Gottfried sum rule and the light-flavor content of the nucleon

Stefano Forte
Istituto Nazionale di Fisica Nucleare, Sezione di Torino, via P. Giuria 1, I-10125 Torino, Italy
(Received 1 April 1992)

We discuss the implications for the light-flavor content of the nucleon of the recent experimental results of the New Muon Collaboration (NMC) on the Gottfried sum rule. We perform a phenomenological analysis of the data and of the symmetries of the quark sea in the nucleon. We show that a discrepancy between the data and naive parton model expectations as seen in the NMC data is also indicated by measurements of the nucleon σ term. We compute the isotriplet value of the σ term through the flavor-singlet scale Ward identity, and we use the result to fix the light-flavor content of the nucleon. We discuss the possibility of reproducing the experimental results in various theoretical frameworks, both within effective models of the nucleon and on the basis of perturbative or nonperturbative QCD.

PACS number(s): 13.60.Hb, 11.30.Hv, 11.50.Li, 12.40.Aa

I. INTRODUCTION

A recent experimental determination of the proton-neutron difference of structure functions F_2, performed by the New Muon Collaboration (NMC) [1], has pointed out yet one more instance of disagreement between naive parton model expectations and deep-inelastic scattering data. As is well known [2], the first moment of $F_2(x)$ in the parton model is equal to the sum of the electric charges squared, e_i^2, of all partons in the target:

$$\int_0^1 dx \frac{F_2(x)}{x} = \sum_i e_i^2 n_i \ , \tag{1}$$

$$n_i = \int_0^1 dx \, [q_i(x) + \bar{q}_i(x)] \ , \tag{2}$$

where $q_i(x)$ $[\bar{q}_i(x)]$ is the momentum distribution of the ith quark [antiquark] parton flavor. Notice that n_i indicates the total number of quarks *plus* antiquarks of flavor i [3]; as we shall discuss extensively in the sequel, this is a scale-dependent quantity $[n_i = n_i(Q^2)]$.

The difference of the first moments (1) measured with a proton and a neutron target,

$$S_G \equiv \int_0^1 dx \frac{F_2^p(x) - F_2^n(x)}{x} \ , \tag{3}$$

gives a handle on the isospin asymmetry of partons in the nucleon. If one computes the difference (3) by including only valence quarks in the sum over partons (1), one gets

$$S_G(\text{valence}) = e_u^2 - e_d^2 = \tfrac{1}{3} \ , \tag{4}$$

the difference of the square charges of the up and down quarks. Equation (4) is known as the Gottfried sum rule. This is usually justified [2] by assuming that the sea, being isotopically neutral, cannot contribute to a proton-neutron difference; previous data, affected by large experimental uncertainties, supported this conclusion. The NMC data [1] instead lead to a value

$$S_G = 0.24 \pm 0.016 \ . \tag{5}$$

This value is obtained at $Q^2 = 4$ GeV^2 by extrapolating to the full $0 \le x \le 1$ range the value

$$S_G[0.004 \le x \le 0.8] = 0.227 \pm 0.021 \tag{5$'$}$$

obtained from the measured x region. The extrapolation to small x is done by assuming Regge behavior. The neutron structure function is obtained, neglecting shadowing, as the difference between deuteron and proton structure functions.

Both the extrapolation and the neglect of shadowing may be questioned. On the one hand, one may obtain from the data values of S_G which agree with the sum rule (4) if one assumes Regge behavior to set in at a much smaller x value than usually taken [4]. Then, the small-x behavior of the valence-quark distributions can be determined from the data, and in particular, if the distributions are assumed to rise more steeply with x than predicted by Regge behavior (which would then set in, if at all, only for extremely small x), the contribution from the small-x region to S_G may be made larger, possibly reconciling the data with the naive parton model prediction (4).

On the other hand, because shadowing is a small-x effect, it may potentially give a large contribution to the first moment (1). If shadowing is taken into account, the neutron structure function is larger than the deuteron-proton difference of structure functions, thus yielding a yet smaller value of S_G [Eq. (3)]. Two computations of the shadowing correction to S_G [Eq. (3)] have appeared in the literature [5,6], in rough agreement with each other. The correction to S_G [Eq. (5$'$)] for the measured range of x is computed to be $\Delta S_G = -0.043$ [5] or $\Delta S_G = -0.026$ [6], the difference being due to details in the parametrization of the diffraction dissociation cross section which is used to estimate the amount of shadowing. In Ref. [5] a further correction $\Delta S_G = -0.038$ is estimated to apply to the unmeasured $x < 0.004$ region. It is perhaps worth noticing that these sizable corrections are obtained from a rather small value of the total shadowing [about 3–4 % of the (virtual) γ-D cross section] due to the fact that the effect is concentrated at small x (typically $x \sim 0.01$) [7].

Thus a conservative estimate of the shadowing correction to S_G [Eq. (5)] leads to the approximate value

$$S_G \approx 0.2 \ . \tag{5''}$$

Such a small value would be accordingly harder to reconcile with the sum rule (4), even by modifying the small-x extrapolation. Resolution of these problems rests ultimately with experiment [4]; in the sequel we shall assume the result of Ref. [1], as given in Eq. (5), to be correct, although we shall also take the determination (5'') into account.

Then, if we accept the experimental result, we must conclude that the light-flavor content of the nucleon is very different from the content usually assumed according to parton model expectations. This conclusion is rather striking both from the phenomenological and from the theoretical point of view. On the one hand, conventional parton intuition is embodied in all the phenomenological parametrizations of parton distributions which are used in analyzing experimental data. On the other hand, a dramatic violation of the light-flavor symmetry of sea-quark distributions presents us with a theoretical puzzle which we would like to understand in terms of QCD or at least effective models of the nucleon. It is the purpose of this paper to address these issues by trying to establish as accurately as possible what the experimental result (5) entails for the light-flavor content of the nucleon, and to see how this fits in the framework of QCD and of effective models of the nucleon.

In Sec. II we shall discuss the SU(2) light-flavor symmetry structure of the nucleon sea quarks at various energy scales. In Sec. III we shall show that whereas Eq. (5) alone is not sufficient to fix uniquely the up and down content of the nucleon, this may be done, within the parton model, through the knowledge of a different, experimentally accessible quantity, namely, the nucleon σ term. In Sec. IV we shall determine the isotriplet matrix element of the latter (i.e., the difference of its proton and neutron values) by means of a Ward identity satisfied by the trace of the energy-momentum tensor in QCD, and we shall use the result to perform a phenomenological analysis of the pattern of flavor-symmetry violation in the nucleon sea. Finally, in Sec. V we shall discuss the flavor content of the nucleon in the framework of the Skyrme and bag model, and we shall see how perturbative and nonperturbative mechanisms for sea-quark generation compare to the data. Conclusions are drawn in Sec. VI.

II. THE SU(2) SYMMETRIES OF THE NUCLEON SEA

The Gottfried sum rule (4) is a parton model sum rule, and not a consequence of QCD (unlike other parton model sum rules) in that it does not follow from a conservation law (such as the Adler sum rule), nor from the symmetry properties of an operator (a current) whose matrix elements are related to the quantity measured by the sum rule (as the Bjorken sum rule). In particular, there is no conservation principle to protect the value (4) computed from valence partons alone; indeed, radiation of a quark-antiquark pair of flavor i by a gluon produces a contribution of 2 units to n_i [Eq. (2)]. These problems are exacerbated by the fact that (as we shall discuss in detail in the next section) there exists no leading-twist operator whose matrix elements measure S_G [Eq. (3)], or more generally the first moments (1). Thus, one must assume that the parton picture makes sense in the first place in order for the quantity (3) to have any physical meaning. This is the point of view that we shall take in the sequel of this paper. Of course, one could take the opposite point of view [8], that parton model sum rules are meaningful only to the extent that they follow from the operator-product expansion applied to deep-inelastic scattering, and conclude that the discrepancy between the parton prediction (4) and the experimental value (5) indicates that a parton model computation of S_G is not meaningful. In view of the wide success of the QCD-improved parton model, we feel that this conclusion is unwarranted.

Then, the Gottfried sum rule (4) is a consequence of a specific assumption on the separation of valence and sea contributions to the first moment (1), namely, that sea contributions cancel in the difference (3). This is true only if the nucleon sea satisfies two different SU(2) symmetries. Using the parton determination of the first moment of F_2 [Eq. (1)] in the definition (3) of S_G leads to

$$S_G = (e_u^2 n_u + e_d^2 n_d)|_{I=1} = e_u^2(n_u^p - n_u^n) + e_d^2(n_d^p - n_d^n) \ , \tag{6}$$

where $I=1$ indicates the isotriplet matrix element, i.e., the proton-neutron difference; u and d label up and down quarks, while n and p label neutron and proton matrix elements; and we assumed that the content of strange and heavier flavors in the proton and neutron is the same (we shall come back on this assumption in Sec. IV). Equation (6) may be conveniently rewritten as

$$\begin{aligned} S_G &= \tfrac{1}{2}[(e_u^2 + e_d^2)(n_u + n_d) + (e_u^2 - e_d^2)(n_u - n_d)]|_{I=1} \\ &= \tfrac{1}{2}[\tfrac{5}{9}(n_u + n_d) + \tfrac{1}{3}(n_u - n_d)]|_{I=1} \ . \end{aligned} \tag{7}$$

If we assume isospin symmetry [$SU(2)_I$, henceforth], then the coefficient of $\frac{5}{9}$ in Eq. (7) vanishes because it is the isotriplet matrix element of an isosinglet quantity. In terms of parton content $SU(2)_I$ implies

$$n_u^p = n_d^n, \quad n_d^p = n_u^n \ , \tag{8}$$

which entails

$$S_G^{SU(2)_I} = \tfrac{1}{2}[\tfrac{1}{3}(n_u - n_d)]|_{I=1} \tag{9}$$

$$= \tfrac{1}{3}\{1 + [n_u^p(\text{sea}) - n_d^p(\text{sea})]\} \ , \tag{10}$$

where in Eq. (10) we have separated the valence and sea contributions to the sum rule. The valence-quark component of the nucleon satisfies by definition $SU(2)_I$ [thus it never contributes to the coefficient of $\frac{5}{9}$ in Eq. (7)]; hence, Eq. (8) may be viewed as a requirement on the nucleon sea.

S_G reduces to its valence value (4) if one assumes the sea to satisfy both $SU(2)_I$ and a further SU(2) symmetry: namely,

$$n_u^p = n_d^p, \quad n_u^n = n_d^n \ , \tag{11}$$

which we shall call Q spin, i.e., isospin at the quark (rather than nucleon) level, or $SU(2)_Q$. If we assume $SU(2)_Q$ alone, and not $SU(2)_I$, then the coefficient of $\frac{1}{3}$ in the sum rule (7) reduces to its valence value, but there might be still a contribution from the sea in the coefficient of $\frac{5}{9}$:

$$S_g^{SU(2)_Q}=\tfrac{1}{3}+\tfrac{5}{9}(n_u^p-n_u^n)\ . \qquad (12)$$

The most general parametrization of S_G [Eq. (7)] is thus

$$S_G=\tfrac{1}{3}(1+\Delta Q)+\tfrac{5}{9}\Delta I\ , \qquad (13)$$

where ΔQ and ΔI denote, respectively, the violation of $SU(2)_Q$ and $SU(2)_I$:

$$\begin{aligned}\Delta Q &\equiv \tfrac{1}{2}[n_u(\text{sea})-n_d(\text{sea})]|_{I=1}=\tfrac{1}{2}(n_u-n_d)|_{I=1}-1\ ,\\ \Delta I &\equiv \tfrac{1}{2}(n_u+n_d)|_{I=1}\ .\end{aligned} \qquad (14)$$

The experimental value (5) may be explained by assuming a violation of either, or both, of the symmetries $SU(2)_I$ [Eq. (8)] and $SU(2)_Q$ [Eq. (11)]; there is no *a priori* reason to favor either symmetry, apart from the naive expectation that $SU(2)_I$ should be better established. Physically, the experimental value (5) may be understood as a consequence of the fact that the flavor asymmetry of the nucleon sea is opposite to that of the valence quarks, thus leading to a negative ΔQ, or of the fact that the proton sea is smaller than the neutron sea, thus leading to a negative ΔI.

It should be noticed that, whereas the first moment of any quark distribution [Eq. (1)] is ill defined, it is quite conceivable that the quantities introduced in Eq. (14) are all well defined, and generally nonzero. This can be understood as follows. Quark distributions are expected to behave as $q(x)\sim 1/x$ as $x\to 0$ on the basis of Regge behavior (see, e.g., Ref. [2]). Thus $q(x)$ admits an expansion $q(x)=q_0/x+q_1+q_2x+\cdots$ in the neighborhood of zero, and ΔQ and ΔI are finite if q_0 satisfies Eq. (11). Now, the value of q_0 is controlled by the Pomeron contribution to the structure function $F_2(x)$ as $x\to 0$; since the Pomeron is charge-conjugation even and carries zero isospin it follows that the quark distributions satisfy both $SU(2)_Q$ and $SU(2)_I$ as $x\to 0$. However, no constraint is imposed by the small-x behavior on the higher-order coefficients, which may not satisfy Eq. (11). This leads to finite but nonzero values of $\Delta I, \Delta Q$.

In view of the possibility of understanding the source of possible isospin violations, it is interesting to discuss the scale dependence of the various terms in Eqs. (7) and (13). Indeed, the two terms on the right-hand side (RHS) of Eq. (7) are both scale dependent; also, their scale dependence is not the same. This means that the pattern of SU(2)-symmetry breaking depends on scale. Now, the two combinations $n_u\pm n_d$ which determine the values of ΔI and ΔQ [Eq. (14)] renormalize both multiplicatively (see, e.g., Ref. [9]). It follows that if ΔI is nonzero, its value is scale dependent, whereas if it vanishes at some scale, then it remains zero at all scales. The value of ΔQ, instead, cannot be zero at all scales because the quantity which is renormalized multiplicatively is $1+\Delta Q$; otherwise stated, the value of ΔQ is scale independent only if $\Delta Q=-1$ [i.e., if the $SU(2)_Q$ asymmetry of sea quarks exactly compensates that of valence quarks].

Furthermore, the evolution of ΔI takes place at one loop, whereas that of ΔQ only starts at two loops, because ΔI evolves as the singlet quark number and ΔQ as the nonsinglet one. At one loop, evolution takes place through emission of one gluon; thus, it does not change the nonsinglet number, although the singlet number may evolve due to gluon admixture. At two loops a quark can emit a quark-antiquark pair through gluon radiation; this determines the evolution of the nonsinglet number because the emission kernel is different according to whether or not the final-state quark-antiquark pair has the same flavor as the starting quark (due to Fermi statistics) [10].

The evolution of ΔI is thus simply obtained by taking the first moment of the singlet Altarelli-Parisi equation, i.e.,

$$\frac{d}{dt}(n_u+n_d)|_{I=1}=\frac{\alpha_s}{2\pi}\left[A_1^{qq}(n_u+n_d)|_{I=1}+\frac{1}{f}A_1^{qG}2n_g|_{I=1}\right]\ , \qquad (15)$$

where A_1 are first moments of splitting functions, f is the number of flavors, and n_g is the number of gluons, i.e., the first moment of the gluon density, $n_g=\int_0^1 G(x)dx$ (we follow the notation and conventions of Ref. [9]). At one loop $A_1^{qq}=0$. Because the gluon distribution is expected to behave at small x as the quark distribution, the isotriplet n_g does not diverge (although the isosinglet does); then, if we assume that the isotriplet gluon number does not vanish either, the evolution of ΔI is nontrivial.

Complete integration of Eq. (15) would require simultaneous diagonalization of the corresponding equation for n_g. However, this is unfeasible because the evolution equation for the gluon distribution is inconsistent with the expected behavior $G(x)\sim 1/x$; rather, it predicts a stronger growth at small x which violates unitarity, indicating a failure of the leading-logarithm approximation under which the evolution equation is derived (see, e.g., Ref. [11]). This is reflected by the fact that the first moments of the gluon splitting functions diverge, thus leading to an infinite value of the anomalous dimension of the quark number, too. At very small x, though, the Altarelli-Parisi equations should be replaced by nonlinear equations [12], which are not amenable to the standard treatment in that they do not factorize upon taking moments and can in general only be integrated numerically.

Nevertheless, we may compute the evolution of ΔI from Eq. (15) if we assume the isospin violation in the first moments of the gluon distributions to be proportional to that of the quark distributions, i.e.,

$$(n_u+n_d)_{I=1}=2\lambda n_g|_{I=1}\ . \qquad (16)$$

Then Eq. (15) can be integrated immediately:

$$\begin{aligned}&\Delta I(Q^2)=\Delta I(Q_0^2)\exp(\lambda d_1^{qG}s)\ ,\\ &d_1^{qG}=\frac{1}{2\pi fb}A_1^{qG},\quad s=\ln\left[\frac{\ln(Q^2/\Lambda^2)}{\ln(Q_0^2/\Lambda^2)}\right]\ ,\end{aligned} \qquad (17)$$

where b is the lowest-order coefficient of the β function [9]. With $f=3$ we get $d_1^{qG}=0.15$. Unless the value of λ in Eq. (16) is exceedingly large, this leads to a moderate dependence of ΔI on scale. For example, taking $\lambda=1$,

$$\Delta I(4\ \mathrm{GeV}^2)=1.06\Delta I(1\ \mathrm{GeV}^2)\ . \quad (18)$$

Of course, a large value of ΔI could be obtained due to strong scale dependence in the nonperturbative region.

The two-loop evolution of ΔQ, instead, is given by [9,10]

$$1+\Delta Q(Q^2)=\left[1-\frac{\gamma_1^{(2)}}{b}[\alpha_s(Q^2)-\alpha_s(Q_0^2)]\right]\times[1+\Delta Q(Q_0^2)]\ , \quad (19)$$

where b is as in Eq. (17) and the value of the two-loop anomalous dimension of the nonsinglet first moment of quark distributions is [9,10] $\gamma_1^{(2)}=-8.1\times10^{-3}$. Because the anomalous dimension is very small, the Q^2 dependence of ΔQ is negligible for practical purposes; for instance, if $\Delta Q(Q_0^2)=0$ the asymptotic value $\Delta Q(\infty)$ attained as $Q^2\to\infty$ is just $\Delta Q(\infty)=-0.01\alpha_s(Q_0^2)$ [13]. Also,

$$\Delta Q(4\ \mathrm{GeV}^2)=0.999\Delta Q(1\ \mathrm{GeV}^2)-0.001\ . \quad (20)$$

Thus, QCD evolution does lead to negative ΔQ, as required to explain the data; unfortunately the effect is negligibly small.

Whereas the Q^2 dependence of both the light-flavor symmetry-violation parameters ΔQ and ΔI is too weak to account for the violation of the Gottfried sum rule, it may be strong enough to be measured experimentally. In particular, the scale dependence of the ratio of structure functions F_2^p/F_2^n may be a sensitive probe of the values of ΔQ and ΔI. Such a dependence is seen in the data of Ref. [1]; more precise measurements over a wider range of Q^2 could provide independent evidence for the violation of $\mathrm{SU}(2)_I$ and $\mathrm{SU}(2)_Q$, and could allow disentangling the relative amount of violation of the two symmetries.

III. THE σ TERM AS A FLAVOR PROBE

The discussion in the previous section shows that, at a fixed value of Q^2, it is possible to disentangle the amount of violation of $\mathrm{SU}(2)_Q$ and $\mathrm{SU}(2)_I$ only through a knowledge of the experimental value of a different linear combination of n_i [Eq. (2)]. This is nontrivial because there is actually no leading-twist operator associated with the first moment (1) of the structure function F_2; i.e., the total quark number (2) cannot be expressed as the forward nucleon matrix element of a twist-2 operator. This is a consequence of the fact that at leading twist the nth moment of a structure function in the scaling limit is proportional to the forward matrix element of the spin-n contribution to the operator-product expansion. However, an operator with odd (even) spin is odd (even) under charge conjugation, while the matrix element which yields the electroproduction structure function $F_2(x)=x\sum_i e_i^2[q_i(x)+\bar{q}_i(x)]$ is charge-conjugation even. It follows that there are no odd-spin contributions to this matrix element, and the odd moments [including the first moment (1)] may only be constructed by analytic continuation from the even ones (see Ref. [10]).

However, there does exist a (higher-twist) operator whose matrix elements are related to the total number of quarks plus antiquarks n_i [Eq. (2)], namely, the fermion bilinear $\bar{\psi}\psi$, which has charge-conjugation properties opposite to those of the charge $\psi^\dagger\psi$ and is twist 3. The identification of the nucleon matrix elements of this operator with the quark number (up to a universal multiplicative renormalization constant) has been suggested in Ref. [14] on phenomenological grounds. Indeed, expanding $\bar{\psi}\psi$ in a plane-wave basis (and using Bjorken-Drell notation) gives

$$\bar{\psi}\psi=\int d^3k\frac{m}{E}\left[\sum_s\hat{N}(\mathbf{k},s)+\sum_{s,s'}\hat{M}(\mathbf{k},s,s')\right]\ ; \quad (21)$$

$$\hat{N}(\mathbf{k},s)=b^\dagger(\mathbf{k},s)b(\mathbf{k},s)+d^\dagger(\mathbf{k},s)d(\mathbf{k},s)\ , \quad (22)$$

$$\hat{M}(\mathbf{k},s,s')=A(\mathbf{k},s,s')b^\dagger(\mathbf{k},s)d^\dagger(-\mathbf{k},s')+\mathrm{H.c.}\ , \quad (23)$$

$$A(\mathbf{k},s,s')=e^{2iEt}2s'u^\dagger(0,s)\frac{\sigma\cdot\mathbf{k}}{m}u(0,s')\ . \quad (24)$$

The operator $\hat{N}$ [Eq. (22)] is recognized as that which counts the number of quarks plus antiquarks; hence, up to the factor of m/E (due to the fact that $\bar{\psi}\psi$ is a Lorentz scalar) the matrix element of the first term on the RHS of Eq. (21) is equal to the total quark plus antiquark number. The operator $\hat{M}$ [Eq. (23)], instead, is nondiagonal in quark number; it creates or annihilates a quark-antiquark pair, thereby generating transitions between different Fock components of the nucleon wave function.

It follows that the identification of the operator (21) with the quark number is correct to the extent that the contribution of the operator $\hat{M}$ [Eq. (23)] to its matrix elements is negligible; the factor m/E may then be approximated by a constant. These approximations are widely used in model computations (see, e.g., Refs. [15,16]) and seem to be phenomenologically successful (see Refs. [14–17]). On phenomenological grounds, we may thus use the operator $\bar{\psi}\psi$ [Eq. (21)] as a probe of the quark content of the nucleon, as suggested in Ref. [14], and assume its forward nucleon matrix elements to be proportional to the quark number [18]. Notice that the constant of proportionality is scale dependent: the operator $\bar{\psi}\psi$ [Eq. (21)] evolves as an inverse mass (see, e.g., Ref. [17]), thus, its evolution starts at one loop, whereas the (nonsinglet) evolution of the quark number n_i [Eq. (2)] begins at two loops. In the next section we shall see that independent evidence for the smallness of the nondiagonal contribution to the matrix elements of $\bar{\psi}\psi$ in the isotriplet case is provided by a Ward identity satisfied by this operator.

We are thus left with the problems of determining the constant of proportionality between the operator $\bar{\psi}_i\psi_i$ [Eq. (21)] and n_i [Eq. (2)] and that of finding a measurable linear combination of $\bar{\psi}_i\psi_i$ when the flavor index i takes different values. It turns out that both problems can be solved at once by considering the combination of operators $\bar{\psi}_i\psi_i$ constructed in analogy with Eq. (1), with the quark square charges replaced by quark masses, i.e., the nucleon σ term (see Refs. [14,17] for reviews):

$$\sigma\equiv\sum_{i=u,d}m_i\bar{\psi}_i\psi_i\ , \quad (25)$$

which is just the light-quark mass term of the QCD Hamiltonian. The rationale for this choice is that, on the one hand, the nucleon matrix elements of σ [Eq. (25)] are measurable, while, on the other hand, we may take advantage of the fact that σ is scale invariant because $m\bar{\psi}\psi$ is. Then, we may absorb the constant of proportionality in the quark masses, i.e., *define* [14] the quark mass to be the (dimensionful) coefficient of proportionality between the operator $m\bar{\psi}\psi$ and the quark number. The values of the quark masses determined in this guise are (at the nucleon scale) [14] $m_u \simeq 4$ MeV, $m_d \simeq 7$ MeV, in good agreement with the more precise determination, found through utterly different techniques [17],

$$m_u = 5.1 \pm 1.5 \text{ MeV}, \quad m_d = 8.9 \pm 2.6 \text{ MeV} . \tag{26}$$

This suggests that we may take $m/E \simeq 1$ in Eq. (21) to an accuracy of about 20%, comparable to that of the determination (26). Thus, we shall henceforth assume phenomenologically that (to the same accuracy) the nucleon matrix element of the operator (21) is

$$\langle N|\bar{\psi}_i\psi_i|N\rangle \simeq \hat{n}_i , \tag{27}$$

where $\hat{n}_i$ is the quark number n_i [Eq. (1)] at the nucleon scale $Q^2 = 1$ GeV^2. We shall comment in the next section on the sensitivity of our results to the absolute value of the quark masses (i.e., on the uncertainty in the determination of the coefficient of proportionality between $\bar{\psi}\psi$ and n_i) [19].

We are thus led to define the quantity

$$\sigma_G \equiv \sigma|_{I=1} = (m_u\hat{n}_u + m_d\hat{n}_d)|_{I=1} . \tag{28}$$

Notice that the values of the masses in Eq. (28) should be fixed without using any information on the experimental value of σ_G, i.e., without knowledge of isospin-violating mass splittings in the nucleon sector; otherwise Eq. (28) would be an identity, which does not carry any information.

Equation (28) may be treated as in Eqs. (6)–(13) with the replacement $e_i^2 \to m_i$. In particular, the analogue of Eqs. (7) and (13) are now

$$\sigma_G = \tfrac{1}{2}[(m_u + m_d)(\hat{n}_u + \hat{n}_d) + (m_u - m_d)(\hat{n}_u - \hat{n}_d)]|_{I=1} \tag{29}$$

$$= (m_u - m_d)[1 + \Delta Q(1 \text{ GeV})] + (m_u + m_d)\Delta I(1\text{GeV}). \tag{30}$$

If we evolve ΔI and ΔQ to the same scale, through the evolution equations (17) and (20), then Eq. (30), together with Eq. (13), allows us to solve for the two unknowns ΔQ and ΔI, namely, the amount of violation of SU(2)$_Q$ and SU(2)$_I$, respectively. In practice, given the weakness of the evolution (18),(20) [as compared to the uncertainty in Eqs. (26) and (27)], we may take ΔI and ΔQ to be constants.

We may now check qualitatively the consistency of our picture by verifying that the violation of SU(2)$_I$ or SU(2)$_Q$, which is obtained from Eq. (13), is also displayed by Eq. (30). Indeed, if $\Delta I = \Delta Q = 0$, then σ_G reduces to its valence determination, namely,

$$\sigma_G(\text{valence}) = m_u - m_d \approx -3 \text{ MeV} , \tag{31}$$

which is the parton model prediction for this quantity, just as Eq. (4) is the parton prediction for S_G. Now, a rough estimate [17] of σ_G [Eq. (28)] can be obtained by observing that, by definition, σ_G is the zeroth-order perturbative contribution to the proton-neutron mass difference due to the presence of the light-quark mass term (25) in the QCD Hamiltonian. The proton-neutron mass difference, after subtraction of the electromagnetic contribution, is [17]

$$M_p - M_n = -2.05 \pm 0.30 \text{ MeV} . \tag{32}$$

Thus,

$$\sigma_G \approx -2 \text{ MeV} , \tag{33}$$

which shows a disagreement with the naive prediction (31) comparable in magnitude to the violation of the Gottfried sum rule (4): in both cases the experimental value is about $\frac{2}{3}$ of the valence-quark contribution.

This provides independent evidence for a violation of SU(2)$_I$ or SU(2)$_Q$ and seems to disfavor explanations of the experimental results (5) based on mechanisms that reduce the value of S_G while leaving σ_G unaffected. One such mechanism, suggested in Ref. [20], consists of assuming that deep-inelastic scattering should not be taken to be entirely incoherent at the quark level, but rather that scattering on quark pairs may be coherent; i.e., that effective diquark constituents contribute to the cross section. Clearly, this effect does not modify the sum rule (31), and seems thus to be disfavored by the data.

However, a quantitative computation of the sizes of ΔQ and ΔI, even to the accuracy to which we expect Eq. (27) to hold, requires a more accurate determination of the experimental value of σ_G. We shall do this in the next section.

IV. THE σ TERM AND THE SCALE WARD IDENTITY

The most direct way of measuring the nucleon matrix element of σ [Eq. (25)] is to relate it to pion-nucleon scattering amplitudes (see, e.g., Ref. [17]). Then, an experimental determination of σ_G [Eq. (28)] can be simply obtained by subtracting the proton and neutron σ terms [Eq. (25)]. This leads [17,21] again to the value of σ_G given by Eq. (33). Unfortunately, the σ term, which is typically $\sigma \simeq 35$–45 MeV, is determined in pion-nucleon scattering only up to an accuracy of about 20% [21]; thus it is unclear whether the value (33) determined in this way is significant at all. On the other hand, there is no way of estimating the error in the zeroth-order perturbative computation which led to Eq. (33).

We may, however, do better by observing that classically the σ term (25) is equal to the divergence of the Noether current for scale transformations; i.e., it equals the trace of the energy-momentum tensor. In the quantized theory, this implies a Ward identity that allows relating its matrix elements to the mass of physical states, as we shall now prove.

To this purpose, define the dilation current [22] $j_D^\mu = x_\nu T^{\mu\nu}$. This is the Noether current for scale transformations; its divergence is equal to the trace of the energy-momentum tensor $T^{\mu\nu}$. In the quantized theory, the trace of the energy-momentum tensor satisfies on shell the operator equation [23]

$$\partial_\mu j^\mu_D = T^\mu{}_\mu = (1+\gamma_m)\sum_i m_i \bar{\psi}_i \psi_i + \frac{\beta(\alpha_s)}{4\alpha_s} G_a^{\mu\nu} G^a_{\mu\nu} , \quad (34)$$

where the sum runs over all quark flavors, $G_a^{\mu\nu}$ is the gluon field strength, $\gamma_m = -\mu(\partial m/\partial\mu)$ is the mass anomalous dimension, and β is the β function for the strong coupling [24] α_s. All operators appearing in Eq. (34) are renormalized and normal ordered. The last term on the RHS of Eq. (34) is due to the conformal anomaly, and the quantum correction γ_m is present because the anomaly term and the mass term in Eq. (34) are not separately scale invariant, while the energy-momentum tensor is (up to surface terms) [23]. Taking matrix elements of Eq. (34) generates the scale Ward identities of QCD.

On the other hand, the nucleon matrix element of the energy-momentum tensor may be parametrized by two form factors:

$$\langle N(k_2)|T_{\mu\nu}|N(k_1)\rangle = 2k_\mu k_\nu F_N(q^2) + (q^2 g_{\mu\nu} - q_\mu q_\nu) G_N(q^2) , \quad (35)$$

with $k = \frac{1}{2}(k_1+k_2)$ and $q = k_2 - k_1$. Identification of T^{00} with the Hamiltonian density fixes $F_N(0) = 1/(2M_N)$, where M_N is the mass of the given nucleon state [25]. Because of the absence of a scalar Goldstone boson, $G_N(q^2)$ has no pole at $q^2=0$, and the forward matrix element of the trace of $T^{\mu\nu}$ is

$$\langle N(k)|T^\mu{}_\mu|N(k)\rangle = M_N . \quad (36)$$

Thus, the isotriplet nucleon matrix element of Eq. (34) is

$$\left[\sum_i (1+\gamma_m) m_i \bar{\psi}_i \psi_i + \frac{\beta(\alpha)}{4\alpha_s} G_a^{\mu\nu} G^a_{\mu\nu} \right] \Bigg|_{I=1} = M_p - M_n . \quad (37)$$

Assuming as usual the content of strange and heavier quarks to be isosinglet (we shall come back to this assumption at the end of this section), we get

$$(1+\gamma_m)\sigma_G = M_p - M_n - \left[\frac{\beta(\alpha_s)}{4\alpha_s} G_a^{\mu\nu} G^a_{\mu\nu} \right] \Bigg|_{I=1} , \quad (38)$$

which is the desired Ward identity.

Although the gluon operator on the RHS of Eq. (38) (the scalar gluon condensate) is obviously flavor singlet, we cannot conclude immediately that its isotriplet matrix element vanishes since it is well known [26] that its *pseudoscalar* counterpart, i.e., the condensate $\epsilon^{\mu\nu\rho\sigma} G^a_{\mu\nu} G^a_{\rho\sigma}$, has a large isotriplet nucleon matrix element. This is, however, not the case in the scalar sector. To understand this, it is convenient to rewrite the matrix element on the LHS of Eq. (37) in terms of one-particle-irreducible physical couplings, analogously to what is done in the pseudoscalar case in order to derive [27] the isosinglet Goldberger-Treiman relation, which is closely related to Eq. (37).

To this end, we define a generating functional

$$W(D_\mu, S_i) = \ln\left\langle \exp\left[i \left[D_\mu j^\mu_D + \sum_i S_i \Phi_i \right] \right] \right\rangle , \quad (39)$$

where the set of fields $\Phi_i = (N, \bar{N}, \phi_a = \bar{\psi}\lambda_a\psi,\ \phi_0 = \bar{\psi}\psi,\ Q = [\beta(\alpha_s)/4\alpha_s] G_a^{\mu\nu} G_a^{\mu\nu})$ includes the nucleon, the scalar mesons, and the gluon condensate. Also, we define the effective action obtained by Legendre transformation with respect to the fields Φ_i (but not j^μ_D):

$$\Gamma(D^\mu, \Phi_i^{\rm cl}) = W(D^\mu, S_i) - \Phi_i^{\rm cl} S_i, \quad \Phi_i^{\rm cl} = \frac{\delta W}{\delta S^a} . \quad (40)$$

The functional Γ (called the Zumino effective action) generates diagrams which are one-particle irreducible with respect to $\Phi_i^{\rm cl}$.

Now, the nucleon matrix element of the Ward identity (34) in terms of W [Eq. (39)] reads (at zero momentum transfer and setting all sources to zero)

$$\partial_\mu \frac{\delta^3 W}{\delta D_\mu \delta S_N \delta S_{\bar{N}}} = (1+\gamma_m)\sum_i m_i \frac{\delta^3 W}{\delta S_i \delta S_N \delta S_{\bar{N}}} + \frac{\delta^3 W}{\delta\Theta\, \delta S_N \delta S_{\bar{N}}} , \quad (41)$$

where the sum runs over the sources for the quark bilinears obtained as linear combinations of the flavor-diagonal scalar mesons, S_N are the sources for the nucleons, and Θ is the source for the anomaly. On the other hand, in terms of Γ, the Ward identity is

$$\partial_\mu \frac{\delta^3\Gamma}{\delta D_\mu \delta N\, \delta\bar{N}} = -\frac{\delta\Gamma}{\delta\phi_0^{\rm cl}\delta N\, \delta\bar{N}} \delta_D \phi_0^{\rm cl} , \quad (42)$$

where $\delta_D \phi_a^{\rm cl} = -\phi_a^{\rm cl}$ is the variation upon dilatation of the fermion bilinears, and when setting sources to zero $\Phi_a^{\rm cl} = \langle \Phi_a \rangle$, so that only the singlet contribution survives.

We may compare Eq. (41) with Eq. (42) if we recall [27] that the LHS of Eq. (42) is related to that of Eq. (41) by

$$\begin{aligned} \partial_\mu \frac{\delta^3 W}{\delta D_\mu \delta S_N \delta S_{\bar{N}}} &= S_{\bar{N}\bar{N}} \left[\partial_\mu \frac{\delta^3\Gamma}{\delta D_\mu \delta N\, \delta\bar{N}} + (\partial_\mu S_{D_\mu\phi_0}) \frac{\delta^3\Gamma}{\delta\phi_0^{\rm cl}\delta N\, \delta\bar{N}} + (\partial_\mu S_{D_\mu Q}) \frac{\delta^3\Gamma}{\delta Q\, \delta N\, \delta\bar{N}} \right] S_{NN} \\ &= S_{\bar{N}\bar{N}} \left[-\frac{\delta\Gamma}{\delta\phi_0^{\rm cl}\delta N\, \delta\bar{N}} \langle\phi_0\rangle + (\partial_\mu S_{D_\mu\phi_0}) \frac{\delta^3\Gamma}{\delta\phi_0^{\rm cl}\delta N\, \delta\bar{N}} + (\partial_\mu S_{D_\mu Q}) \frac{\delta^3\Gamma}{\delta Q\, \delta N\, \delta\bar{N}} \right] S_{NN} , \end{aligned} \quad (43)$$

where $S_{\Phi_i\Phi_j} \equiv \delta^2 W/\delta S_i \delta S_j = \langle T(\Phi_i \Phi_j)\rangle$ are the various full propagators, and in the last step we used Eq. (42).

If there was no anomaly (for example, for a conformal theory), then the last terms on the RHS of Eqs. (41) and (43) would both be missing, and we would identify the two surviving contributions on the RHS of Eq. (43) as the one-meson-irreducible (OMI) and one-meson-reducible (OMR) contributions to the expectation value of σ, respectively. However, in the presence of the anomaly the OMR terms on the RHS of Eq. (43) both vanish at zero momentum transfer because the anomaly removes the scalar Goldstone boson and only the first term survives, i.e., that given by Eq. (42). In this case, using Eq. (41) in Eq. (43) we get

$$(1+\gamma_m)\sum_i m_i \frac{\delta^3 W}{\delta S_i \delta S_N \delta S_{\bar N}} + \frac{\delta^3 W}{\delta\Theta\, \delta S_N \delta S_{\bar N}} = S_{\bar N \bar N} \frac{\delta \Gamma}{\delta\phi_0^{\rm cl} \delta N\, \delta \bar N} \langle \phi_0 \rangle S_{NN} \ . \quad (44)$$

It follows that the contribution of the anomaly to the LHS of Eq. (44) removes the OMR part of the expectation value of σ in order to yield the OMI coupling on the RHS. That is, taking the isotriplet matrix element,

$$\left[\frac{\beta(\alpha_s)}{4\alpha_s} G_a^{\mu\nu} G_{\mu\nu}^a\right]\Bigg|_{I=1} = -(1+\gamma_m)\left[\sum_i m_i \bar\psi_i \psi_i\right]\Bigg|_{I=1}^{\rm OMR} . \quad (45)$$

Now, the one-particle-reducible matrix element of the σ term is dominated by the diagrams where the σ term couples directly to the meson states with the same quantum numbers. In the pseudoscalar sector (where a similar relation holds [27]) this in turn is dominated by the pion pole; it follows [26] that the OMR contribution to the pseudoscalar analogue of the σ term, namely, $\sigma_5 = m_u \bar u \gamma_5 u + m_d \bar d \gamma_5 d$, has an isotriplet component of order

$$\sigma_5|_{I=1}^{\rm OMR} = \frac{m_u - m_d}{m_u + m_d} \sigma_5^3 \Bigg|_{I=1}^{\rm OMR} , \quad (46)$$

where $\sigma_5^3 = m_u \bar u \gamma_5 u - m_d \bar d \gamma_5 d$. According to the pseudoscalar analogue of Eq. (45), this implies that the *pseudoscalar* gluon condensate has a large isotriplet component. In the scalar sector, instead, there is no dominance of the triplet mesons over the singlet ones; thus the isospin violation of the OMR contribution to the σ term is of the order of the isotriplet component of σ_G [Eq. (28)] at the relevant scale [26], and Eq. (46) is replaced by

$$\sigma_G^{\rm OMR} \sim \frac{m_u - m_d}{M_N} \sigma_G^{3\ \rm OMR} ,$$
$$\sigma_G^3 = (m_u \bar u \gamma u - m_d \bar d d)|_{I=1} \ . \quad (47)$$

By Eq. (45), it follows that the isotriplet gluon condensate's contribution to Eq. (38) vanishes up to corrections of order $(m_u - m_d)/M_N$. We conclude that

$$\sigma_G = \frac{1}{1+\gamma_m^0}(M_p - M_n) \quad (48)$$

with an accuracy of 0.5%, where γ_m^0 is the value of γ_m at the nucleon scale $Q^2 = 1\ {\rm GeV}^2$. This shows that the first-order perturbative estimate (33), up to the quantum correction due to the anomalous dimension γ_m, is protected by the scale Ward identity and holds in the full theory.

It is interesting to compare the decomposition of σ_G into its OMI and OMR portions to that of the operator $\bar\psi\psi$ [Eq. (21)] into its diagonal (22) and nondiagonal (23) parts (in quark number). It is clear that the OMR part of σ_G comes entirely from the matrix elements of the nondiagonal operator $\hat M$ [Eq. (23)]. Thus, the conclusion [Eq. (47)] that the former quantity is small agrees with the assumption we made in the previous section (on phenomenological grounds) that the latter contribution to $\bar\psi\psi$ is negligible, and provides theoretical evidence in favor of it as long as we assume that the matrix elements of $\hat M$ [Eq. (24)] are dominated by meson exchange.

Notice, furthermore, that, independently of whether or not the OMR part of σ_G is small, because of Eq. (45), it is always true that

$$\hat\sigma_G = \frac{1}{1+\gamma_m^0}(M_p - M_n) , \quad (49)$$

where $\hat\sigma_G$ is the OMI part of σ_G. But then, because all OMR contributions to σ_G come from matrix elements of the nondiagonal operator $\hat M$ [Eq. (23)], $\hat\sigma_G$ is a better approximation than σ_G itself to the quark number, which is identified with matrix elements of the diagonal operator $\hat N$ [Eq. (22)]. In particular, if we assume that the matrix elements of σ are dominated by coupling through meson states, then $\hat\sigma$ is always proportional to the quark number, regardless of whether or not the matrix elements of $\hat N$ are small. Thus, even if Eq. (27) were not true because of a large contribution of the nondiagonal operator $\hat M$ [Eq. (23)] to $\bar\psi\psi$, we could still approximate n_i with matrix elements of $\hat\sigma$, which is all we need in order to proceed.

These results can be checked in any model of the nucleon where the relevant quantities may be computed explicitly, like chiral models. In these models (as we shall see in the next section in the case of the Skyrme model) $\Delta I = 0$; hence, we can express the quantity S_G [Eq. (13)] in terms of $\hat\sigma_G$, which in turn is related through Eq. (49) to the (strong) nucleon mass splitting:

$$S_G = \frac{1}{r}\frac{\hat\sigma_g}{3(m_u - m_d)} = \frac{1}{r}\frac{1}{3(m_u - m_d)}\frac{M_p - M_n}{1+\gamma_m^0} , \quad (50)$$

where we have restored the constant of proportionality r between $\hat\sigma$ and the quark number,

$$\hat\sigma = r \sum_i m_i \hat n_i \ . \quad (51)$$

The constant was previously set equal to 1, however in specific models it may differ significantly from unity. Equation (50) is exact in all models where $\Delta I = 0$, and the coupling of composite operators (such as $\bar\psi\psi$) to the nucleon is dominated by meson exchange. Indeed, Eq. (51) has been obtained in Ref. [28] by explicit computation in a chiral quark-meson model.

We may now use the two-loop value [29,17] $\gamma_m = 0.27$ and the value (32) of the nucleon strong mass splitting to get

$$\sigma_G = -1.61 \pm 0.24 \text{ MeV} . \quad (52)$$

Notice that the quantum correction is actually quite sizable. Equation (52) thus confirms the qualitative estimate at the end of the previous section, and its implied conclusion that the violation of $SU(2)_I$ or $SU(2)_Q$ which is displayed by the NMC measurement also shows up in the value of the isotriplet σ term.

We can finally get a quantitative determination of the symmetry violation by solving for ΔI and ΔQ Eqs. (30) and (13) with the determinations (5) and (47) of S_G and σ_G and the values of the quark masses Eq. (26). Using the value (5), uncorrected for shadowing, we get

$$\Delta Q = -0.38 \pm 0.19 , \quad (53)$$

$$\Delta I = 0.057 \pm 0.094 . \quad (54)$$

This means that the naive expectation is confirmed: Q spin is violated by over 30%, whereas isospin is a much better symmetry, violated only up to a few percent. If we use instead the value of S_G corrected to account for shadowing [Eq. (5″) with the experimental error (5)], then

$$\Delta Q = -0.46 \pm 0.20 , \quad (55)$$

$$\Delta I = 0.034 \pm 0.096 , \quad (56)$$

which implies an even larger violation of Q spin, almost of order 50%. In order to see whether a violation of isospin is really required, we may check whether the value $\Delta I = 0$ (i.e., no isospin violation) is consistent with the data. If we take $\Delta I = 0$ by assumption, then we can predict S_G using Eq. (50) (with $r = 1$). This gives $S_G = 0.14 \pm 0.075$, to be compared with the experimental results [Eqs. (5) and (5″)]. Even though the evidence is not conclusive, a small but nonzero value of ΔI is suggested by the data.

The error in Eqs. (53)–(56) has been determined from the experimental errors in the determination of S_G, σ_G, and the quark and nucleon mass splittings. On top of that, there is the extra theoretical uncertainty in the determination of the value of the coefficient r [Eq. (51)], which we assumed in Eq. (27) to be $r \simeq 1$ to about 20% accuracy, on phenomenological grounds. The uncertainty in the determination of r can be viewed, according to the definition [14] of the quark masses given in the previous section, as an uncertainty in the absolute value of the quark mass while the mass ratios are being held fixed.

Even though the uncertainty on the determination of the quark masses is the main source of error in Eqs. (53)–(56), the values of ΔI and ΔQ turn out to be rather stable under variations of the absolute value of the masses. In particular, if we assume that the approximation introduced in Eq. (27) has introduced a further uncertainty of 20% on the absolute value of the masses (i.e., if we allow rescaling the masses up or down by 20%), this only increases the value of the error in Eqs. (53) and (54) to $\sigma(\Delta Q) = 0.20$ and $\sigma(\Delta I) = 0.099$, and in Eqs. (55) and (56) to $\sigma(\Delta Q) = 0.22$ and $\sigma(\Delta I) = 0.10$. Even if we assume that the masses may be off by up to 50%, which would mean that the proportionality constant r [Eq. (51)] which we took to be 1 is actually $0.5 \le r \le 1$, we get $\sigma(\Delta Q) = 0.22$ and $\sigma(\Delta I) = 0.10$ in Eqs. (53) and (54) and $\sigma(\Delta Q) = 0.31$ and $\sigma(\Delta I) = 0.12$ in Eqs. (55) and (56). We conclude that the theoretical error in the determination of r does not substantially affect our conclusions. Notice that all these errors are certainly overestimated since $r \le 1$, whereas we allowed for rescaling the masses in both directions.

If we relax the assumption that strange and heavier quark components of the nucleon sea be isosinglet, we get extra contributions to both S_G [Eq. (13)] and σ_G [Eq. (28)] from $\Delta n_i = n_i^p - n_i^n$, where i is a heavy-quark flavor. Although now we can no longer solve for ΔQ, ΔI, and Δn_i simultaneously, we can get a feeling for the expected values of Δn_i. Suppose, for example, $\Delta n_s \neq 0$. Because $m_s = 25(m_u + m_d)/2$, this gives a very large contribution to σ_G. Thus, unless we imagine rather unlikely scenarios with large violations of all symmetries that miraculously cancel each other, the value (24) of σ_G leads to the estimate $\Delta n_s < 1\%$; hence, Δn_s does not contribute significantly to S_G. It follows that an isotriplet strange-quark contribution cannot modify the value of ΔQ [Eqs. (35) or (37)] by more than a few percent, although it could alter significantly ΔI. This is *a fortiori* true for heavier flavors: for example the charm quark could at most provide a 0.1% contribution to ΔQ, and so forth.

V. THE FLAVOR STRUCTURE IN EFFECTIVE MODELS OF THE NUCLEON AND IN QCD

We may now discuss whether and how the symmetry violations (53)–(56) in the nucleon sea can be reproduced within various approaches. Our aim is to see whether the pattern of symmetry violation found in the previous section is compatible with our understanding of the nucleon; detailed model computations will be reserved for forthcoming publications. First, we shall consider some popular effective models of the nucleon; then we shall discuss whether the results of model computations can be reproduced in a more fundamental approach.

Let us consider first the bag model (see, e.g., Ref. [30]). The expectation value of the operator $\bar{\psi}\psi$ in a nucleon state can be computed directly [15,16]: it receives a valence contribution from quarks inside the bag and a sea contribution due to the fact that inside the bag $\langle \bar{\psi}\psi \rangle = 0$, whereas outside the bag the vacuum is not the perturbative one and yields a nonvanishing chiral condensate. The sea contribution to n_i in the bag model is thus (to leading order in α_s)

$$n_i(\text{sea}) = -\frac{1}{r} \langle \bar{\psi}_i \psi_i \rangle V_B , \quad (57)$$

where the expectation value is to be taken in the nonperturbative QCD vacuum, while V_B is the bag volume. In Eq. (57) we have restored the proportionality constant r [Eq. (51)], which in this model can be significantly smaller than 1; typically [31] $r \sim 0.5$.

The sea component generated through Eq. (57) cannot violate Q spin without violating isospin. Either V_B is not isospin invariant (i.e., the size of the proton is not the same as that of the neutron), in which case n_i [Eq. (57)] violates isospin only, or the vacuum chiral condensate is not SU(2) invariant (i.e., $\langle \bar{u}u \rangle \neq \langle \bar{d}d \rangle$), in which case $SU(2)_Q$ is violated, but $SU(2)_I$ is also violated by the same amount. Now, the chiral condensate is usually assumed to be SU(2) symmetric, and can be determined using current algebra (see, e.g., Ref. [17]) as $\langle \bar{u}u \rangle = \langle \bar{d}d \rangle = -(225\pm 25\ \text{MeV})^3$. A violation of $SU(2)_I$ of, say, 4% (i.e., $\Delta I \sim 0.04$), as in Eq. (38), can then be obtained, if the nucleon radius is $R_N \simeq 1$ fm, by taking $R_p - R_n = 0.002$ fm (even taking $r=1$). Such a tiny isospin violation is certainly compatible with present data on electron-nucleus scattering; as a matter of fact, isospin violation of nucleon radii up to about 0.2 fm in either direction is compatible with the data (see Ref. [32] and references therein), and if we did not know about the σ term, in this model we could easily reproduce the entire violation of the sum rule (4) from violation of isospin.

Thus, although we can easily reproduce the violation of $SU(2)_I$ required by the data [Eqs. (53)–(56)], it does not seem possible to obtain the required violation of $SU(2)_Q$. Because in our approach the problem of determining the flavor content of the sea has been related, through Eq. (49), to that of computing the proton-neutron mass difference, this can be seen as a manifestation of the well-known [31] difficulty of reproducing this difference correctly in pure bag models.

Let us turn now to the Skyrme model. Here there are no explicit quark degrees of freedom; however, we may identify quark operators with soliton operators that carry the same quantum numbers, i.e., transform in the same way under the flavor SU(N) group. Hence, in the Skyrme model [15]

$$\langle N|\bar{\psi}X\psi|N\rangle = Z[\mathrm{Tr}(XU_N + U_N^\dagger X^\dagger) - \mathrm{Tr}(X+X^\dagger)] , \tag{58}$$

where X is a flavor SU(N) matrix, U_N is the nucleon Skyrmion, Z is a renormalization constant with the dimensions of $(\text{mass})^3$, and Tr indicates both trace on flavor indices and integration over all space. The last term on the RHS corresponds to the subtraction of the vacuum expectation value of the given operator X. We may fix the value of Z by requiring the vacuum condensate per unit volume to take its physical value:

$$\langle \bar{u}u \rangle = Z\,\mathrm{tr}(X_u + X_u^\dagger) = -(225\ \text{MeV})^3 , \tag{59}$$

where tr is the ordinary trace and X_u is the SU(N) matrix $X_u = \mathrm{diag}(1,0,\ldots,0)$, i.e., $Z = -\frac{1}{2}(225\ \text{MeV})^3$.

The explicit values of n_i depend now on the version of the Skyrme model which is used. In the SU(3) model we have [15]

$$r n_u^p[\mathrm{SU}(3)] = \langle p|\bar{u}u|p\rangle = \tfrac{12}{10}k , \tag{60}$$

$$r n_d^p[\mathrm{SU}(3)] = \langle p|\bar{d}d|p\rangle = \tfrac{11}{10}k , \tag{61}$$

where the constant k depends on the details of the Skyrmion field, and the values for the neutron expectation values are found using $SU(2)_I$ [Eq. (8)], which is exact in this model. If the SU(2) uncranked soliton (see, e.g., Ref. [33]) is $U_0 = \exp[i\boldsymbol{\sigma}\cdot\hat{\mathbf{x}}F(x)]$, then

$$k = \frac{16\pi}{3} Z \int_0^\infty dx\, x^2[\cos F(x) - 1] . \tag{62}$$

The integration can be performed with an explicit numeric solution for the Skyrme profile $F(x)$. Using the numerical values tabulated in Ref. [33] to determine k [Eq. (62)] we get, from Eqs. (60)–(62),

$$r(n_u^p - n_d^p) = 0.15 , \tag{63}$$

whereas, from Ref. [34],

$$r(n_u^p - n_d^p) = 0.35 . \tag{64}$$

Since these results are very model dependent (for instance, they are very sensitive to the value of the coupling of the quartic Skyrme term), Eqs. (63) and (64) should be taken as order-of-magnitude estimates. Recalling that in this model $\Delta I = 0$, and taking the value of r (which cannot be calculated directly as there are no explicit quarks) to be the same as in the bag model (according to the Cheshire cat principle), i.e., $r=0.5$, we get, from Eq. (14),

$$\Delta Q = -(0.3-0.7) , \tag{65}$$

where the two extreme values correspond to Eqs. (63) and (64). We conclude that in the SU(3) Skyrme model the violation of $SU(2)_Q$ is rather large, even more than required by the data on the Gottfried sum rule.

The SU(3) model, however, is known to predict an unusually large value of $\langle N|\bar{s}s|N\rangle$ [15], which seems to contradict recent experimental data [21]. We may thus turn to the SU(2) model, which gives generally a better description of the data. In this model [15],

$$n_u^p[\mathrm{SU}(2)] = n_d^p[\mathrm{SU}(2)] = n_u^n[\mathrm{SU}(2)] = n_d^n[\mathrm{SU}(2)] , \tag{66}$$

which implies that $SU(2)_I$ is again exact, whereas

$$\Delta Q = -1 , \tag{67}$$

i.e., $SU(2)_Q$ is violated by 100%. The same is true for models where the light quarks are viewed as rotational excitations of an SU(2) Skyrmion, while strange quarks are vibrational excitations; these models are known [34] to give a good description of SU(3) breaking in $\langle N|\bar{\psi}_i\psi_i|N\rangle$.

The extreme result (67) is certainly phenomenologically untenable, since it would entail the vanishing of the proton-neutron mass difference. However, isospin noninvariant terms may be added to the SU(2) Skyrme Lagrangian as subleading terms in a $1/N_f$ expansion [35,31]. For example, such terms would be induced in an effective Lagrangian describing the coupling of Skyrmions to vector mesons [31,36]. With these terms it is possible to obtain satisfactory fits of the proton-neutron mass difference from $SU(2)_Q$ violation [31], at the expense of introducing many additional parameters in the model. Since, according to Eqs. (53)–(56), ΔQ provides the bulk of the violation of the Gottfried sum rule (4), the experimental value of the latter [Eq. (5)] can be easily accommodated in these models. In this picture, the isotriplet σ

terms behave in the Skyrme model in a way similar to the much discussed isosinglet axial charge [37]: it vanishes in the pure Skyrme model, and it receives contributions at higher orders in $1/N_f$.

The facts that in a pure bag model only $SU(2)_I$ can be violated, and that pure Skyrme models display only violation of $SU(2)_Q$ (generally larger than required by the data) suggest considering hybrid models. It is clear that in the most naive hybrid model one can easily reproduce the data (53)–(56) by simply weighing ΔQ [Eqs. (65) or (67)] by the percentage of the baryon number (or of the nucleon's moment of inertia [31]) carried by the Skyrme field, adjusting the bag radius in order to get the correct value of the weighted ΔQ, and then reproducing the required ΔI by a small isospin dependence of V_B, as discussed above. Indeed, less crude hybrid models are known to reproduce to good accuracy the proton-neutron mass difference [36], as well as [16] the SU(3) violation of $\langle N|\bar{\psi}_i\psi_i|N\rangle$. The strong dependence of the SU(2) violation on the bag radius seems to hold in more refined models, too [36].

Having ascertained that a large violation of $SU(2)_Q$ is actually predicted by the Skyrme model, and that the experimental pattern of symmetry violation can be reproduced in hybrid chiral models, it is natural to ask whether known perturbative or nonperturbative mechanisms for sea-quark generation can explain the large SU(2)-symmetry violation which is displayed by both experimental data and model calculations.

Within perturbative QCD there is little more to say than the evolution equations (18) and (20). Namely, if we take, at a relatively small scale (where, however, we still trust perturbation theory), $\Delta Q = 0$ and ΔI of a few percent, then ΔI is hardly scale dependent, whereas a tiny ΔQ is generated that is negligible for all practical purposes. An important by-product of the computation [10] of the scale dependence of ΔQ is the explicit proof that the Pauli principle does not yield any contribution to ΔQ. The latter has been suggested time and again recently [38,39] as a possible explanation of the result (5) (an excess of up quarks in the valence component would suppress the up component of the sea quarks, and conversely), but the explicit computation of Ref. [10] shows that this effect, contrary to the naive expectation, has (due to interference) the wrong sign, in addition to being negligibly small.

Since all the evolution equations discussed so far hold in the limit of vanishing quark masses, there remains the possibility that isospin breaking in the quark masses may play a role; this is also the only possibility of finding a violation of $SU(2)_I$, which is necessarily exact in the chiral limit. This possibility is currently under investigation [40]; otherwise perturbative QCD does not provide a natural explanation of the data, and we must invoke some nonperturbative mechanism instead.

Among nonperturbative mechanisms for the violation of $SU(2)_Q$, pion cloud effects have been repeatedly suggested. Some of these [28,38] are based on an incomplete counting of the contributions to S_G: in these references, Eq. (7) is written as

$$S_G = \tfrac{1}{3} + \tfrac{2}{3}(n_{\bar{u}}^p - n_{\bar{d}}^p) , \qquad (68)$$

where $n_{\bar{u}}^p$ ($n_{\bar{d}}^p$) is the number of $\bar{u}$ ($\bar{d}$) antiquarks in the proton. Equation (68) follows from Eq. (7) assuming $SU(2)_I$ and

$$n_{\bar{u}}^p(\text{sea}) = \tfrac{1}{2} n_u^p(\text{sea}), \quad n_{\bar{d}}^p(\text{sea}) = \tfrac{1}{2} n_d^p(\text{sea}) \qquad (69)$$

[in which case it reduces to Eq. (10)]. In Eq. (69) n_u, n_d are, as usual, the total numbers of u,d quarks plus antiquarks. It is then observed that, for an assembly of pions,

$$N_{\pi^+} - N_{\pi^-} = n_{\bar{u}}^p - n_{\bar{d}}^p , \qquad (70)$$

which, if used in Eq. (68), would express the violation of the sum rule in terms of the asymmetry in pion content. Of course, however, pions violate Eq. (69) unless $N_{\pi^+} - N_{\pi^-} = 0$; thus we cannot make this substitution: either $N_{\pi^+} = N_{\pi^-}$ or the $q\bar{q}$ sea, due to Eq. (69), exactly compensates the pion contribution to S_G.

Indeed, because quarks and antiquarks are counted with the same sign in n_i [Eq. (2)] any pion always gives a vanishing contribution to S_G, or generally to $n_u - n_d$. If, however, we assume that *all* sea quarks condense into pions (i.e., there are no free sea quarks, but only sea quarks bound into pions), then either $N_{\pi^+} = N_{\pi^-}$ or some sea antiquarks must be bound into a pion with a valence quark. Under this rather strong assumption the expression [Eqs. (68) and (70)] of S_G is correct. In general, however, it is unclear that this assumption is true; if it is false, those pions whose quark should be counted in the valence component contribute to S_G [Eq. (68)] through Eq. (70), but in addition there is an undetermined quark contribution to S_G whose sign and magnitude cannot be estimated, thus making the applicability of the results of Refs. [28,38] to the explanation of the experimental result (5) dubious.

However, a pion contribution to the violation of the Gottfried sum rule can be obtained [39,41], provided one assumes that at least some part of the nucleon sea is generated through transitions where a pion is radiated by a nucleon. A proton then would favor the transition where a π^+ is created, namely, $p \to n + \pi^+$, over that where a π^- is created, namely, $p \to \Delta^{++} + \pi^-$ (and conversely for a neutron) because of the nucleon-Δ mass difference. This, spelling out the quark content, is seen to favor the production of $d\bar{d}$ pairs over $u\bar{u}$ pairs, thus producing a violation of ΔQ with the right sign. The problem with this kind of explanation is that one would expect the contribution to ΔQ thus generated to be of the order of $\Delta Q \sim (M_N - M_\Delta)/(M_\Delta + M_N)$, i.e., $\Delta Q \sim -0.14$; explicit calculations [41] support this conclusion. This seems to be too small to account for the data.

A final option which may be considered is nonperturbatively induced quark-quark interactions. The simplest example of these are instanton-induced interactions. It is immediately clear [42] that these lead to a contribution to ΔQ which is qualitatively in the right direction, basically for the same reason why they lead to a cancellation of the isosinglet axial charge of the nucleon [43], namely, the instanton-induced effective quark-quark interaction ('t Hooft interaction) couples N_f quark-antiquark pairs of different flavors. Thus, this interaction can only generate a down (up) $q\bar{q}$ content of an up (down) quark; i.e., it always gives a negative contribution to ΔQ. A quantitative investigation of this mechanism is currently in progress.

VI. CONCLUSION

In this paper, we have shown that the recent measurement of the Gottfried sum rule by the NMC implies a substantial violation of the $SU(2)_Q$ symmetry of the nucleon sea [Eq. (11)]. Independent evidence for this effect is provided by the value of the isotriplet nucleon σ term, which we were able to determine through the scale Ward identity. Combining these two results allows us to fix to good accuracy the pattern of SU(2) violation of the nucleon sea, and suggests that the large violation of $SU(2)_Q$ is accompanied by a small violation of isospin.

These results have a considerable phenomenological import, since the exactness of the symmetries $SU(2)_Q$ and $SU(2)_I$ [Eqs. (8) and (11)] is always assumed in phenomenological parametrizations of the nucleon structure functions (see, e.g., Ref. [44]). Allowing for a violation of these symmetries would considerably alter our picture of the nucleon structure functions. It would be very important to get independent measurements of the symmetry-violation parameters.

From the theoretical viewpoint, the violation of isospin and Q spin challenges our understanding of the nucleon. Although this kind of effect may be reproduced in models of the nucleon such as the Skyrme model [which allows for violation of $SU(2)_Q$], bag models [which allow for violation of $SU(2)_I$], or hybrid chiral models (where both symmetries can be violated), it is hard to get a fundamental understanding of their origin. Perturbative QCD evolution almost certainly cannot produce the required effects, and even traditional nuclear physics, such as pion radiation, are insufficient to produce effects as large as the observed ones.

It would seem that, once more, precision deep-inelastic scattering measurements are giving us a hint of physics that belongs to the nonperturbative QCD domain which still remains so elusive to theoretical investigation.

ACKNOWLEDGMENTS

I thank W. Alberico, M. Arneodo, B. Badełek, R. Ball, M. Bos, N. Nikolaev, M. Nowak, M. Rho, and E. Predazzi for discussions; M. Chemtob, P. Gensini, and J. Kwieciński for correspondence; and especially M. Anselmino for his continued interest in this work and a critical reading of the manuscript.

[1] P. Amaudruz *et al.*, Phys. Rev. Lett. **66**, 560 (1991).
[2] See, e.g., E. Leader and E. Predazzi, *Gauge Theories and the New Physics* (Cambridge University Press, Cambridge, England, 1982).
[3] In the sequel, unless otherwise stated explicitly, n_i shall always denote a sum of numbers of quarks plus antiquarks.
[4] A. D. Martin, W. J. Stirling, and R. G. Roberts, Phys. Lett. B **252**, 653 (1990); S. D. Ellis and W. J. Stirling, *ibid.* **256**, 258 (1991).
[5] V. R. Zoller, Phys. Lett. B **279**, 145 (1992).
[6] B. Badełek and J. Kwieciński, Nucl. Phys. **B370**, 278 (1992).
[7] A rough estimate of the shadowing correction to S_G has also been given by S. Kumano and J. T. Londergan [Phys. Rev. D **44**, 717 (1991)], with the conclusion that the correction is negligibly small. This result is due to the fact that in this reference the correction to the neutron structure function is estimated by multiplying a phenomenological shadowing factor by the experimental value of the neutron structure function F_2^n from Ref. [1]. However, in Ref. [1] the neutron structure function is obtained as the difference between the deuteron and the proton structure functions. This is correct only under the assumption that there is no shadowing; therefore, the procedure of Kumano and Londergan appears to be inconsistent. In particular, in this reference the ratio $R = F_2^p/F_2^n$ is determined from the (uncorrected) data of Ref. [1] through a phenomenological parametrization such that R tends to 1 from above as $x \to 0$. However, the explicit computations of Refs. [5,6] show that, after the shadowing correction is applied, while R is larger than 1 at large x, it becomes smaller than 1 for $x \sim 0.005-0.01$, and must thus tend to 1 from below as $x \to 0$. The assumption that there is no such crossover is compatible with the data of Ref. [1] only if shadowing is negligibly small; the result of Kumano and Londergan appears thus to be a consequence of this assumption.
[8] G. Preparata, P. G. Ratcliffe, and J. Soffer, Phys. Rev. Lett. **66**, 687 (1991).
[9] G. Altarelli, Phys. Rep. **81**, 1 (1982).
[10] D. A. Ross and C. T. Sachrajda, Nucl. Phys. **B149**, 497 (1979).
[11] B. Badełek *et al.*, "Small x Physics," DESY Report No. 91-124, 1991 (unpublished).
[12] A. H. Mueller and J. Qiu, Nucl. Phys. **B268**, 427 (1986).
[13] Even though the values of the anomalous dimension given in Refs. [9,10] agree with each other, there appears to be an arithmetic error in Ref. [10] in the computation of the coefficient of α in the evolution equation. This leads to disagreement with the result of Ref. [9], which we follow here.
[14] S. Weinberg, in *A Festschrift for I. I. Rabi,* edited by L. Motz (New York Academy of Sciences, New York, 1977).
[15] J. F. Donoghue and C. R. Nappi, Phys. Lett. **168B**, 105 (1986).
[16] R. L. Jaffe, Phys. Rev. D **21**, 3215 (1980).
[17] J. Gasser and H. Leutwyler, Phys. Rep. **87**, 77 (1982).
[18] The reason for the phenomenological success of these approximations can be traced to the fact that the matrix elements of $\hat{M}$ [Eq. (23)] vanish at zero momentum transfer in the naive parton model. In the QCD-improved parton model the leading nonzero twist contributions to $\hat{M}$ appear at the same order in m/E (i.e., they vanish in the chiral limit) and higher order in α_s; hence, they are strongly suppressed. Furthermore, use of popular parametrizations of the quark distributions suggests that the error made by replacing m/E by a constant (of order 1) is small. A detailed discussion of the relation between the

operator $\bar{\psi}\psi$ and the quark plus antiquark number is given by M. Anselmino and S. Forte, Torino University Report No. DFTT 92/6, 1992 (unpublished).

[19] Equation (27), as it stands, can only hold for nonsinglet combinations of quark numbers n_i; otherwise, its RHS is certainly divergent, as discussed in the previous section. In such a case, it could be derived only by assuming a suitably subtracted nonsinglet first moment (1). A detailed discussion is given by R. L. Jaffe and C. H. Llewellyn-Smith, Phys. Rev. D **7**, 2506 (1973). Here we shall consider only nonsinglet combinations, which are finite.

[20] M. Anselmino and E. Predazzi, Phys. Lett. B **254**, 203 (1991).

[21] P. Gensini, Nuovo Cimento A **102**, 75 (1989); J. Gasser, H. Leutwyler, and M. E. Sainio, Phys. Lett. B **253**, 252 (1991).

[22] V. De Alfaro, S. Fubini, G. Furlan, and C. Rossetti, *Currents in Hadron Physics* (North-Holland, Amsterdam, 1973); M. A. Shifman, Usp. Fiz. Nauk **157**, 561 (1989) [Sov. Phys. Usp. **32**, 289 (1989)].

[23] J. C. Collins, A. Duncan, and S. D. Joglekar, Phys. Rev. D **16**, 438 (1977).

[24] We define the β function as in Ref. [9]. Notice that this definition differs from that used in Ref. [23].

[25] We normalize spinor fields to 1 rather than to $2M$ as in Refs. [17,22].

[26] D. J. Gross, S. B. Treiman, and F. Wilczek, Phys. Rev. D **19**, 2188 (1979).

[27] G. M. Shore and G. Veneziano, Phys. Lett. B **244**, 75 (1990); Nucl. Phys. **B381**, 23 (1992).

[28] J. Stern and G. Clément, Phys. Lett. B **264**, 426 (1991); Nucl. Phys. **A539**, 685 (1992).

[29] R. Tarrach, Nucl. Phys. **B183**, 384 (1981).

[30] R. L. Jaffe, in *Pointlike Structures Inside and Outside Hadrons,* edited by A. Zichichi (Plenum, New York, 1977).

[31] P. Jain *et al.*, Phys. Rev. D **40**, 855 (1989).

[32] T. W. Donnelly, J. Dubach, and I. Sick, Nucl. Phys. **A503**, 589 (1989).

[33] M. Chemtob, Nuovo Cimento A **89**, 381 (1985).

[34] J. P. Blaizot, M. Rho, and N. Scoccola, Phys. Lett. B **209**, 27 (1988). The profile function in this reference is determined by the numerical solution of a slightly different model; it is, however, not expected to be very different [M. Rho (private communication)].

[35] E. Guadagnini, Nucl. Phys. **B236**, 35 (1984).

[36] B.-Y. Park and M. Rho, Phys. Lett. B **220**, 7 (1989).

[37] S. J. Brodsky, J. Ellis, and M. Karliner, Phys. Lett. B **214**, 75 (1990).

[38] E. M. Henley and G. A. Miller, Phys. Lett. B **251**, 453 (1990).

[39] A. Signal, A. W. Schreiber, and A. W. Thomas, Mod. Phys. Lett. A **6**, 271 (1991).

[40] M. Anselmino and S. Forte (in preparation).

[41] S. Kumano, Phys. Rev. D **43**, 59 (1991).

[42] S. Forte, Acta Phys. Pol. **B22**, 1063 (1991).

[43] S. Forte and E. V. Shuryak, Nucl. Phys. **B357**, 153 (1991).

[44] P. N. Harriman, A. D. Martin, W. J. Stirling, and R. G. Roberts, Phys. Rev. D **42**, 798 (1990).