a p-surface $\mathscr{P}(\lambda^1, \ldots, \lambda^p)$ in n-dimensional space. From this show that the following computational rule (also given in Box 4.1) works: (1) substitute the equation for the surface,

$$x^k = x^k(\lambda^1, \ldots, \lambda^p),$$

into $\boldsymbol{\alpha}$ and collect terms in the form

$$\boldsymbol{\alpha} = a(\lambda^1, \ldots, \lambda^p)\,\boldsymbol{d}\lambda^1 \wedge \cdots \wedge \boldsymbol{d}\lambda^p;$$

(2) integrate

$$\int \boldsymbol{\alpha} = \int \ldots \int a(\lambda^1, \ldots, \lambda^p)\, d\lambda^1 \ldots d\lambda^p$$

using the elementary definition of integration.

Exercise 4.10. WHITAKER'S CALUMOID, OR, THE LIFE OF A LOOP

Take a closed loop, bounding a 2-dimensional surface $\mathcal{S}$. It entraps a certain flux of ***Faraday*** $\boldsymbol{\Phi}_F = \int_S \boldsymbol{F}$ ("magnetic tubes") and a certain flux of ***Maxwell*** $\boldsymbol{\Phi}_M = \int_S {}^*\boldsymbol{F}$ ("electric tubes").

(a) Show that the fluxes $\boldsymbol{\Phi}_F$ and $\boldsymbol{\Phi}_M$ depend only on the choice of loop, and not on the choice of the surface $\mathcal{S}$ bounded by the loop, if and only if $\boldsymbol{dF} = \boldsymbol{d}{}^*\boldsymbol{F} = 0$ (no magnetic charge; no electric charge). *Hint:* use generalized Stokes theorem, Boxes 4.1 and 4.6.

(b) Move the loop in space and time so that it continues to entrap the same two fluxes. Move it forward a little more here, a little less there, so that it continues to do so. In this way trace out a 2-dimensional surface ("calumoid"; see E. T. Whitaker 1904) $\mathcal{P} = \mathcal{P}(a, b)$; $x^\mu = x^\mu(a, b)$. Show that the elementary bivector in this surface, $\boldsymbol{\Sigma} = \partial\mathcal{P}/\partial a \wedge \partial\mathcal{P}/\partial b$ satisfies $\langle \boldsymbol{F}, \boldsymbol{\Sigma} \rangle = 0$ and $\langle {}^*\boldsymbol{F}, \boldsymbol{\Sigma} \rangle = 0$.

(c) Show that these differential equations for $x^\mu(a, b)$ can possess a solution, with given initial condition $x^\mu = x^\mu(a, 0)$ for the initial location of the loop, if $\boldsymbol{dF} = 0$ and $\boldsymbol{d}{}^*\boldsymbol{F} = 0$ (no magnetic charge, no electric charge).

(d) Consider a static, uniform electric field $\boldsymbol{F} = -E_x \boldsymbol{dt} \wedge \boldsymbol{dx}$. Solve the equations, $\langle \boldsymbol{F}, \boldsymbol{\Sigma} \rangle = 0$ and $\langle {}^*\boldsymbol{F}, \boldsymbol{\Sigma} \rangle = 0$ to find the equation $\mathcal{P}(a, b)$ for the most general calumoid. [*Answer:* $y = y(a)$, $z = z(a)$, $x = x(b)$, $t = t(b)$.] Exhibit two special cases: (i) a calumoid that lies entirely in a hypersurface of constant time [loop moves at infinite velocity; analogous to super-light velocity of point of crossing for two blades of a pair of scissors]; (ii) a calumoid whose loop remains forever at rest in the t, x, y, z Lorentz frame.

Exercise 4.11. DIFFERENTIAL FORMS AND HAMILTONIAN MECHANICS

Consider a dynamic system endowed with two degrees of freedom. For the definition of this system as a Hamiltonian system (special case: here the Hamiltonian is independent of time), one needs (1) a definition of canonical variables (see Box 4.5) and (2) a knowledge of the Hamiltonian H as a function of the coordinates q^1, q^2 and the canonically conjugate momenta p_1, p_2. To derive the laws of mechanics, consider the five-dimensional space of p_1, p_2, q^1, q^2, and t, and a curve in this space leading from starting values of the five coordinates (subscript A) to final values (subscript B), and the value

$$I = \int_A^B p_1\, \boldsymbol{d}q^1 + p_2\, \boldsymbol{d}q^2 - H(p, q)\, \boldsymbol{dt} = \int_A^B \boldsymbol{\omega}$$

of the integral I taken along this path. The difference of the integral for two "neighboring" paths enclosing a two-dimensional region $\mathcal{S}$, according to the theorem of Stokes (Boxes 4.1 and 4.6), is

$$\delta I = \oint_{\partial S} \boldsymbol{\omega} = \int_S \boldsymbol{d\omega}.$$

The principle of least action (principle of "extremal history") states that the representative point of the system must travel along a route in the five-dimensional manifold (route with tangent vector $d\mathcal{P}/dt$) such that the variation vanishes for this path; i.e.,

$$\boldsymbol{d\omega}(\ldots, d\mathcal{P}/dt) = 0$$

(2-form $\boldsymbol{d\omega}$ with a single vector argument supplied, and other slot left unfilled, gives the 1-form in 5-space that must vanish). This fixes only the direction of $d\mathcal{P}/dt$; its magnitude can be normalized by requiring $\langle \boldsymbol{dt}, d\mathcal{P}/dt \rangle = 1$.

(a) Evaluate $\boldsymbol{d\omega}$ from the expression $\boldsymbol{\omega} = p_j\, \boldsymbol{d}q^j - H\, \boldsymbol{dt}$.

(b) Set $d\mathcal{P}/dt = \dot{q}^j(\partial\mathcal{P}/\partial q^j) + \dot{p}_j(\partial\mathcal{P}/\partial p_j) + \dot{t}(\partial\mathcal{P}/\partial t)$, and expand $\boldsymbol{d\omega}(\ldots, d\mathcal{P}/dt) = 0$ in terms of the basis $\{\boldsymbol{d}p_j, \boldsymbol{d}q^k, \boldsymbol{dt}\}$.

Box 4.5 METRIC STRUCTURE AND HAMILTONIAN OR "SYMPLECTIC STRUCTURE" COMPARED AND CONTRASTED

	Metric structure	*Symplectic structure*
1. Physical application	Geometry of spacetime	Hamiltonian mechanics
2. Canonical structure	$(\ldots\cdot\ldots) =$ "$\boldsymbol{ds}^2$" $= -\boldsymbol{dt}\otimes\boldsymbol{dt} + \boldsymbol{dx}\otimes\boldsymbol{dx} + \boldsymbol{dy}\otimes\boldsymbol{dy} + \boldsymbol{dz}\otimes\boldsymbol{dz}$	$\boldsymbol{\Theta} = \boldsymbol{d}p_1 \wedge \boldsymbol{d}q^1 + \boldsymbol{d}p_2 \wedge \boldsymbol{d}q^2$
3. Nature of "metric"	Symmetric	Antisymmetric
4. Name for given coordinate system and any other set of four coordinates in which metric has same form	Lorentz coordinate system	System of "canonically" (or "dynamically") conjugate coordinates
5. Field equation for this metric	$R_{\mu\nu\alpha\beta} = 0$ (zero Riemann curvature; flat spacetime)	$\boldsymbol{d\Theta} = 0$ ("closed 2-form"; condition automatically satisfied by expression above).
6. The four-dimensional manifold	Spacetime	Phase space
7. Coordinate-free description of the structure of this manifold	***Riemann*** $= 0$	$\boldsymbol{d\Theta} = 0$
8. Canonical coordinates distinguished from other coordinates (allowable but less simple)	Make metric take above form (item 2)	Make metric take above form (item 2)

(c) Show that this five-dimensional equation can be written in the 4-dimensional phase space of $\{q^j, p_k\}$ as

$$\boldsymbol{\Theta}(\ldots, d\mathscr{P}/dt) = \boldsymbol{d}H,$$

where $\boldsymbol{\Theta}$ is the 2-form defined in Box 4.5.

(d) Show that the components of $\boldsymbol{\Theta}(\ldots, d\mathscr{P}/dt) = \boldsymbol{d}H$ in the $\{q^j, p_k\}$ coordinate system are the familiar Hamilton equations. Note that this conclusion depends only on the form assumed for $\boldsymbol{\Theta}$, so that one also obtains the standard Hamilton equations in any other phase-space coordinates $\{\bar{q}^j, \bar{p}_k\}$ ("canonical variables") for which

$$\boldsymbol{\Theta} = \boldsymbol{d}\bar{p}_1 \wedge \boldsymbol{d}\bar{q}^1 + \boldsymbol{d}\bar{p}_2 \wedge \boldsymbol{d}\bar{q}^2.$$

Exercise 4.12. SYMMETRY OPERATIONS AS TENSORS

We define the meaning of square and round brackets enclosing a set of indices as follows:

$$V_{(\alpha_1\ldots\alpha_p)} \equiv \frac{1}{p!}\Sigma V_{\alpha_{\pi_1}\ldots\alpha_{\pi_p}}, \qquad V_{[\alpha_1\ldots\alpha_p]} \equiv \frac{1}{p!}\Sigma(-1)^{\pi} V_{\alpha_{\pi_1}\ldots\alpha_{\pi_p}}.$$

Box 4.6 BIRTH OF STOKES' THEOREM

Central to the mathematical formulation of electromagnetism are the theorems of Gauss (taken up in Chapter 5) and Stokes. Both today appear together as one unity when expressed in the language of forms. In earlier times the unity was not evident. Everitt (1970) recalls the history of Stokes' theorem: "The Smith's Prize paper set by [G. C.] Stokes [Lucasian Professor of Mathematics] and taken by Maxwell in [February] 1854 . . .

> 5. Given the centre and two points of an ellipse, and the length of the major axis, find its direction by a geometrical construction.
>
> 6. Integrate the differential equation
>
> $$(a^2 - x^2)\,dy^2 + 2xydydx + (a^2 - y^2)\,dx^2 = 0.$$
>
> Has it a singular solution?
>
> 7. In a double system of curves of double curvature, a tangent is always drawn at the variable point P; shew that, as P moves away from an arbitrary fixed point Q, it must begin to move along a generating line of an elliptic cone having Q for vertex in order that consecutive tangents may ultimately intersect, but that the conditions of the problem may be impossible.
>
> 8. If X, Y, Z be functions of the rectangular co-ordinates x, y, z, dS an element of any limited surface, l, m, n the cosines of the inclinations of the normal at dS to the axes, ds an element of the bounding line, shew that
>
> $$\iint \left\{ l\left(\frac{dZ}{dy} - \frac{dY}{dz}\right) + m\left(\frac{dX}{dz} - \frac{dZ}{dx}\right) + n\left(\frac{dY}{dx} - \frac{dX}{dy}\right)\right\} dS$$
> $$= \int \left(X\frac{dx}{ds} + Y\frac{dy}{ds} + Z\frac{dz}{ds}\right) ds,$$
>
> the differential coefficients of X, Y, Z being partial, and the single integral being taken all round the perimeter of the surface

marks the first appearance in print of the formula connecting line and surface integrals now known as Stokes' theorem. This was of great importance to Maxwell's development of electromagnetic theory. The earliest explicit proof of the theorem appears to be that given in a letter from Thomson to Stokes dated July 2, 1850." [Quoted in Campbell and Garnett (1882), pp. 186–187.]

Here the sum is taken over all permutations π of the numbers 1, 2, . . . , p, and $(-1)^\pi$ is $+1$ or -1 depending on whether the permutation is even or odd. The quantity V may have other indices, not shown here, besides the set of p indices $\alpha_1, \alpha_2, \ldots, \alpha_p$, but only this set of indices is affected by the operations described here. The numbers $\pi_1, \pi_2, \ldots, \pi_p$ are the numbers 1, 2, . . . , p rearranged according to the permutation π. (Cases $p = 2, 3$ were treated in exercise 3.12.) We therefore have machinery to convert any rank-p tensor with components $V_{\alpha_1 \ldots \alpha_p}$ into a new tensor with components

$$[\boldsymbol{Alt}(\boldsymbol{V})]_{\mu_1 \ldots \mu_p} = V_{[\mu_1 \ldots \mu_p]}.$$

Since this machinery $\boldsymbol{Alt}$ is linear, it can be viewed as a tensor which, given suitable arguments $\boldsymbol{u}, \boldsymbol{v}, \ldots, \boldsymbol{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \ldots, \boldsymbol{\gamma}$ produces a number

$$u^\mu v^\nu \ldots w^\lambda \alpha_{[\mu} \beta_\nu \ldots \gamma_{\lambda]}.$$

(a) Show that the components of this tensor are

$$(\boldsymbol{Alt})_{\beta_1 \ldots \beta_p}{}^{\alpha_1 \ldots \alpha_p} = (p!)^{-1}\, \delta^{\alpha_1 \ldots \alpha_p}_{\beta_1 \ldots \beta_p}$$

(Note: indices of δ are almost never raised or lowered, so this notation leads to no confusion.)

where

$$\delta^{\alpha_1 \ldots \alpha_p}_{\beta_1 \ldots \beta_p} = \begin{cases} +1 \text{ if } (\alpha_1, \ldots, \alpha_p) \text{ is an even permutation of } (\beta_1, \ldots, \beta_p), \\ -1 \text{ if } (\alpha_1, \ldots, \alpha_p) \text{ is an odd permutation of } (\beta_1, \ldots, \beta_p), \\ \;\;0 \text{ if (i) any two of the } \alpha\text{'s are the same}, \\ \;\;0 \text{ if (ii) any two of the } \beta\text{'s are the same}, \\ \;\;0 \text{ if (iii) the } \alpha\text{'s and } \beta\text{'s are different sets of integers.} \end{cases}$$

Note that the demonstration, and therefore these component values, are correct in any frame.

(b) Show for any "alternating" (i.e., "*completely antisymmetric*") tensor $A_{\alpha_1 \ldots \alpha_p} = A_{[\alpha_1 \ldots \alpha_p]}$ that

$$\begin{aligned} \frac{1}{p!} A_{\alpha_1 \ldots \alpha_p} \delta^{\alpha_1 \ldots \alpha_p \beta_1 \ldots \beta_q}_{\gamma_1 \ldots \gamma_p \gamma_{p+1} \ldots \gamma_{p+q}} &= \sum_{\alpha_1 < \alpha_2 < \ldots < \alpha_p} A_{\alpha_1 \ldots \alpha_p} \delta^{\alpha_1 \ldots \alpha_p \beta_1 \ldots \beta_q}_{\gamma_1 \ldots \ldots \ldots \gamma_{p+q}} \\ &\equiv A_{|\alpha_1 \ldots \alpha_p|} \delta^{\alpha_1 \ldots \alpha_p \beta_1 \ldots \beta_q}_{\gamma_1 \ldots \ldots \ldots \gamma_{p+q}}. \end{aligned}$$

The final line here introduces the convention that a summation over indices enclosed between vertical bars includes only terms with those indices in increasing order. Show, consequently or similarly, that

$$\delta^{\alpha_1 \ldots \alpha_p \beta_1 \ldots \beta_q}_{\gamma_1 \ldots \ldots \ldots \gamma_{p+q}} \delta^{\mu_1 \ldots \mu_q}_{|\beta_1 \ldots \beta_q|} = \delta^{\alpha_1 \ldots \alpha_p \mu_1 \ldots \mu_q}_{\gamma_1 \ldots \ldots \ldots \gamma_{p+q}}.$$

(c) *Define* the exterior ("wedge") product of any two alternating tensors by

$$(\boldsymbol{\alpha} \wedge \boldsymbol{\beta})_{\lambda_1 \ldots \lambda_{p+q}} = \delta^{\mu_1 \ldots \mu_p\, \nu_1 \ldots \nu_q}_{\lambda_1 \ldots \lambda_p \lambda_{p+1} \ldots \lambda_{p+q}} \alpha_{|\mu_1 \ldots \mu_p|} \beta_{|\nu_1 \ldots \nu_q|};$$

and similarly

$$(\boldsymbol{U} \wedge \boldsymbol{V})^{\lambda_1 \ldots \lambda_{p+q}} = \delta^{\lambda_1 \ldots \lambda_p \lambda_{p+1} \ldots \lambda_{p+q}}_{\mu_1 \ldots \mu_p \nu_1 \ldots \nu_q} U^{|\mu_1 \ldots \mu_p|} V^{|\nu_1 \ldots \nu_q|}.$$

Show that this implies equation (3.45b). Establish the associative law for this product rule by showing that

$$
\begin{aligned}
&[(\boldsymbol{\alpha} \wedge \boldsymbol{\beta}) \wedge \boldsymbol{\gamma}]_{\sigma_1 \ldots \sigma_{p+q+r}} \\
&= \delta^{\lambda_1 \ldots \lambda_p \mu_1 \ldots \mu_q \nu_1 \ldots \nu_r}_{\sigma_1 \ldots \sigma_{p+q+r}} \alpha_{|\lambda_1 \ldots \lambda_p|} \beta_{|\mu_1 \ldots \mu_q|} \gamma_{|\nu_1 \ldots \nu_r|} \\
&= [\boldsymbol{\alpha} \wedge (\boldsymbol{\beta} \wedge \boldsymbol{\gamma})]_{\sigma_1 \ldots \sigma_{p+q+r}};
\end{aligned}
$$

and show that this reduces to the 3-form version of Equation (3.45c) when $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ are all 1-forms.

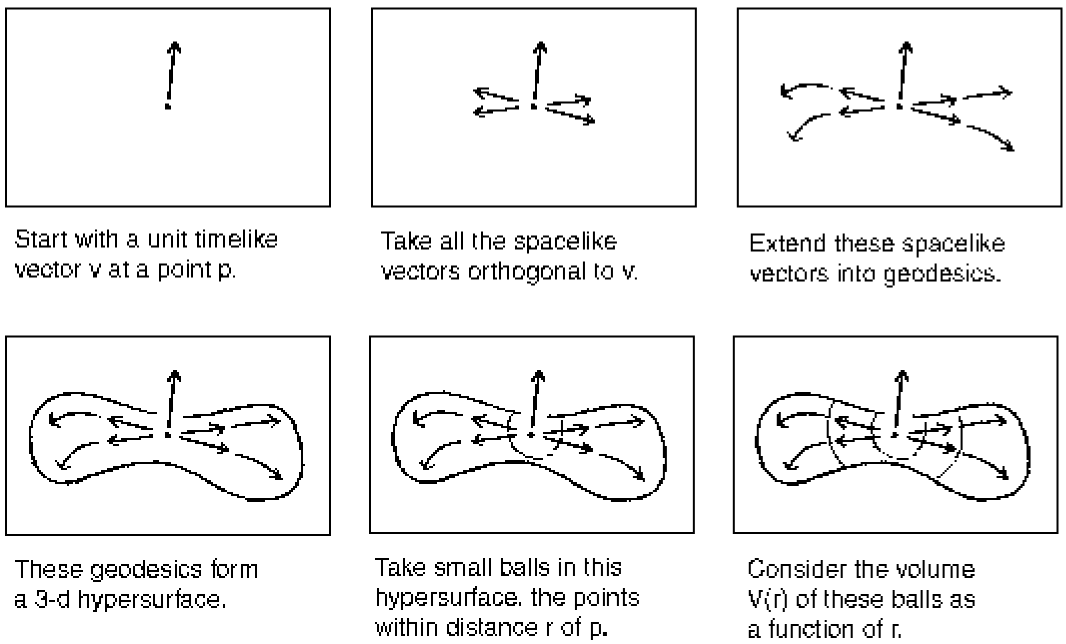
(d) Derive the following formula for the components of the exterior product of p vectors

$$
\begin{aligned}
(\boldsymbol{u}_1 \wedge \boldsymbol{u}_2 \wedge \cdots \wedge \boldsymbol{u}_p)^{\alpha_1 \ldots \alpha_p} &= \delta^{\alpha_1 \ldots \alpha_p}_{\mu \ldots \nu} (u_1)^\mu \ldots (u_p)^\nu \\
&= p! u_1{}^{[\alpha_1} u_2{}^{\alpha_2} \ldots u_p{}^{\alpha_p]} \\
&= \delta^{\alpha_1 \alpha_2 \ldots \alpha_p}_{1\ 2\ \ldots\ p} \det [(u_\mu)^\lambda].
\end{aligned}
$$

Einstein's Equation in Pictures

Matthew Frank

March 15, 2002



Einstein's equation says that energy is the curvature of space. What does this mean? In terms of the above pictures, it can be expressed as:

$$16\pi\,(\text{energy density as measured by } v) = \lim_{r\to 0}\frac{15}{r^2}\left(1-\frac{V(r)}{\frac{4}{3}\pi r^3}\right)$$

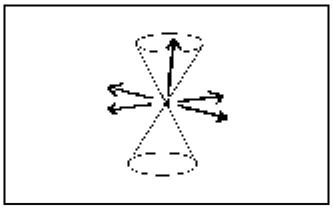
for each unit timelike vector v. (And that's it!)

How To Use This Paper

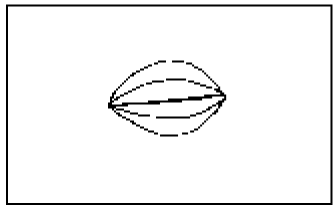
I hope that both people new to and people familiar with general relativity will read this paper. The section on preliminaries is intended primarily for those new to general relativity; I hope that physical intuition will carry people most of the way through that section, but mathematicians may find it useful to know that the four-dimensional space-time metric has signature -+++. The section on comparisons with other formulations and the appendices are intended primarily for those already familiar with general relativity; I hope that these people will appreciate the novelties of this approach.

Preliminaries

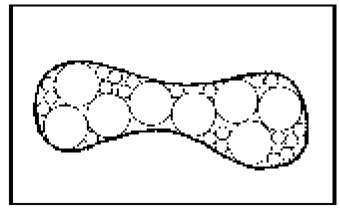
The background to this is that general relativity treats space-time as a four-dimensional manifold with a metric and an energy tensor. Since space-times are four-dimensional, space-time diagrams conventionally omit one dimension.



Timelike vectors are the ones that go inside a light cone; spacelike vectors, outside.



A geodesic is a path whose length is insensitive to small perturbations.



Volumes can be estimated by filling them with small, nearly Euclidean spheres.

It is the metric which makes the pictures and the right hand side of Einstein's equation meaningful. The metric determines (among other things) the possible paths of light rays, and these form a light cone. The vectors which point outside the light cone are called spacelike; the vectors on the light cone are called null. The vectors which point inside the light cone are called timelike; they represent the possible directions of inertial observers, and are often identified with observers. (This is treated in more detail in chapter 6 of Taylor and Wheeler's Spacetime Physics.)

The metric also determines distances and lengths of curves. A geodesic is a locally straightest path, a path γ_0 such that for any continuous variation of curves γ_u, the derivative (d/du)(length of γ_u) vanishes at u=0. Note that any vector can be followed into a geodesic path, as indicated in the third picture for Einstein's equation.

With the notion of distance comes a notion of volume: there is a unique volume function such that, for all sufficiently small r and all points p, $(4/3)\pi(r^3-r^4) < V(B(p,r)) < (4/3)\pi(r^3+r^4)$, where B(p,r) is the ball of radius r around p. (This is proved in the appendix.) In fact these volumes do not deviate from Euclidean volumes at the r^4 level at all, and the r^5 deviation from Euclidean volume (as on the right hand side of Einstein's equation) measures the curvature of the manifold at the point p. However, these r^4 inequalities are strict enough to give a procedure for estimating volumes to within an arbitrarily small factor of $1\pm\varepsilon$: simply divide the region into countably many balls of radius at most ε and sum their Euclidean volumes.

It is the energy tensor which makes the left-hand side of the equation meaningful. This concept of "energy density in the direction of v" or "local energy as measured by the observer v" is best illustrated by examples. In a vacuum, the energy in any direction is always 0. An electromagnetic field has total energy $(E \cdot E - B \cdot B)/8\pi$ where E and B are the vectors for the electric and magnetic fields as measured by the observer v. A perfect fluid is observed by v to have energy $(\rho + P)(u \cdot v)^2 - P$, where ρ is the density of the fluid, P is its pressure, and u is its direction.

Having this at hand makes it possible to go through an example of Einstein's equation in detail; this may in particular help clarify the dimensionality of the vector spaces and manifolds. Consider Minkowsi space (the space-time of special relativity), coordinatized as (x,y,z,t). Also consider an observer at the origin moving in the t-direction. For this observer, the spacelike vectors are all the vectors which have no t-component. The geodesics form the 3-d hypersurface t=0. The ball of radius r is all points (x,y,z,0) with $x^2 + y^2 + z^2 < r^2$, and it has volume exactly $(4/3)\pi r^3$. Hence the curvature of the hypersurface is 0, as it should be since Minkowski space is a vacuum.

Comparisons with the usual statement of Einstein's equation

The statement of Einstein's equation here makes precise "energy is the curvature of space", while the usual statement makes precise "energy-momentum is the curvature of space-time". The right hand side of the statement here gives the scalar curvature of the indicated hypersurface, and that hypersurface is the natural "space" associated to the given observer.

It is possible to do calculations that are guided by these pictures rather than by the standard Christoffel symbols or differential forms. The key is to coordinatize several things and represent them as power series in the distance from the initial point: first the geodesics from that point, then the metric of the resulting hypersurface, then the volumes of the geodesic balls in it. The curvature is proportional to the r^5 term in the power series for the volume. Even if the metric is only C^3 and not analytic, there are enough meaningful terms in these power series to allow this calculation of the curvature. I have used these techniques to rederive the Schwarzschild and Robertson-Walker solutions in this format. Unfortunately, the calculations by this method require calculations much longer than the usual ones, even when all of them are automated in Mathematica.

This statement of Einstein's equation is equivalent to the usual one. (Indeed, it is very close to the statement of Einstein's equation given by Misner, Thorne, and Wheeler on p. 515.) Unfortunately, while the usual statement is clean, and these pictures are clean, proving the equivalence of the two is somewhat messy. The following is a sketch of a proof in three steps, using geometrized units c=G=1.

- First, the usual statement $8\pi T_{ab} = G_{ab}$ is equivalent to the claim that $8\pi T_{ab}\, v^a\, v^b = G_{ab}\, v^a\, v^b$ for all unit timelike vectors v (where $T_{ab}\, v^a\, v^b$ is what is referred to above as the energy density in the inertial frame of v). This equivalence is a matter of linear algebra, using the facts that T and G are symmetric tensors and that the unit timelike vectors from a spanning set for the space of all tangent vectors.
- Second, $G_{ab}\, v^a\, v^b$ is half the scalar curvature of the indicated hypersurface; this is a special case of the Gauss-Codazzi equations without the terms for extrinsic curvature, because the extrinsic curvature of the hypersurface vanishes at p. These

equations are discussed in Wald, sec. 10.2, and the appendix proves the vanishing of the extrinsic curvature can also be proved using the machinery of that section.
- Third, the scalar curvature of the hypersurface is given by the limit of $(15/r^2)(1-V(r)/V_{Euc}(r))$ as r goes to 0; for this, see Cartan, sec. 234.

Conclusion

Two advantages of this presentation of Einstein's equation may be obvious:

- It is very pictorial.
- It requires much less of the standard mathematical apparatus: no curvature tensors (almost no tensors at all), and no parallel transport / derivative operators / affine connections.

Let me also call attention to a few advantages which may not be obvious.

- It may be easier to appraise the standard presentation of Einstein's equation given another presentation as different as this one; the standard presentation may seem less geometrically compelling but computationally not so bad by comparison. (For another alternative presentation, see Baez.)
- This presentation brings the geometry of general relativity closer to the ideal of a synthetic differential geometry set out by Herbert Busemann. (That was some of my inspiration for this project.)
- Most optimistically, this presentation of Einstein's equation (or slight variants) may be meaningful in physical theories which do not treat space-time as a 4-dimensional Lorentzian manifold.

In any case, I will be happy if this helps people to understand Einstein's equation, or gives pleasure to those who already do.

Bibliography and Acknowledgements

- John Baez (2001), The Meaning of Einstein's Equation, gr-qc/0103044
- Herbert Busemann (1955), Geometry of Geodesics.
- Elie Cartan (1928), Lecons sur la Geometrie des Espaces de Riemann; translated by James Glazebrook (1985) as Geometry of Riemannian Spaces.
- Charles Misner, Kip Thorne, and John Wheeler (1973), Gravitation.
- Edwin Taylor and John Wheeler (1992), Spacetime Physics.
- Robert Wald (1984), General Relativity.
- my website, http://zaphod.uchicago.edu/~mfrank/GR/einpics.html, for further appendices with the Mathematica code refered to above.

Thanks to: several people at the University of Chicago and at the 1999 Notre Dame conference on history and foundations of general relativity, an October 1999 audience at the Chicago-area philosophy of physics group, John Beem, and Stephen Wolfram.

Appendix on Riemannian volumes

Let B(p,r) be the ball of radius r around p. Then the precise claim is:

For any compact Riemannian manifold, the usual volume is the unique countably additive measure such that for all sufficiently small r and all points p,
$(4/3)\pi(r^3-r^4) < V(B(p,r)) < (4/3)\pi(r^3-r^4)$.

Proof of existence: By the result of Cartan, the limit as r approaches 0 of
$(15/r^2)(1 - V(B(p,r))/(4/3)\pi r^3)$ is equal to the curvature of the manifold at p;
in particular it exists and is finite. Hence $f(p,r)=[V(B(p,r)) - (4/3)\pi r^3] / [(4/3)\pi r^4]$ is a continuous function of p and r which is 0 when r is 0. Since the manifold is compact, there is some δ such that for all r<δ, |f(p,r)|<1; this yields the claimed inequality.

Proof of uniqueness: Say V and V' are both volume functions satisfying the above inequalities. Then, for any S and any ε, we can cover S by countably many balls of radius at most min(δ,ε) with non-overlapping interiors. The boundaries of these balls will be of measure 0 according to both V and V'. Hence V(S) is the sum of V of the balls, and likewise for V'. For each ball B, V(B) and V'(B) both differ from Euclidean volume by within a factor of 1±ε, so V(B) and V'(B) differ from each other by within a factor of $(1\pm\varepsilon)^2$. Hence V(S) and V'(S) differ from each other by within a factor of $(1\pm\varepsilon)^2$ for each ε, and so V(S)=V'(S). QED.

Appendix on the curvature of the hypersurface

This uses abstract index notation and several results from Wald. Let *v* be the original timelike vector at *P*, and let ***H*** be the corresponding spacelike hypersurface. Let n^a be a vector field (including *v*) of unit normals to ***H***, and extend it beyond ***H*** in such a way that its integral curves are unit geodesics; this is useful in defining the extrinsic curvature K_{ab}.

Wald defines K_{ab} as $\nabla_a n_b$. At *P*, $K_{ab}=0$ since its contraction with any bivector $y^a z^b$ is 0.
Proof: $n^a \nabla_a n_b=0$ because the integral curves of n^a are geodesics; $n^b \nabla_a n_b = \nabla_a(n^b n_b)/2 = 0$ because n^a is of unit length. So it suffices to consider spatial vectors y^a and z^b. For any vector w^a orthogonal to n^a, there is a geodesic vector field including w^a tangent to ***H***; and for any two such vector fields, their Lie bracket is also tangent to ***H***. Hence
$[y^b, z^b]\, n_b=0$; using the orthogonality of y^b and z^b with n_b, this may be rewritten as
$y^{[a} z^{b]} \nabla_a n_b= 0$. So it suffices to consider symmetric bivectors $y^{(a} z^{b)}$, which may in turn be reduced to those of the form $w^a w^b$. For these also, $w^a w^b \nabla_a n_b = w^a \nabla_a (w^b n_b) = 0$, where the first equality is because w^a is geodesic and the second because $w^b n_b=0$. QED.

Now $h_{ab}=g_{ab} + n_a n_b$ is the metric on ***H***, and $h^a{}_b h^{bc} = h^{ac}$.
At *P,* Wald 10.2.23 may be written without the terms for *K* as:
${}^{(3)}R_{abcd} = h_a{}^f h_b{}^g h_c{}^k h_d{}^j R_{fgkj}$. Contracting both sides with $h^{ac} h^{bd}$ we get
${}^{(3)}R = h^{fk} h^{gj} R_{fgkj}$, which Wald 10.2.29 shows equal to $2\, G_{ac}\, n^a n^c$.
In other words we have ${}^{(3)}R/2 = G_{ac}\, n^a n^c$, as claimed in the text of the paper.

The meaning of Einstein's equation

John C. Baez[a)]
Department of Mathematics, University of California, Riverside, California 92521

Emory F. Bunn[b)]
Physics Department, University of Richmond, Richmond, Virginia 23173

(Received 20 August 2004; accepted 3 December 2004)

This is a brief introduction to general relativity, designed for both students and teachers of the subject. While there are many excellent expositions of general relativity, few adequately explain the geometrical meaning of the basic equation of the theory: Einstein's equation. Here we give a simple formulation of this equation in terms of the motion of freely falling test particles. We also sketch some of the consequences of this formulation and explain how it is equivalent to the usual one in terms of tensors. Finally, we include an annotated bibliography of books, articles, and websites suitable for the student of relativity.
[DOI: 10.1119/1.1852541]

I. INTRODUCTION

General relativity explains gravity as the curvature of space–time. It's all about geometry. The basic equation of general relativity is called Einstein's equation. In units where $c=8\pi G=1$, it says

$$G_{\alpha\beta}=T_{\alpha\beta}. \tag{1}$$

It looks simple, but what does it mean? Unfortunately, the beautiful geometrical meaning of this equation is a bit hard to find in most treatments of relativity. There are many nice popularizations that explain the philosophy behind relativity and the idea of curved space–time, but most of them don't get around to explaining Einstein's equation and showing how to work out its consequences. There are also more technical introductions which explain Einstein's equation in detail—but here the geometry is often hidden under piles of tensor calculus.

This is a pity, because there is an easy way to express the whole content of Einstein's equation in plain English. After a suitable prelude, one can summarize it in a single sentence! One needs a lot of mathematics to derive all the consequences of this sentence, but we can work out *some* of its consequences quite easily.

In what follows, we start by outlining some differences between special and general relativity. Next we give a verbal formulation of Einstein's equation. Then we derive a few of its consequences concerning tidal forces, gravitational waves, gravitational collapse, and the big bang cosmology. In an appendix we explain why our verbal formulation is equivalent to the usual one in terms of tensors. This article is mainly aimed at those who teach relativity, but except for an appendix, we have tried to make it accessible to students. We conclude with a bibliography of sources to help teach the subject.

II. PRELIMINARIES

Before stating Einstein's equation, we need a little preparation. We assume the reader is somewhat familiar with special relativity—otherwise general relativity will be too hard. But there are some big differences between special and general relativity, which can cause immense confusion if neglected.

In special relativity, we cannot talk about *absolute* velocities, but only *relative* velocities. For example, we cannot sensibly ask if a particle is at rest, only whether it is at rest relative to another particle. The reason is that in this theory, velocities are described as vectors in four-dimensional space–time. Switching to a different inertial coordinate system can change which way these vectors point relative to our coordinate axes, but not whether two of them point the same way.

In general relativity, we cannot even talk about *relative* velocities, except for two particles at the same point of space–time—that is, at the same place at the same instant. The reason is that in general relativity, we take very seriously the notion that a vector is a little arrow sitting at a particular point in space–time. To compare vectors at different points of space–time, we must carry one over to the other. The process of carrying a vector along a path without turning or stretching it is called "parallel transport." When space–time is curved, the result of parallel transport from one point to another depends on the path taken, which is a direct consequence of a curved space–time. Thus it is ambiguous to ask whether two particles have the same velocity vector unless they are at the same point of space–time.

It is hard to imagine the curvature of four-dimensional space–time, but it is easy to see it on a two-dimensional surface, like a sphere. The sphere fits nicely in three-dimensional flat Euclidean space, so we can visualize vectors on the sphere as "tangent vectors." If we parallel transport a tangent vector from the north pole to the equator by going straight down a meridian, we get a different result than if we go down another meridian and then along the equator as shown in Fig. 1.

Because of the analogy to vectors on the surface of a sphere, in general relativity vectors are usually called "tangent vectors." However, it is important not to take this analogy too seriously. Our curved space–time need not be embedded in some higher-dimensional flat space–time for us to understand its curvature, or the concept of a tangent vector. The mathematics of tensor calculus is designed to let us handle these concepts "intrinsically"—i.e., working solely

[a)]Electronic mail: baez@math.ucr.edu
[b)]Electronic mail: ebunn@richmond.edu

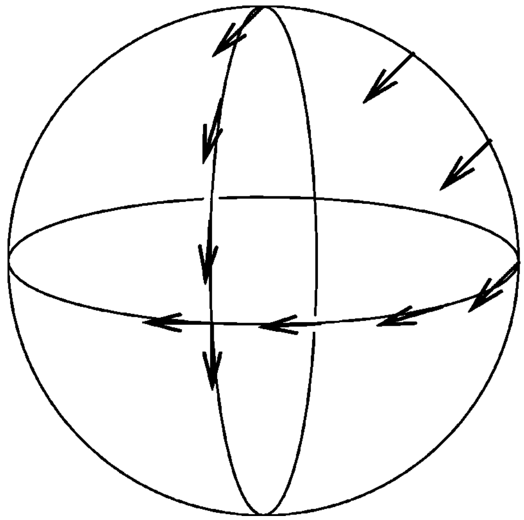


Fig. 1. Two ways to parallel transport a tangent vector from the north pole to a point on the equator of a sphere.

within the four-dimensional space–time in which we find ourselves. This is one reason tensor calculus is so important in general relativity.

In special relativity we can think of an inertial coordinate system, or ''inertial frame,'' as being defined by a field of clocks, all at rest relative to each other. In general relativity this makes no sense, since we can only unambiguously define the relative velocity of two clocks if they are at the same location. Thus the concept of inertial frame, so important in special relativity, is *banned* from general relativity!

If we are willing to put up with limited accuracy, we can still talk about the relative velocity of two particles in the limit where they are very close, since curvature effects will then be very small. In this approximate sense, we can talk about a ''local'' inertial coordinate system. However, we must remember that this notion makes perfect sense only in the limit where the region of space–time covered by the coordinate system goes to zero in size.

Einstein's equation can be expressed as a statement about the relative acceleration of very close test particles in free fall. Let us clarify these terms a bit. A ''test particle'' is an idealized point particle with energy and momentum so small that its effects on space–time curvature are negligible. A particle is said to be in ''free fall'' when its motion is affected by no forces except gravity. In general relativity, a test particle in free fall will trace out a ''geodesic.'' This means that its velocity vector is parallel transported along the curve it traces out in space–time. A geodesic is the closest thing there is to a straight line in curved space–time.

This is easier to visualize in two-dimensional space rather than four-dimensional space–time. A person walking on a sphere ''following their nose'' will trace out a geodesic—that is, a great circle. Suppose two people stand side-by-side on the equator and start walking north, both following geodesics. Though they start out walking parallel to each other, the distance between them will gradually start to shrink, until finally they bump into each other at the north pole. If they didn't understand the curved geometry of the sphere, they might think a ''force'' was pulling them together.

In general relativity gravity is not really a ''force,'' but just a manifestation of the curvature of space–time. Note it is not the curvature of space, but of *space–time* that is involved. The distinction is crucial. If you toss a ball, it follows a parabolic path. This is far from being a geodesic in *space*. Space is curved by the Earth's gravitational field, but it is certainly not so curved as all that! The point is that while the ball moves a short distance in space, it moves an enormous distance in *time*, because one second equals about 300 000 km in units where $c=1$. Thus, a slight amount of space–time curvature can have a noticeable effect.

III. EINSTEIN'S EQUATION

To state Einstein's equation in simple English, we need to consider a round ball of test particles that are all initially at rest relative to each other. As we have seen, this is a sensible notion only in the limit where the ball is very small. If we start with such a ball of particles, it will, to second order in time, become an ellipsoid as time passes. This should not be too surprising, because any linear transformation applied to a ball gives an ellipsoid, and any transformation can be approximated by a linear one to first order. Here we get a bit more: the relative velocity of the particles starts out being zero, so to first order in time the ball does not change shape at all: the change is a second-order effect.

Let $V(t)$ be the volume of the ball after a proper time t has elapsed, as measured by the particle at the center of the ball. Then Einstein's equation says:

$$\left.\frac{\ddot{V}}{V}\right|_{t=0} = -\frac{1}{2}\begin{pmatrix} \text{flow of } t\text{-momentum in } t \text{ direction} + \\ \text{flow of } x\text{-momentum in } x \text{ direction} + \\ \text{flow of } y\text{-momentum in } y \text{ direction} + \\ \text{flow of } z\text{-momentum in } z \text{ direction} \end{pmatrix} \tag{2}$$

where these flows are measured at the center of the ball at time zero, using local inertial coordinates. These flows are caused by all particles and fields. They form the diagonal components of a 4×4 matrix T called the ''stress-energy tensor.'' The components $T_{\alpha\beta}$ of this matrix say how much momentum in the α direction is flowing in the β direction through a given point of space–time, where $\alpha,\beta=t,x,y,z$. The flow of t-momentum in the t-direction is just the energy density, often denoted ρ. The flow of x-momentum in the x-direction is the ''pressure in the x direction'' denoted P_x, and similarly for y and z. It takes a while to figure out why pressure is really the flow of momentum, but it is eminently worth doing. Most texts explain this fact by considering the example of an ideal gas.

In any event, we may summarize Einstein's equation as follows:

$$\left.\frac{\ddot{V}}{V}\right|_{t=0} = -\frac{1}{2}(\rho + P_x + P_y + P_z). \tag{3}$$

This equation says that positive energy density and positive pressure curve space–time in a way that makes a freely falling ball of point particles tend to shrink. Since $E=mc^2$ and we are working in units where $c=1$, ordinary mass density counts as a form of energy density. Thus a massive object will make a swarm of freely falling particles at rest around it start to shrink. In short: *gravity attracts*.

We promised to state Einstein's equation in plain English, but have not done so yet. Here it is:

Given a small ball of freely falling test particles initially at rest with respect to each other, the rate at which it begins to shrink is proportional to its volume times: the energy density at the center of the ball, plus the pressure in the x direction at that point, plus the pressure in the y direction, plus the pressure in the z direction.

One way to prove this is by using the Raychaudhuri equation, discussions of which can be found in the textbooks by Wald[17] and by Ciufolini and Wheeler[25] cited in the bibliography. But an elementary proof can also be given starting from first principles, as we show in the Appendix.

The reader who already knows some general relativity may be somewhat skeptical that all of Einstein's equation is encapsulated in this formulation. After all, Einstein's equation in its usual tensorial form is really a bunch of equations: the left and right sides of Eq. (1) are 4×4 matrices. It is hard to believe that the single Eq. (3) captures all that information. It does, though, as long as we include one bit of fine print: to get the full content of the Einstein equation from Eq. (3), we must consider small balls with *all possible* initial velocities—i.e., balls that begin at rest in all possible local inertial reference frames.

Before we begin, it is worth noting an even simpler formulation of Einstein's equation that applies when the pressure happens to be the same in every direction:

> Given a small ball of freely falling test particles initially at rest with respect to each other, the rate at which it begins to shrink is proportional to its volume times: the energy density at the center of the ball plus three times the pressure at that point.

This version is only sufficient for ''isotropic'' situations: that is, those in which all directions look the same in some local inertial reference frame. But, since the simplest models of cosmology treat the universe as isotropic—at least approximately, on large enough distance scales—this is all we shall need to derive an equation describing the big bang!

IV. SOME CONSEQUENCES

The formulation of Einstein's equation we have given is certainly not the best for most applications of general relativity. For example, in 1915 Einstein used general relativity to correctly compute the anomalous precession of the orbit of Mercury and also the deflection of starlight by the Sun's gravitational field. Both these calculations would be very hard starting from Eq. (3); they really call for the full apparatus of tensor calculus. However, we can easily use our formulation of Einstein's equation to get a qualitative—and sometimes even quantitative—understanding of *some* consequences of general relativity. We have already seen that it explains how gravity attracts. We sketch a few other consequences below.

A. Tidal forces, gravitational waves

Let $V(t)$ be the volume of a small ball of test particles in free fall that are initially at rest relative to each other. In the vacuum there is no energy density or pressure, so $\ddot{V}|_{t=0}=0$, but the curvature of space–time can still distort the ball. For example, suppose you drop a small ball of instant coffee when making coffee in the morning. The grains of coffee closer to the earth accelerate toward it a bit more, causing the ball to start stretching in the vertical direction. However, as the grains all accelerate toward the center of the earth, the ball also starts being squashed in the two horizontal directions. Einstein's equation says that if we treat the coffee grains as test particles, these two effects cancel each other when we calculate the second derivative of the ball's volume, leaving us with $\ddot{V}|_{t=0}=0$. It is a fun exercise to check this using Newton's theory of gravity!

This stretching/squashing of a ball of falling coffee grains is an example of what people call ''tidal forces.'' As the name suggests, another example is the tendency for the ocean to be stretched in one direction and squashed in the other two by the gravitational pull of the moon.

Gravitational waves are another example of how space–time can be curved even in the vacuum. General relativity predicts that when any heavy object wiggles, it sends out ripples of space–time curvature which propagate at the speed of light. This is far from obvious starting from our formulation of Einstein's equation! It also predicts that as one of these ripples of curvature passes by, our small ball of initially test particles will be stretched in one transverse direction while being squashed in the other transverse direction. From what we have already said, these effects must precisely cancel when we compute $\ddot{V}|_{t=0}$.

Hulse and Taylor won the Nobel prize in 1993 for careful observations of a binary neutron star which is slowly spiraling down, just as general relativity predicts it should, as it loses energy by emitting gravitational radiation.[27,28] Gravitational waves have not been *directly* observed, but there are a number of projects under way to detect them.[29–32] For example, the LIGO project will bounce a laser between hanging mirrors in an L-shaped detector, to see how one leg of the detector is stretched while the other is squashed. Both legs are 4 km long, and the detector is designed to be sensitive to a 10^{-18} m change in length of the arms.

B. Gravitational collapse

One remarkable feature of this equation is the pressure term, which says that not only energy density but also pressure causes gravitational attraction. This may seem to violate our intuition that pressure makes matter want to expand! Here, however, we are talking about *gravitational* effects of pressure, which are undetectably small in everyday circumstances. To see this, let's restore the factors of c and G. Also, let's remember that in ordinary circumstances most of the energy is in the form of rest energy, so we can write the energy density ρ as $\rho_m c^2$, where ρ_m is the ordinary mass density:

$$\left.\frac{\ddot{V}}{V}\right|_{t=0} = -\frac{4\pi G}{c^4}(\rho_m c^2 + P_x + P_y + P_z). \tag{4}$$

On the human scale all of the terms on the right are small, since G is very small and c is very big. (Gravity is a weak force!) Furthermore, the pressure terms are much smaller than the mass density term, since the former has an extra c^2.

There are a number of important situations in which ρ does not dominate over P. For example, in a neutron star, which is held up by the degeneracy pressure of the neutronium it consists of, pressure and energy density contribute comparably to the right-hand side of Einstein's equation. Moreover, above a mass of about two solar masses a nonrotating neutron star will inevitably collapse to form a black hole, thanks in part to the gravitational attraction caused by pressure.

C. The big bang

Starting from our formulation of Einstein's equation, we can derive some basic facts about the big bang cosmology. Let us assume the universe is not only expanding but also homogeneous and isotropic. The expansion of the universe is vouched for by the redshifts of distant galaxies. The other assumptions also seem to be approximately correct, at least when we average over small-scale inhomogeneities such as stars and galaxies. For simplicity, we will imagine the universe is homogeneous and isotropic even on small scales.

An observer at any point in such a universe would see all objects receding from her. Suppose that, at some time $t=0$, she identifies a small ball B of test particles centered on her. Suppose this ball expands with the universe, remaining spherical as time passes because the universe is isotropic. Let $R(t)$ stand for the radius of this ball as a function of time. The Einstein equation will give us an equation of motion for $R(t)$. In other words, it will say how the expansion rate of the universe changes with time.

It is tempting to apply Eq. (3) to the ball B, but we must take care. This equation applies to a ball of particles that are initially at rest relative to one another—that is, one whose radius is not changing at $t=0$. However, the ball B is expanding at $t=0$. Thus, to apply our formulation of Einstein's equation, we must introduce a second small ball of test particles that are at rest relative to each other at $t=0$.

Let us call this second ball B', and call its radius as a function of time $r(t)$. Since the particles in this ball begin at rest relative to one another, we have

$$\dot{r}(0)=0. \tag{5}$$

To keep things simple, let us also assume that at $t=0$ both balls have the exact same size:

$$r(0)=R(0). \tag{6}$$

Equation (3) applies to the ball B', since the particles in *this* ball are initially at rest relative to each other. Since the volume of this ball is proportional to r^3, and using Eq. (5), the left-hand side of Eq. (3) becomes simply

$$\left.\frac{\ddot{V}}{V}\right|_{t=0}=\left.\frac{3\ddot{r}}{r}\right|_{t=0}. \tag{7}$$

Since we are assuming the universe is isotropic, we know that the various components of pressure are equal: $P_x=P_y=P_z=P$. Einstein's equation, Eq. (3), thus says that

$$\left.\frac{3\ddot{r}}{r}\right|_{t=0}=-\frac{1}{2}(\rho+3P). \tag{8}$$

We would much prefer to rewrite this expression in terms of R rather than r. Fortunately, we can do this. At $t=0$, the two spheres have the same radius: $r(0)=R(0)$. Furthermore, the second derivatives are the same: $\ddot{r}(0)=\ddot{R}(0)$. This follows from the equivalence principle, which says that, at any given location, particles in free fall do not accelerate with respect to each other. At the moment $t=0$, each test particle on the surface of the ball B is right next to a corresponding test particle in B'. Since they are not accelerating with respect to each other, the observer at the origin must see both particles accelerating in the same way, so $\ddot{r}(0)=\ddot{R}(0)$. It follows that we can replace r with R in the above equation, obtaining

$$\left.\frac{3\ddot{R}}{R}\right|_{t=0}=-\frac{1}{2}(\rho+3P). \tag{9}$$

We derived this equation for a very small ball, but in fact it applies to a ball of any size. This is because, in a homogeneous expanding universe, the balls of all radii must be expanding at the same fractional rate. In other words, $\ddot{R}/R$ is independent of the radius R, although it can depend on time. Also, there is nothing special in this equation about the moment $t=0$, so the equation must apply at all times. In summary, therefore, the basic equation describing the big bang cosmology[36–41] is

$$\frac{3\ddot{R}}{R}=-\frac{1}{2}(\rho+3P), \tag{10}$$

where the density ρ and pressure P can depend on time but not on position. Here we can imagine R to be the separation between any two "galaxies."

To go further, we must make more assumptions about the nature of the matter filling the universe. One simple model is a universe filled with pressureless matter. Until recently, this was thought to be an accurate model of our universe. Setting $P=0$, we obtain

$$\frac{3\ddot{R}}{R}=-\frac{\rho}{2}. \tag{11}$$

If the energy density of the universe is mainly due to the mass in galaxies, "conservation of galaxies" implies that $\rho R^3=k$ for some constant k. This gives

$$\frac{3\ddot{R}}{R}=-\frac{k}{2R^3} \tag{12}$$

or

$$\ddot{R}=-\frac{k}{6R^2}. \tag{13}$$

Amusingly, this is the same as the equation of motion for a particle in an attractive $1/R^2$ force field. In other words, the equation governing this simplified cosmology is the same as the Newtonian equation for what happens when you throw a ball vertically upwards from the earth! This is a nice example of the unity of physics. Since "whatever goes up must come down—unless it exceeds escape velocity," the solutions of this equation look roughly like those shown in Fig. 2.

In other words, the universe started out with a big bang! It will expand forever if its current rate of expansion is sufficiently high compared to its current density, but it will recollapse in a "big crunch" otherwise.

D. The cosmological constant

The simplified big bang model just described is inaccurate for the very early history of the universe, when the pressure of radiation was important. Moreover, recent observations seem to indicate that it is seriously inaccurate even in the present epoch. First of all, it seems that much of the energy density is not accounted for by known forms of matter. Still more shocking, it seems that the expansion of the universe may be accelerating rather than slowing down! One possibility is that the energy density and pressure are nonzero even for the vacuum. For the vacuum to not pick out a preferred

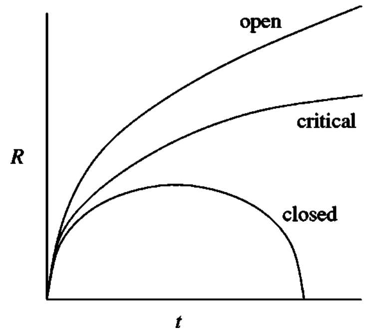


Fig. 2. The size of the universe as a function of time in three scenarios: open (where it expands forever), closed (where it recollapses), and critical (where it expands forever, but just barely).

notion of "rest," its stress-energy tensor must be proportional to the metric. In local inertial coordinates this means that the stress-energy tensor of the vacuum must be

$$T=\begin{pmatrix} \Lambda & 0 & 0 & 0 \\ 0 & -\Lambda & 0 & 0 \\ 0 & 0 & -\Lambda & 0 \\ 0 & 0 & 0 & -\Lambda \end{pmatrix}, \tag{14}$$

where Λ is called the "cosmological constant." This amounts to giving empty space an energy density equal to Λ and pressure equal to $-\Lambda$, so that $\rho+3P$ for the vacuum is -2Λ. Here pressure effects dominate because there are more dimensions of space than of time! If we add this cosmological constant term to Eq. (10), we get

$$\frac{3\ddot{R}}{R}=-\frac{1}{2}(\rho+3P-2\Lambda), \tag{15}$$

where ρ and P are the energy density and pressure due to matter. If we treat matter as we did before, this gives

$$\frac{3\ddot{R}}{R}=-\frac{k}{2R^3}+\Lambda. \tag{16}$$

Thus, once the universe expands sufficiently, the cosmological constant becomes more important than the energy density of matter in determining the fate of the universe. If $\Lambda>0$, a roughly exponential expansion will then ensue. This seems to be happening in our universe now.[35]

E. Spatial curvature

We have emphasized that gravity is due not just to the curvature of space, but of *space–time*. In our verbal formulation of Einstein's equation, this shows up in the fact that we consider particles moving forwards in time and study how their paths deviate in the space directions. However, Einstein's equation also gives information about the curvature of space. To illustrate this, it is easiest to consider not an expanding universe but a static one.

When Einstein first tried to use general relativity to construct a model of the entire universe, he assumed that the universe must be static—although he is said to have later described this as "his greatest blunder." As we did in the previous section, Einstein considered a universe containing ordinary matter with density ρ, no pressure, and a cosmological constant Λ. Such a universe can be static—the galaxies can remain at rest with respect to each other—only if the right-hand side of Eq. (15) is zero. In such a universe, the cosmological constant and the density must be carefully "tuned" so that $\rho=2\Lambda$. It is tempting to conclude that space–time in this model is just the good old flat Minkowski space–time of special relativity. In other words, one might guess that there are no gravitational effects at all. After all, the right-hand side of Einstein's equation was tuned to be zero. This would be a mistake, however. It is instructive to see why.

Remember that Eq. (3) contains all the information in Einstein's equation only if we consider all possible small balls. In all of the cosmological applications so far, we have applied the equation only to balls whose centers were at rest with respect to the local matter. It turns out that only for such balls is the right-hand side of Eq. (3) zero in the Einstein static universe.

To see this, consider a small ball of test particles, initially at rest relative to each other, that is moving with respect to the matter in the universe. In the local rest frame of such a ball, the right-hand side of Eq. (3) is nonzero. For one thing, the pressure due to the matter no longer vanishes. Remember that pressure is the flux of momentum. In the frame of our moving sphere, matter is flowing by. Also, the energy density goes up, both because the matter has kinetic energy in this frame and because of Lorentz contraction. The end result, as the reader can verify, is that the right-hand side of Eq. (3) is negative for such a moving sphere. In short, although a stationary ball of test particles remains unchanged in the Einstein static universe, our moving ball shrinks!

This has a nice geometric interpretation: the geometry in this model has spatial curvature. As we noted in Sec. II, on a positively curved surface such as a sphere, initially parallel lines converge toward one another. The same thing happens in the three-dimensional space of the Einstein static universe. In fact, the geometry of space in this model is that of a three-sphere. Figure 3 illustrates what happens.

One dimension is suppressed in this figure, so the two-dimensional spherical surface shown represents the three-dimensional universe. The small shaded circle on the surface represents our tiny ball of test particles, which starts at the equator and moves north. The sides of the sphere approach each other along the dashed geodesics, so the sphere shrinks in the transverse direction, although its diameter in the direction of motion does not change.

As an exercise, readers who want to test their understanding can fill in the mathematical details in this picture and determine the radius of the Einstein static universe in terms of the density. Here are step-by-step instructions:

- Imagine an observer moving at speed v through a cloud of stationary particles of density ρ. Use special relativity to determine the energy density and pressure in the observer's rest frame. Assume for simplicity that the observer is moving fairly slowly, and thus keep only the lowest-order nonvanishing term in a power series in v.
- Apply Eq. (3) to a sphere in this frame, including the contribution due to the cosmological constant (which is the same in all reference frames). You should find that the volume of the sphere decreases with $\ddot{V}/V\propto-\rho v^2$ to leading order in v.
- Suppose that space in this universe has the geometry of a large three-sphere of radius R_U. Show that the radii in the directions transverse to the motion start to shrink at a rate

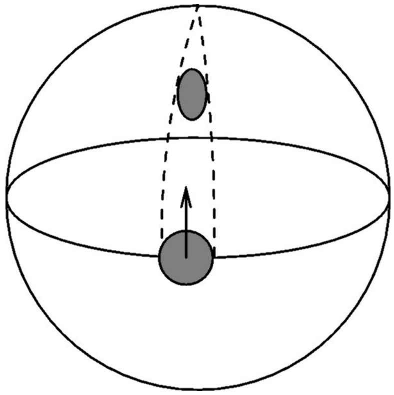


Fig. 3. The motion of a ball of test particles in a spherical universe.

given by $(\ddot{R}/R)|_{t=0} = -v^2/R_U^2$. (If, like most people, you are better at visualizing two-spheres than three-spheres, do this step by considering a small circle moving on a two-sphere, as shown above, rather than a small sphere moving on a three-sphere. The result is the same.)

- Since our little sphere is shrinking in two dimensions, its volume changes at a rate $\ddot{V}/V = 2\ddot{R}/R$. Use Einstein's equation to relate the radius R_U of the universe to the density ρ.

The final answer is $R_U = \sqrt{2/\rho}$, as you can find in standard textbooks.

Spatial curvature like this shows up in the expanding cosmological models described earlier in this section as well. In principle, the curvature radius can be found from our formulation of Einstein's equation by similar reasoning in these expanding models. However, such a calculation is extremely messy. Here the apparatus of tensor calculus comes to our rescue.[16,17]

ACKNOWLEDGMENT

E.F.B. is supported by National Science Foundation Grant No. 0233969.

APPENDIX A: THE MATHEMATICAL DETAILS

To see why Eq. (3) is equivalent to the usual formulation of Einstein's equation, we need a bit of tensor calculus. In particular, we need to understand the Riemann curvature tensor and the geodesic deviation equation. For a detailed explanation of these, the reader must turn to some of the texts in the bibliography.[16,17,21–23] Here we briefly sketch the main ideas.

When space–time is curved, the result of parallel transport depends on the path taken. To quantify this notion, pick two vectors u and v at a point p in space–time. In the limit where $\epsilon \rightarrow 0$, we can approximately speak of a "parallelogram" with sides ϵu and ϵv. Take another vector w at p and parallel transport it first along ϵv and then along ϵu to the opposite corner of this parallelogram. The result is some vector w_1. Alternatively, parallel transport w first along ϵu and then along ϵv. The result is a slightly different vector, w_2 as shown in Fig. 4. The limit

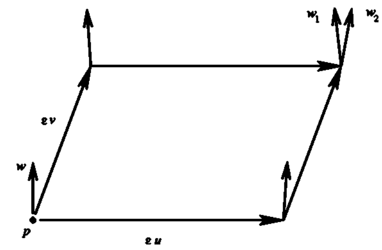


Fig. 4. Parallel transporting a vector w from one corner of a parallelogram to the opposite corner in two ways: up and then across, giving w_1, or across and then up, giving w_2.

$$\lim_{\epsilon \rightarrow 0} \frac{w_2 - w_1}{\epsilon^2} = R(u,v)w \tag{A1}$$

is well-defined, and it measures the curvature of space–time at the point p. In local coordinates we can write it as

$$R(u,v)w = R^{\alpha}_{\beta\gamma\delta} u^{\beta} v^{\gamma} w^{\delta}, \tag{A2}$$

where as usual we sum over repeated indices. The quantity $R^{\alpha}_{\beta\gamma\delta}$ is called the "Riemann curvature tensor."

We can use this tensor to compute the relative acceleration of nearby particles in free fall if they are initially at rest relative to one another. Consider two freely falling particles at nearby points p and q. Let v be the velocity of the particle at p, and let ϵu be the vector from p to q. Since the two particles start out at rest relative to one other, the velocity of the particle at q is obtained by parallel transporting v along ϵu.

Now let us wait a short while. Both particles trace out geodesics as time passes, and at time ϵ they will be at new points, say p' and q'. The point p' is displaced from p by an amount ϵv, so we get a little parallelogram, exactly as in the definition of the Riemann curvature as shown in Fig. 5.

Next let us compute the new relative velocity of the two particles. To compare vectors we must carry one to another using parallel transport. Let v_1 be the vector we get by taking the velocity vector of the particle at p' and parallel transporting it to q' along the top edge of our parallelogram. Let v_2 be the velocity of the particle at q'. The difference $v_2 - v_1$ is the new relative velocity. Figure 6 shows a picture of the whole situation. The vector v is depicted as shorter than ϵv for purely artistic reasons.

It follows that over this passage of time, the average relative acceleration of the two particles is $a = (v_2 - v_1)/\epsilon$. By Eq. (A1),

$$\lim_{\epsilon \rightarrow 0} \frac{v_2 - v_1}{\epsilon^2} = R(u,v)v, \tag{A3}$$

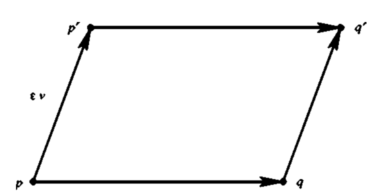


Fig. 5. Freely falling particles at p and q trace out geodesics taking them to p' and q'.

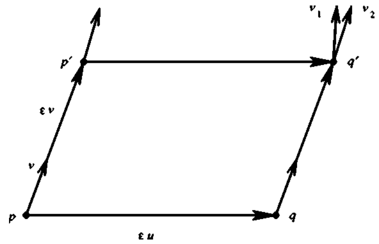


Fig. 6. Parallel transporting the velocity vector of the particle at p' to the point q' gives the vector v_1. The velocity vector of the particle at q' is v_2.

so

$$\lim_{\epsilon\to 0}\frac{a}{\epsilon}=R(u,v)v. \tag{A4}$$

This is called the "geodesic deviation equation." From the definition of the Riemann curvature it is easy to see that $R(u,v)w=-R(v,u)w$, so we can also write this equation as

$$\lim_{\epsilon\to 0}\frac{a^\alpha}{\epsilon}=-R^\alpha_{\beta\gamma\delta}v^\beta u^\gamma v^\delta. \tag{A5}$$

Using this equation we can work out the second time derivative of the volume $V(t)$ of a small ball of test particles that start out at rest relative to each other. As we mentioned earlier, to second order in time the ball changes to an ellipsoid. Furthermore, since the ball starts out at rest, the principal axes of this ellipsoid don't rotate initially. We can therefore adopt local inertial coordinates in which, to second order in t, the center of the ball is at rest and the three principal axes of the ellipsoid are aligned with the three spatial coordinates. Let $r^j(t)$ represent the radius of the jth axis of the ellipsoid as a function of time. If the ball's initial radius is ϵ, then

$$r^j(t)=\epsilon+\tfrac{1}{2}a^j t^2+O(t^3),$$

or in other words,

$$\lim_{t\to 0}\frac{\ddot r^j}{r^j}=\lim_{t\to 0}\frac{a^j}{\epsilon}.$$

Here the acceleration a^j is given by Eq. (A5), with u being a vector of length ϵ in the jth coordinate direction and v being the velocity of the ball, which is a unit vector in the time direction. In other words,

$$\lim_{t\to 0}\frac{\ddot r^j(t)}{r^j(t)}=-R^j_{\beta j\delta}v^\beta v^\delta=-R^j_{tjt}.$$

No sum over j is implied in the above expression.

Because the volume of our ball is proportional to the product of the radii, $\ddot V/V\to\Sigma_j \ddot r^j/r^j$ as $t\to 0$,

$$\lim_{V\to 0}\frac{\ddot V}{V}\bigg|_{t=0}=-R^\alpha_{t\alpha t}, \tag{A6}$$

where now a sum over α is implied. The sum over α can range over all four coordinates, not just the three spatial ones, since the symmetries of the Riemann tensor demand that $R^t_{ttt}=0$.

The right-hand side is minus the time-time component of the "Ricci tensor"

$$R_{\beta\delta}=R^\alpha_{\beta\alpha\delta}. \tag{A7}$$

That is,

$$\lim_{V\to 0}\frac{\ddot V}{V}\bigg|_{t=0}=-R_{tt} \tag{A8}$$

in local inertial coordinates where the ball starts out at rest.

In short, the Ricci tensor says how our ball of freely falling test particles starts changing in volume. The Ricci tensor only captures some of the information in the Riemann curvature tensor. The rest is captured by something called the "Weyl tensor," which says how any such ball starts changing in shape. The Weyl tensor describes tidal forces, gravitational waves and the like.

Now, Einstein's equation in its usual form says

$$G_{\alpha\beta}=T_{\alpha\beta}. \tag{A9}$$

Here the right side is the stress-energy tensor, while the left side, the "Einstein tensor," is just an abbreviation for a quantity constructed from the Ricci tensor:

$$G_{\alpha\beta}=R_{\alpha\beta}-\tfrac{1}{2}g_{\alpha\beta}R^\gamma_\gamma. \tag{A10}$$

Thus Einstein's equation really says

$$R_{\alpha\beta}-\tfrac{1}{2}g_{\alpha\beta}R^\gamma_\gamma=T_{\alpha\beta}. \tag{A11}$$

This implies

$$R^\alpha_\alpha-\tfrac{1}{2}g^\alpha_\alpha R^\gamma_\gamma=T^\alpha_\alpha, \tag{A12}$$

but $g^\alpha_\alpha=4$, so

$$-R^\alpha_\alpha=T^\alpha_\alpha. \tag{A13}$$

Plugging this into Eq. (A11), we get

$$R_{\alpha\beta}=T_{\alpha\beta}-\tfrac{1}{2}g_{\alpha\beta}T^\gamma_\gamma. \tag{A14}$$

This is an equivalent version of Einstein's equation, but with the roles of R and T switched! The good thing about this version is that it gives a formula for the Ricci tensor, which has a simple geometrical meaning.

Equation (A14) will be true if any one component holds in all local inertial coordinate systems. This is a bit like the observation that all of Maxwell's equations are contained in Gauss's law and $\nabla\cdot B=0$. Of course, this is only true if we know how the fields transform under change of coordinates. Here we assume that the transformation laws are known. Given this, Einstein's equation is equivalent to the fact that

$$R_{tt}=T_{tt}-\tfrac{1}{2}g_{tt}T^\gamma_\gamma \tag{A15}$$

in every local inertial coordinate system about every point. In such coordinates we have

$$g=\begin{pmatrix}-1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix} \tag{A16}$$

so $g_{tt}=-1$ and

$$T^\gamma_\gamma=-T_{tt}+T_{xx}+T_{yy}+T_{zz}. \tag{A17}$$

Equation (A15) thus says that

$$R_{tt}=\tfrac{1}{2}(T_{tt}+T_{xx}+T_{yy}+T_{zz}). \tag{A18}$$

By Eq. (A8), this is equivalent to

$$\lim_{V\to 0}\frac{\ddot{V}}{V}\bigg|_{t=0} = -\frac{1}{2}(T_{tt}+T_{xx}+T_{yy}+T_{zz}). \qquad \text{(A19)}$$

As promised, this is the simple, tensor-calculus-free formulation of Einstein's equation.

APPENDIX B: REFERENCES

We provide an annotated bibliography of material on relativity that we have found particularly helpful for students.

1. WEBSITES

There is a lot of material on general relativity available online. Most of it can be found starting from here:

1. **Relativity on the World Wide Web**, C. Hillman, http://math.ucr.edu/home/baez/relativity.html

The beginner will especially enjoy the many gorgeous websites aimed at helping one visualize relativity. There are also books available for free online, such as this:

2. **Lecture Notes on General Relativity**, S. M. Carroll, http://pancake.uchicago.edu/~carroll/notes/

The free online journal *Living Reviews in Relativity* is an excellent way to learn more about many aspects of relativity. One can access it at:

3. **Living Reviews in Relativity**, http://www.livingreviews.org

Part of learning relativity is working one's way through certain classic confusions. The most common are dealt with in the ''Relativity and Cosmology'' section of this site:

4. **Frequently Asked Questions in Physics**, edited by D. Koks, http://math.ucr.edu/home/baez/physics/

2. NONTECHNICAL BOOKS

Before diving into the details of general relativity, it is good to get oriented by reading some less technical books. Here are four excellent ones written by leading experts on the subject:

5. **General Relativity from A to B**, R. Geroch (University of Chicago Press, Chicago, 1981).
6. **Black Holes and Time Warps: Einstein's Outrageous Legacy**, K. S. Thorne (Norton, New York, 1995).
7. **Gravity from the Ground Up: An Introductory Guide to Gravity and General Relativity**, B. F. Schutz (Cambridge U. P., Cambridge, 2003).
8. **Space, Time, and Gravity: the Theory of the Big Bang and Black Holes**, R. M. Wald (University of Chicago Press, Chicago, 1992).

3. SPECIAL RELATIVITY

Before delving into general relativity in a more technical way, one must get up to speed on special relativity. Here are two excellent texts for this:

9. **Introduction to Special Relativity**, W. Rindler (Oxford U. P., Oxford, 1991).
10. **Space−time Physics: Introduction to Special Relativity**, E. F. Taylor and J. A. Wheeler (Freeman, New York, 1992).

4. INTRODUCTORY TEXTS

When one is ready to tackle the details of general relativity, it is probably good to start with one of these textbooks:

11. **Introducing Einstein's Relativity**, R. A. D'Inverno (Oxford U. P., Oxford, 1992).
12. **Gravity: An Introduction to Einstein's General Relativity**, J. B. Hartle (Addison-Wesley, New York, 2002).
13. **Introduction to General Relativity**, L. Hughston and K. P. Tod (Cambridge U. P., Cambridge, 1991).
14. **A First Course in General Relativity**, B. F. Schutz (Cambridge U. P., Cambridge, 1985).
15. **General Relativity: An Introduction to the Theory of the Gravitational Field**, H. Stephani (Cambridge U. P., Cambridge, 1990).

5. MORE COMPREHENSIVE TEXTS

To become an expert on general relativity, one really must tackle these classic texts:

16. **Gravitation**, C. W. Misner, K. S. Thorne, and J. A. Wheeler (Freeman, New York, 1973).
17. **General Relativity**, R. M. Wald (University of Chicago Press, Chicago, 1984).

Along with these textbooks, you'll want to do lots of problems! This book is a useful supplement:

18. **Problem Book in Relativity and Gravitation**, A. Lightman and R. H. Price (Princeton U. P., Princeton, 1975).

6. EXPERIMENTAL TESTS

The experimental support for general relativity up to the early 1990s is summarized in:

19. **Theory and Experiment in Gravitational Physics**, Revised ed., C. M. Will (Cambridge U. P., Cambridge, 1993).

A more up-to-date treatment of the subject can be found in:

20. ''The Confrontation between General Relativity and Experiment,'' C. M. Will, **Living Reviews in Relativity 4** (2001). Available online at http://www.livingreviews.org/lrr-2001-4

7. DIFFERENTIAL GEOMETRY

The serious student of general relativity will experience a constant need to learn more tensor calculus—or in modern terminology, ''differential geometry.'' Some of this can be found in the texts listed above, but it is also good to read mathematics texts. Here are a few:

21. **Gauge Fields, Knots and Gravity**, J. C. Baez and J. P. Muniain (World Scientific, Singapore, 1994).
22. **An Introduction to Differentiable Manifolds and Riemannian Geometry**, W. M. Boothby (Academic, New York, 1986).
23. **Semi-Riemannian Geometry with Applications to Relativity**, B. O'Neill (Academic, New York, 1983).

8. SPECIFIC TOPICS

The references above are about general relativity as a whole. Here are some suggested starting points for some of the particular topics touched on in this article.

a. The meaning of Einstein's equation

Feynman gives a quite different approach to this in:

24. **The Feynman Lectures on Gravitation**, R. P. Feynman *et al.* (Westview, Boulder, CO, 2002).

His approach focuses on the curvature of space rather than the curvature of space−time.

b. The Raychaudhuri equation

This equation, which is closely related to our formulation of Einstein's equation, is treated in some standard textbooks, including the one by Wald mentioned above. A detailed discussion can be found in

25. **Gravitation and Inertia**, I. Ciufolini and J. A. Wheeler (Princeton U. P., Princeton, 1995).

c. Gravitational waves

Here are two nontechnical descriptions of the binary pulsar work for which Hulse and Taylor won the Nobel prize:

27. **"The Binary Pulsar: Gravity Waves Exist," C. M. Will, Mercury, Nov-Dec 1987, pp. 162–174.**
28. **"Gravitational Waves from an Orbiting Pulsar," J. M. Weisberg, J. H. Taylor, and L. A. Fowler, Sci. Am., Oct 1981, pp. 74–82.**
Here is a review article on the ongoing efforts to directly detect gravitational waves:
29. "Detection of Gravitational Waves," J. Lu, D. G. Blair, and C. Zhao, Rep. Prog. Phys., **63**, 1317–1427 (2000).
Some present and future experiments to detect gravitational radiation are described here:
30. **LIGO Laboratory Home Page**, http://www.ligo.caltech.edu/
31. **The Virgo Project**, http://www.virgo.infn.it/
32. **Laser Interferometer Space Antenna**, http://lisa.jpl.nasa.gov/

d. Black holes

Astrophysical evidence that black holes exist is summarized in:

33. "Evidence for Black Holes," M. C. Begelman, Science **300**, 1898–1903 (2003).
A less technical discussion of the particular case of the supermassive black hole at the center of our Milky Way Galaxy can be found here:
34. **The Black Hole at the Center of Our Galaxy**, F. Melia (Princeton U. P., Princeton, 2003).

e. Cosmology

There are lots of good popular books on cosmology. Since the subject is changing rapidly, pick one that is up to date. At the time of this writing, we recommend:

35. **The Extravagant Universe: Exploding Stars, Dark Energy, and the Accelerating Cosmos**, R. P. Kirshner (Princeton U. P., Princeton, 2002).
A good online source of cosmological information is:
36. **Ned Wright's Cosmology Tutorial**, http://www.astro.ucla.edu/~wright/cosmolog.htm
The following cosmology textbooks are arranged in increasing order of technical difficulty:
37. **Cosmology: The Science of the Universe**, 2nd ed., E. Harrison (Cambridge U. P., Cambridge, 2000).
38. **Cosmology: a First Course**, M. Lachièze-Rey (Cambridge U. P., Cambridge, 1995).
39. **Principles of Physical Cosmology**, P. J. E. Peebles (Princeton U. P., Princeton, 1993).
40. **The Early Universe**, E. W. Kolb and M. S. Turner (Addison–Wesley, New York, 1990).
41. **The Large-Scale Structure of Space–time**, S. W. Hawking and G. F. R. Ellis (Cambridge U. P., Cambridge, 1975).